DEVELOPMENT OF A NON-IMMUNOLOGICAL SYSTEM FOR THE STUDY OF THE CELLULAR LOCALIZATION OF BRCA1 GENE PRODUCT IN LIVING CELLS

A THESIS SUBMITTED TO
THE DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS
AND THE INSTITUTE OF ENGINEERING AND SCIENCE
OF BİLKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE

BY
TOLGA ÇAĞATAY
AUGUST 1997

DEVELOPMENT OF A NON-IMMUNOLOGICAL SYSTEM FOR THE STUDY OF THE CELLULAR LOCALIZATION OF *BRCA1* GENE PRODUCT IN LIVING CELLS

A THESIS SUBMITTED TO
THE DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS
AND
THE INSTITUTE OF ENGINEERING AND SCIENCE OF
BİLKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE

By

TOLGA ÇAĞATAY

August, 1997

To my parents Nesrin & Erhan Çağatay
and
uncle Namık Türkmenoğlu

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Işık G. Yuluğ

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Mehmet Öztürk

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Meral Özgüç

Approved for the Institute of Engineering and Science

Prof. Dr. Mehmet Baray, Director of Institute of Engineering and Science

ABSTRACT

DEVELOPMENT OF A NON-IMMUNOLOGICAL SYSTEM FOR

THE STUDY OF THE CELLULAR LOCALIZATION OF *BRCA1* GENE

PRODUCT IN LIVING CELLS

TOLGA ÇAĞATAY
M. S. in Molecular Biology and Genetics
Supervisor: Assist. Prof. Işık G. Yuluğ
August 1997, 92 Pages

BRCA1, is a familial breast and ovarian cancer susceptibility gene that has been cloned and shown to be either lost or mutated in families with breast and ovarian cancer. BRCA1, has been postulated to encode a tumor suppressor, a protein that acts as a negative regulator of tumor growth. To explore the biological function of BRCA1, several studies have been performed for the identification of cellular localization of BRCA1 gene product. Results obtained from these immunofluorescent/ immunohistochemical studies generated two opposing views, cytoplasmic localization versus nuclear localization. Here, we describe a non-immunological system employing the Eukaryotic Green fluorescent Protein (EGFP) tag for the study of the cellular localization of BRCA1 gene product in living cells.

Proteins carrying the green fluorescent protein (GFP) of *Aequorea victoria* provide a powerful system to analyze protein expression and targeting in living cells. Fusion proteins containing the GFP tag are therefore valuable tools to analyze nuclear trafficking in living cells. Here, we reporte the use of a mutant GFP, namely Eukaryotic Green Fluorescent Protein (EGFP), as a marker for the protein import into mammalian nuclei. We have analyzed the behavior of a protein domain of the BRCA1, that contains five putative nuclear localization signals (NLSs), *in vivo* using a chimera constructed from this polypeptide and the EGFP. This *in vivo* studies showed that EGFP was distributed uniformly throughout the cytoplasm and the nucleus. When EGFP was fused to NLSs containing domain of the BRCA1 protein, fluorescent was predominantly detected in the nucleus, showing that these potential NLSs consensus sequences may destinate the full-lengh BRCA1 producy into the nucleus of mammalian cell. This study has also shown that EGFP can be used as a potential fluorescent tag for visualization of gene expression and cellular protein localization in living cells.

ÖZET

BRCA1 GEN ÜRÜNÜNÜN

HÜCRE İÇİNDEKİ LOKALİZASYONUNU İNCELENMEK İÇİN

NON-İMMÜNOLOJİK BİR SİSTEMİN GELİŞTİRİLMESİ

TOLGA ÇAĞATAY
Yüksek Lisans Tezi, Moleküler Biyoloji ve Genetik Bölümü
Tez Yöneticisi: Yardımcı Doçent. Dr. Işık G. Yuluğ
Ağustos 1997, 92 sayfa

Ailesel meme ve ovaryum kanserinden sorumlu olan *BRCA1* geni klonlanmış ve meme ile ovaryum kanseri olan ailelerde genin ya mütasyona uğradığı yada kaybolduğu gösterilmiştir. *BRCA1* geninin, tümör büyümesinde negatif düzenleyici olarak rol alan bir tümör baskılayıcı proteini kodladığı ileri sürülmüştür. BRCA1'ın biyolojik işlevinin incelenmesi için BRCA1 gen ürününün hücre içi yerinin belirlenmesini amaçlayan bazı çalışmalar yapılmıştır. Bu immünoflöresan/ immünohistokimyasal çalışmalardan elde edilen sonuçlar gen ürününün sitoplasmada veya hücre çekirdeğinde olduğuna dair iki karşıt görüş ortaya çıkarmıştır. Bizde, canlı hücrede immünolojik olmayan bir sistemde çalışarak BRCA1 gen ürününün hücre içerisindeki yerini tanımlıyoruz.

Aequerea victoria' nın yeşil flöresan protein (GFP) taşıyan proteinler, protein sentezin ve hedeflenmesinin canlı hücre içinde analizi için güçlü bir sistem sağlarlar. Bu yüzden GFP içeren füzyon proteinler canlı hücrede çekirdek trafiğini analiz etmede değerli bir araçtırlar. Bu çalışmada ökaryotik yeşil flöresan protein (EGFP) olarak bilinen bir çeşit mutant GFP' nin, memeli hücre çekirdeğine taşınan bir proteininin işaretlenmesindeki kullanılımı rapor edilmiştir. BRCA1 proteinin beş adet çekirdek lokalizasyon sinyalini (NLSs) içeren parçasının EGFP ile birleştirilmesiyle yapılan kimerik proteinin canlı hücre içerisindeki davranışı incelenmiştir. Yapılan in vivo incelemenin sonucunda, EGFP'nin tek başına sentezlendiğinde sitoplazmaya ve çekirdeğe eşit bir şekilde dağıldığı gözlenmiştir. EGFP'nin BRCA1 proteinin NLSs içeren parçasıyla birleştirildiğinde flöresan sinyal dominant bir biçimde hücrenin çekirdeğinde tespit edilmesi ise bu NLSs sekanslarinin tam uzunliktaki BRCA1 proteinini hücre çekirdeğinde lokalize edebilecegini gösterilmiştir. Ayrıca bu çalışmada

ACKNOWLEDGMENT

It is my pleasure to express my deepest gratitude to my advisor Assist. Prof. Işık G. Yuluğ for her supervision to my graduate study. Without her excellent logic and knowledge none of this work could have been produced.

I wish to express my thanks to Prof. Mehmet Öztürk who made it possible to have an excellent working environment in Turkey and thanks to his help.

My thanks to Dr. Ergün Pınarbaşı and Dr. Aydın Yuluğ for their advice and discussions on my project. Many thanks to Assist. Prof. Marie D. Ricciardone, M. S. Hilal Özdağ and biologist Birsen Cevher for automated sequencing study of the Nls4 construct.

I appreciate moral support by research assistants of Department of Molecular Biology and Genetics in Bilkent University, especially to Tuba, Şafak , Hilal, Cemaliye, Reşat and biologist Lütfiye Mesci for their endurance toward any trouble that I caused in course of close interactions. Thanks to my laboratory partners Kezi, Berna and Emre for their reciprocal warm feelings.

Very special thanks to my university friends and friends from İstanbul who gave me further encouragement and moral during my thesis.

Finally, my ultimate thanks to my family for always giving their unconditioned interest and support.

TABLE OF CONTENTS

SIGNATURE PAGE		ii
ABSTRACT		iii
ÖZET		iv
ACKNOWLEDGMENTS		v
TABLE OF CONTENTS		vi
LIST OF FIGURES		x
LIST OF TABLES		xii
ABBREVIATIONS		xiii
CHAPTER 1.	INTRODUCTION	1
1.1	Hereditary Breast and Ovarian Cancer and BRCA1 (Breast Cancer 1)	1
1.2	Linkage analysis and cloning of *BRCA1* gene	1
1.3	Genotype-phenotype correlation and mutations in the *BRCA1* gene	2
1.4	Structural analysis of *BRCA1* gene	3
1.5	Structural and functional analysis of the BRCA1 protein	6
1.6	Subcellular localization of *BRCA1* gene product	10
1.7	Practical approaches to the structural and functional analysis of a protein	12
1.8	AIM AND STRATEGY	14

CHAPTER 2. MATERIALS AND METHODS 16

2.1 MATERIALS 16

2.1.1 Reagents 16

2.1.2 Bacterial strains 16

2.1.3 Enzymes 17

2.1.4 Nucleic acids 17

2.1.5 Oligonucleotides 17

2.1.6 Electrophoresis and photography 18

2.1.7 Tissue culture reagents and cell lines 18

2.2 SOLUTION AND MEDIA 19

2.2.1 General solutions 19

2.2.2 Microbiological media and antibiotics 19

2.2.3 Tissue culture solutions 20

2.3 METHODS 21

2.3.1 General methods 21

2.3.1.1 Transformation of *E.coli* 21

2.3.1.2 Growth and storage of bacterial strains 22

2.3.1.3 Plasmid DNA preparation 22

2.3.1.4 Extraction and precipitation of DNA 23

2.3.1.5 Quantification and qualification of nucleic acids 24

2.3.1.6 Restriction enzyme digestion of DNA 25

2.3.1.7 Agarose gel electrophoresis 25

2.3.2 Computer analysis of DNA sequences 26

2.3.3 Construction of the pEGFP-N2-BRCA1$_{(83-2436)}$ eukaryotic expression vector 27

2.3.3.1 Eukaryotic Green Fluorescent Protein-N-Terminal protein fusion vector, pEGFP-N2 27

2.3.3.2 pLXSN-BRCA1 vector 29

2.3.3.3 Contruction of vector encoding a 2316 bp fragment of BRCA1 fused with the N-nerminal of the EGFP 30

2.3.3.4 Klenow treatment of Nls 31

2.3.4 Automated sequencing 32

2.3.5 Tissue culture techniques 32

2.3.5.1 Cell line 32

2.3.5.2 Growth conditions 34

2.3.5.3 Cryopreservation of cell line 34

2.3.5.4 Transfection of eukaryotic cells using electroporation 35

2.3.5.5 Fluorescent signal detection 36

CHAPTER 3. RESULTS 38

3.1. Computer analysis of the *BRCA1* sequence 38

3.1.1 Computer analysis of the BRCA1 protein sequence 38

3.1.2 Computer analysis of the pLXSN-BRCA1 construct 39

3.2 Qualification of the pLXSN-BRCA1 and pEGFP-N2 40

3.3 Endonuclease digestion of the pLXSN-BRCA1 and pEGFP-N2 41

3.3.1 *Eco*47 III and *Kpn* I double digestion of pEGFP 41

3.3.2 *Kpn* I and *Ehe* I double digestion of pLXSN-BRCA1 42

3.4 Construction of the pEGFP-BRCA1$_{83-2436)}$ eukaryotic expression vector 43

3.4.1 Ligation reaction 43

3.4.2 Selection of positive colonies after transformation 44

3.4.3 Klenow treatment of the Nls4 construct 46

3.4.4 Automated sequencing of *Kpn* I junction of the $BRCA1_{(83-2436)}$-EGFP fusion sequence 48

3.5 Expression Analysis of the pEGFP-N2 vector and the pEGFP-$BRCA1_{(83-2436)}$ Construct in Eukaryotic System 51

3.5.1 Transfection of MCF-7 by Electroporation 51

3.5.2 Expression analysis of the pEGFP-N2 vector in living cells 52

3.5.3 Monitoring the expression and the cellular localization of EGFP-$BRCA1_{(83-2436)}$ fusion protein 54

3.5.4 Effect of EGFP and EGFP-$BRCA1_{(83-2436)}$ fusion protein overexpression in MCF-7 cell line 58

CHAPTER 5. DISCUSSION 60

REFERENCES 64

APPENDICES

Appendix 1 72

Appendix 2 76

Appendix 3 79

Appendix 4 92

LIST OF FIGURES

Figure. 1	The human BRCA1 cDNA as described by Miki *et al* (1994)	3
Figure 2.	Restriction map and multiple cloning site of pEGFP-N2.	28
Figure 3	*Not* I / *Eco*R I double digest profile of pEGFP-N2 and mini-prep plasmid DNA of n1, n2, n3 colonies	40
Figure 4	*Sac* I digestion profile of the original PLXSN-BRCA1 and mini-prep plasmid DNA of the IT3 colony on 0.8% agarose gel	41
Figure 5	*Eco*47 III and *Ehe* I double digestion profile of pEGFP-N2 vector on 0.8% agarose gel	42
Figure 6	*Kpn* I and *Ehe* I digestion profile of pLXSN-BRCA1	43
Figure 7	Ethidium bromide intensities of the $BRCA1_{(83-2436)}$ fragment (insert) and the pEGFP-N2 (vector) dots under ultraviollet illuminator	44
Figure 8	*Eco*R I digestion profile of the positive colonies	45
Figure 9	*Not* I / *Eco*RI double digest profile of mini-prep plasmid DNA isolated from the Nls3, NL4 and the control ligation colony2	45
Figure 10	The recombinant pEGFP-N2 construct, Nls4 (7076 bp)	46
Figure 11	Schematic representation of the Klenow treatment and self ligation steps in the construction of pEGFP-$BRCA1_{(83-2436)}$	47
Figure 12	*Kpn*I digestion profile the Klenow treated Nls^{Kpn-} colonies on 8% agarose gel	48
Figure 13	The sequencing result of *Kpn* I junction	49
Figure 14	Phase-contrast microscopy appearance of MCF-7 cells	51

Figure 15	Unfixed MCF-7 cells transfected with EGFP-N2	53
Figure 16	Unfixed MCF-7 cells transfected with EGFP-BRCA1$_{(83-2436)}$	55
Figure 17	Fixed MCF-7 cells transfected with EGFP-BRCA1$_{(83-2436)}$	56
Figure 18.	Effect of EGFP and EGFP-BRCA1$_{(83-2436)}$ expression on the viability of MCF-7 cells	59

LIST OF TABLES

Table 1	Characteristics of MCF-7 cell line	33
Table 2	Results of the PSORT computer analysis for BRCA1 sequence	39
Table 3	Percent survival of MCF-7 cells that over express EGFP	58
Table 4	Percent survival of MCF-7 cells that over express the EGFP-BRCA1$_{(83-2436)}$	58

ABBREVIATIONS

ATP	adenine triphosphate
bp	base pair
cDNA	complementary DNA
cm	centimeter
dATP	adenosine deoxyribonucleoside triphosphate
dCTP	cytosine deoxyribonucleoside triphosphate
dGTP	guanosine deoxyribonucleoside triphosphate
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
dTTP	thymine deoxyribonucleoside triphosphate
EDTA	ethylenediaminetetra-acetic acid
EGFP	eukaryotic green fluorescent protein
FCS	foetal calf serum
FITC	fluorescein isothiocyanate
GFP	green fluorescent protein
xg	gravity
g	gram
h	hour
H33258	the fluorochrome dye H33258
kb	kilobase
kV	kilovolt
LB	Luria-Bertani medium
M	molar
ml	milliliter
min	minute
mRNA	messenger RNA
ms	millisecond
OD600	optical density at 600 nm
PBS	phosphate buffered saline
rpm	revolution per minute

RNA	ribonucleic acid
RNase	ribonuclease
SDS	sodium dodecyl sulphate
sec	second
TAE	tris/acetic acid/EDTA buffer
Tris	2-amino-2-[hydroxymethyl]- 1,3 propandiol
U	unit
V	volt
v/v	volume for volume
w/v	weight for volume
w/w	weight for weight
μF	microfaraday

CHAPTER 1. INTRODUCTION

1.1. Hereditary Breast and Ovarian Cancer and *BRCA1* (Breast Cancer 1) Gene

Breast cancer is the most frequent malignancy in women. The total lifetime risk for developing a breast cancer in the general population is estimated to be 10% (Wooster *et al.*, 1995). Breast cancer has been estimated to be one of the most common hereditary malignant diseases. Before the identification of *BRCA1* (BReast CAncer susceptibility gene 1), pedigrees have been used in breast cancer risk estimations. Familial cancer clinic studies have identified that a woman who has breast cancer case(s) in her first degree relatives (mother, father, aunts, uncles, etc.) has an elevated risk of disease and that the younger the age of diagnosis in her relatives, the higher the risk. The families showed the epidemiological characteristics of familial, versus sporadic breast cancer, younger age at diagnosis, frequent bilateral disease, and frequent occurrence of disease among men. Epidemiological studies have also shown that genetic, hormonal and environmental factors have a role in the etiology of breast cancer (Langston *et al.* ,1996).

1.2. Linkage analysis and cloning of the BRCA1 gene

Skolnick *et al.* (1990) identified a region on the long arm of chromosome 17 by linkage analysis; the17q21 region appeared to contain a gene for inherited

susceptibility to breast cancer in families with early-age onset of breast cancer (Hall *et al.*,1990). Subsequently, several gene-hunting studies were performed to find a candidate gene for familial breast cancer within this region. Finally in 1994, the *BRCA1* gene that affects the tumorogenesis of breast and ovarian cancer or both, has been identified from a large, genetically defined 17q21 region by positional cloning (Miki *et al.* ,1994, Brown *et al.* , 1995, Harshman *et al.*, 1995, Tonin *et al.* , 1995).

1.3. Genotype-phenotype correlation and mutations in the BRCA1 gene

Inherited mutations in the *BRCA1* gene in female carriers have been implicated in a predisposition to breast cancer with 87% lifetime risk, and families who have breast/ovarian cancer in their family history have a 44% risk of developing breast cancer. Furthermore female carriers have a 4-fold increased risk of colon cancer, while male carriers face a 3-fold increased risk of developing prostate cancer (Durocher *et al.* 1996, Futreal *et al.* 1994).

Cytogenic studies have shown that the most common genetic abnormality in breast cancer (as in most tumors) is loss of heterozygosity (LOH) besides gene amplifications. The frequency of loss of heterozygosity in the *BRCA1* region is between 40% - 80% among sporadic breast cancer and varies between 30% - 60% in sporadic ovarian cancer cases (Rowell *et al.* ,1994). More than 70 distinct germ-line mutations have already been identified through the screening of the *BRCA1* gene (Friedman *et al.* ,1995, Berman *et al* . , 1996, Eisinger *et al.* ,1996, Inoue *et al.* ,1996, Hogervorst *et al.* ,1995 Struewing *et al.* , 1995). Penetrance of *BRCA1* is incomplete

and depends on both age and gender so not all carriers of germ-line mutation will develop a breast cancer.

Alterations in the *BRCA1* gene can be listed as frame shift mutations, non-sense mutations and splice or regulatory region alterations. These mutations account for approximately 85% of the cumulative *BRCA1* mutations, while the remainder are due to missense mutations (Miki *et al.* ,1994). More than 75% of these mutations result in the truncation of the *BRCA1* protein.

1.4. Structural analysis of the BRCA1 gene

Miki *et al.* (1994) elucidated the structure of the *BRCA1* gene . The *BRCA1* gene spreads over approximately 100 kb of the long arm of chromosome 17. The genomic structure of the *BRCA1* gene is composed of 24 exons, 22 of which encode a 7.8 kb mRNA (Figure 1).

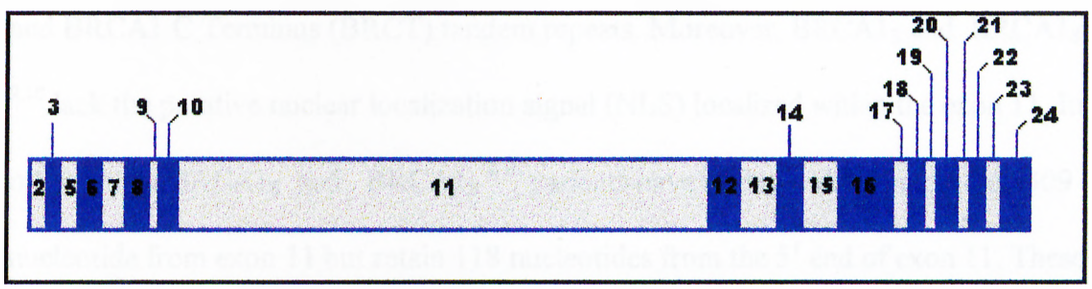


Figure 1. The human BRCA1 cDNA as described by Miki *et al* (1994).

Alternative splicing has been observed both in normal and malignant mammary tissue and placenta, but the significance of this is not known (Xu *et al.* ,1995). The 7.8 kb mRNA is abundant in the mammary gland and placenta and also in the testis and thymus, but *BRCA1* expression is not restricted to these tissues; it is also expressed in lymphocytes and hepatocytes although at a very low level

The *BRCA1* gene encodes a 7.8 kb mRNA transcript and this transcript encodes a 1863-amino acids protein, so approximately 73.2 % of the complete transcript is the coding sequence.

Lu *et al.* (1996) have reported four splice variants of the *BRCA1* in nonmalignant and tumor-derived breast cells by sequence analysis of reverse transcribed, PCR-amplified trancripts: the full-length *BRCA1* (*BRCA1*$_L$, 7.8 kb), the internally deleted sequence leading to a protein lacking amino acids between 264 and 1366 of *BRCA1*$_L$ (*BRCA1*$_S$, 4.4 kb), and two very minor variants lacking exons 9 and 10, referred to as *BRCA1*$_L^{-9,10}$ and *BRCA1*$_S^{-9,10}$. All variants contain the N-terminal RING motif, the C-terminal acidic activation domain (Jensen *et al.*, 1996) and BRCA1 C Terminus (BRCT) tandem repeats. Moreover, BRCA1$_S$ and BRCA1$_S^{-9,10}$ lack the putative nuclear localization signal (NLS) localized within the exon 11. In other words BRCA1$_S$ and BRCA1$_S^{-9,10}$ variants have in-frame deletion of the 3309 nucleotide from exon 11 but retain 118 nucleotides from the 5' end of exon 11. These splice variants are found on polysomes and are predicted to encode 80-85 kDa BRCA1-derived proteins beside the full-length *BRCA1* gene product (Lu *et al.*, 1996).

The *BRCA1* gene appears to be conserved in mammals, however the presence of the *BRCA1* gene in the genome of other species is one of the open questions should be handled. The mouse *Brca1* gene, which maps on chromosome 11 and specifically on the 11D region, has 75% identity of coding sequence with human *BRCA1* sequence at the nucleotide level and 56% identity at the predicted amino acids sequence.

Multiple BRCA1 proteins, approximately 245, 185-220, 160, , 100, 52, and 38 kD in size, have been identified in both human and mouse cell lines by immunohistochemical methods (Chen *et al.* ,1995, Rao *et al.* ,1996) and all these proteins are phosphoproteins. It is not known whether they are the isoform of BRCA1 protein or its related proteins. These conflicting results arise from the usage of different immunofluorescent and immunohistochemical methods.

The wild-type *BRCA1* allele is often lost in cancers that arise within breast cancer families, presumably leaving the cell without any functional *BRCA1* protein. It has been shown that when chromosomal loss is defined in breast and ovarian tumors from patients who carry the *BRCA1* predisposition alleles, the wild-type copy of the gene is constantly lost while the mutant allele is kept (Munn *et al.* ,1996, Neuhausen *et al.* , 1994 and 1996). This is the familiar pattern expected from the loss of function of a tumor suppressor gene. Therefore, the gene product of *BRCA1* is thought to be a tumor suppressor gene. The tumor suppressor function of the *BRCA1* gene product has been demonstrated by inhibition of endogenous *BRCA1* expression with antisense

RNA in mouse fibroblast cells that resulted in neoplastic transformation (Rao *et al.* ,1996), and by expression of wild-type *BRCA1* gene in breast cancer cell lines which resulted in growth retardation and tumor inhibition (Holt *et al.* , 1996).

1.5. Structural and functional analysis of the BRCA1 protein

Conceptual translation of the *BRCA1* cDNA (Genbank accession no: HSU14689) reveals an open reading frame (ORF) beginning at nucleotide 119 and encoding a protein of 1863 amino acids (Appendix 1). According to analysis done with a computer based peptide analysis program(*PSORT* - protein analysis program, *http://psort.nibb.ac.jp*), the BRCA1 protein seems to have five nuclear targeting sequences at the positions of amino acids 502, 503 , 504, 603 and 650 with 0.7000/1 certainty and the protein has no obvious membrane spanning regions or N-terminal signal sequence. The *BRCA1* protein is highly charged; 5% of the total protein sequence is composed of negatively charged amino acid residues, while approximately 11% is positively charged. The excess negative charge is particularly concentrated near the C-terminus (ExPASy-*ProtParam* : Protein primary structure analysis computer program, *http://expasy.hcuge.ch/sprot/protparam.html.*).

Studies on the physiological function of the BRCA1 protein and its role in breast and ovarian carcinogenesis have accelerated with the identification of the mouse homologue of breast cancer associated gene, *Brca1*. Specific and dynamic expression of *BRCA1* protein during differentiation and embryogenesis was studied on animal models and an absolute requirement for *BRCA1* protein has been

demonstrated. Differentiation and proliferation of mouse mammary epithelial cells is directly correlated with the level of BRCA1 expression(Lane *et al.* ,1995) and it has been reported that inactivation of *Brca1* gene in the mouse embryo (mutant *Brca*1 $^{-/-}$) resulted in either neural tube defects at E9.5 (Gowen *et al.*, 1996) or the failure of differentiation and formation of the egg cylinder (Liu *et al,* 1996). Interestingly, the one woman homozygous for a germ-line *BRCA1* mutation who was identify by Boyd lead to an expectation that a *BRCA1*$^{-/-}$ or *Brca1*$^{-/-}$ mutant would be viable (Boyd *et al.* , 1995). A new biological function of the *BRCA1* protein as an inducer of apoptosis has been speculated in a recent report in which it was shown that lack/decreased level of functional *BRCA1* protein results in a decreased response to apoptotic stresses in mouse fibroblast cell lines and human breast cancer cell lines (Shao *et al.* ,1996).

Despite the accumulated data, the function of *BRCA1* in either normal development or tumorigenesis remains unknown. So far, several searches for functional domains in the *BRCA1* sequence have come up with the discovery of four different conceptual consensus sequences. These functional motifs are an N-terminal RING-finger (C3HC4 type zing finger) or A-box domain, a C-terminal acidic blob domain (Miki *et al.* ,1994, Futreal *et al.* ,1994), a granin consensus at the central region of the protein (Jensen *et al.* ,1996), and a globular domain within the C-terminal of the protein called BRCT that contains an analogous region of a human p53 binding protein1 (53BP1; Koonin *et al.* ,1996).

Beyond the determination of these sequences by computer sequence homology and alignment programs, Jensen *et al.* (1996) have shown that the BRCA1 protein is secreted and present in breast milk, and that it shares many biochemical characteristics of the granin-1 family of proteins such as heat stability, acidity and vesicle localization (Jensen, *et al,* 1996).There are also reports that support the granin feature of the BRCA1 protein as a regulated secretory protein. These reports show that *BRCA1* is upregulated during pregnancy (Lane *et al.* ,1995) and its expression is induced by estrogen (Gudas *et al.* ,1995). However, the recent article by Hakem *et al.* (1996) has suggested that *BRCA1* protein may act as a transcriptional regulator. They have shown that homozygous *BRCA1* mutant-mice embryos (*BRCA1* $^{-/-}$) die before 7.5 days of embryogenesis due to the reduced cell proliferation coupled with the decreased expression of cyclin E (one of the key components of the G1/S transition of cell cycle) and mdm-2 (a negative regulator of p53 activity), and they also noticed significantly increased expression of a cyclin-dependent kinase inhibitor, p21 (Hakem *et al.* ,1996).

Another important evidence about the physiological function of BRCA1 protein has recently come from the study of Wu *et al.* (1996) They identified a novel protein that interacts with BRCA1 under both *in vitro* and *in vivo* conditions. Yeast and mammalian two hybrid system, and immunoprecipitation analysis have shown that this BRCA1-associated RING domain (BARD1) protein specifically forms a stable heterodimer complex through binding with the BRCA1 1-184 amino acid residues where the cystein-rich RING motif is found. Further molecular and biological analysis of BARD1 protein has shown that BARD1 is transcribed from chromosome 2 and

two major BRAD1 transcripts have been observed in several breast and ovarian cancer cell lines (e.g., ZR-75, T-47D, BT-483). BARD1 protein contains an N-terminal RING motif, three tandem ankyrin repeats and a C-terminal sequence with significant homology to the BRCT domains of the C-terminus of BRCA1. More interestingly. it has also been shown that the C61G and C64G missense mutations of *BRCA1*, which are directly related to breast cancer susceptibility, prevent the formation of the BRCA1/BARD1 complex (Wu *et al.* , 1996).

The C-terminal region of BRCA1 has a highly conserved stretch in both mouse and human *BRCA1* sequences (Abel *et al.*, 1995). Different germ-line mutations (Ala- 1708→ Glu, Gln- 1765 C+, Met- 1775→ Arg and Try- 1858→ stop) have been reported in patients with breast or ovarian cancer (Futreal *et al.*, 1994, Langston *et al.*, 1996, Gayther *et al.*, 1995 and Serova *et al.*, 1996). Monteiro *et al.* have investigated whether the C-terminal region of BRCA1 is able to activate transcription by using both mammalian and yeast two- hybrid systems and they have reported that the C-terminal acidic region spreading the exons16- 24 (aa 1560- 1863) is able to activate transcription, and the region comprising exons 21- 24 (aa 1760-1863) is the smallest region required for sufficient transactivation function of BRCA1 (Monteiro *et al.*, 1996).

Studies for the identification of the biochemical and biological functions of BRCA1, which is a large gene with probably many functional domains, have found entirely different properties of the gene. Two recent striking data come from Scully *et al.* (1997). In their first study they reported that Brca1 binds to human Rad51

(homologous of bacterial RecA), a protein that is involved in the integrity of the genome. The human Rad1 had an ability promote ATP-dependent homologous pairing and strand-transfer *iv vitro*, but precise function(s) of the mammalian Rad1 protein is not clear yet. BRCA1 immunostaining displayed distinct nuclear foci during the S phase of the cell cycle, like the human Rad51 protein. These two proteins were colocalized *in vivo* on the synaptonemal complexes (junction between meiotic chromosomes, necessary for homologous recombination) in meiotic cells, and were found on asynapsed (axial) elements of human synaptonemal complexes. A coimmunoprecipitation study revealed that BRCA1 residues 758-1064 are required for formation of Rad51-containing complexes *in vitro* (Scully *et al.*, 1997).

The second study by Scully *et al.* (1997) proposed that the BRCA1 protein product is a transcriptional factor which processes RNA polymerase II holoenzyme (pol II)-bound protein, and it has also been shown that the C-terminal 11 amino acid residues of BRCA1 which have an identified role in trans-activation function (Monteiro *et al.*, 1996 and Chapman *et al.*, 1996) are important for holoenzyme binding,

1.6. Subcellular localization of the BRCA1 gene product

One of the hot topics in *BRCA1* studies is the identification of the *BRCA1* protein function. As cellular localization can provide clues about protein function, subcellular localization of the BRCA1 protein is studied intensively. Two conflicting points of view about BRCA1 subcellular localization have been generated by Chen *et*

al. (1996) and Jensen *et al.* (1996) so far. Results obtained by Chen *et al.* have indicated that while BRCA1 protein in normal cells is destined to the nucleus in breast and ovarian tumor cell lines the *BRCA1* protein is aberrantly localized to the cytoplasm (Chen *et al.* , 1995, Chen *et al.* , 1996). Subsequently, Scully *et al.* (1996) suggested that nuclear distribution of the BRCA1 is a general characteristic of the cell, regardless of whether it is normal or not (Scully *et al.* , 1996). Their work seems to agree that BRCA1 is a 220 kD protein localized in the nuclei of breast epithelial cells.

Jensen *et al.* (1996) have recently described results that contravenes substantially those of Chen *et al.* (1996) and Scully *et al.* (1996). They defined BRCA1 as a 190 kD granin-1 type protein which is localized to membrane vesicles regardless of the cell type (Jensen *et al.* , 1996).

More recently, in an attempt to characterize the subcellular localization of BRCA1 in more detail, Wilson *et al.* (1997) have studied the subcellular localization of both full-length BRCA1 (5.7 kb) and an exogenous BRCA1 spliced variant, namely BRCA1-Δ11b (previously described as $BRCA1_S$ by Lu *et al.* , 1996).Their results have shown that full-length BRCA1 protein product was a nuclear protein, whereas BRCA1-Δ11b was localized to the cytoplasm. They also reported that unlike the full-length protein, over expression of the protein encoded by splice variant did not appear to be toxic to the cell. Another interesting result of Wilson *et al.* is that the expression of BRCA1-Δ11b was reduced or completely absent in several breast and ovarian tumor cell lines (Wilson *et al.*, 1997).

All these subcellular localization analyses were carried out by using commercially available variety of antibodies raised against the C-terminal, N-terminal and exon 11 of *BRCA1* proteins. Chen *et al.* (1996) , Scully *et al.* (1996) , Jensen *et al.* (1996) and Wilson *et al.* (1997) have used the same antibodies such as C-20, N-25, D-20, I-20 but surprisingly their results were different. Since they employed either immunoflourescent or immunohistochemical methods in their analysis, the cross-reactivity capacity of the antibodies, the genotype and the phenotype of the host cell line, and features of the specimen fixation might have resulted in artifacts that give rise to a false-positive results. Likewise, C-20 antibody has the problem of cross-reactivity since it recognizes both human epidermal growth factor receptor (EGFR) and HER2 as well as the *BRCA1* protein (Wilson *et al.* , 1996). Therefore the specificity problem of antibodies is the most important disadvantage of immunoscreening techniques in the clinical and molecular analysis of the BRCA1 protein.

1.7. Practical approaches to the structural and functional analysis of a protein

Several methods are available to monitor gene activity and protein distribution within cells. These include the formation of fusion proteins with coding sequences for a reporter gene such as, firefly or bacterial luciferase, and β-galactosidase(Old *et al,* 1994 , Brown , 1993) . Because such methods require exogenously added substrates or cofactors, they are of limited use with living tissue. Another application called the epitope tagging of protein is the combination of fusion protein and

immunofluoresence or immunohistochemical methods. Tagging of a known sequence of 8-12 amino acids , called an epitope, (for example; flag myc) in frame with the sequence of protein of interest overcomes the need for antibodies specific for each studied protein. Tagged wild-type or mutant form of the sequence of interest can be introduced into the cultured cell to allow identification of *in vivo* localization and functional analysis of a whole protein or specific domains (Pierre, 1996).

Nowadays the most commonly used technology is to tag the protein of interest with green fluorescent protein (GFP) which is the bioluminescent protein of jellyfish *Aequorea victoria*. GFP can either be used in immunofluorescence (or -histochemical methods) as an epitope or can be directly visualized with the fluorescence microscope (Peters *et al* , 1995, Amsterdam *et al* , 1995, Webb C. D. , 1995, Olson *et al* , 1995, Ogawa *et al*, 1995, Wang *et al* , 1994). Wild-type GFP is a protein of 238 amino acids having a chromophore structure that absorbs blue light (395-470 nm) and emits green light (509 nm). Chromophore formation is not species-specific and occurs either through the use of ubiquitous cellular components or by autocatalysis (Chalfie *et al.* ,1994). GFP has several ideal characteristics over the other tag proteins as it requires neither an additional gene from *Aequorea victoria* nor exogenous additional substrate and cofactor. GFP persists after formaldehyde treatment in specimen fixation required analysis, and the most important characteristic is that GFP combined with the protein of interest preserves both the fluorescence of GFP and all the targeting and function of the protein of interest (Chalfie *et al* , 1994, Cubitt *et al* , 1995 and Olson *et al* , 1995)

1.8 AIM AND STRATEGY

This project aims to establish a non immunological model system that permits kinetic studies of subcellular localization of BRCA1 protein in a living eukaryotic system. In order to set up such a system, we used a novel genetic reporter system that employs the green fluorescence protein under the fluorescence microscope. When expressed in prokaryotic or eukaryotic cells and illuminated by blue light, GFP yields a bright green fluorescence. Additionally, detection of GFP can be performed with living tissues instead of fixed sample. The use of GFP in these capacities provides a "fluorescent tag" on the protein, which allows for *in vivo* localization of protein.

In the course of this project 2316 bp coding sequence of the putative tumor suppressor *BRCA1* gene was cloned into pEGFP-N2, Eukaryotic Green Fluorescence Protein -N terminal fusion protein vector The *BRCA1* fragment codes for amino acids of 1 to 772 of the full-length BRCA1 protein. This region harbors the 5 nuclear localization signal (NLSs) patterns localized in N-terminal region at position of amino acids 502, 503, 504, 609 and 650, and the RING finger domain at position of amino acids between 25 to 64. The chimera constructed from this peptide and EGFP was tranfected into MCF-7 breast cancer cells to observe where the 2316 bp fragment destinate the BRCA1 protein. The strategy was as follows:

- the nucleotide and the conceptual amino acid sequence of BRCA1 was analyzed with computer programs to determine the protein localization sites on *BRCA1*

- the sequence that contains the NLSs ($BRCA1_{(83-2436)}$) was isolated from the full-length *BRCA1* construct (pLXSN-BRCA1) by restriction endonuclease digestions;
- the isolated DNA fragment was cloned into pEGFP-N2 eukaryotic expression vector to create a fusion protein containing the fluorescent GFP at the C-terminus of the BRCA1 protein which was a valuable tool to analyze subcellular trafficking in living cells;
- pEGFP-N2 vector and the new construct, pEGFP-$BRCA1_{(83-2436)}$, were transfected into MCF-7 cell line by electroporation;
- subcellular localization of the $BRCA1_{(83-2436)}$-GFP fusion protein was monitored with fluorescent microscopy.

CHAPTER 2. MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Reagents

All laboratory chemicals were analytical grade from Sigma Biosciences Chemical Company Ltd. (St. Louis, MO, U.S.A) with the following exceptions: Tris-base was from Stratagene (La Jolla, CA, U.S.A). Ethanol was from Delta Kim Sanayi ve Ticaret A.S (Turkey). Midi-prep kit and Qiaex kit (for recovery and extraction of DNA from agarose gel) were from Qiagen (Chatsworth, CA, U.S.A). Tryptone and yeast extract was obtained from Gibco, BRL Life Technology Inc. (Gaithersburgs, MD, U.S.A). Agar, ampicillin were from Sigma. Kanamycin was from Appligene-Oncor (Illkirch, France).

2.1.2 Bacterial strain

The bacterial strain used in this work was:

***E. coli*, DH5α:** F-, (f80dÊ(*lac*Z)M15), *recA1*, *end*A1, *gyr*A96, *thi*1, *hsd*R17, (r-km-k), *sup*E44,*rel*A1, *deo*R, Ê(*lac*ZYA-ar gF)U169

2.1.3 Enzymes

Restriction endonucleases, Klenow fragment of *E. Coli* DNA polymerase I and T4 DNA Ligase were purchased from MBI FERMENTAS Inc. (NY, U.S.A). The *Kpn* I endonuclease that was obtained from Stratagene GmbH (Heidelberg, Germany). DNase free RNase was from Promega (Madison, WI, U.S.A)

2.1.4 Nucleic acids

DNA molecular weight standard was supplied by Gibco BRL. Ultrapure deoxyribonucleotides were from Boehringer Mannheim GmbH (Mannheim, Germany). Eukaryotic cloning and expression vector pEGFP-N2 (C-terminal protein fusion vector) was purchased from CLONTECH Laboratories, Inc. (CA,U.S.A). Retroviral expression vector containing the *BRCA1* cDNA (nucleotides 85-5711), pLXSN-BRCA1 was a gift from Dr. Tim Crook (Institute of Cancer Research Haddow Lab., Surrey, England).

2.1.5 Oligonucleotides

The sequencing-primer used for cycle sequencing reactions was synthesized in the Beckman Oligo 1000M DNA synthesizer (Beckman Instruments Inc. CA. U.S.A) at the Bilkent University, Faculty of Science, Department of Molecular Biology and Genetics, (Ankara, Turkey).

2.1.6 Electrophoresis and Photography

Electrophoresis grade agarose was supplied from Sigma Biosciences Chemical Company Ltd. (St. Louis, MO, U.S.A). Horizontal electrophoresis apparatuses were from Stratagene (Heidelberg, Germany) and E-C Apparatus Corporation (Florida, U.S.A). The power supply Power-PAC300 was from Bio Rad Laboratories (CA, U.S.A). Imagemaker used in agarose gel profile visualizing was from Herolab except for the video graphic printer/UP-890CE and UPP-110 paper which were obtained from Sony Corporation (Japan).

2.1.7 Tissue culture reagents and cell lines

Dulbecco's modified Eagle's medium (DMEM), fetal calf serum was obtained from Sigma Biosciences Chemical Company Ltd. (St. Louis, MO, U.S.A). L-glutamine, gentamycin, calcium and magnesium-free phosphate buffered saline (PBS) were obtained from Gibco BRL. Penicillin / Streptomycin mixture was from Biological Industries (Haemel, Israel). Tissue culture flasks, petri dishes, 15 ml polycarbonate centrifuge tubes with lids and cryotubes were purchased from Costar Corp. (Cambridge, England).

MCF-7 (ATTC no: HTB-22), human breast epithelial pleural effusion cell line was provided by Prof. Mehmet Öztürk.

2.2 SOLUTIONS AND MEDIA

2.2.1 General solutions

1X Tris-acetic acid-EDTA (TAE):	40mM Tris-acetate, 1mM EDTA
Ethidium bromide:	10 mg/ml in water (stock solution), 30 ng/ml (working solution)
1x Gel loading buffer:	0.25% bromophenol blue, 0.25% xylene cyanol, 50% glycerol, 1mM EDTA

Solutions for plasmid DNA isolation :

Solution I	50 mM Glucose, 25 mM Tris.Cl, pH 8.0, 10 M EDTA. Sterilize in autoclave.
Solution II	0.2 N NaOH, 1% (wt/vol) SDS
Solution III	3 M Potassium acetate, pH 4.8

2.2.2 Microbiological media and antibiotics

Luria-Bertani medium (LB)	*Per liter:* 10 g bacto-tryptone, 5 g bacto-yeast extract, 10 g NaCl. For LB agar plates, add 15 g/L bacto agar.
Glycerol stock solution	65% glycerol, 0.1 M $MgSO_4$, 0.025 M Tris.Cl, pH 8.0

Ampicillin	100 mg/ml solution in double-distilled water, sterilized by filtration and stored at -20°C (stock solution). 100 μg/ml (working solution)
Kanamycin	300 mg/ml solution in double-distilled water, sterilized by filtration and stored at -20°C (stock solution). 30 μg/ml (working solution)

2.2.3 Tissue culture solutions:

H33258 fluorochromo dye	1 mg/ml solution in double-distilled water and stored at -20 °C. 300 μg/ml (working solution).
4% paraformaldehyde	0.04 g/ml solution in PBS. Stored at 4°C
DMEM	*For 500 ml* DMEM: 2mM-Fetal calf serum, 100 U/ml Penicillin, 50 mg/ml Streptomycinand and 1 mM L-Glutmanine. Stored at -4 °C.

2.3 METHODS

2.3.1 General methods

2.3.1.1 Transformation of *E.coli*

Transformation of plasmid DNA into *E.coli* was achieved by using calcium chloride method. The following procedure is based on Ausubel *et al.* (1991).

Preparation of competent cell

500 µl of DH5α glycerol stock solution was inoculated into 5 ml of LB medium containing selective agent and cells were grown at 37°C, shaking at 200 rpm to an optical density at 590 nm (OD_{590}) of 0.4 (approximately for 3 h). 1.5 ml of growing cells were centrifuged at 13,000 rpm for 1 min at 4°C and gently resuspended in 500 µl ice-cold 50 mM $CaCl_2$. After preparation, competent cells were used within 24 h or stored at -80°C for future use.

Transformation

Competent cells were suspended with 500 µl ice-cold 50 mM $CaCl_2$ and centrifuged at 13,000 rpm for 1 min at 4°C. The pellet was resuspended gently in 100 µl of ice-cold 50 mM $CaCl_2$. 1 µl plasmid (1ng/µl) was mixed with the competent cells and incubated on ice for 30 min. The competent cells were heat shocked at 42°C for 90 seconds and the cells were then incubated on ice for 2 min. 1 ml of LB medium was added onto competent cells and incubated at 37°C, 200 rpm for 1 h to allow the expression of antibiotic resistance gene before plating. After the incubation, 200 µl of transformation mixture was plated onto LB agar plates containing 100 µg/ml ampicillin or 30 µg/ml kanamycin to provide a selection for positive colonies

carrying the newly introduced antibiotic resistance gene via transformed plasmid and incubated at 37°C overnight for the selection of antibiotic resistant tranformants.

2.3.1.2 Growth and storage of bacterial strains

A single bacterial colony picked from either an agar plate or a loopfull of bacterial glycerol stock was inoculated into 5 ml LB broth in 15 ml screw capped tubes. The tubes were incubated at 200 rpm at 37°C overnight in a rotator-incubator.

Bacterial strains were stored at -70°C in LB medium containing 50% bacterial glycerol stock solution for long term storage. Recombinant clones were stored under the same condition in media containing the appropriate antibiotic. Strains were maintained as isolated colonies on LB agar plates at 4°C for short term storage.

Bacterial strain used in this study is defined in section 2.1.2.

2.3.1.3 Plasmid DNA preparation

Small scale isolation of plasmid DNA (mini-prep)

This protocol is based on the alkaline lysis method of Birnboim and Doly (1979).

The transformant bacteria strain containing the plasmid of interest was grown in 5 ml LB medium containing 100 μg/ml ampicillin at 37°C, while shaking at 200 rpm overnight. 1.5 ml culture was pelleted in 1.5 ml microfuge at 13,000 rpm for 2 min. After removal of supernatant, the cell were resuspended in 100 μl ice-cold Solution I and stored at room temperature for 5 min. Freshly prepared 200 μl of Solution II was added and mixed in by inverting the tube very gently and then placed on ice for 5 min. Bacterial chromosomal DNA and cell debris were precipitated by

the addition of 150 µl of Solution III. The mixture was then stored on ice for 5 min and centrifuged for 5 min at 13,000 rpm, 4°C to pellet the host DNA and proteins. Supernatant was transferred into a new eppendorf tube, mixed with 800 µl ice-cold absolute ethanol and the mixture was incubated at -20°C for an hour . The plasmid was recovered by centrifugation at 13,000 rpm for 15 min at room temperature. The pellet was washed with 300 µl 70% ethanol and centrifuged for 15 minutes at 13,000 rpm, at room temperature. Ethanol was aspirated and the pellet was dried under vacuum. The pellet was resuspended in 20-30 µl sterile distilled H_2O containing 10 µg/ml RNase A and incubated at 37°C for an hour. The sample was stored at 4°C for short-term or at -20°C for long-term. This procedure yields approximately 1-1.5 µg of DNA.

Purification of plasmid DNA using the Qiagen Kit

The Qiagen 100 kit was used for large scale isolation of pure plasmid DNA. The method is based on the "midi-prep" instructions supplied with the QIAGEN Plasmid Midi Kit (Cat. No. 12145) by Qiagen (Germany).

This procedure yields approximately 60-150 µg of plasmid DNA for 100 ml initial LB culture.

2.3.1.4 Extraction and precipitation of DNA

Extraction and precipitation of DNA from aqueous solution were achieved by using phenol extraction and ethanol precipitation methods.

Phenol extraction

The DNA solution was mixed with an equal volume of 25:24:1 phenol/chloroform/isoamylalcohol and vortexed vigorously. The aqueous and organic phases were separated by centrifugation at 13,000 rpm for 2 Min. The top (aqueous) phase was transferred to a new tube. In order to improve the recovery of DNA (especially in cases where DNA concentration is < 1 μg), the organic phase was re-extracted with 100 μl double-distilled H_2O and the second extract was pooled with the first extract.

Ethanol precipitation

DNA solution or the aqueous phase collected from phenol extraction was mixed with $^1/_{10}$ volume of 3 M sodium acetate, pH 5.2 and mixed by vortexing briefly. After the addition of 2 volume of ice-cold absolute ethanol, the sample was left at -20°C for an hour. The pellet was recovered by centrifugation at 13,000 rpm for 20 min and washed by 1 ml of 70% ethanol. The pellet was air-dried and resuspended in 30-50 μl sterile distilled water.

2.3.1.5 Quantification and Qualification of Nucleic Acids

Concentrations and purity of the double stranded nucleic acids (plasmid DNAs, restriction endonuclease fragments and constructs) and oligonucleotides were determined by using the Beckman Instruments Du Series 600 Spectrophotometer software programs (ds DNA and Oligo DNA Short methods) on the Beckman Spectrophotometer Du 640 (Beckman Instruments Inc. CA. U.S.A).

When the new plasmid was available, 500ng of the plasmid DNA was transformed into *E.coli* DH5α strain to obtain permanent stock. First, transformants

were selected according to their characteristic antibiotic resistance (kanamycin resistance for pEGFP-N2 and ampicillin resistance for pLXSN-BRCA1). In order to confirm the presence of the transformed plasmid in the positive colonies, mini-prep plasmid DNA isolation was performed . After the digestion with proper restriction endonuclease(s), their restriction endonuclease maps were compared to known profile of the original plasmid stock.

2.3.1.6 Restriction enzyme digestion of DNA

Restriction enzyme digestions were routinely performed in 10-70 µl reaction volumes and typically 2-10 µg DNA were used. Reactions were carried out with the appropriate reaction buffer and conditions according to manufacturer's recommendations.

Digestion of DNA with two different restriction was performed in the same reaction buffer to provide the optimal condition for both restriction enzymes.

If no single reaction buffer could be found to satisfy the buffer requirements of both enzymes, the reactions were achieved sequentially. First, DNA was digested with one of the enzymes completely and then the digested DNA was recovered by ethanol precipitation (section 2.3.1.4) followed by digestion with the second enzyme.

2.3.1.7 Agarose gel electrophoresis of DNA

DNA fragments were fractionated by horizontal electrophoresis by using standard buffers and solutions. DNA fragments less than 1 kb were generally separated on 1.0 % agarose gel, those greater than 1 kb (up to 11 kb) were separated on 0.8 % agarose gels.

Agarose gels were completely dissolved in 1x TAE electrophoresis buffer to required percentage in microwave and ethidium bromide was added to final concentration of 30 μg/ml. The DNA samples were mixed with one volume loading buffer and loaded onto gels. The gel was run in 1x TAE at different voltage and time depending on the size of the fragments at room temperature.

Nucleic acids were visualized under ultraviolet light (long wave, 340 nm) and Standard DNA size marker, 1 kb DNA ladder, was used to estimate the fragment sizes. Fragment sizes of the 1 kb DNA ladder were as follows:
12.2, 11.2, 10.2, 9.2, 8.1, 7.1, 6.1, 5.1, 4.1, 3.1, 2.0, 1.6, 1.0, 0.5, 0.4, 0.3 and 0,2 kb

Extraction of DNA fragments from agarose gel

DNA fragments were extracted from agarose gels by using the QIAEX II (150) gel extraction kit according to the manufacturer's instructions.

Gel purification with the QIAEX II kit yields 60-70 % recovery of DNA fragments between 1.0 kb to 6.0 kb in 10-20 μl volume.

2.3.2 Computer analysis of DNA sequences

Restriction endonuclease maps of the plasmid DNAs and the *BRCA1* cDNA were analyzed by using The WebCutter program (designed by Max Heiman, 1995, maxwell@minerva.cis.yale.edu) available for free and public use at http://www.medkem.gu.se/cutter and http://firstmarket.com/firstmarket/cutter.

Oligonucleotide for sequencing was designed by using the primer programs "Primer Designer- Version 2.0 (Scientific and educational software, 1990-91)" and

"Amplify for analyzing PCR experiment (Bill Engels, 1992, University of Wisconsin, Genetics, Madison,U.S.A,WREngels@mace.wisc.edu)".

The annealing temperature for a sequencing primer was calculated using the Tm determination program provided by The Alces WWW Server, Virtual Genome Center (VCG) at http://alces.med.umn.edu/rawtm.html (stew@lenti.med.umn.edu).

Protein sorting signal analysis of the BRCA1 protein was done by using the *PSORT*, Server for Analyzing and Predicting Protein Sorting Signals Coded in Amino Acid Sequence, version 6.3 (WWW) program (*http://psort.nibb.ac.jp)* Kenta Nakai, Osaka University (nakai@nibb.ac.jp).

2.3.3. Construction of the pEGFP-N2-BRCA1(83 - 2436) eukaryotic expression vector

2.3.3.1 Eukaryotic Green Fluorescence Protein-N-Terminal protein fusion vector, pEGFP-N2

pEGFP-N2 (ClonTech) vector encodes a variant of the *Aequorea victotia* green fluorescent protein (GFP) that has been optimized for brighter fluorescence and high expression in mammalian cells. pEGFP-N2 allows genes cloned into the multiple cloning site (MCS) upstream of the EGFP coding sequences to be expressed as fusions to the N-terminus of EGFP. The unmodified vector will express EGFP in mammalian cells. The Genebank sequence of the pEGFP-N2 vector was given in Appendix 2.

pEGFP-N2 vector was used for cloning of the 2353 base pairs (bp) long fragment of the *BRCA1* gene. The vector DNA was double digested with the *Eco*47

III and *Kpn* I restriction enzymes. Map ad the multiple cloning site of the vector was given in figure 2.

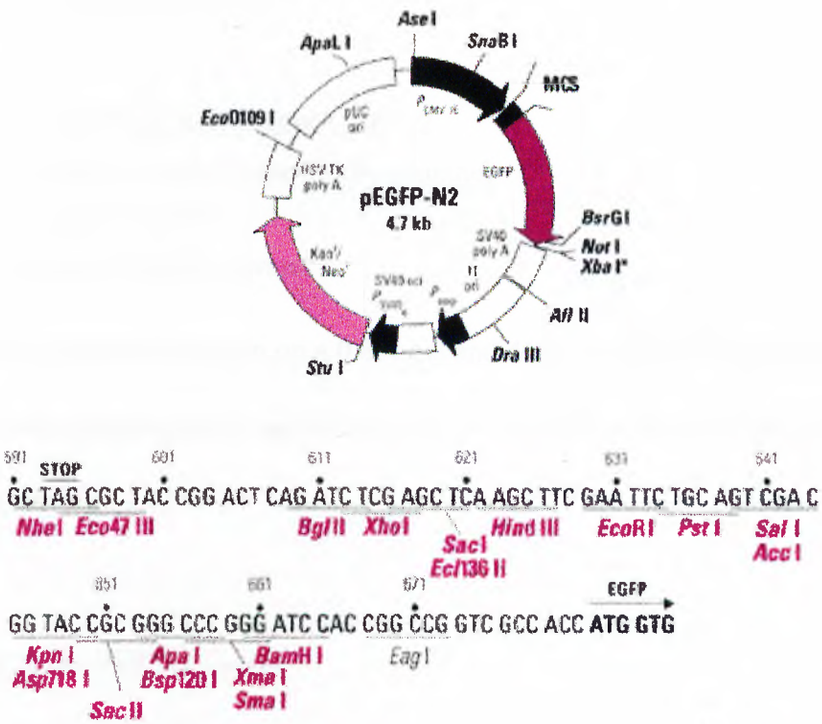


Figure 2. Restriction map and multiple cloning site of pEGFP-N2. (Unique restriction sites are in color or bold.) The *Not* I site follows the EGFP stop codon. The *Nhe* I site cannot be used for fusions since it contains an in-frame stop codon. The *Xba* I site (*) is methylated in the DNA provided by CLONTECH. If you wish to digest the vector with this enzyme, you will need to transform the vector into a *dam*$^-$ host and make fresh DNA.

Double digestion of pEGFP-N2 vector was performed in two sequential reactions. The following reagents were added into an eppendorf tube in order for the first digestion reaction:

5.52 µl pEGFP-N2 (10 µg)
2 µl 10x Green Buffer (MBI)
2 µl 10 mg/ml BSA (0.1 µg/µl final concentration)
1 µl *Eco*47 III (10U/µl)
0.48 µl of sterile distilled water

The reaction was incubated at 37°C for 4 h. The digestion mixture was run on a 0.8% gel and the 6188 bp *Eco*47 III fragment was isolated from agarose gel and used for the second digestion reaction:

20 µl 4737 bp linearized pEGFP-N2
3 µl 10x Optimal buffer #1 (Stratagene)
1 µl *Kpn* I (25 U)
5 µl sterile distilled water

The digestion mixture was run on a 0.8% gel and 4684 bp *Kpn* I fragment was isolated from agarose gel and recovered in 10 µl dH_2O by using the QIAEX gel extraction kit.

2.3.3.2 pLXSN-BRCA1 vector

The 2353 bp fragment of *BRCA1*, namely $BRCA1_{(83-2436)}$, that includes the 83 - 2436 nucleotides of *BRCA1* and the ATG start codon at position 120 nucleotide was prepared by digesting the retroviral expression vector, pLXSN-BRCA1 (Holt *et al.*, 1996) containing the full length of BRCA1 with *Ehe* I and *Kpn* I restriction endonucleases. The double digestion profile of the pLXSN-BRCA1 was visualized on 0.8% agarose and the 2389 bp fragment ($BRCA1_{83-2436}$) was purified from agarose gel by using the QIAEX gel extraction kit.

Double digestion of pLXSN-BRCA1 was performed in single reaction mixture within the common buffer Y (MBI). The following reagents were added into an eppendorf tube in order for *Ehe* I and *Kpn* I digestion reaction of pLXSN-BRCA1;

5 µl pLXSN-BRCA1 (10 µg)
2 µl 10x Yellow Buffer (MBI)
0.2 µl 10 mg/ml BSA (0.1 mg/ml final concentration)
2 µl *Ehe* I (10 U)
1 µl *Kpn* I (25 U)
9.5 µl sterile distilled water

The mixture was incubated at 37°C for 4 h

The digested DNA was run on a 0.8% agarose gel and the 2389 bp pLXSN-BRCA1 fragment (BRCA1 $_{(83-2436)}$) was isolated from agarose gel and eluted in 10 µl sterile distilled water.

2.3.3.3 Construction of vector encoding a 2316 bp fragment of BRCA1 fused with the N-terminal of the EGFP

The molar ratio of 1:3 (vector:insert) was used to clone the 2389 bp BRCA1$_{(83-2436)}$ into 4684 bp pEGFP-N2. The reaction conditions were as follows:

2 µl pEGFP-N2 4684 bp fragment
5 µl BRCA1 2389 bp fragment
1 µl 10x Ligation buffer
2 µl T4 DNA ligase (2 Weiss units/ml)

The digested vector itself was used as a control for the ligation reaction. The ligation reaction for the control was as follows:

2 µl pEGFP-N2 (vector, 270 ng)
5 µl sterile distilled water
1 µl 10x Ligation buffer
2 µl T4 DNA ligase (2 Weiss units/ml)

Both reactions were incubated at room temperature for 4h and transformed into *E.coli* DH5α strain as described in the section 2.3.1.2. The recombinant colonies were picked and analyzed. Large scale plasmid DNA preparation was performed to of the positive recombinants (named Nls) as described in the section 2.3.1.3.

2.3.3.4 Klenow treatment of Nls

Isolated Nls DNAs were digested with *Kpn* I and ethanol precipitated. The precipitated DNA was treated with Klenow fragment to remove the 4 bases of 3'-protruding ends which were created by *Kpn* I digestion. The following reagents were added in order:

15.5 µl *Kpn* I digested Nls (12 µg)
1 µl 10 mM dTTP
1 µl 10 mM dATP
1 µl 10 mM dCTP
1 µl 10 mM dGTP
2.3 µl 10x Klenow incubation buffer
0.5 µl Klenow fragment (10 U/ml)
0.7 µl sterile distilled water

After 30 min of incubation at 37°C, Klenow enzyme was heat inactivated at 75°C for 10 min. The DNA sample was run on 0.8% agarose gel and extracted from the gel by using the QIAEX gel extraction kit.

Klenow treated and purified DNA was self ligated to create the pEGFP-$BRCA1_{(83-2436)}$ construct. The following reagents were added to the ligation reaction:

16 µl Klenow treated Nls DNA
2 µl 10x Ligation buffer
2 µl T4 DNA ligase (2 Weiss units/ml)

Reaction was incubated at room temperature for 4 h and transformed into DH5α as described in the section 2.3.1.2. The positive transformants that carry the pEGFP-BRCA1$_{(83-2436)}$ construct were subjected to large scale plasmid DNA isolation as described in the section 2.3.1.3.

2.3.4 Automated DNA sequencing

Sequencing of the 3'-ligation junction of the pEGFP-BRCA1$_{(83-2436)}$ construct was performed at Bilkent University, Department of Molecular Biology and Genetics, (Ankara, Turkey).

Midi-prep DNA of pEGFP-BRCA1$_{(83-2436)}$ was linearized with *Not* I digestion and sequenced by using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin Elmer, U.S.A) on the ABI PRISM 377 Automated DNA Sequencer (Perkin Elmer, U.S.A). Following reverse sequencing primer was used:

TC-102 (reverse): 5'-TCG ACC AGG ATG GGG CA-3'

2.3.5 Tissue culture techniques

2.3.5.1 Cell line

MCF-7 tissue was used as a model cell line in the eukaryotic expression studies. The characteristics of the MCF-7 was obtained from http://www.attc.org web site (Table 1).

Table 1: Characteristics of MCF-7 cell line

ATCC Number:	HTB-22
Name:	MCF7
Tissue:	Mammary gland; breast; adenocarcinoma; carcinoma; pleural effusion; cancer
Species:	Human; 69 year old; female; Caucasian
Receptors:	Estrogen
HeLaMarkers:	no
Oncogene:	wnt7h +
Karyotype:	The stemline chromosome numbers ranged from hypertriploidy to hypotetraploidy, with the 2S component occurring at 1%. There were 29 to 34 marker chromosomes per S metaphase, of which 24 to 28 markers occurred in at least 30% of cells, and generally one large submetacentric (M1) and 3 large subtelocentric (M2, M3, and M4) markers were recognizable in over 80% of metaphases. No DM were detected. Chromosome No. 20 was nullisomic and X was disomic
Morphology:	Epithelial
AntigenExp:	Blood Type O; Rh+
Products:	Insulin like growth factor binding proteins (IGFBP) BP-2; BP-4; BP-5
Growth:	Monolayer
References:	J. Natl. Cancer Inst. 51:1409-1416, 1973; Cancer Res. 43:2831-2835, 1983; Science 230:943-945, 1985; Cancer Res. 50:2997-3001, 1990; Cancer Res. 53:5193-5198, 1993; Int. J. Cancer 55:453-458, 1993
Medium:	Minimum essential medium Eagle with 2 mM L-glutamine and Earle's BSS adjusted to contain 1.5 g/L sodium bicarbonate, 0.1 mM non-essential amino acids and 1.0 mM sodium pyruvate and supplemented with 0.01 mg/ml bovine insulin, 90%; fetal bovine serum, 10%.
FluidRenewal:	2 to 3 times weekly
SplitRatio:	A ratio of 1:3 to 1:6 is recommended
Comments:	The MCF7 line retains several characteristics of differentiated mammary epithelium including ability to process estradiol via cytoplasmic estrogen receptors and the capability of forming domes; the line may harbor B or C type virus genomes and should be handled as a potentially biohazards agent; contains the Tx-4 oncogene; growth of MCF7 cells is inhibited by tumor necrosis factor alpha (TNF alpha); secretion of IGFBP's can be modulated by treatment with anti-estrogens.

2.3.5.2 Growth conditions

Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FCS, 1 mM glutamine and penicillin and streptomycin (50 mg/ml) was used to culture the MCF-7. The cells were incubated in at 37°C in an incubator with an atmosphere of 5% CO_2 in air.

The cells were passaged before reaching confluence. The growth medium was aspirated and the cells were washed once with calcium and magnesium-free PBS . Trypsin was added to the flask to remove the monolayer cells from the surface. The fresh medium was added and the suspension was pipetted gently to disperse the cells. The cells were transferred to either fresh petri dishes or fresh flasks using different dilutions (from 1:2 to 1:10) depending on requirements.

All media and solutions used for culture were kept at 4°C (except stock solutions) and warmed to 37°C before use.

2.3.5.3 Cryopreservation of cell lines

Exponentially growing cells were harvested by trypsinisation and neutralized with growth medium. The cells were counted and precipitated at 1500 rpm for 5 min. The pellet was suspended in a freezing solution containing 10% DMSO, 20% FCS and 70% DMEM at a concentration of $4x10^6$ cells/ml. 1 ml of this solution was

placed into 1 ml screw cap cryotubes. The tubes were left at -70°C overnight. The next day, the tubes were transferred into the liquid nitrogen storage tank.

When frozen stocks were recovered from liquid nitrogen, the tubes were incubated at 37°C waterbath. When the solution was thawed, the cells were transferred into a 15 ml centrifuge tube and 10 ml fresh DMEM was added gradually. The sample was centrifuged at 1500 rpm for 5 min. The supernatant was aspirated and the precipitated cells were resuspended with 5 ml fresh DMEM and transferred into 25 cm^2 flask.

2.3.5.4 Transfection of eukaryotic cells using electroporation

The MCF-7 cells were plated into a 75 cm^2 flask the day before the electroporation to obtain 60-90% confluence on the day of transfection. 30 μg supercoiled the pEGFP-BRCA1$_{(83-2436)}$ construct and the pEGFP-N2 plasmid were ethanol precipitated and washed with 70% ethanol. The samples were dried in the sterile hood and dissolved in 20 μl sterilized distilled water.

Sterilization of glass coverslips

In a tissue culture hood, coverslips were placed into 95% ethanol. The excess ethanol was removed from the coverslips and flamed to sterilized them. The coverslips were then placed into 12 multi-well tissue-culture dishes.

Electroporation

The MCF-7 cells were harvested by trypsinisation and washed twice with ice-cold 1X calcium-magnesium free PBS. Harvested cells were counted by using a haemocytometer and centrifuged at 1500 rpm for 5 min. The pellet was resuspended in 800 ml ice-cold 1X calcium-magnesium free PBS at a density of $4x10^6$ cells/ml. 30 μg of the supercoiled pEGFP-$BRCA1_{(83-2436)}$ construct or pEGFP-N2 plasmid were added into the cell suspension and mixed well. The DNA-cell mixture was transferred to 0.4 cm electroporation cuvette and incubated on ice for 10 min. The samples were then transferred to the BioRad Gene Pulser (BioRad) and placed into the chamber. The cells were electroporated at 950 mF, 0.22 kV/cm (t=19-22 ms). The cuvettes were then incubated on ice for 10 min and the cells were transferred into a tube containing complete DMEM at a density of $2x10^5$ cells/ml. $1x10^5$ cells were plated in 12 multi-well culture dishes containing coverslips and allowed to incubate at 37°C.

2.3.5.5 Fluorescence signal detection

The coverslips were carefully removed from the culture dishes, placed onto slides and washed gently with 1XPBS solution twice. The slides were immediately examined by fluorescent microscopy. The fluorescence signal was detected with FITC filter set (Filter I: 450-490 nm for GFP and Filter II: BP-365 nm for H33258 staining),(ZEISS). The FITC signal was visualized using a Zeiss MC80/Axioskop fluorescent microscope camera system.

Expression of green or blue fluorescence was examined 24 hours after the transfection within 12 h intervals.

Fixation of cells for staining

After 24 h incubation, culture medium was removed by aspiration and the coverslip was washed twice with PBS. The coverslips were placed onto the slides and 2 ml of freshly made 4% paraformaldehyde solution was directly applied. The coverslips were incubated at room temperature for 30 min at dark. The coverslips were then washed twice with 1XPBS and 500 µl of 1/100 dilution of fluorochrome dye H33258 (1mg/ml) was applied. Samples were incubated at room temperature for 10 min at dark and then washed with 1XPBS twice. The excess PBS around the edges of the coverslip was removed with a clean tissue and then the samples were examined by fluorescent microscopy.

CHAPTER 3. RESULTS

3.1. Computer analysis of the *BRCA1* sequence

3.1.1 Computer analysis of the BRCA1 protein sequence

The human BRCA1 protein sequence (Genbank accession no: HSU 14680) was analyzed by using PSORT computer program to determine the possible protein localization sites within the protein sequences (Appendix 1).

The result of the PSORT analysis of BRCA1 protein is given in Table 2. The PSORT determines the candidate localization-sites for prediction as listed below:

Cytoplasm, mitochondria (outer membrane, intermembrane space, inner membrane and matrix space), microbody (peroxisome), nucleus, endoplasmic reticulum (lumen and membrane), Golgi body, lysosome, plasma membrane and outsite. At the end of the analysis the conclusive prediction, i.e. the top five probable localization sites with their certainty factors (ranging between 0.00 and 1.00) is given finally. The data obtained from the PSORT has shown that the most probable protein targeting sequences were nuclear localization sequences (NLSs). This NLSs were localized to amino acids residues between 502 to 650. So we decided to clone a 2316 bp fragments of *BRCA1* cDNA into 5' end of the EGFP sequence.

Table 2: The results of the PSORT computer analysis for BRCA1 sequence

Targeting Site	Consensus Sequence on BRCA1	Total Certainty
Nucleus	KRKR, amino acid position: 502	0.70
	RKRR, amino acid position: 503	
	KRRP, amino acid position: 504	
	NRLRRKS, amino acid position: 609[‡]	
	KKKK, amino acid position: 650	

[‡]Consensus sequence already defined by Chen *et al*, 1995

3.1.2 Computer analysis of the pLXSN-BRCA1 construct

The restriction endonuclease map of the pLXSN-BRCA1 construct was obtained by using the WebCutter computer program to find two single cutter endonucleases which create a fragment including the NLSs conserved sequences on the *BRCA1* cDNA sequence and also allow to directional cloning of the fragment into the pEGFP-N2 eukaryotic expression vector.

According to the restriction map of the pLXSN-BRCA1 construct (Appendix 3), double digestion with the *Ehe* I and *Kpn* I restriction endonucleases was used for digestion of both pLXSN-BRCA1 construct to obtain the $BRCA1_{(83-2436)}$ fragment containing the five putative nuclear localization signals (NLS) sequences and a highly conserved N-terminal RING finger domain (C3HC4 zinc-finger domain) (Bienstock *et al.* (1996).

3.2 Qualification of the pLXSN-BRCA1 and pEGFP-N2

The undigested- pEGFP-N2 vector isolated from kanamycin resistant transformants was run on 0.8 % agarose gel. After this primary selection, plasmid isolated from candidate colonies were double digested with *Not* I / *Eco*R I endonucleases for further characterization of plasmid and subjected to 0.8 % agarose gel electrophoresis where *Not* I / *Eco*R I digested original pEGFP-N2 plasmid was used as control. The expected 3061 bp and 776 bp fragments were only detected in mini-prep plasmid DNA of the n1 colony (Figure 3, lane 5). *Not* I / *Eco*R I double digestion profile is shown at Figure 3.

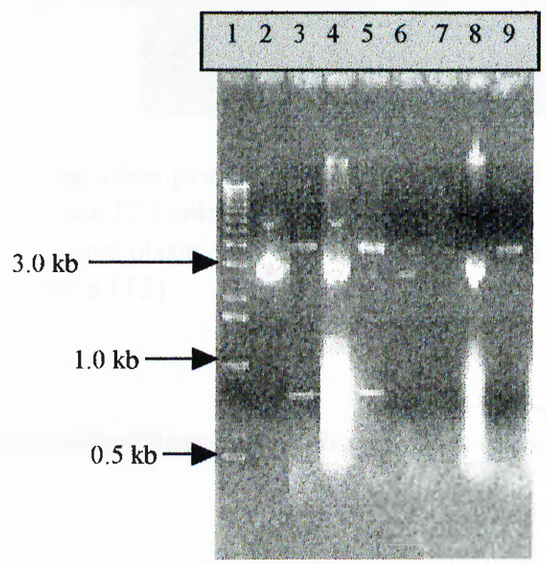


Figure 3. *Not* I / *Eco*RI double digest profile of pEGFP-N2 and mini-prep plasmid DNA of n1, n2, n3 colonies.
(*1:* 1 kb DNA ladder, *2:* undigested pEGFP-N2, *3:* *Not* I / *Eco*RI digested pEGFP-N2, *4:* undigested n1, *5:* *Not* I / *Eco*RI digested n1, *6:* undigested n2, *7:* *Not* I / *Eco*RI digested n2, *8:* undigested n3, *9:* *Not* I / *Eco*RI digested n3)

The same procedure was applied to ampicillin resistant pLXSN-BRCA1 transformants and *Sac* I endonuclease was used to analyze the restriction endonuclease profile of candidate colony, IT3. Expected fragment sizes for *Sac* I digestion profile of the original pLXSN-BRCA1 plasmid were 5707 bp, 3124 bp and 2713 bp. Results are shown at Figure 4.

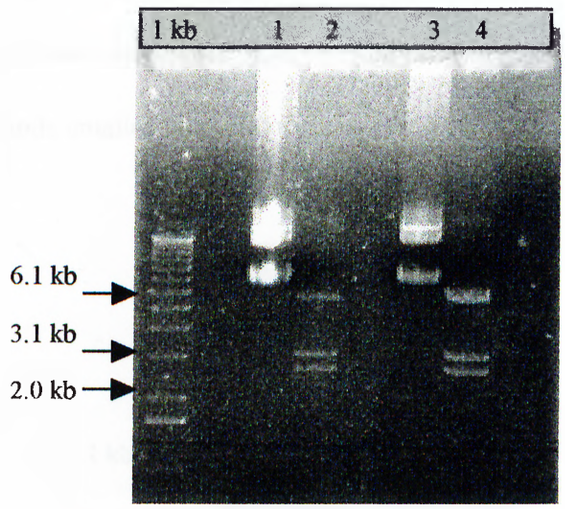


Figure 4. *Sac* I digestion profile of the original PLXSN-BRCA1 and mini-prep plasmid DNA of the IT3 colony on 0.8% agarose gel.
(*1:* undigested original plasmid, *2:* *Sac* I digested original plasmid, *3:*undigested IT3 and *4:* *Sac* I digested IT3)

3.3. Endonuclease digestion of the pLXSN-BRCA1 and pEGFP-N2

3.3.1 *Eco*47 III and *Kpn* I double digestion of pEGFP-N2 vector

Two step double digestion was performed. The pEGFP-N2, was first digested with *Eco*47 III restriction endonuclese. The expected fragment size after *Eco*47 III

digestion was; 4737 bp (linearized vector). The 4737 bp fragment was isolated from gel in 20 µl sterile distilled water.

The isolated 4737 bp fragment was used for the second, *Kpn* I, digestion. Expected fragments were 4684 bp and 53 bp respectively. The 4684 bp fragment was isolated from the gel and resuspended in sterile distilled water with 135 ng/µl final concentration (Figure 5). However, the 53 bp fragment could not be observed on 0.8% agrose gel since the percentage and the run time of gel was not suitable to observe the bands smaller than 500 bp.

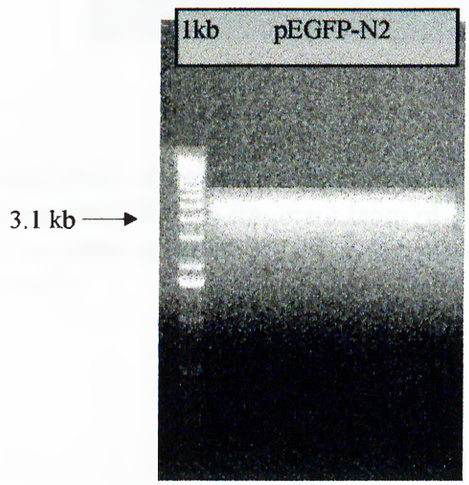


Figure 5. *Eco*47 III and *Ehe* I double digestion profile of pEGFP-N2 vector on 0.8% agarose gel.

3.3.2 *Kpn* I and *Ehe* I double digestion of pLXSN-BRCA1

Single reaction mixture was prepared for *Kpn* I and *Ehe* I digestion of the pLXSN-BRCA1. The common Yellow Buffer (MBI) was used. The expected fragment sizes after *Kpn* I and *Ehe* I digestion were; 3837, 3124, 2389, 1175. 985 and

34 bp (Figure 6). The 2389 bp fragment that contains the five NLSs were isolated from the gel in sterile distilled water witha final concentration of 68 ng/µl .

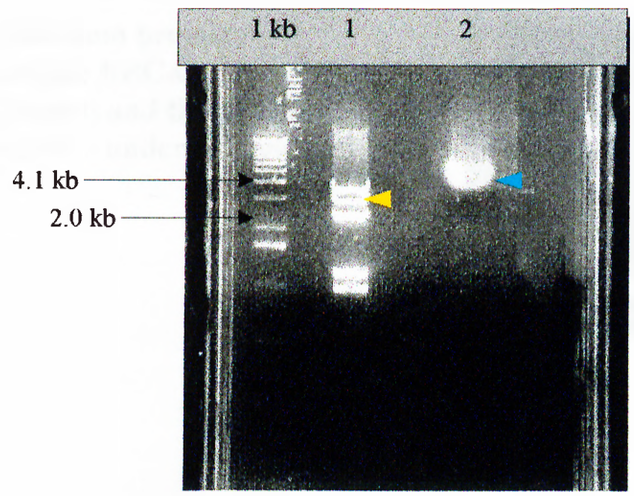


Figure 6. *Kpn* I and *Ehe* I digestion profile of pLXSN-BRCA1.
Yellow arrow indicates the 2389 bp fragment from pLXSN-BRCA1 digestion and cyan arrow shows the 4684 bp *Eco*47 III / *Kpn* I digested pEGFP-N2. (*1:* pLXSN-BRCA1, *2:* pEGFP-N2)

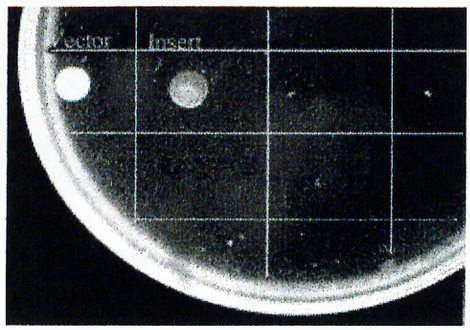
3.4 Construction of the pEGFP-BRCA1$_{(83-2436)}$ eukaryotic expression vector

3.4.1 Ligation reaction

The BRCA1$_{(83-2436)}$ fragment was cloned into pEGFP-N2 vector as described in section 2.3.3.3 and "no insert" control ligation was performed to test the self ligation of cut vector. The vector:insert molar ratio was determined according to the data obtained from both the quantitative analysis of insert (BRCA1$_{(83-2436)}$) and

vector (pEGFP-N2) as described in section 2.3.1.5 and ethidium bromide intensities dotted on agarose plate that was illuminated by ultraviolet light (Figure 7).

Figure 7. Ethidium bromide intensities of the BRCA1$_{(83 - 2436)}$ fragment (insert) and the pEGFP-N2 (vector) dotts under ultraviolet illuminator.



3.4.2 Selection of positive colonies after transformation

After transformation of recombinant pEGFP-N2 into *E.*coli DH5α strain, transformants were subjected to kanamycin selection overnight. The following day, 17 colonies from ligation plate and 20 colonies from "no insert" ligation control plate were obtained. DNAs were isolated from all colonies in the ligation plate and one colony from the control plate. All samples were digested with *Eco*R I restriction endonuclease (Figure 8). The mini-prep plasmid DNA of candidate Nls3, Nls4 and control ligation colony2 colonies were further characterized by *Eco*R I / *Not* I double digestion analysis. Result is shown at Figure 9. The Nls4 colony was selected for the following steps of the cloning study and glycerol stocks was prepared.

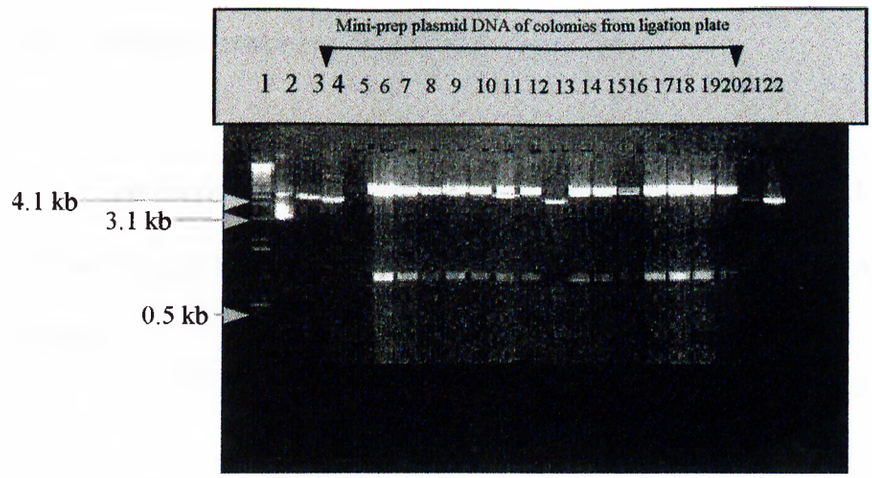


Figure 8. *Eco*R I digestion profile of the positive colonies (Lane number 4 to 20). *1:* 1 kb DNA ladder, *2:* undigested pEGFP-N2 vector, *3:* *Eco*R I digested pEGFP-N2 *Lane 4 - 20:* *Eco*R I digested Nls DNA samples (samples 1 to 17) *21:* *Eco*R I digested control ligation colony1, *22:* *Eco*R I digested control ligation colony2.

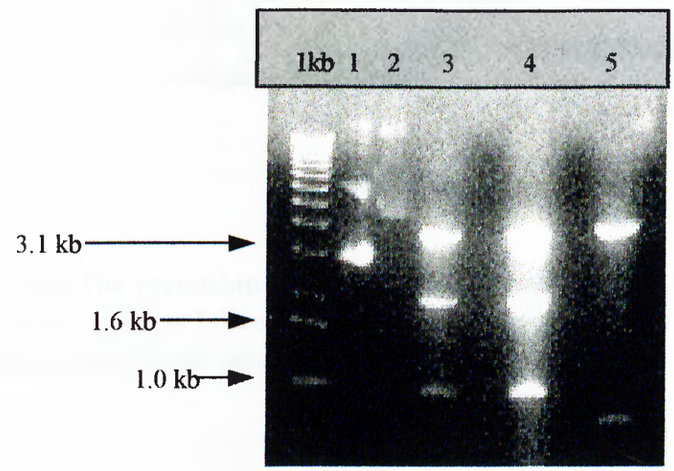


Figure 9. *Not* I / *Eco*RI double digest profile of mini-prep plasmid DNA isolated from the Nls3, Nls4 and the control ligation colony2.
(*1:* undigested pEGFP-N2, *2:* linearized pEGFP-N2 with *Not* I digest, *3:* *Not* I / *Eco*RI digested Nls3, *4:* *Not* I / *Eco*RI digested Nls4, *5:* *Not* I / *Eco*RI digested control ligation colony2.)

3.4.3 Klenow treatment of the Nls4 construct

The recombinant pEGFP-N2 construct containing the BRCA1$_{(83 - 24 36)}$, named as Nls4, (Figure 10) was isolated from the Nls4 colonies by using QIAGEN Plasmid Midi kit.

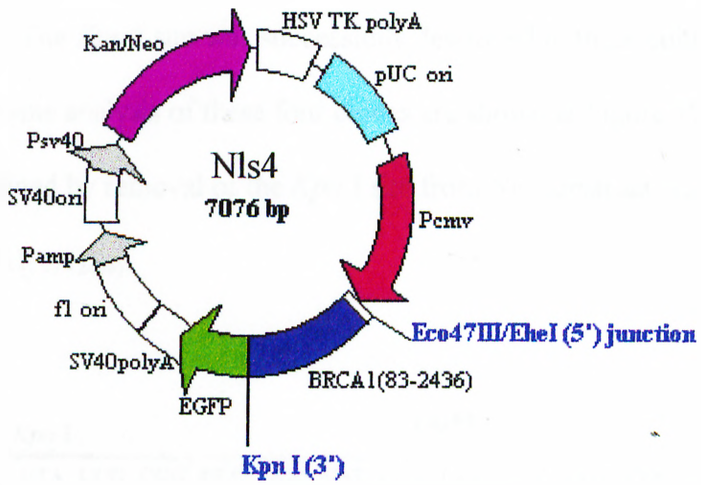


Figure 10. The recombinant pEGFP-N2 construct, Nls4 (7076 bp).
The *Eco*47 III / *Kpn* I digested BRCA1$_{(83 - 2436)}$ fragment was ligated into *Ehe* I / *Kpn* I digested pEGFP-N2 vector.

The Nls4 construct was subjected to the Klenow treatment. After ligation of BRCA1$_{(83 - 2436)}$ fragment into pEGFP-N2 vector, the *Kpn* I junction between 3'-terminus of BRCA1$_{(83 - 2436)}$ and 5'- terminus of EGFP had a frame shift in the transcription of BRCA1$_{(83 - 2436)}$-EGFP fusion protein. Therefore, the Nls4 construct was digested with *Kpn* I and treated with Klenow enzyme to remove 4 bases from the 5'-protruding end of the BRCA1 coding sequence.

Following the Klenow treatment, the blunt ends of the Nls4 construct was ligated and transformed into DH5α competant cells. Schematic representation of the Klenow treatment and self-ligation steps of the construct is given at Figure 11.

The transformants were grown on kanamycin plate and four positive colonies were obtained (Nls1^{Kpn-}, Nls2^{Kpn-}, Nls3^{Kpn-} and Nls4^{Kpn-}). Mini-prep DNAs were prepared and digested with *Kpn* I enzyme to check the presence or absence of *Kpn* I sites in the constructs. Two colonies (Nls3^{Kpn-} and Nls4^{Kpn-}) can not be digested with *Kpn* I enzyme. The *Kpn* I site was successfully destroyed in these colonies. The *Kpn* I restriction enzyme analysis of these four clones are shown at Figure 12. The new construct obtained by removal of the *Kpn* I site from Nls construct was named as pEGFP-BRCA1$_{(83-2436)}$.

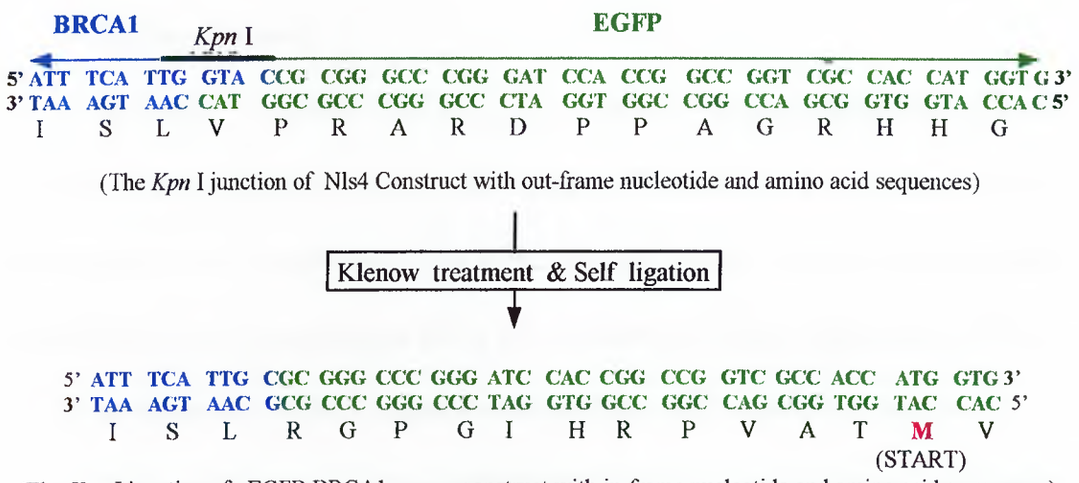


Figure 11. Schematic representation of the Klenow treatment and self ligation steps in the construction of pEGFP-BRCA1$_{(83-2436)}$.
Blue letters indicate the sequence of the BRCA1$_{(83-2436)}$ fragment and green letters represent the sequence of EGFP. Capital letters in black are used to show the amino acid encoded sequences and red 'M' denotes the start codon of the EGFP.

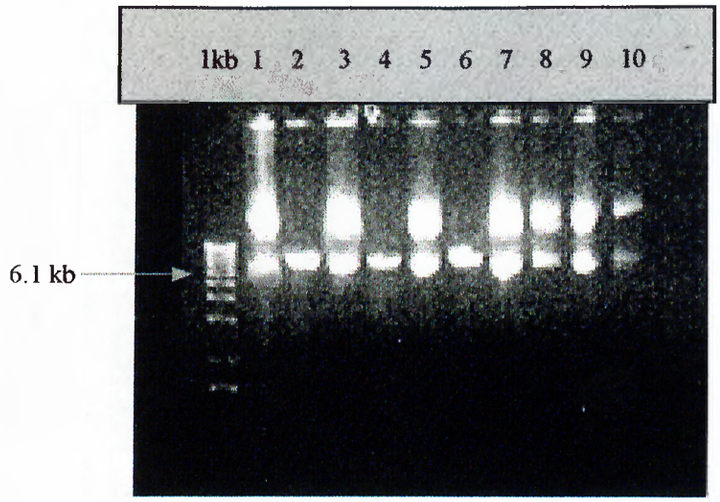


Figure 12. *Kpn*I digestion profile the Klenow treated Nls^{Kpn-} colonies on 8% agarose gel.
(*1:* undigested Nls4 without Klenow treatment, *2: Kpn* I digested Nls4 without Klenow treatment, *3:* undigested Nls1^{Kpn-}, *4: Kpn* I digested Nls1^{Kpn-}, *5:* undigested Nls2^{Kpn-}, *6: Kpn* I digested Nls2^{Kpn-}, *7:* undigested Nls3^{Kpn-}, *8: Kpn* I digested Nls3^{Kpn-}, *9:* undigested Nls4^{Kpn-}, *10: Kpn* I digested Nls4^{Kpn-}).

3.4.4 Automated sequencing of *Kpn* I junction of the BRCA1$_{(83-2436)}$-EGFP fusion sequence

The Nls4^{Kpn-}, pEGFP-BRCA1$_{(83-2436)}$ construct was partially sequenced with TC102 sequencing primer by using Perkin Elmer 377 Automated Sequencer. TC102 primers lies on the 3'-upstream of the *Kpn* I junction, within the vector sequence.The pEGFP-BRCA1$_{(83-2436)}$ contruct DNA was prepared by Qiagen midi-prep kit. The DNA was linearized by *Not* I digestion and used for cycle sequencing reaction.

The partial sequence of the pEGFP-BRCA1$_{(83-2436)}$ construct was analyzed by Sequence Navigator computer program and shown that Klenow treatment had removed the *Kpn* I site successfully (Figure 13) . The actual sequencing result is given in Appendix 4.

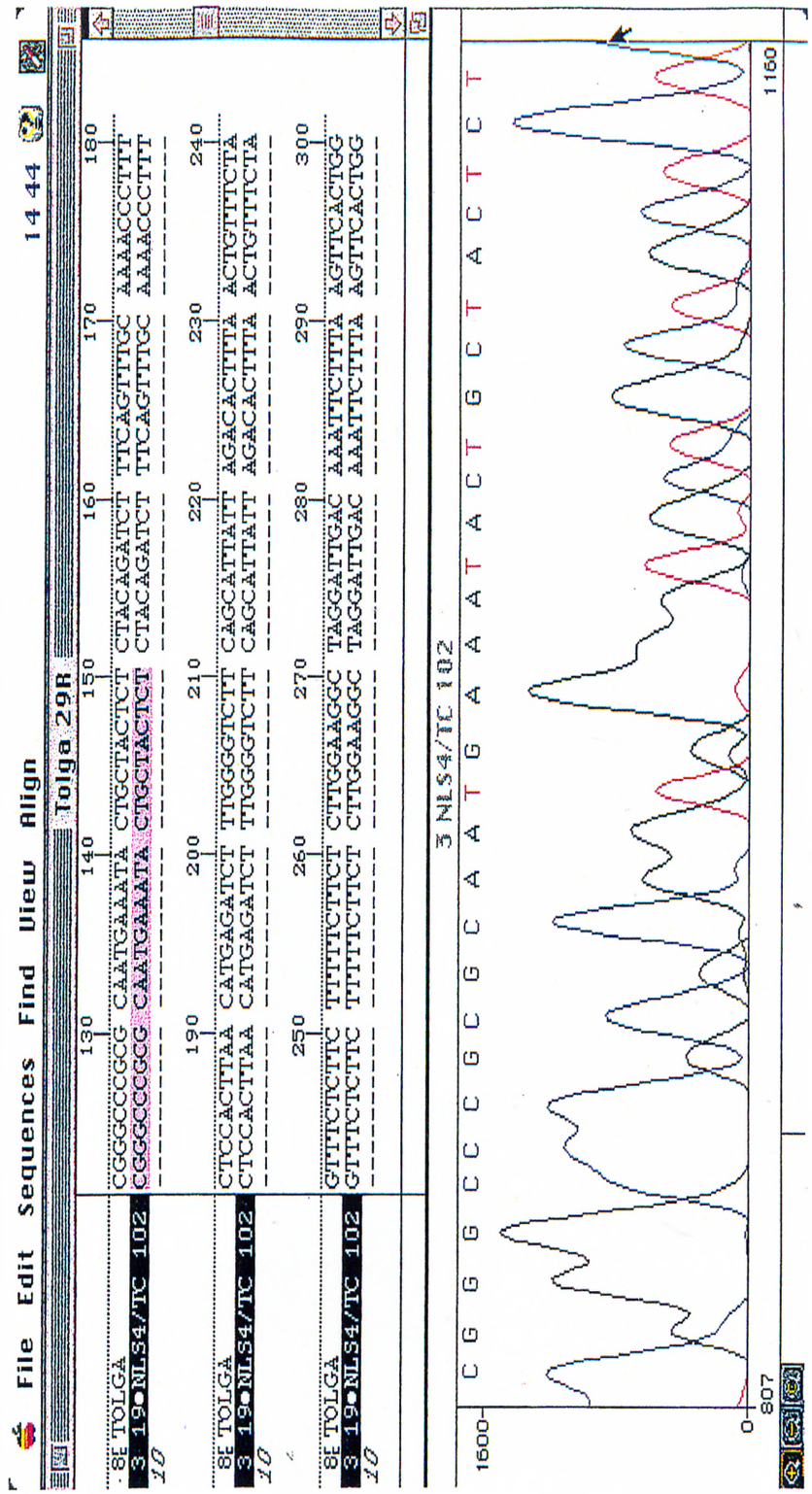


Figure 13. The sequencing result of *Kpn* I junction.

The given electropherogram shows the cyan highlighter box including the removed *Kpn* I junction at nucleotide 129. (*8E TOLGA*: the reference sequence, *3 19•NLS4/TC 102*: the sequence of the Nls4^{Kpn-} plasmid DNA).

In summary, this cloning strategy which was confirmed by DNa sequencing resulted in the in-frame fusion of the carboxy terminus of 772 amino acid fragment of BRCA1 protein with at the carboxy terminus of EGFP. This vector, namely pEGFP-BRCA1$_{(83-2436)}$, contains a Kozak consensus translation initiation site and immediate early promoter of human cytomegalovirus (CMV) to increase the translation efficiency in eukaryotic cells.

3.5 Expression Analysis of the pEGFP-N2 vector and the pEGFP-BRCA1$_{(83-2436)}$ Construct in Eukaryotic System

The MCF-7 human breast adenocarcinoma cell line was used for all transfection experiments (Figure 14) . MCF-7 breast cancer cell line has very low expression of *BRCA1* mRNA and BRCA1 protein (Jensen *at al*, 1996).

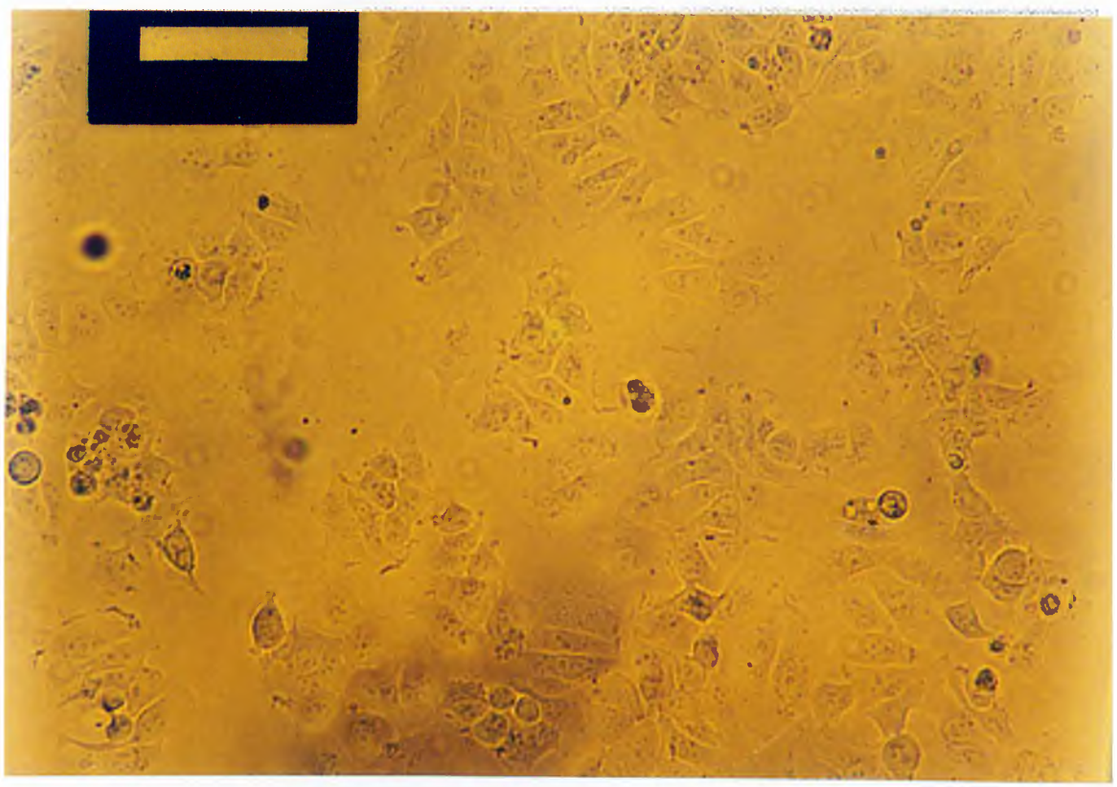


Figure 14. Phase-contrast microscopy apperance of MCF-7 cells (200x magnification)

3.5.1 Transfection of MCF-7 by Electroporation

The pEGFP-BRCA1$_{(83-2436)}$ construct and pEGFP-N2 were transfected into MCF-7 by electroporation method (Section 2.3.5.4) . The efficiency of the transfection was determinated by counting the cells that express EGFP which emits

green fluorescent signal versus the cells having non fluorescent signal under fluorescence microscope. The transformation efficiency for the pEGFP-N2 electroporation was 52.69% and for the pEGFP-BRCA1$_{(83-2436)}$ construct electroporation was 48.87%.

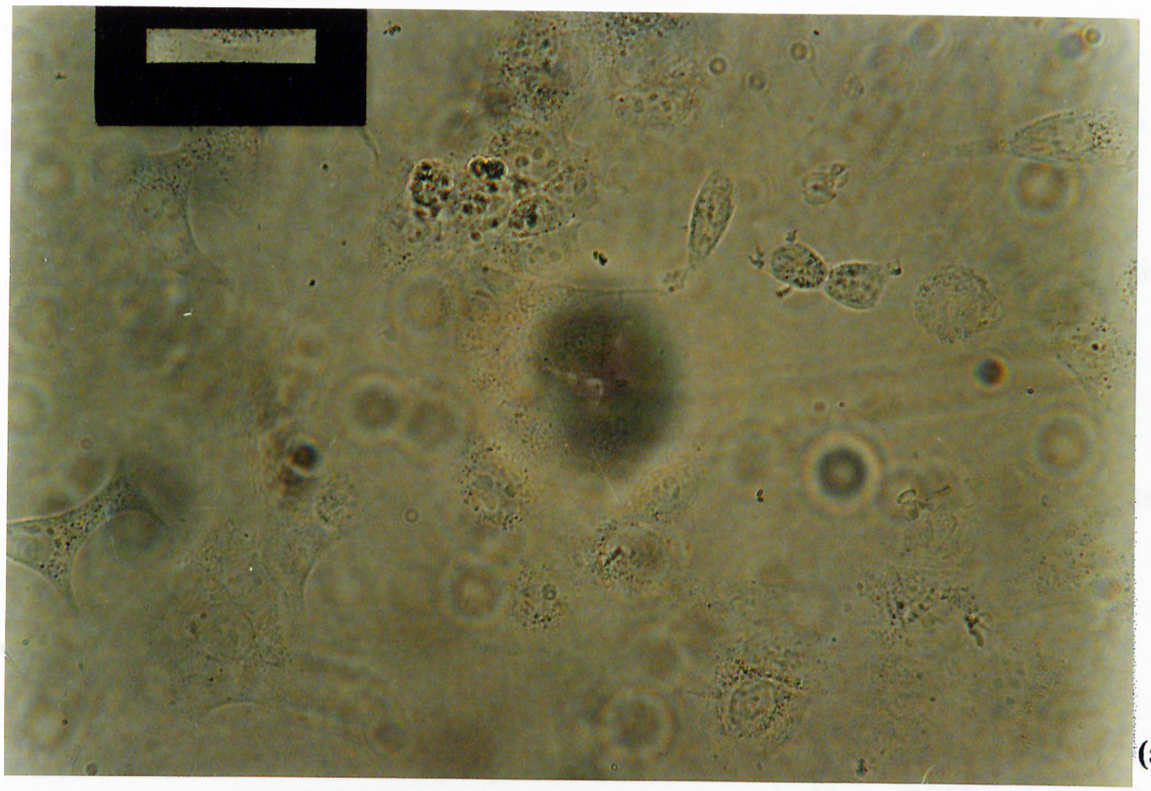
3.5.2. Expression analysis of the EGFP vector in living cells

The MCF-7 cells were transfected with pEGFP-N2 vector both to asses the transfection efficiency of the cells and to set up the maximal experimental conditions to observe the expression of EGFP.

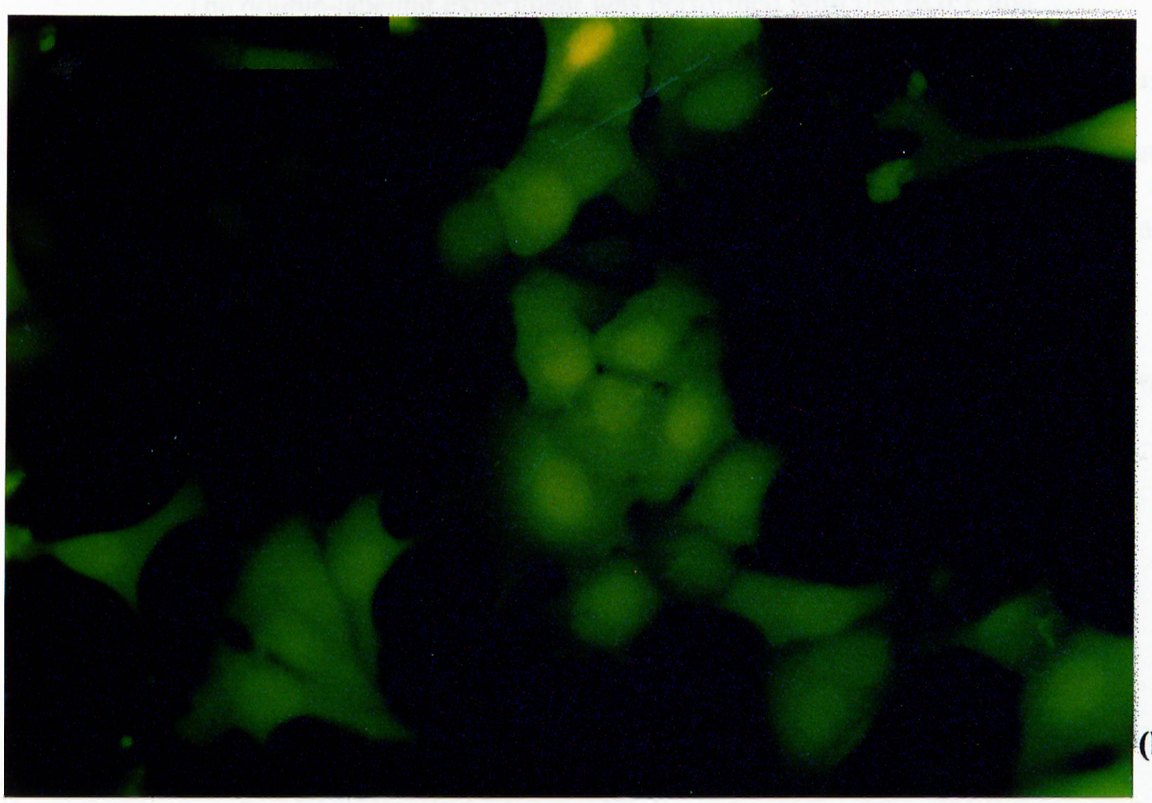
The expression analysis of EGFP was observed 24 h after transfection under the fluorescence microscope with FITC filter set (Filter 1; 450-490 nm). It was shown that the EGFP protein spreads all over the cell (Figure 15). The EGFP protein was very stable and green fluorescence signal could be observed 10 days after the transfection.

Figure 15. Unfixed MCF-7 cells transfected with EGFP-N2 (600x magnification).

(a) Phase-contrast microscopy photograph of the unfixed EGFP-N2 transfected MCF-7. **(b)** Unfixed EGFP-N2 transfected MCF-7 cells were photographed under 450-490 excitation fitted fluorescence microscope.



(a)



(b)

3.5.3 Monitoring the expression and the cellular localization of EGFP-BRCA1$_{(83-2436)}$ fusion protein

Monitoring the cellular localization in living cell

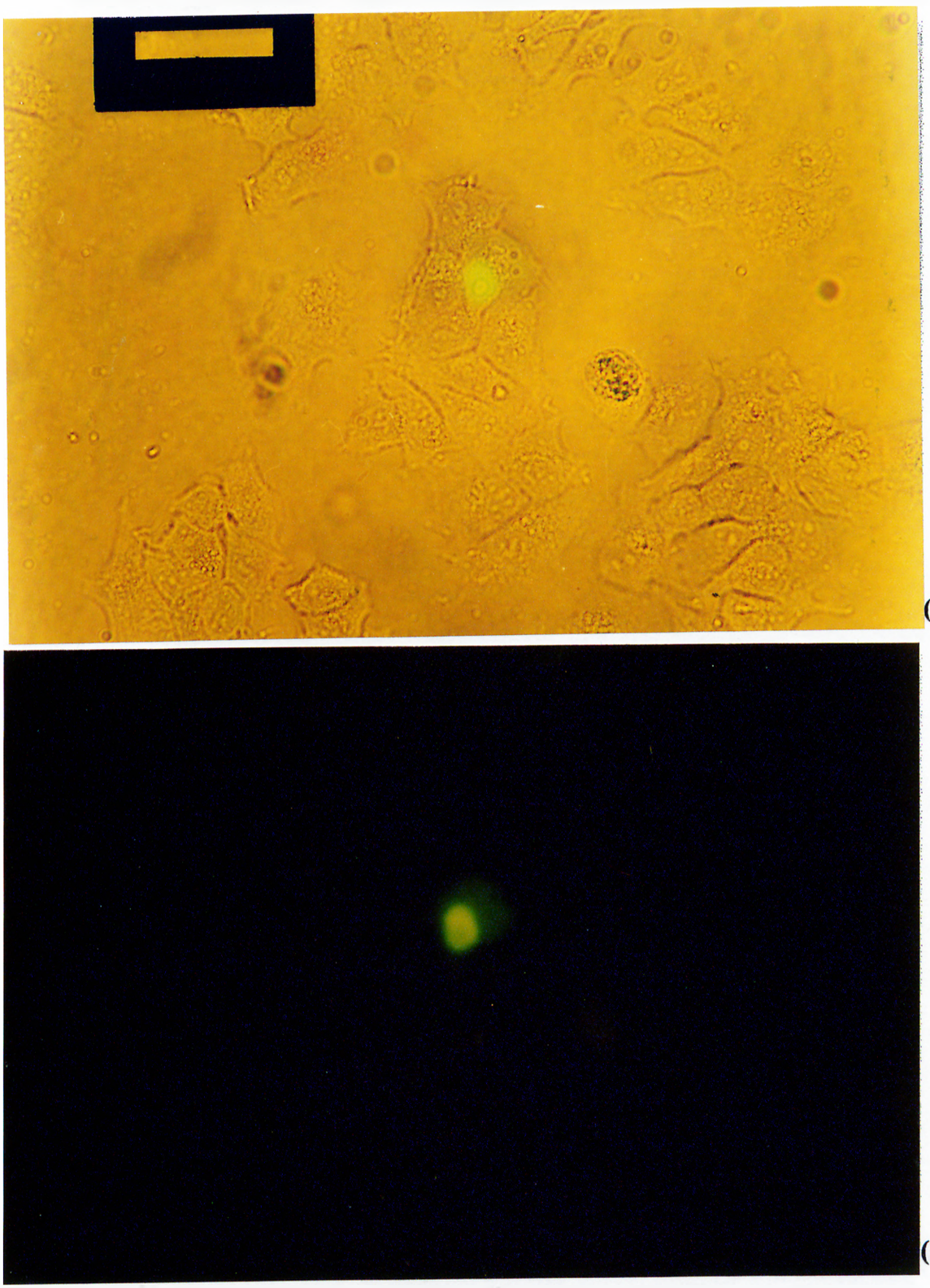
The expression and localization analysis of the EGFP-BRCA1$_{(83-2436)}$ fusion protein in living MCF-7 cells was accomplished with 450-490 excitation fitted fluorescence microscope. The strongest green fluorescent signal was observed within the nuclei of MCF-7 breast cancer cells transfected with EGFP-BRCA1$_{(83-2436)}$ construct. This high fluorescent activity suggested the nuclear localization of the EGFP-BRCA1$_{(83-2436)}$ fusion protein (Figure 16). The expression of the fusion protein started 48 h after the transfection.

Monitoring the cellular localization in fixed cell

The double-labeling experiment was performed with the EGFP and fluorochrome dye H33258 to clarify the nuclear localization of the EGFP-BRCA1$_{(83-2436)}$ fusion protein. The fluorochrome dye H33258 has a blue fluorescent emission at 365 nm and binds to DNA content of the dead cells. The DNA-binding fluorochrome dye H33258 was used to specify the nucleus of the fixed MCF-7 and examined by shifting the FITC fluorescence filter to the BP-365 filter (Filter 2). Results are shown at Figure 17.

Figure 16. Unfixed MCF-7 cells transfected EGFP-BRCA1$_{(83-2436)}$ (400x magnification).

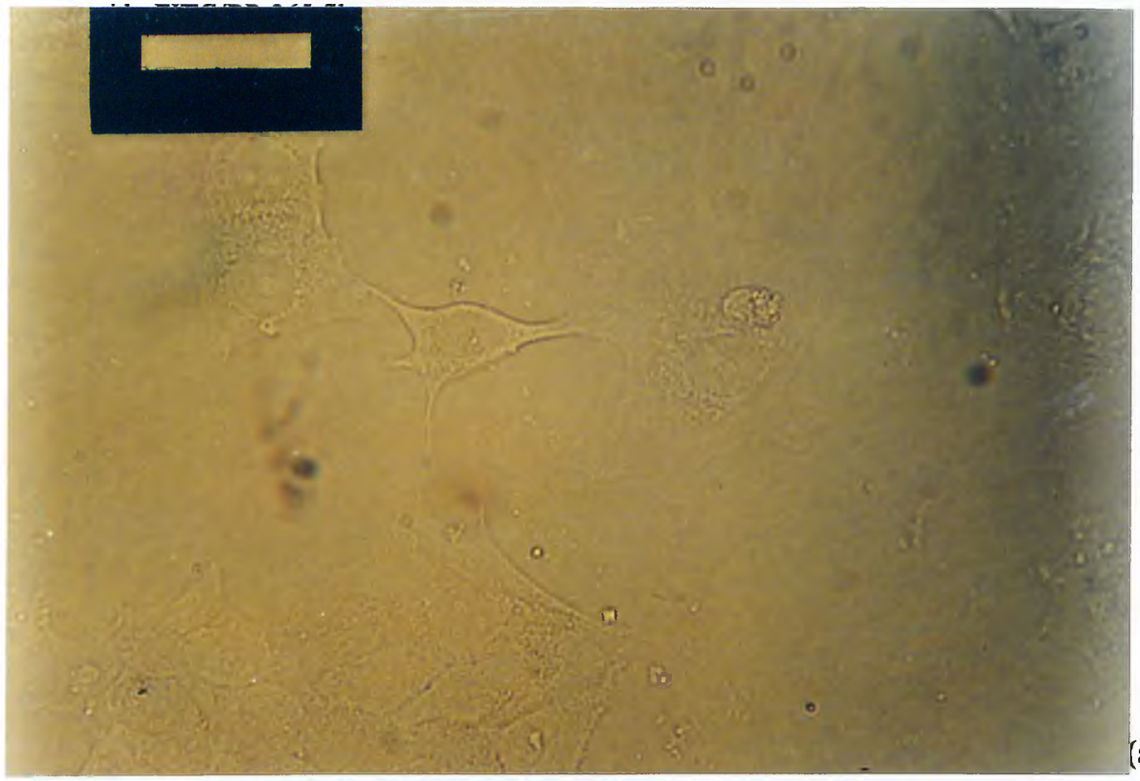
(a) Phase-contrast microscopy photograph of the unfixed MCF-7 cells transfected with EGFP-BRCA1$_{(83-2436)}$. **(b)** Unfixed MCF-7 cells transfected with EGFP-BRCA1$_{(83-2436)}$. were photographed under 450-490 excitation fitted fluorescence microscope.



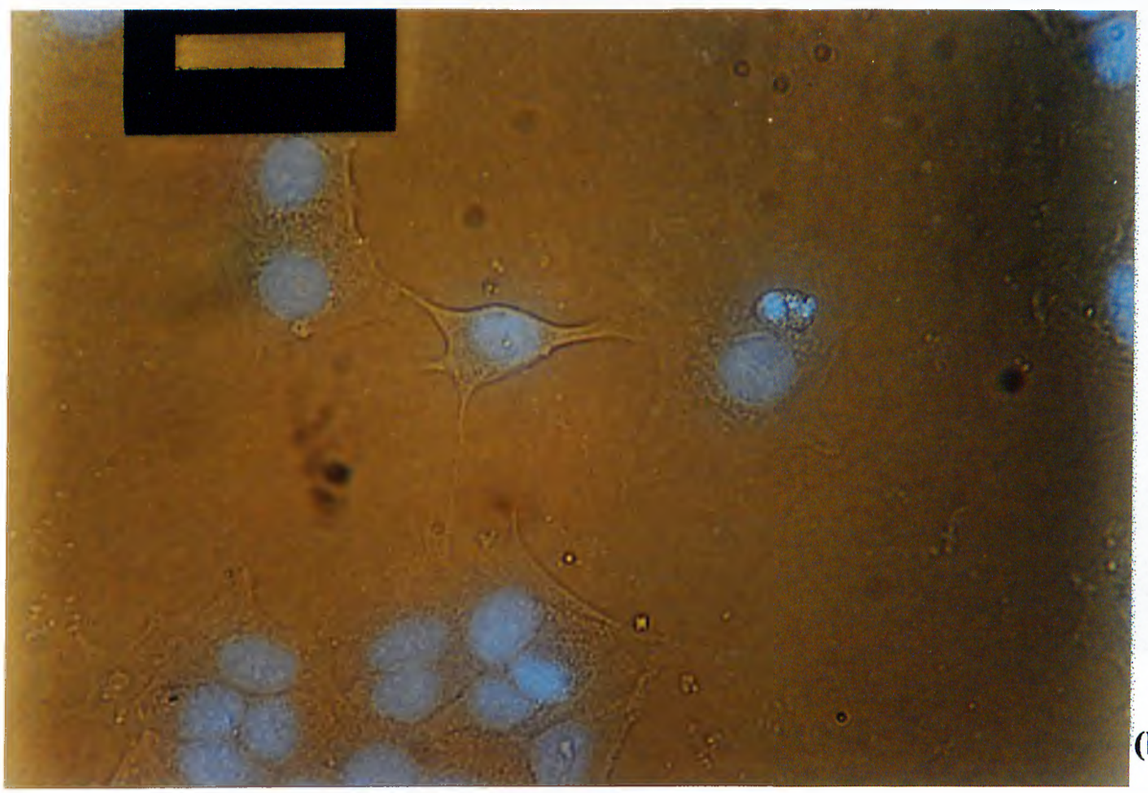
(a)

(b)

Figure 17. Fixed MCF-7 cells transfected EGFP-BRCA1$_{(83-2436)}$ (400x magnification). **(a)** Phase-contrast microscopy photograph of the fixed MCF-7 cells transfected with EGFP-BRCA1$_{(83-2436)}$. **(b)** Apperance of the fixed MCF-7 cells transfected with EGFP-BRCA1$_{(83-2436)}$ under both visible light and 450-490 excitation fitted fluorescence microscope **(c)** Fixed MCF-7 cells transfected with EGFP-BRCA1$_{(83-2436)}$ were photographed under 450-490 excitation fitted fluorescence microscope. **(d)** Fixed MCF-7 cells transfected with EGFP-BRCA1$_{(83-2436)}$ were stained with H33258 and photographed under fluorescence microscope

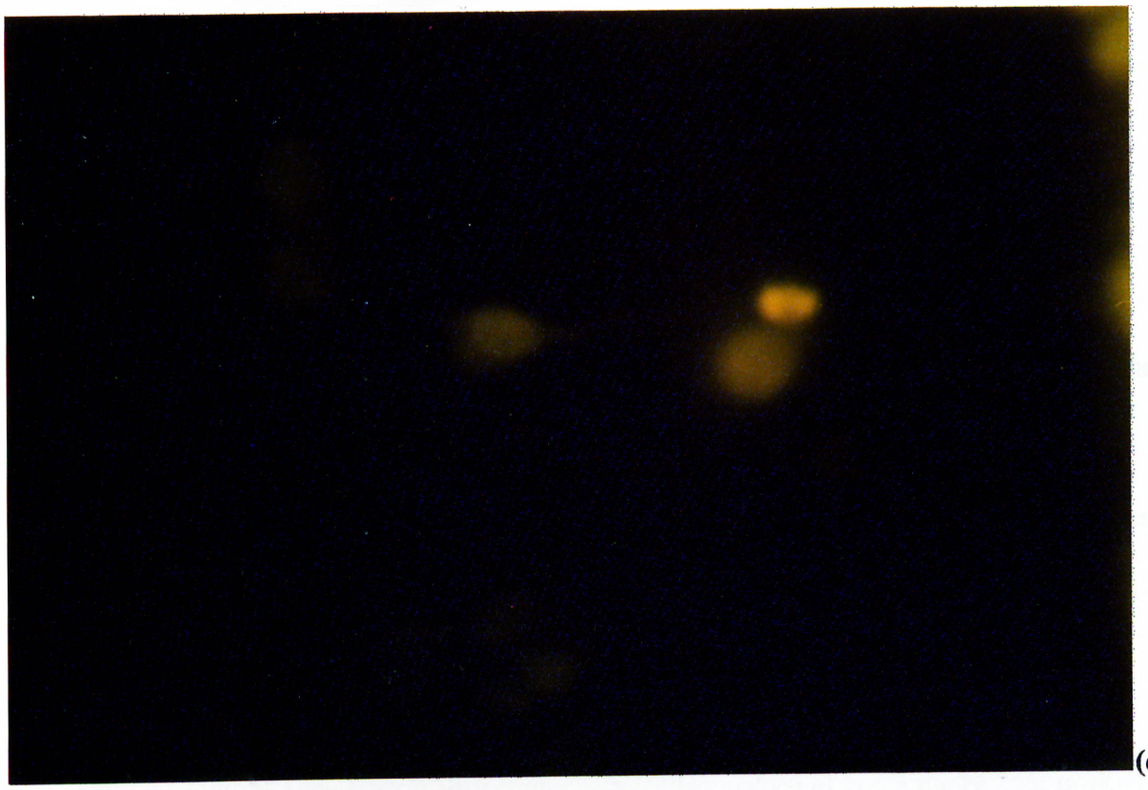


(a)

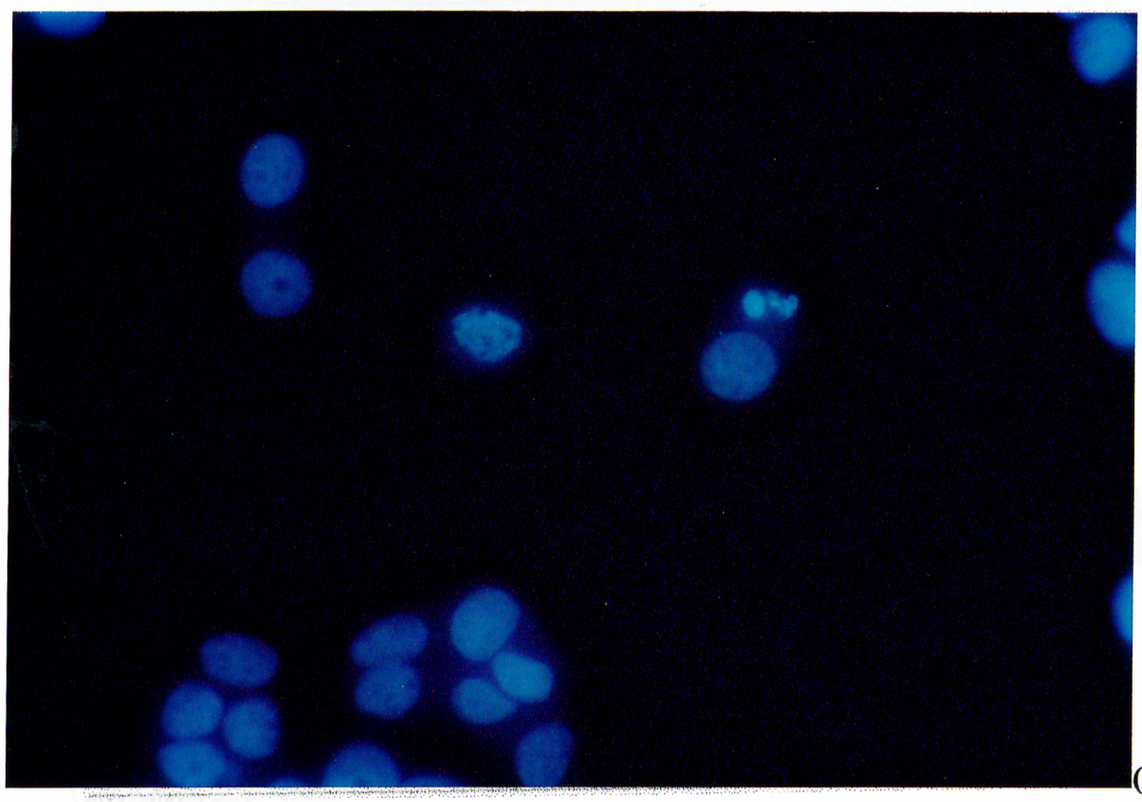


(b)

Figure 17 (cont'd). (c), (d) Fixed MCF-7 cells transfected EGFP-BRCA1$_{(83-2436)}$



(c)



(d)

3.5.4 Effect of EGFP and EGFP-BRCA1$_{(83-2436)}$ fusion protein over expression in MCF-7 cell line

The EGFP and EGFP-BRCA1$_{(83-2436)}$ expressing cells were monitored through 96 h within a 12h intervals and observation was started at 24 h after transfection to determine the percentage of the cell survival. Cell counting was performed as mentioned in section 3.5.1. Over expression of the EGFP in MCF-7 has no significant toxic effect on cells and cell counting data were given in Table 3. However, a clear decrease in percent survival was observed in MCF-7 cells that over express the EGFP-BRCA1$_{(83-2436)}$ (Table 4). The total % of the death within a 96 h period after transfection was 51.19% (Figure 18).

Counting time after transfection	% survival of green fluorescent positive cells over express the EGFP
24 h	52.69
36 h	50.61
48 h	51.23
60 h	50.78
72 h	49.89
84 h	49.81
96 h	49.67

Table 3. Percent (%) survival of MCF-7 cells that over express EGFP.
(Percent (%) survival was calculated as the ratio of green fluorescent positive cells to total cells within a unit area).

Counting time after transfection	% survival of green fluorescent positive cells over express the EGFP-BRCA1$_{(83-2436)}$
24 h	48.97
36 h	40.87
48 h	38.15
60 h	31.72
72 h	30.25
84 h	25.47
96 h	25.08

Table 3. Percent (%) survival of MCF-7 cells that over express the EGFP-BRCA1$_{(83-2436)}$. (Percent (%) survival was calculated as the ratio of green fluorescent positive cells to total cells within a unit area).

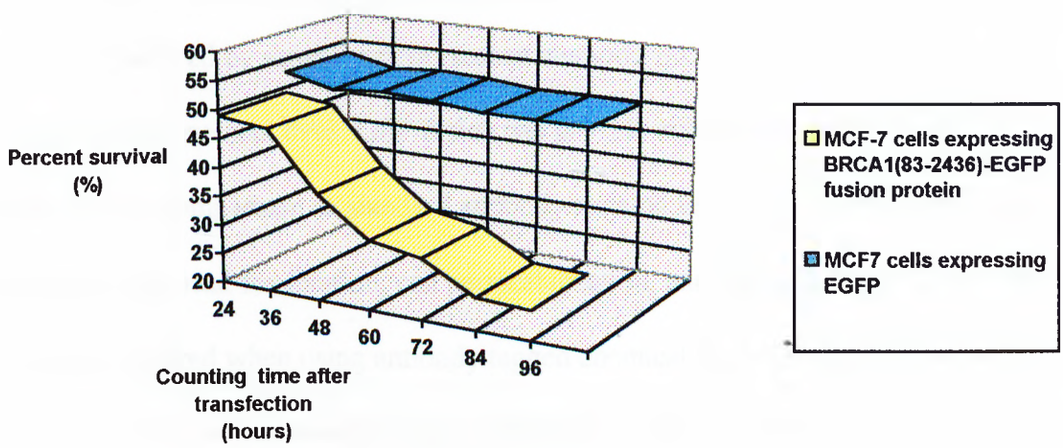


Figure 18. Effect of EGFP and EGFP-BRCA1$_{(83-2436)}$ expression on viability of MCF-7 cells.

CHAPTER 4. DISCUSSION

The green fluorescent protein (GFP) of the jellyfish *Aequeorea Victotia* has attracted widespread interest since the discovery that its chromophore is generated by the autocatalytic post-translational cyclization. This permits fusion of the DNA sequence of GFP with that of any protein whose expression and transport can then be readily monitored by sensitive fluorescence methods without the need to add exogenous fluorescent dyes. GFP fusion can provide enhanced sensitivity and resolution in comparison to standard antibody staining techniques and the GFP tag eliminates the need for fixation, cell permeabilization, and antibody incubation steps normally required when using antibody tagged chemical fluorophores. In this study, Eukaryotic Green Fluorescent Protein (EGFP) was used instead of wild type GFP. EGFP encodes a protein which has a single, red-shifted excitation peak and fluoresces with about 35 times more intensity than wild type GFP. The coding sequence of the EGFP contains more than 190 silent base changes that correspond to human codon-usage preferences for maximal translation efficiency in mammalian cells.

Results represented here demonstrate that the EGFP, encoded by the Eukaryotic Green Fluorescence Protein-N-Terminal protein fusion vector, is an effective reporter for studying the subcellular localization of any protein, especially for

the proteins which are newly characterized and have no highly specific antibodies. We have used this vector to analyze the effect of the five native and putative nuclear targeting sequences of *BRCA1* gene product on the celluar localization BRCA1.

Our utilization of EGFP for understanding the importance of these five NLS in targeting of whole BRCA1 protein has supported the previously reported concept that the whole BRCA1 protein is a nuclear protein (Chen *et al.*, 1995 and 1996, Scully *et al.*, 1996, Wilson *et al.*, 1997, Ruffner *et al.*, 1997, Coene *et al.*, 1997). The fusion protein, EGFP-BRCA1$_{(83-2436)}$, carrying the 2389 bp fragment of *BRCA1* cDNA sequence, which contains the four SV40 like NLSs and one steroid hormone like NLS, mostly accumulated and was retained in the nucleus as evident by its co-localization with nuclear DNA that was stained with fluorochrome dye H33258. By contrast, EGFP lacking the BRCA1$_{(83-2436)}$ fragment at its N-terminus was found both in the nucleus and the cytoplasm. This is to be expected ,because the PSORT analysis of EGFP was shown that there is no defined or evident eukaryotic protein sorting sequences in the sequence of EGFP protein and there are also published data that confirm the uniform distribution of EGFP all over the eukaryotic cell (Ogawa *et al.*,1995, Olson *et al.*, 1995, Levy *et al.*, 1996, Zolotukhin *et al.*, 1996).

The EGFP-BRCA1$_{(83-2436)}$ transfectants showed no major morphological alterations compared to the parental MCF-7 cells. Here, we also report that there was a notable decrease in the percent survival of the MCF-7 breast cancer cell line transfected with EGFP-BRCA1$_{(83-2436)}$. In order to investigate whether this toxic effect could be due to over expression of the exogenous EGFP, we analyzed the endotoxic effect of the mere EGFP expression in MCF-7 cells. The results showed that overexpression of the mere EGFP had no toxic effect in MCF-7 cells.

One explanation for these observations might be the instability of fusion protein coding plasmid, which leds to loss of of the chimera construct after several cell divisions. The half-life of the fusion protein might be another criteria that should be taken into account to explain the decrease in positive fluorescent signal. On the other hand, if we contemplate this decrease as a function of a toxic affect of the fusion protein, we can say that the 2316 bp coding sequence of the putative tumor suppressor *BRCA1*, that contains the potential structural cystein rich RING finger domain, is likely either to cause a growth inhibition as stated by Holt *et al.* (1996) or to induce a apoptosis in MCF-7 cells (Shao *et al.*, 1996). Nevertheless, even to reach a such conclusions additional studies are necessary to define the role of the potential functional domains of the *BRCA1* gene product.

Perspectives

The system described here could be used for efficient characterization and expression of genes in cells of mammalian origin. Visualizing intracellular protein localization as done with the EGFP-BRCA1$_{(83-2436)}$ should also allow experiments to be designed to address important biological questions. For example, an immediate application would be the study of subcellular localization and trafficking of the full-length *BRCA1* gene product and its spliced variants.

In this study we demonstrate that the NLSs consensus sequences of the BRCA1 protein might target the EGFP-BRCA1$_{(83-2436)}$ fusion protein into the nucleus. We should also provide evidence supporting the expression of the EGFP-BRCA1$_{(83-2436)}$ fusion protein by using either immunohistochemical or immunohistofluorescence technique, suggesting that the EGFP was successfully fused to the NLSs of the

BRCA1 protein and was targeted into the nucleus of the MCF-7 cells by using antibody against either N-terminal of the BRCA1 protein or EGFP.

Another exciting question would be opened up for new studies based on observations that was reported in the literature (Shao *et al*, 1996, Gowen *et al*, 1996) and in this study: one of the candidate functional domain, a cystein rich RING finger may induce cell death. Besides this suggestion, involvement of the BRCA1 protein in the cellular proliferation process has been recently reported (Hakem *et al*, 1996, Hakem *et al*, 1997 Scully *et al*, (b) 1997 and Scully *et al*, (c) 1997). Taking these new studies into consideration, physiological and biochemical function of BRCA1 protein and its spliced variants in cellular proliferation processes will be studied in different mammalian cell lines that have different genetic contents in terms of candidate proteins that are expected to correlate with the BRCA1 protein.

REFERENCES

Abel K J., Xu J., Yin G., Lyons R.H., Meisler M. H. and Weber B. L. (1995) Mouse *BRCA1*: localization, sequence analysis and identification of evolutionary conserved domains. *Hum. Mol. Genet.* **12**:2265-2273

Amsterdam A., Lin S. and Hopkins N. (1995) The *Aequorea victoria* green fluorescent protein can be used as a reporter in live Zebrafish embryos. *Dev. Biol.* **171**:123-129

Ausubel F. M., Brent R., Kingston R. E., Moore D. D., Seidman J. G., Smith J. A., Struhl K. (1991) *Current protocols in molecular biology*, **Vol. I Section** :1.8.1-1.8.3

Berman D. B., Costala-Wagner J., Schultz D. C., Lynch H. T., Daly M. and Godwin Andrew K. (1996) Two distinct origins of a common *BRCA1* mutation in breast-ovarian cancer families: a genetic study of 15 185delAG-mutation kindreds. *Am. J. Hum. Genet. 58:1166-1176*

Bienstock R. J., Darden T., Wiseman R., Pedersen L. and Barrett J. C. (1996) Molecular Modeling of the amino-terminal zing ring domain of *BRCA1. Cancer* Res. 56:2539-2545

Birnboim, H. C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* **7:** 1513-1523

Boyd M., Harris F., McFarlane R., Davidson H. R., Black D., M (1995) A human *BRCA1* gene knockout. *Nature* **375**:541-542

Brown M. A, Jones K. A, Nicolai H., Bonjardim M., Black D., McFarlane R., De Jong P., Ouirk J., P, Lehrach H. and Solomon A. (1995) Physical mapping , cloning, and identification of genes within a 500-kb region containing *BRCA1. Proc. Natl.Acad. Sci. USA* **92**:4362-4366

Brown T. A. (1993) Gene Cloning. (Second edition). *Chamber & Hall Publication* pp:225-245

Chalfie M., Tu Y., Euskirchen G., Ward W. W. and Prasher D. C. (1994) Green fluorescent protein as a marker for gene expression. *Science* **263**:802-804

Chapman M. K. and Verma I., M. (1996) Transcriptional activation by BRCA1. *Nature* **382**: 678-679.

Chen Y., Chen C.-F., Riley D. J, Allred D. C., Chen P.-.L., Hoff Daniel V., Osborne C. K. and Lee W.-H. (1995) Aberrant subcellular localization of *BRCA1* in breast cancer. *Science* **270**:789-791

Chen Y., Chen P.-L., Riley D. J, Lee W.-H., Allerd D. C. and Osborne C. K. (1996) Response: Localization of *BRCA1* in human breast and ovarian cancer cells. *Science* **272**:125-126

Coene E., Oostveldt van P., Willems K., Emmelo van J. and Christian R. De P. (1997) BRCA1 is localized in cytoplasmic tube-like invaginations in the nucleus. *Nature Genet.* **16**:122-124

Cubitt A. B., Heim R., Adams S. R., Boyd A. E., Gross L. A. and Tsien R. Y. (1995) Understanding, improving and using green fluorescent proteins. *TIBS* **20**:448-455

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

Eisinger F., Stoppa-Lyonnet D., Longy Michel, Kerangueven F., Noguchy T., Bailly C., Vincent-Salomon A., Jacquemier J., Binbaum D. and Sobol H. (1996) Germline mutation at *BRCA1* affected the histoprognostic grade in hereditary breast cancer. *Cancer Res.* **56**:1-4

Friedman L. S., Szabo C. I., Ostermeyer E. A., Dowd P., Butler L., Park T., Lee M. K., Goode E. L., Rowell S. E. and King M.-C. (1995) Novel inherited mutations and variable Expression of *BRCA1* alleles, including the mutation 185delAG in Ashkenazi Jewish families. *Am. J. Hum. Genet.* **57**:1284-1297

Futreal P. A., Liu Q., Shattuck-Eidens D., Cochran C., Harshman K., Tavtigian S., Bennett L. M., Haugen-Strano A., Swensen J., Miki Y., Eddindton K., McClure M., Frye C., Weaver-Feldhaus J., Ding W., Gholami Z., Soderkvist P., Terry L., Jhanwar S., Berchuck A., Iglehart J. D., Marks J., Ballinger D. G., Barrett J. C., Skolnick M.H., Kamb A. and Wiseman R. (1994) *BRCA1* mutations in primary breast and ovarian carcinomas. *Science* **266**:120-122

Gayther S. A., Warren W., Mazoyer S., Russell P. A., Harrington P. A., Chiano M., Seal S., Hamoudi R., Van R. E., Dunning A. M., Love R., Evans G., Easton D., Clayton D., Stratton M. R. and Ponder B. A. J. (1995) Germ-line mutations of *BRCA1* gene in breast and ovarian cancer families provide evidence for a genotype-phenotype evidence. *Nature Genet. 11:428-433*

Gowen L. C., Johnsin A. M., Latour A. M., Sulik K. and Koller B. H. (1996) *Brca1* deficiency results in early embryonic lethality characterized by neuroepithelial abnormalities. *Nature Genet.* **12**:191-194

Gudas J. M, Nyugen H., Li T. and Cowan K. H. (1995) Hormone-dependent regulation of *BRCA1* in human breast cancer cells. *Cancer Res.* **55**:4561-4565

Hakem R., de la Pompa J. L., Sirard C., Mo R., Woo M., Hakem A., Wakeham ., Potter J., Reitmair A., Billia F., Firpo E., Hui C.C., Roberts J., Rossant J. and Mak T. W. (1996) The tumor suppressor gene *BRCA1* is required for embryonic cellular proliferation in the mouse. *Cell* **85**:1009-1023

Hakem R, de la Pompa J. L., Andrew E., Potter J. and Mak T.W. (1997) Partial recue of *BRCA1*$^{5-6}$ early embryonic lethality by *p53* and *p21* null mutation. *Nature Genet.* **16**: 298-302

Hall J. M, Lee M. K., Newman B., Morrow J. E., Anderson L. A., Huey B., King M.-C. (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**:1684-1659

Hogervorst F. B. L., Cornelis R. S., Bout M., Va Vliet M., Oosterwijk J. C., Olmer R., Bakker B., Klijn J. G. M., Vasen H. F. A., Meijers-Heijboer H., Menko F. H., Cornelisse C. J., den Dunnen J. T., Devilee P. and van Ommen G.-J. B. (1995) Rapid detection of *BRCA1* mutations by the protein truncation test. *Nature Genet.* **10**:208-212

Holt J. T., Thompson M. E., Szabo C., Robinson-Benion C., Arteaga C. L, King M.-C. and Jensen R. A. (1996) Growth retardation and tumor inhibition by *BRCA1. Nature Genet. 12:298-302*

Inoue R., Fukutomi T., Ushijima T., Matsumoto Y., Sugimura T.and Nagao M. (1996) Germline mutation of *BRCA1* in Japanese breast cancer families. *Cancer Res. 55:*3521-3524

Jensen R. A., Thompson M. E., Jetton T. L., Szabo C.I., van der Meer R.,Helou B., Tronick S. R., Page D. L., King M.-C. and Holt J. T.(1996) *BRCA1* is secreted and exhibits properties of a granin. *Nature Genet.* **12**:303-308

Jensen R. A., Thompson M. E., Jetton T. L., van der M. R., Helou B.Arteaga C. L., Page D. L., Holt J. T., Tronick S. R., Gown A. M., Skelly M., Schieeltz D., Szabo C. I., King M.-C. (1996) Reply: ...and secreted tomour suppressors. *Nature Genet.* **13**:269-270

Koonin V. E., Altschul S. F. and Brok P. (1996) ...Functional motifs. *Nature Genet. 13:266-267*

Lane T. F., Deng C., Elson A., Lyu M. S., Kozak C. A. and Leder P. (1995) Expression of *BRCA1* is associated with terminal differentiation ectodermally and mesodermally derived tissues in mice. *Genes & Dev.* **9**:2712-2722

Langston A. A., Malone K. E., Thompson J., Daling J. R. and Ostrander E. A.(1996) *BRCA1* mutation in population-based sample of young women with breast cancer. *N. Eng. J. Med.* **334**:137-142

Levy J. P., Muldoon R. R., Zolotukhin S. and Harles J. L. Jr. (1996) Retroviral transfer and expression of a humanized, red-shift green fluoresent protein gene into human tumor cells. *Nature Biotechnology* **14**: 610-614

Liu C.-Y., Flesken-Nikitin A., Li S., Zeng Y. and Lee W.-H. (1996) Inactivation of the mouse *Brca1* gene leads to failure in the morphogenesis of the egg cylinder in early postimplantation development. *Genes & Dev.* **10**:1835-1843

Lu M., Conzen S. D., Cole C. N. and Arrick B. A. Characterization of functional messenger RNA spliced variants of *BRCA1* expressed in nonmailgnant and tumor-derived breast cells (1996) *Cancer Res.* **56**:4578-4581

Miki Y., Swensen J., Shattuck-Eidens D., Futreal P. A., Harshman K., Tavtigian S., Liu Q., Cochran C., Bennett L. M., Ding W., Bell R., Rosenthal J., Hussey C., Tran T., McClure M., Frye C.,Hattier T., Phelps R., Haugen- Strano A., Katcher H., Yakumo K., Gholami Z., Shaffer D., Stone S., Bayer S., Wray C., Bogden R., Dayananth P., Ward J., Tonin P., Narod S., Bristow P. K., Norris F.H., Helvering L., Morrison P., Rosteck Paul, L. M., Barrett J. C.,Lewis C., Nauhausen S., Cannon-Albright L., Goldgar D., Wiseman R., Kamb A. and Skolnick M. H. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* **266**:66-71

Monteiro N.A., August A. and Hanafusa H. (1996) Evidence for a transcriptional activation function of BRCA1 C-terminal region. *Proc. Natl. Acad. Sci. USA* **93**:13595- 13599

Munn K. E., Walker R. A., Menasce L. and Varley J. M. (1996) Allelic imbalance in the region of the *BRCA1* gene in ductal carcinoma in situ of the breast cancer. *British J. Cancer* **73**:636-639

Neuhausen S. L. and Marshall C. J. (1994) Loss of heterozygosity in familial tumors from three BRCA1-linked kindred. *Cancer Res.* **54**:6069-6072

Neuhausen S. L., Mazoyer S., Friedman L., Stratton M., Ofitt K., CaligoA., Tomlinson G., Cannon-Albright L., Bishop T., Kelsell D.,Solomon E., Weber B., Couch F., Struewing J., Tonin P., Durocher F., Narod S., Skolnick M. H., Lenoir G., Serova O.,Ponder B., Stoppa-Lyonnet D., Easton D., King M.-C. and Goldgar D. E. (1996) Haplotype and phenotype analysis of six recurrent *BRCA1* mutations in 61 families: results of an international study. *Am. J. Hum. Genet.* **58**:271-280

Ogawa H., Inouye S., Tsuji F. I., Yasuda K. and Umesono K. (1995) Localization, trafficking, and temperature-dependence of the *Aequorea* green fluorescent protein in cultured vertebrate cells. *Proc. Natl. Acad. Sci. USA* **92**:11899-1903

Olson K. R., McInstosh J. R. and Olmsted J. B. (1995) Analysis of MAP 3 function in living cells using green fluorescent protein (GFP) chimeras. *J. Cell Biol.* **130**:639-650

Peters K. G., Rao P. S., Bell B. S. and Kindman L. A. (1995) Green fluorescent fusion proteins: Powerful tools for monitoring protein expression in live Zebrafish embryos. *Dev. Biol.* **171**:252-257

Pierre P. (1996) Basic Methods in cell biological research. *EMBO Practical Course* pp:42-44

Rao V. N., Shao N., Ahmad M. and Shyman E., Reddy P. (1996) Antisense RNA to the putative tumor suppressor gene *BRCA1* transforms mouse fibroblasts. *Oncogene* **12**:523-528

Rowell S., Newman B., Boyd J. and King M-C. (1994) Inherited predisposition to breast and ovarian cancer. *Am. J. Hum. Genet.* **55**:861-865

Ruffner H. and Verma I. M. (1997) BRCA1 is a cell cycle-regulated nuclear protein. *Proc. Natl. Acad. Sci. USA* **94**:7138-7143

Serova O., Montagna M., Torchard D., Narod S. A., Tonin P., SyllaB., Lynch H. T., Feunteun J. and Lenoie G. M. (1996) A high incidence of *BRCA1* mutations in 20 breast-ovarian cancer families. *Am. J. Hum. Genet.* **58**:42-51

Scully R., Ganesan S., Brown M., De Caprio J. A., Cannistra S. A.,Feunteun J., Schnitt S., Livingston D.M. **(a)** (1996) Location of *BRCA1* in human breast and ovarian cancer cells. *Science* **272**:123-124

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

Scully R., Anderson S. F., Chao D. M., Wei W., Ye L., Young R. A., Livingston D. M. and Parvin J. D. **(c)** (1997) BRCA1 is a component of the RNA polymearse II haloenzyme. *Proc. Natl. Acad. Sci. USA* **94**:5605-5610

Shao N., Chai Y. L., Reddy P. and Rao V. N. (1996) Induction of apoptosis by the tumor suppressor protein *BRCA1*. *Oncogene* **13**:1-7

Skolnick M. H., Cannon-Albright L. A., Goldgar D. E., Ward J. H., Marshall C.J., Schuman G. B., Hogle H., McWhorter W. P., Wright E., Tran T. D., Bishop D. T., Kushner J. P. and Eyre H. J. (1990) Inheritance of proliferative breast disease in breast cancer kindreds. *Science* **250**:1715-1720

Struewing J. P., Brody L. C., Erdos M. R., Kase R. G., Giambarresi T. R., Smith S. A., Collins F. S. and Tucker M.A. (1995) Detection of eight *BRCA1* mutations in 19 breast/ovarian cancer families, including one family with male breast cancer. *Am. J. Hum. Genet.* **57**:1-7

Tonin P., Moslehi R., Green R., Rosen B., Cole D., Boyd N.,Cutler C., Margolese R., Carter R., McGillivray B., Ives E., LabrieF., Gilchrist D., Morgan K., Simard J. and Narod S. A .(1995) Linkage analysis of 26 Canadian breast-ovarian cancer families. *Hum. Genet.* **95**:545-550

Wang S. and Hazeirigg T. (1994) Implications for *bcd* mRNA localization from spatial distribution of exu protein in *Drosophila* oogenesis. *Nature* **369**:400-403

Webb C. D., Decatur A., Teleman A. and Losick R. (1995) Use of green fluorescent protein for visualization of cell-specific gene expression and subcellular protein localization during sporulation in *Bacillus subtilis*. *J. Bacteriol.* **177**:5906-5911

Wilson C. N., Payton M. N., Pekar S. K., Zhang K., Pacifici R. E., Gudas J. L., Thukral S., Calzone F. J., Reese D. M., Slamon D. I. (1996) *BRCA1* protein products: antibody specificity. *Nature Genet.* **13**:264-265

Wilson C.N., Payton M. N., Elliott G. S., Buaas F. W., Cajulis E. E., Grosshans D., Ramos L., Reese D. M., Slamon D. J. and Calzone F. J. (1997) Differential subcellular localization, expression and biological toxicity of BRCA1 and splice variant BRCA1-Δ11b. *Oncogene* **14**: 1-16

Wooster Richard and Stratton Michael R (1995) Breast cancer susceptibility: a complex disease unravels. *TIG* **11**:3-5

Wu L. C., Wang Z. W., Tsan J, T., Spillman M. A., Phung A., Xu X. L., Yang M.-C. W., Hwang L.-X., Bowcock A. M. and Baer R. (1996) Identification of a RING protein that can interact *in vivo* with the *BRCA1* gene product. *Nature Genet.* **14**:430-440

Xu C.-F., Brown M. A., Chambers J. A., Griffiths B., Nicolai H. and Solomon E. (1995) Distinct transcription start sites generate two forms of *BRCA1* mRNA. *Hum. Mol. Genet.* **4**:2259-2264

Zolotukhin S., Potter M., Hauswirth W., Guy J. and Muzyczka N. (1996) A "Humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cell. *J. Virology* **70**: 4646-4654

APPENDICES

APPENDIX 1.

Human BRCA1 cDNA and protein sequence, Genbank HSU14680

LOCUS	HSU14680 5711 bp mRNA PRI 05-AUG-1995
DEFINITION	Human breast and ovarian cancer susceptibility (BRCA1) mRNA, complete cds.
ACCESSION	U14680
NID	g555931
KEYWORDS	-
SOURCE	human.
ORGANISM	Homo sapiensEukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 5711)
AUTHORS	Miki Y., Swensen J., Shattuck-Eidens D., Futreal P.A., Harshman K., Tavtigian S., Liu Q., Cochran C., Bennett L.M., Ding W., Bell R., Rosenthal J., Hussey C., Tran T., McClure M., Frye C., Hattier T., Phelps R., Haugen-Strano A., Katcher H., Yakumo K., Gholami Z., Shaffer D., Stone S., Bayer S., Wray C., Bogden R., Dayananth P., Ward J., Tonin P., Narod S., Bristow P.K., Norris F.H., Helvering L., Morrison P., Rosteck P., Lai M., Barrett J.C., Lewis C., Neuhausen S., Cannon-Albright L., Goldgar D., Wiseman R., Kamb A. and Skolnick M.H.
TITLE	A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1
JOURNAL	Science 266 (5182), 66-71 (1994)
MEDLINE	95025896
REFERENCE	2 (bases 1 to 5711)
AUTHORS	Skolnick,M.H.
TITLE	Direct Submission
JOURNAL	Submitted (14-SEP-1994) Mark H. Skolnick, Myriad Genetics Inc. and the University of Utah, 421 Wakara Way, Suite 201, Salt Lake City, UT 84108, USA
FEATURES	Location/Qualifiers
source	1..5711
	/organism="Homo sapiens"
	/note="For sequence of alternatively spliced exon 4, see GenBank Accession Number U15595"
	/chromosome="17"
	/map="17q21; spans D17S855"
5'UTR	1..119
exon	1..100
	/number=1
exon	101..199
	/number=2
gene	120..5711
	/gene="BRCA1"
CDS	120..5711
	/gene="BRCA1"
	/note="influences susceptibility to breast and ovarian cancer"
	/codon_start=1
	/db_xref="PID:g555932"

/translation="MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFC
KFCMLKLLNQKKGPSQCPLCKNDITKRSLQESTRFSQLVEELLKIICAFQLDTGLEYA
NSYNFAKKENNSPEHLKDEVSIIQSMGYRNRAKRLLQSEPENPSLQETSLSVQLSNLG
TVRTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATYCSVGDQELLQITPQGTRDEISL
DSAKKAACEFSETDVTNTEHHQPSNNDLNTTEKRAAERHPEKYQGSSVSNLHVEPCGT
NTHASSLQHENSSLLLTKDRMNVEKAEFCNKSKQPGLARSQHNRWAGSKETCNDRRTP

STEKKVDLNADPLCERKEWNKQKLPCSENPRDTEDVPWITLNSSIQKVNEWFSRSDEL
LGSDDSHDGESESNAKVADVLDVLNEVDEYSGSSEKIDLLASDPHEALICKSERVHSK
SVESNIEDKIFGKTYRKKASLPNLSHVTENLIIGAFVTEPQIIQERPLTNKLKRKRRP
TSGLHPEDFIKKADLAVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGDSIQNEK
NPNPIESLEKESAFKTKAEPISSSISNMELELNIHNSKAPKKNRLRRKSSTRHIHALE
LVVSRNLSPPNCTELQIDSCSSSEEIKKKKYNQMPVRHSRNLQLMEGKEPATGAKKSN
KPNEQTSKRHDSDTFPELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLETVKV
SNNAEDPKDLMLSGERVLQTERSVESSSISLVPGTDYGTQESISLLEVSTLGKAKTEP
NKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNHSRETSIEMEESELDAQY
LQNTFKVSKRQSFAPFSNPGNAEEECATFSAHSGSLKKQSPKVTFECEQKEENQGKNE
SNIKPVQTVNITAGFPVVGQKDKPVDNAKCSIKGGSRFCLSSQFRGNETGLITPNKHG
LLQNPYRIPPLFPIKSFVKTKCKKNLLEENFEEHSMSPEREMGNENIPSTVSTISRNN
IRENVFKEASSSNINEVGSSTNEVGSSINEIGSSDENIQAELGRNRGPKLNAMLRLGV
LQPEVYKQSLPGSNCKHPEIKKQEYEEVVQTVNTDFSPYLISDNLEQPMGSSHASQVC
SETPDDLLDDGEIKEDTSFAENDIKESSAVFSKSVQKGELSRSPSPFTHTHLAQGYRR
GAKKLESSEENLSSEDEELPCFQHLLFGKVNNIPSQSTRHSTVATECLSKNTEENLLS
LKNSLNDCSNQVILAKASQEHHLSEETKCSASLFSSQCSELEDLTANTNTQDPFLIGS
SKQMRHQSESQGVGLSDKELVSDDEERGTGLEENNQEEQSMDSNLGEAASGCESETSV
SEDCSGLSSQSDILTTQQRDTMQHNLIKLQQEMAELEAVLEQHGSQPSNSYPSIISDS
SALEDLRNPEQSTSEKAVLTSQKSSEYPISQNPEGLSADKFEVSADSSTSKNKEPGVE
RSSPSKCPSLDDRWYMHSCSGSLQNRNYPSQEELIKVVDVEEQQLEESGPHDLTETSY
LPRQDLEGTPYLESGISLFSDDPESDPSEDRAPESARVGNIPSSTSALKVPQLKVAES
AQSPAAAHTTDTAGYNAMEESVSREKPELTASTERVNKRMSMVVSGLTPEEFMLVYKF
ARKHHITLTNLITEETTHVVMKTDAEFVCERTLKYFLGIAGGKWVVSYFWVTQSIKER
KMLNEHDFEVRGDVVNGRNHQGPKRARESQDRKIFRGLEICCYGPFTNMPTDQLEWMV
QLCGASVVKELSSFTLGTGVHPIVVVQPDAWTEDNGFHAIGQMCEAPVVTREWVLDSV
ALYQCQELDTYLIPQIPHSHY"

exon	200..253 /gene="BRCA1" /number=3
exon	254..331 /gene="BRCA1" /number=5
exon	332..420 /gene="BRCA1" /number=6
exon	421..560 /gene="BRCA1" /number=7
exon	561..665 /gene="BRCA1" /number=8
exon	666..712 /gene="BRCA1" /number=9
exon	713..788 /gene="BRCA1" /number=10
exon	789..4215 /gene="BRCA1" /number=11
exon	4216..4302 /gene="BRCA1" /number=12
exon	4303..4476 /gene="BRCA1" /number=13
exon	4477..4603 /gene="BRCA1" /number=14
exon	4604..4794 /gene="BRCA1" /number=15
exon	4795..5105 /gene="BRCA1" /number=16
exon	5106..5193 /gene="BRCA1" /number=17
exon	5194..5273 /gene="BRCA1" /number=18
exon	5274..5310 /gene="BRCA1" /number=19
exon	5311..5396 /gene="BRCA1" /number=20
exon	5397..5451 /gene="BRCA1" /number=21
exon	5452..5526 /gene="BRCA1" /number=22
exon	5527..5586 /gene="BRCA1" /number=23
exon	5587..5711 /gene="BRCA1" /number=24

BASE COUNT 1956 a 1099 c 1274 g 1382 t
ORIGIN

1 agctcgctga gacttcctgg accccgcacc aggctgtggg gtttctcaga taactgggcc
61 cctgcgctca ggaggccttc accctctgct ctgggtaaag ttcattggaa cagaaagaaa
121 tggatttatc tgctcttcgc gttgaagaag tacaaaatgt cattaatgct atgcagaaaa
181 tcttagagtg tcccatctgt ctggagttga tcaaggaacc tgtctccaca aagtgtgacc
241 acatattttg caaattttgc atgctgaaac ttctcaacca gaagaaaggg ccttcacagt
301 gtcctttatg taagaatgat ataaccaaaa ggagcctaca agaaagtacg agatttagtc

361 aacttgttga agagctattg aaaatcattt gtgcttttca gcttgacaca ggtttggagt
421 atgcaaacag ctataatttt gcaaaaaagg aaaataactc tcctgaacat ctaaaagatg
481 aagtttctat catccaaagt atgggctaca gaaaccgtgc caaaagactt ctacagagtg
541 aacccgaaaa tccttccttg caggaaacca gtctcagtgt ccaactctct aaccttggaa
601 ctgtgagaac tctgaggaca aagcagcgga tacaacctca aaagacgtct gtctacattg
661 aattgggatc tgattcttct gaagataccg ttaataaggc aacttattgc agtgtgggag
721 atcaagaatt gttacaaatc acccctcaag gaaccaggga tgaaatcagt ttggattctg
781 caaaaaaggc tgcttgtgaa ttttctgaga cggatgtaac aaatactgaa catcatcaac
841 ccagtaataa tgatttgaac accactgaga agcgtgcagc tgagaggcat ccagaaaagt
901 atcagggtag ttctgtttca aacttgcatg tggagccatg tggcacaaat actcatgcca
961 gctcattaca gcatgagaac agcagtttat tactcactaa agacagaatg aatgtagaaa
1021 aggctgaatt ctgtaataaa agcaaacagc ctggcttagc aaggagccaa cataacagat
1081 gggctggaag taaggaaaca tgtaatgata ggcggactcc cagcacagaa aaaaaggtag
1141 atctgaatgc tgatcccctg tgtgagagaa aagaatggaa taagcagaaa ctgccatgct
1201 cagagaatcc tagagatact gaagatgttc cttggataac actaaatagc agcattcaga
1261 aagttaatga gtggttttcc agaagtgatg aactgttagg ttctgatgac tcacatgatg
1321 gggagtctga atcaaatgcc aaagtagctg atgtattgga cgttctaaat gaggtagatg
1381 aatattctgg ttcttcagag aaaatagact tactggccag tgatcctcat gaggctttaa
1441 tatgtaaaag tgaaagagtt cactccaaat cagtagagag taatattgaa gacaaaatat
1501 ttgggaaaac ctatcggaag aaggcaagcc tccccaactt aagccatgta actgaaaatc
1561 taattatagg agcatttgtt actgagccac agataataca agagcgtccc ctcacaaata
1621 aattaaagcg taaaaggaga cctacatcag gccttcatcc tgaggatttt atcaagaaag
1681 cagatttggc agttcaaaag actcctgaaa tgataaatca gggaactaac caaacggagc
1741 agaatggtca agtgatgaat attactaata gtggtcatga gaataaaaca aaaggtgatt
1801 ctattcagaa tgagaaaaat cctaacccaa tagaatcact cgaaaaagaa tctgctttca
1861 aaacgaaagc tgaacctata agcagcagta taagcaatat ggaactcgaa ttaaatatcc
1921 acaattcaaa agcacctaaa aagaataggc tgaggaggaa gtcttctacc aggcatattc
1981 atgcgcttga actagtagtc agtagaaatc taagcccacc taattgtact gaattgcaaa
2041 ttgatagttg ttctagcagt gaagagataa agaaaaaaaa gtacaaccaa atgccagtca
2101 ggcacagcag aaacctacaa ctcatggaag gtaaagaacc tgcaactgga gccaagaaga
2161 gtaacaagcc aaatgaacag acaagtaaaa gacatgacag cgatactttc ccagagctga
2221 agttaacaaa tgcacctggt tcttttacta agtgttcaaa taccagtgaa cttaaagaat
2281 ttgtcaatcc tagccttcca agagaagaaa aagaagagaa actagaaaca gttaaagtgt
2341 ctaataatgc tgaagacccc aaagatctca tgttaagtgg agaaagggtt ttgcaaactg
2401 aaagatctgt agagagtagc agtatttcat tggtacctgg tactgattat ggcactcagg
2461 aaagtatctc gttactggaa gttagcactc tagggaaggc aaaaacagaa ccaaataaat
2521 gtgtgagtca gtgtgcagca tttgaaaacc ccaagggact aattcatggt tgttccaaag
2581 ataatagaaa tgacacagaa ggctttaagt atccattggg acatgaagtt aaccacagtc
2641 gggaaacaag catagaaatg gaagaaagtg aacttgatgc tcagtatttg cagaatacat
2701 tcaaggtttc aaagcgccag tcatttgctc cgttttcaaa tccaggaaat gcagaagagg
2761 aatgtgcaac attctctgcc cactctgggt ccttaaagaa acaaagtcca aaagtcactt
2821 ttgaatgtga acaaaaggaa gaaaatcaag gaaagaatga gtctaatatc aagcctgtac
2881 agacagttaa tatcactgca ggctttcctg tggttggtca gaaagataag ccagttgata
2941 atgccaaatg tagtatcaaa ggaggctcta ggttttgtct atcatctcag ttcagaggca
3001 acgaaactgg actcattact ccaaataaac atggactttt acaaaaccca tatcgtatac
3061 caccactttt tcccatcaag tcatttgtta aaactaaatg taagaaaaat ctgctagagg
3121 aaaactttga ggaacattca atgtcacctg aaagagaaat gggaaatgag aacattccaa
3181 gtacagtgag cacaattagc cgtaataaca ttagagaaaa tgtttttaaa gaagccagct
3241 caagcaatat taatgaagta ggttccagta ctaatgaagt gggctccagt attaatgaaa
3301 taggttccag tgatgaaaac attcaagcag aactaggtag aaacagaggg ccaaaattga
3361 atgctatgct tagattaggg gttttgcaac ctgaggtcta taaacaaagt cttcctggaa
3421 gtaattgtaa gcatcctgaa ataaaaaagc aagaatatga agaagtagtt cagactgtta
3481 atacagattt ctctccatat ctgatttcag ataacttaga acagcctatg ggaagtagtc
3541 atgcatctca ggtttgttct gagacacctg atgacctgtt agatgatggt gaaataaagg
3601 aagatactag ttttgctgaa aatgacatta aggaaagttc tgctgttttt agcaaaagcg
3661 tccagaaagg agagcttagc aggagtccta gcccttcac ccatacacat ttggctcagg
3721 gttaccgaag aggggccaag aaattagagt cctcagaaga gaacttatct agtgaggatg
3781 aagagcttcc ctgcttccaa cacttgttat ttggtaaagt aaacaatata ccttctcagt

3841 ctactaggca tagcaccgtt gctaccgagt gtctgtctaa gaacacagag gagaatttat
3901 tatcattgaa gaatagctta aatgactgca gtaaccaggt aatattggca aaggcatctc
3961 aggaacatca ccttagtgag gaaacaaaat gttctgctag cttgttttct tcacagtgca
4021 gtgaattgga agacttgact gcaaatacaa acacccagga tcctttcttg attggttctt
4081 ccaaacaaat gaggcatcag tctgaaagcc agggagttgg tctgagtgac aaggaattgg
4141 tttcagatga tgaagaaaga ggaacgggct tggaagaaaa taatcaagaa gagcaaagca
4201 tggattcaaa cttaggtgaa gcagcatctg ggtgtgagag tgaaacaagc gtctctgaag
4261 actgctcagg gctatcctct cagagtgaca ttttaaccac tcagcagagg gataccatgc
4321 aacataacct gataaagctc cagcaggaaa tggctgaact agaagctgtg ttagaacagc
4381 atgggagcca gccttctaac agctaccctt ccatcataag tgactcttct gcccttgagg
4441 acctgcgaaa tccagaacaa agcacatcag aaaaagcagt attaacttca cagaaaagta
4501 gtgaataccc tataagccag aatccagaag gcctttctgc tgacaagttt gaggtgtctg
4561 cagatagttc taccagtaaa aataaagaac caggagtgga aaggtcatcc ccttctaaat
4621 gcccatcatt agatgatagg tggtacatgc acagttgctc tgggagtctt cagaatagaa
4681 actacccatc tcaagaggag ctcattaagg ttgttgatgt ggaggagcaa cagctggaag
4741 agtctgggcc acacgatttg acggaaacat cttacttgcc aaggcaagat ctagagggaa
4801 ccccttacct ggaatctgga atcagcctct tctctgatga ccctgaatct gatccttctg
4861 aagacagagc cccagagtca gctcgtgttg gcaacatacc atcttcaacc tctgcattga
4921 aagttcccca attgaaagtt gcagaatctg cccagagtcc agctgctgct catactactg
4981 atactgctgg gtataatgca atggaagaaa gtgtgagcag ggagaagcca gaattgacag
5041 cttcaacaga aagggtcaac aaaagaatgt ccatggtggt gtctggcctg accccagaag
5101 aatttatgct cgtgtacaag tttgccagaa aacaccacat cactttaact aatctaatta
5161 ctgaagagac tactcatgtt gttatgaaaa cagatgctga gtttgtgtgt gaacggacac
5221 tgaaatattt tctaggaatt gcgggaggaa aatgggtagt tagctatttc tgggtgaccc
5281 agtctattaa agaaagaaaa atgctgaatg agcatgattt tgaagtcaga ggagatgtgg
5341 tcaatggaag aaaccaccaa ggtccaaagc gagcaagaga atcccaggac agaaagatct
5401 tcaggggget agaaatctgt tgctatgggc ccttcaccaa catgcccaca gatcaactgg
5461 aatggatggt acagctgtgt ggtgcttctg tggtgaagga gctttcatca ttcacccttg
5521 gcacaggtgt ccacccaatt gtggttgtgc agccagatgc ctggacagag gacaatggct
5581 tccatgcaat tgggcagatg tgtgaggcac ctgtggtgac ccgagagtgg gtgttggaca
5641 gtgtagcact ctaccagtgc caggagctgg acacctacct gataccccag atcccccaca
5701 gccactactg a

APPENDIX 2.

Cloning vector pEGFP-N2 with enhanced green fluorescent protein gene, complete sequence. Genebank U57608

LOCUS CVU57608 4737 bp DNA circular SYN 14-SEP-1996
DEFINITION Cloning vector pEGFP-N2 with enhanced green fluorescent protein gene, complete sequence.
ACCESSION U57608
NID g1543069
KEYWORDS
SOURCE Cloning vector pEGFP-N2.
ORGANISM Cloning vector pEGFP-N2 artificial sequence; cloning vectors.
REFERENCE 1 (bases 1 to 4737) AUTHORS Cormack,B.P., Valdivia,R. and Falkow,S. TITLE FACS-optimized mutants of the green fluorescent protein (GFP) JOURNAL Gene (1996) In press
REFERENCE 2 (bases 1 to 4737) AUTHORS Kitts,P.A. TITLE pEGFP-N2 complete sequence JOURNAL Unpublished
REFERENCE 3 (bases 1 to 4737) AUTHORS Kitts,P.A. TITLE Direct Submission JOURNAL Submitted (06-MAY-1996) P.A. Kitts, CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303-4230, USA
COMMENT This vector can be obtained from CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303-4230, USA. To place an order call (415) 424-8222 or (800) 662-2566, extension 1. International customers, please contact your local distributor. For technical information, call (415) 424-8222 or (800) 662-2566, extension 3. This sequence has been compiled from information in the sequence databases, published literature and other sources, together with partial sequences obtained by CLONTECH. If you suspect there is an error in this sequence, please contact CLONTECH's Technical Service Department at (415) 424-8222 or (800) 662-2566, extension 3 or E-mail TECH@CLONTECH.COM.FEATURES

FEATURES Location/Qualifiers
source 1..4737
/organism="Cloning vector pEGFP-N2"
CDS 683..1402
/gene="egfp"
/codon_start=1
/product="Enhanced Green Fluorescent Protein"
/db_xref="PID:g1373322"
/transl_table=11
/translation="MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKL
PVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFE
GDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQL
ADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDH MVLLEFVTAAGITLGMDELYK"
gene 683..1402
/gene="egfp"
CDS 2633..3427
/codon_start=1
/product="neomycin phosphotransferase"
/db_xref="PID:g1373323"
/transl_table=11
/translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGRPVLFVKTDL
SGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDLLSSHLAPAEKVSI
MADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQGLAPAELFARLKA

SMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRYQDIALATRDIAEELGGEWAD
RFLVLYGIAAPDSQRIAFYRLLDEFF"

BASE COUNT 1140 a 1287 c 1248 g 1062 t
ORIGIN

1 tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg
61 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt
121 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
181 atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc
241 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
301 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac
361 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg
421 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg
481 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt
541 acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcc gctagcgcta
601 ccggactcag atctcgagct caagcttcga attctgcagt cgacggtacc gcgggcccgg
661 gatccaccgg ccggtcgcca ccatggtgag caagggcgag gagctgttca ccggggtggt
721 gcccatcctg gtcgagctgg acggcgacgt aaacggccac aagttcagcg tgtccggcga
781 gggcgagggc gatgccacct acggcaagct gaccctgaag ttcatctgca ccaccggcaa
841 gctgcccgtg ccctggccca ccctcgtgac caccctgacc tacggcgtgc agtgcttcag
901 ccgctacccc gaccacatga agcagcacga cttcttcaag tccgccatgc ccgaaggcta
961 cgtccaggag cgcaccatct tcttcaagga cgacggcaac tacaagaccc gcgccgaggt
1021 gaagttcgag ggcgacaccc tggtgaaccg catcgagctg aagggcatcg acttcaagga
1081 ggacggcaac atcctggggc acaagctgga gtacaactac aacagccaca acgtctatat
1141 catggccgac aagcagaaga acggcatcaa ggtgaacttc aagatccgcc acaacatcga
1201 ggacggcagc gtgcagctcg ccgaccacta ccagcagaac accccatcg gcgacggccc
1261 cgtgctgctg cccgacaacc actacctgag cacccagtcc gccctgagca aagaccccaa
1321 cgagaagcgc gatcacatgg tcctgctgga gttcgtgacc gccgccggga tcactctcgg
1381 catggacgag ctgtacaagt aaagcggccg cgactctaga tcataatcag ccataccaca
1441 tttgtagagg ttttacttgc tttaaaaaac ctcccacacc tccccctgaa cctgaaacat
1501 aaaatgaatg caattgttgt tgttaacttg tttattgcag cttataatgg ttacaaataa
1561 agcaatagca tcacaaattt cacaaataaa gcatttttt cactgcattc tagttgtggt
1621 ttgtccaaac tcatcaatgt atcttaaggc gtaaattgta agcgttaata ttttgttaaa
1681 attcgcgtta aatttttgtt aaatcagctc atttttaac caataggccg aaatcggcaa
1741 aatcccttat aaatcaaaag aatagaccga gatagggttg agtgttgttc cagtttggaa
1801 caagagtcca ctattaaaga acgtggactc caacgtcaaa gggcgaaaaa ccgtctatca
1861 gggcgatggc ccactacgtg aaccatcacc ctaatcaagt tttttggggt cgaggtgccg
1921 taaagcacta aatcggaacc ctaaagggag cccccgattt agagcttgac ggggaaagcc
1981 ggcgaacgtg gcgagaaagg aagggaagaa agcgaaagga gcgggcgcta gggcgctggc
2041 aagtgtagcg gtcacgctgc gcgtaaccac cacacccgcc gcgcttaatg cgccgctaca
2101 gggcgcgtca ggtggcactt ttcggggaaa tgtgcgcgga acccctattt gtttatttttt
2161 ctaaatacat tcaaatatgt atccgctcat gagacaataa ccctgataaa tgcttcaata
2221 atattgaaaa aggaagagtc ctgaggcgga aagaaccagc tgtggaatgt gtgtcagtta
2281 gggtgtggaa agtccccagg ctccccagca ggcagaagta tgcaaagcat gcatctcaat
2341 tagtcagcaa ccaggtgtgg aaagtcccca ggctccccag caggcagaag tatgcaaagc
2401 atgcatctca attagtcagc aaccatagtc ccgcccctaa ctccgcccat cccgccccta
2461 actccgccca gttccgccca ttctccgccc catggctgac taatttttt tatttatgca
2521 gaggccgagg ccgcctcggc ctctgagcta ttccagaagt agtgaggagg ctttttgga
2581 ggcctaggct tttgcaaaga tcgatcaaga gacaggatga ggatcgtttc gcatgattga
2641 acaagatgga ttgcacgcag gttctccggc cgcttgggtg gagaggctat tcggctatga
2701 ctgggcacaa cagacaatcg gctgctctga tgccgccgtg ttccggctgt cagcgcaggg
2761 gcgcccggtt ctttttgtca agaccgacct gtccggtgcc ctgaatgaac tgcaagacga
2821 ggcagcgcgg ctatcgtggc tggccacgac gggcgttcct tgcgcagctg tgctcgacgt
2881 tgtcactgaa gcgggaaggg actggctgct attgggcgaa gtgccggggc aggatctcct
2941 gtcatctcac cttgctcctg ccgagaaagt atccatcatg gctgatgcaa tgcggcggct
3001 gcatacgctt gatccggcta cctgcccatt cgaccaccaa gcgaaacatc gcatcgagcg
3061 agcacgtact cggatggaag ccggtcttgt cgatcaggat gatctggacg aagagcatca
3121 ggggctcgcg ccagccgaac tgttcgccag gctcaaggcg agcatgcccg acggcgagga

3181 tctcgtcgtg acccatggcg atgcctgctt gccgaatatc atggtggaaa atggccgctt
3241 ttctggattc atcgactgtg gccggctggg tgtggcggac cgctatcagg acatagcgtt
3301 ggctacccgt gatattgctg aagagcttgg cggcgaatgg gctgaccgct tcctcgtgct
3361 ttacggtatc gccgctcccg attcgcagcg catcgccttc tatcgccttc ttgacgagtt
3421 cttctgagcg ggactctggg gttcgaaatg accgaccaag cgacgcccaa cctgccatca
3481 cgagatttcg attccaccgc cgccttctat gaaaggttgg gcttcggaat cgttttccgg
3541 gacgccggct ggatgatcct ccagcgcggg gatctcatgc tggagttctt cgcccaccct
3601 agggggaggc taactgaaac acggaaggag acaataccgg aaggaacccg cgctatgacg
3661 gcaataaaaa gacagaataa aacgcacggt gttgggtcgt ttgttcataa acgcggggtt
3721 cggtcccagg gctggcactc tgtcgatacc ccaccgagac cccattgggg ccaatacgcc
3781 cgcgtttctt ccttttcccc accccacccc ccaagttcgg gtgaaggccc agggctcgca
3841 gccaacgtcg gggcggcagg ccctgccata gcctcaggtt actcatatat actttagatt
3901 gatttaaaac ttcatttta atttaaaagg atctaggtga agatcctttt tgataatctc
3961 atgaccaaaa tcccttaacg tgagttttcg ttccactgag cgtcagaccc cgtagaaaag
4021 atcaaaggat cttcttgaga tcctttttt ctgcgcgtaa tctgctgctt gcaaacaaaa
4081 aaaccaccgc taccagcggt ggtttgtttg ccggatcaag agctaccaac tctttttccg
4141 aaggtaactg gcttcagcag agcgcagata ccaaatactg tccttctagt gtagccgtag
4201 ttaggccacc acttcaagaa ctctgtagca ccgcctacat acctcgctct gctaatcctg
4261 ttaccagtgg ctgctgccag tggcgataag tcgtgtctta ccgggttgga ctcaagacga
4321 tagttaccgg ataaggcgca gcggtcgggc tgaacggggg gttcgtgcac acagcccagc
4381 ttggagcgaa cgacctacac cgaactgaga tacctacagc gtgagctatg agaaagcgcc
4441 acgcttcccg aagggagaaa ggcggacagg tatccggtaa gcggcagggt cggaacagga
4501 gagcgcacga gggagcttcc agggggaaac gcctggtatc tttatagtcc tgtcgggttt
4561 cgccacctct gacttgagcg tcgatttttg tgatgctcgt caggggggcg gagcctatgg
4621 aaaaacgcca gcaacgcggc ctttttacgg ttcctggcct tttgctggcc ttttgctcac
4681 atgttctttc ctgcgttatc ccctgattct gtggataacc gtattaccgc catgcat

APPENDIX 3.

Restriction endonuclase map of the pLXSN-BRCA1 eukaryotic expression vector.

pLXSN-BRCA1 (11544 base pairs)

```
GAATTGCTAGCAATTGCTAGCAATTGCTAGCAATTCATACCAGATCACCGAAAACTGTCCTCCAAATGTG  Base pairs
CTTAACGATCGTTAACGATCGTTAACGATCGTTAAGTATGGTCTAGTGGCTTTTGACAGGAGGTTTACAC  1 to 70

TCCCCCTCACACTCCCAAATTCGCGGGCTTCTGCCTCTTAGACCACTCTACCCTATTCCCCACACTCACC  Base pairs
AGGGGGAGTGTGAGGGTTTAAGCGCCCGAAGACGGAGAATCTGGTGAGATGGGATAAGGGGTGTGAGTGG  71 to 140

GGAGCCAAAGCCGCGGCCCTTCCGTTTCTTTGCTTTTGAAAGACCCCACCCGTAGGTGGCAAGCTAGCTT  Base pairs
CCTCGGTTTCGGCGCCGGGAAGGCAAAGAAACGAAAACTTTCTGGGGTGGGCATCCACCGTTCGATCGAA  141 to 210

AAGTAACGCCACTTTGCAAGGCATGGAAAAATACATAACTGAGAATAGAAAAGTTCAGATCAAGGTCAGG  Base pairs
TTCATTGCGGTGAAACGTTCCGTACCTTTTTATGTATTGACTCTTATCTTTTCAAGTCTAGTTCCAGTCC  211 to 280

AACAAAGAAACAGCTGAATACCAAACAGGATATCTGTGGTAAGCGGTTCCTGCCCCGGCTCAGGGCCAAG  Base pairs
TTGTTTCTTTGTCGACTTATGGTTTGTCCTATAGACACCATTCGCCAAGGACGGGGCCGAGTCCCGGTTC  281 to 350

AACAGATGAGACAGCTGAGTGATGGGCCAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGGCTCGGG  Base pairs
TTGTCTACTCTGTCGACTCACTACCCGGTTTGTCCTATAGACACCATTCGTCAAGGACGGGGCCGAGCCC  351 to 420

GCCAAGAACAGATGGTCCCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGTGAATCATCAGATGTTTCCA  Base pairs
CGGTTCTTGTCTACCAGGGGTCTACGCCAGGTCGGGAGTCGTCAAAGATCACTTAGTAGTCTACAAAGGT  421 to 490

GGGTGCCCCAAGGACCTGAAAATGACCCTGTACCTTATTTGAACTAACCAATCAGTTCGCTTCTCGCTTC  Base pairs
CCCACGGGGTTCCTGGACTTTTACTGGGACATGGAATAAACTTGATTGGTTAGTCAAGCGAAGAGCGAAG  491 to 560

TGTTCGCGCGCTTCCGCTCTCCGAGCTCAATAAAAGAGCCCACAACCCCTCACTCGGCGCGCCAGTCTTC  Base pairs
ACAAGCGCGCGAAGGCGAGAGGCTCGAGTTATTTTCTCGGGTGTTGGGGAGTGAGCCGCGCGGTCAGAAG  561 to 630

CGATAGACTGCGTCGCCCGGGTACCCGTATTCCCAATAAAGCCTCTTGCTGTTTGCATCCGAATCGTGGT  Base pairs
GCTATCTGACGCAGCGGGCCCATGGGCATAAGGGTTATTTCGGAGAACGACAAACGTAGGCTTAGCACCA  631 to 700

CTCGCTGTTCCTTGGGAGGGTCTCCTCTGAGTGATTGACTACCCACGACGGGGGTCTTTCATTTGGGGGC  Base pairs
GAGCGACAAGGAACCCTCCCAGAGGAGACTCACTAACTGATGGGTGCTGCCCCCAGAAAGTAAACCCCCG  701 to 770

TCGTCCGGGATTTGGAGACCCCTGCCCAGGGACCACCGACCCACCACCGGGAGGTAAGCTGGCCAGCAAC  Base pairs
AGCAGGCCCTAAACCTCTGGGGACGGGTCCCTGGTGGCTGGGTGGTGGCCCTCCATTCGACCGGTCGTTG  771 to 840

TTATCTGTGTCTGTCCGATTGTCTAGTGTCTATGTTTGATGTTATGCGCCTGCGTCTGTACTAGTTAGCT  Base pairs
AATAGACACAGACAGGCTAACAGATCACAGATACAAACTACAATACGCGGACGCAGACATGATCAATCGA  841 to 910

AACTAGCTCTGTATCTGGCGGACCCGTGGTGGAACTGACGAGTTCTGAACACCCGGCCGCAACCCTGGGA  Base pairs
TTGATCGAGACATAGACCGCCTGGGCACCACCTTGACTGCTCAAGACTTGTGGGCCGGCGTTGGGACCCT  911 to 980

GACGTCCCAGGGACTTTGGGGGCCGTTTTTGTGGCCCGACCTGAGGAAGGGAGTCGATGTGGAATCCGAC  Base pairs
CTGCAGGGTCCCTGAAACCCCCGGCAAAAACACCGGGCTGGACTCCTTCCCTCAGCTACACCTTAGGCTG  981 to 1050

CCCGTCAGGATATGTGGTTCTGGTAGGAGACGAGAACCTAAAACAGTTCCCGCCTCCGTCTGAATTTTTG  Base pairs
GGGCAGTCCTATACACCAAGACCATCCTCTGCTCTTGGATTTTGTCAAGGGCGGAGGCAGACTTAAAAAC  1051 to 1120

CTTTCGGTTTGGAACCGAAGCCGCGCGTCTTGTCTGCTGCAGCGCTGCAGCATCGTTCTGTGTTGTCTCT  Base pairs
GAAAGCCAAACCTTGGCTTCGGCGCGCAGAACAGACGACGTCGCGACGTCGTAGCAAGACACAACAGAGA  1121 to 1190
```

GTCTGACTGTGTTTCTGTATTTGTCTGAAAATTAGGGCCAGACTGTTACCACTCCCTTAAGTTTGACCTT Base pairs
CAGACTGACACAAAGACATAAACAGACTTTTAATCCCGGTCTGACAATGGTGAGGGAATTCAAACTGGAA 1191 to 1260

AGGTCACTGGAAAGATGTCGAGCGGATCGCTCACAACCAGTCGGTAGATGTCAAGAAGAGACGTTGGGTT Base pairs
TCCAGTGACCTTTCTACAGCTCGCCTAGCGAGTGTTGGTCAGCCATCTACAGTTCTTCTCTGCAACCCAA 1261 to 1330

ACCTTCTGCTCTGCAGAATGGCCAACCTTTAACGTCGGATGGCCGCGAGACGGCACCTTTAACCGAGACC Base pairs
TGGAAGACGAGACGTCTTACCGGTTGGAAATTGCAGCCTACCGGCGCTCTGCCGTGGAAATTGGCTCTGG 1331 to 1400

TCATCACCCAGGTTAAGATCAAGGTCTTTTCACCTGGCCCGCATGGACACCCAGACCAGGTCCCCTACAT Base pairs
AGTAGTGGGTCCAATTCTAGTTCCAGAAAAGTGGACCGGGCGTACCTGTGGGTCTGGTCCAGGGGATGTA 1401 to 1470

CGTGACCTGGGAAGCCTTGGCTTTTGACCCCCCTCCCTGGGTCAAGCCCTTTGTACACCCTAAGCCTCCG Base pairs
GCACTGGACCCTTCGGAACCGAAAACTGGGGGGAGGGACCCAGTTCGGGAAACATGTGGGATTCGGAGGC 1471 to 1540

CCTCCTCTTCCTCCATCCGCCCCGTCTCTCCCCCTTGAACCTCCTCGTTCGACCCCGCCTCGATCCTCCC Base pairs
GGAGGAGAAGGAGGTAGGCGGGGCAGAGAGGGGGAACTTGGAGGAGCAAGCTGGGGCGGAGCTAGGAGGG 1541 to 1610

TTTATCCAGCCCTCACTCCTTCTCTAGGCGCCGGAATTCGTTAAGTCGAGTCTAGAGGATCCCCCCTCTG Base pairs
AAATAGGTCGGGAGTGAGGAAGAGATCCGCGGCCTTAAGCAATTCAGCTCAGATCTCCTAGGGGGGAGAC 1611 to 1680

CTCTGGGTAAAGTTCATTGGAACAGAAAGAAATGGATTTATCTGCTCTTCGCGTTGAAGAAGTACAAAAT Base pairs
GAGACCCATTTCAAGTAACCTTGTCTTTCTTTACCTAAATAGACGAGAAGCGCAACTTCTTCATGTTTTA 1681 to 1750

GTCATTAATGCTATGCAGAAAATCTTAGAGTGTCCCATCTGTCTGGAGTTGATCAAGGAACCTGTCTCCA Base pairs
CAGTAATTACGATACGTCTTTTAGAATCTCACAGGGTAGACAGACCTCAACTAGTTCCTTGGACAGAGGT 1751 to 1820

CAAAGTGTGACCACATATTTTGCAAATTTTGCATGCTGAAACTTCTCAACCAGAAGAAAGGGCCTTCACA Base pairs
GTTTCACACTGGTGTATAAAACGTTTAAAACGTACGACTTTGAAGAGTTGGTCTTCTTTCCCGGAAGTGT 1821 to 1890

GTGTCCTTTATGTAAGAATGATATAACCAAAAGGAGCCTACAAGAAAGTACGAGATTTAGTCAACTTGTT Base pairs
CACAGGAAATACATTCTTACTATATTGGTTTTCCTCGGATGTTCTTTCATGCTCTAAATCAGTTGAACAA 1891 to 1960

GAAGAGCTATTGAAAATCATTTGTGCTTTTCAGCTTGACACAGGTTTGGAGTATGCAAACAGCTATAATT Base pairs
CTTCTCGATAACTTTTAGTAAACACGAAAAGTCGAACTGTGTCCAAACCTCATACGTTTGTCGATATTAA 1961 to 2030

TTGCAAAAAAGGAAAATAACTCTCCTGAACATCTAAAAGATGAAGTTTCTATCATCCAAAGTATGGGCTA Base pairs
AACGTTTTTTCCTTTTATTGAGAGGACTTGTAGATTTTCTACTTCAAAGATAGTAGGTTTCATACCCGAT 2031 to 2100

CAGAAACCGTGCCAAAAGACTTCTACAGAGTGAACCCGAAAATCCTTCCTTGCAGGAAACCAGTCTCAGT Base pairs
GTCTTTGGCACGGTTTTCTGAAGATGTCTCACTTGGGCTTTTAGGAAGGAACGTCCTTTGGTCAGAGTCA 2101 to 2170

GTCCAACTCTCTAACCTTGGAACTGTGAGAACTCTGAGGACAAAGCAGCGGATACAACCTCAAAAGACGT Base pairs
CAGGTTGAGAGATTGGAACCTTGACACTCTTGAGACTCCTGTTTCGTCGCCTATGTTGGAGTTTTCTGCA 2171 to 2240

CTGTCTACATTGAATTGGGATCTGATTCTTCTGAAGATACCGTTAATAAGGCAACTTATTGCAGTGTGGG Base pairs
GACAGATGTAACTTAACCCTAGACTAAGAAGACTTCTATGGCAATTATTCCGTTGAATAACGTCACACCC 2241 to 2310

AGATCAAGAATTGTTACAAATCACCCCTCAAGGAACCAGGGATGAAATCAGTTTGGATTCTGCAAAAAAG Base pairs
TCTAGTTCTTAACAATGTTTAGTGGGGAGTTCCTTGGTCCCTACTTTAGTCAAACCTAAGACGTTTTTTC 2311 to 2380

GCTGCTTGTGAATTTTCTGAGACGGATGTAACAAATACTGAACATCATCAACCCAGTAATAATGATTTGA Base pairs
CGACGAACACTTAAAAGACTCTGCCTACATTGTTTATGACTTGTAGTAGTTGGGTCATTATTACTAAACT 2381 to 2450

ACACCACTGAGAAGCGTGCAGCTGAGAGGCATCCAGAAAAGTATCAGGGTAGTTCTGTTTCAAACTTGCA Base pairs
TGTGGTGACTCTTCGCACGTCGACTCTCCGTAGGTCTTTTCATAGTCCCATCAAGACAAAGTTTGAACGT 2451 to 2520

TGTGGAGCCATGTGGCACAAATACTCATGCCAGCTCATTACAGCATGAGAACAGCAGTTTATTACTCACT Base pairs
ACACCTCGGTACACCGTGTTTATGAGTACGGTCGAGTAATGTCGTACTCTTGTCGTCAAATAATGAGTGA 2521 to 2590

AAAGACAGAATGAATGTAGAAAAGGCTGAATTCTGTAATAAAAGCAAACAGCCTGGCTTAGCAAGGAGCC Base pairs
TTTCTGTCTTACTTACATCTTTTCCGACTTAAGACATTATTTTCGTTTGTCGGACCGAATCGTTCCTCGG 2591 to 2660

AACATAACAGATGGGCTGGAAGTAAGGAAACATGTAATGATAGGCGGACTCCCAGCACAGAAAAAAAGGT Base pairs
TTGTATTGTCTACCCGACCTTCATTCCTTTGTACATTACTATCCGCCTGAGGGTCGTGTCTTTTTTTCCA 2661 to 2730

AGATCTGAATGCTGATCCCCTGTGTGAGAGAAAAGAATGGAATAAGCAGAAACTGCCATGCTCAGAGAAT Base pairs
TCTAGACTTACGACTAGGGGACACACTCTCTTTTCTTACCTTATTCGTCTTTGACGGTACGAGTCTCTTA 2731 to 2800

CCTAGAGATACTGAAGATGTTCCTTGGATAACACTAAATAGCAGCATTCAGAAAGTTAATGAGTGGTTTT Base pairs
GGATCTCTATGACTTCTACAAGGAACCTATTGTGATTTATCGTCGTAAGTCTTTCAATTACTCACCAAAA 2801 to 2870

CCAGAAGTGATGAACTGTTAGGTTCTGATGACTCACATGATGGGGAGTCTGAATCAAATGCCAAAGTAGC Base pairs
GGTCTTCACTACTTGACAATCCAAGACTACTGAGTGTACTACCCCTCAGACTTAGTTTACGGTTTCATCG 2871 to 2940

TGATGTATTGGACGTTCTAAATGAGGTAGATGAATATTCTGGTTCTTCAGAGAAAATAGACTTACTGGCC Base pairs
ACTACATAACCTGCAAGATTTACTCCATCTACTTATAAGACCAAGAAGTCTCTTTTATCTGAATGACCGG 2941 to 3010

AGTGATCCTCATGAGGCTTTAATATGTAAAAGTGAAAGAGTTCACTCCAAATCAGTAGAGAGTAATATTG Base pairs
TCACTAGGAGTACTCCGAAATTATACATTTTCACTTTCTCAAGTGAGGTTTAGTCATCTCTCATTATAAC 3011 to 3080

AAGACAAAATATTTGGGAAAACCTATCGGAAGAAGGCAAGCCTCCCCAACTTAAGCCATGTAACTGAAAA Base pairs
TTCTGTTTTATAAACCCTTTTGGATAGCCTTCTTCCGTTCGGAGGGGTTGAATTCGGTACATTGACTTTT 3081 to 3150

TCTAATTATAGGAGCATTTGTTACTGAGCCACAGATAATACAAGAGCGTCCCCTCACAAATAAATTAAAG Base pairs
AGATTAATATCCTCGTAAACAATGACTCGGTGTCTATTATGTTCTCGCAGGGGAGTGTTTATTTAATTTC 3151 to 3220

CGTAAAAGGAGACCTACATCAGGCCTTCATCCTGAGGATTTTATCAAGAAAGCAGATTTGGCAGTTCAAA Base pairs
GCATTTTCCTCTGGATGTAGTCCGGAAGTAGGACTCCTAAAATAGTTCTTTCGTCTAAACCGTCAAGTTT 3221 to 3290

AGACTCCTGAAATGATAAATCAGGGAACTAACCAAACGGAGCAGAATGGTCAAGTGATGAATATTACTAA Base pairs
TCTGAGGACTTTACTATTTAGTCCCTTGATTGGTTTGCCTCGTCTTACCAGTTCACTACTTATAATGATT 3291 to 3360

TAGTGGTCATGAGAATAAAACAAAAGGTGATTCTATTCAGAATGAGAAAAATCCTAACCCAATAGAATCA Base pairs
ATCACCAGTACTCTTATTTTGTTTTCCACTAAGATAAGTCTTACTCTTTTTAGGATTGGGTTATCTTAGT 3361 to 3430

CTCGAAAAAGAATCTGCTTTCAAAACGAAAGCTGAACCTATAAGCAGCAGTATAAGCAATATGGAACTCG Base pairs
GAGCTTTTTCTTAGACGAAAGTTTTGCTTTCGACTTGGATATTCGTCGTCATATTCGTTATACCTTGAGC 3431 to 3500

AATTAAATATCCACAATTCAAAAGCACCTAAAAAGAATAGGCTGAGGAGGAAGTCTTCTACCAGGCATAT Base pairs
TTAATTTATAGGTGTTAAGTTTTCGTGGATTTTTCTTATCCGACTCCTCCTTCAGAAGATGGTCCGTATA 3501 to 3570

TCATGCGCTTGAACTAGTAGTCAGTAGAAATCTAAGCCCACCTAATTGTACTGAATTGCAAATTGATAGT Base pairs
AGTACGCGAACTTGATCATCAGTCATCTTTAGATTCGGGTGGATTAACATGACTTAACGTTTAACTATCA 3571 to 3640

TGTTCTAGCAGTGAAGAGATAAAGAAAAAAAAAGTACAACCAAATGCCAGTCAGGCACAGCAGAAACCTAC Base pairs
ACAAGATCGTCACTTCTCTATTTCTTTTTTTTCATGTTGGTTTACGGTCAGTCCGTGTCGTCTTTGGATG 3641 to 3710

AACTCATGGAAGGTAAAGAACCTGCAACTGGAGCCAAGAAGAGTAACAAGCCAAATGAACAGACAAGTAA Base pairs
TTGAGTACCTTCCATTTCTTGGACGTTGACCTCGGTTCTTCTCATTGTTCGGTTTACTTGTCTGTTCATT 3711 to 3780

AAGACATGACAGCGATACTTTCCCAGAGCTGAAGTTAACAAATGCACCTGGTTCTTTTACTAAGTGTTCA Base pairs
TTCTGTACTGTCGCTATGAAAGGGTCTCGACTTCAATTGTTTACGTGGACCAAGAAAATGATTCACAAGT 3781 to 3850

AATACCAGTGAACTTAAAGAATTTGTCAATCCTAGCCTTCCAAGAGAAGAAAAAGAAGAGAAACTAGAAA Base pairs
TTATGGTCACTTGAATTTCTTAAACAGTTAGGATCGGAAGGTTCTCTTCTTTTTCTTCTCTTTGATCTTT 3851 to 3920

CAGTTAAAGTGTCTAATAATGCTGAAGACCCCAAAGATCTCATGTTAAGTGGAGAAAGGGTTTTGCAAAC Base pairs
GTCAATTTCACAGATTATTACGACTTCTGGGGTTTCTAGAGTACAATTCACCTCTTTCCCAAAACGTTTG 3921 to 3990

TGAAAGATCTGTAGAGAGTAGCAGTATTTCATTGGTACCTGGTACTGATTATGGCACTCAGGAAAGTATC Base pairs
ACTTTCTAGACATCTCTCATCGTCATAAAGTAACCATGGACCATGACTAATACCGTGAGTCCTTTCATAG 3991 to 4060

TCGTTACTGGAAGTTAGCACTCTAGGGAAGGCAAAAACAGAACCAAATAAATGTGTGAGTCAGTGTGCAG Base pairs
AGCAATGACCTTCAATCGTGAGATCCCTTCCGTTTTTGTCTTGGTTTATTTACACACTCAGTCACACGTC 4061 to 4130

CATTTGAAAACCCCAAGGGACTAATTCATGGTTGTTCCAAAGATAATAGAAATGACACAGAAGGCTTTAA Base pairs
GTAAACTTTTGGGGTTCCCTGATTAAGTACCAACAAGGTTTCTATTATCTTTACTGTGTCTTCCGAAATT 4131 to 4200

GTATCCATTGGGACATGAAGTTAACCACAGTCGGGAAACAAGCATAGAAATGGAAGAAAGTGAACTTGAT Base pairs
CATAGGTAACCCTGTACTTCAATTGGTGTCAGCCCTTTGTTCGTATCTTTACCTTCTTTCACTTGAACTA 4201 to 4270

GCTCAGTATTTGCAGAATACATTCAAGGTTTCAAAGCGCCAGTCATTTGCTCCGTTTTCAAATCCAGGAA Base pairs
CGAGTCATAAACGTCTTATGTAAGTTCCAAAGTTTCGCGGTCAGTAAACGAGGCAAAAGTTTAGGTCCTT 4271 to 4340

ATGCAGAAGAGGAATGTGCAACATTCTCTGCCCACTCTGGGTCCTTAAAGAAACAAAGTCCAAAAGTCAC Base pairs
TACGTCTTCTCCTTACACGTTGTAAGAGACGGGTGAGACCCAGGAATTTCTTTGTTTCAGGTTTTCAGTG 4341 to 4410

TTTTGAATGTGAACAAAAGGAAGAAAATCAAGGAAAGAATGAGTCTAATATCAAGCCTGTACAGACAGTT Base pairs
AAAACTTACACTTGTTTTCCTTCTTTTAGTTCCTTTCTTACTCAGATTATAGTTCGGACATGTCTGTCAA 4411 to 4480

AATATCACTGCAGGCTTTCCTGTGGTTGGTCAGAAAGATAAGCCAGTTGATAATGCCAAATGTAGTATCA Base pairs
TTATAGTGACGTCCGAAAGGACACCAACCAGTCTTTCTATTCGGTCAACTATTACGGTTTACATCATAGT 4481 to 4550

AAGGAGGCTCTAGGTTTTGTCTATCATCTCAGTTCAGAGGCAACGAAACTGGACTCATTACTCCAAATAA Base pairs
TTCCTCCGAGATCCAAAACAGATAGTAGAGTCAAGTCTCCGTTGCTTTGACCTGAGTAATGAGGTTTATT 4551 to 4620

ACATGGACTTTTACAAAACCCATATCGTATACCACCACTTTTTCCCATCAAGTCATTTGTTAAAACTAAA Base pairs
TGTACCTGAAAATGTTTTGGGTATAGCATATGGTGGTGAAAAAGGGTAGTTCAGTAAACAATTTTGATTT 4621 to 4690

TGTAAGAAAAATCTGCTAGAGGAAAACTTTGAGGAACATTCAATGTCACCTGAAAGAGAAATGGGAAATG Base pairs
ACATTCTTTTTAGACGATCTCCTTTTGAAACTCCTTGTAAGTTACAGTGGACTTTCTCTTTACCCTTTAC 4691 to 4760

AGAACATTCCAAGTACAGTGAGCACAATTAGCCGTAATAACATTAGAGAAAATGTTTTTAAAGAAGCCAG Base pairs
TCTTGTAAGGTTCATGTCACTCGTGTTAATCGGCATTATTGTAATCTCTTTTACAAAAATTTCTTCGGTC 4761 to 4830

CTCAAGCAATATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGGCTCCAGTATTAATGAAATAGGTTCC Base pairs
GAGTTCGTTATAATTACTTCATCCAAGGTCATGATTACTTCACCCGAGGTCATAATTACTTTATCCAAGG 4831 to 4900

AGTGATGAAAACATTCAAGCAGAACTAGGTAGAAACAGAGGGCCAAAATTGAATGCTATGCTTAGATTAG Base pairs
TCACTACTTTTGTAAGTTCGTCTTGATCCATCTTTGTCTCCCGGTTTTAACTTACGATACGAATCTAATC 4901 to 4970

GGGTTTTGCAACCTGAGGTCTATAAACAAAGTCTTCCTGGAAGTAATTGTAAGCATCCTGAAATAAAAAA Base pairs
CCCAAAACGTTGGACTCCAGATATTTGTTTCAGAAGGACCTTCATTAACATTCGTAGGACTTTATTTTTT 4971 to 5040

GCAAGAATATGAAGAAGTAGTTCAGACTGTTAATACAGATTTCTCTCCATATCTGATTTCAGATAACTTA Base pairs
CGTTCTTATACTTCTTCATCAAGTCTGACAATTATGTCTAAAGAGAGGTATAGACTAAAGTCTATTGAAT 5041 to 5110

GAACAGCCTATGGGAAGTAGTCATGCATCTCAGGTTTGTTCTGAGACACCTGATGACCTGTTAGATGATG Base pairs
CTTGTCGGATACCCTTCATCAGTACGTAGAGTCCAAACAAGACTCTGTGGACTACTGGACAATCTACTAC 5111 to 5180

GTGAAATAAAGGAAGATACTAGTTTTGCTGAAAATGACATTAAGGAAAGTTCTGCTGTTTTTAGCAAAAG Base pairs
CACTTTATTTCCTTCTATGATCAAAACGACTTTTACTGTAATTCCTTTCAAGACGACAAAAATCGTTTTC 5181 to 5250

CGTCCAGAAAGGAGAGCTTAGCAGGAGTCCTAGCCCTTTCACCCATACACATTTGGCTCAGGGTTACCGA Base pairs
GCAGGTCTTTCCTCTCGAATCGTCCTCAGGATCGGGAAAGTGGGTATGTGTAAACCGAGTCCCAATGGCT 5251 to 5320

AGAGGGGCCAAGAAATTAGAGTCCTCAGAAGAGAACTTATCTAGTGAGGATGAAGAGCTTCCCTGCTTCC Base pairs
TCTCCCCGGTTCTTTAATCTCAGGAGTCTTCTCTTGAATAGATCACTCCTACTTCTCGAAGGGACGAAGG 5321 to 5390

AACACTTGTTATTTGGTAAAGTAAACAATATACCTTCTCAGTCTACTAGGCATAGCACCGTTGCTACCGA Base pairs
TTGTGAACAATAAACCATTTCATTTGTTATATGGAAGAGTCAGATGATCCGTATCGTGGCAACGATGGCT 5391 to 5460

GTGTCTGTCTAAGAACACAGAGGAGAATTTATTATCATTGAAGAATAGCTTAAATGACTGCAGTAACCAG Base pairs
CACAGACAGATTCTTGTGTCTCCTCTTAAATAATAGTAACTTCTTATCGAATTTACTGACGTCATTGGTC 5461 to 5530

GTAATATTGGCAAAGGCATCTCAGGAACATCACCTTAGTGAGGAAACAAAATGTTCTGCTAGCTTGTTTT Base pairs
CATTATAACCGTTTCCGTAGAGTCCTTGTAGTGGAATCACTCCTTTGTTTTACAAGACGATCGAACAAAA 5531 to 5600

CTTCACAGTGCAGTGAATTGGAAGACTTGACTGCAAATACAAACACCCAGGATCCTTTCTTGATTGGTTC Base pairs
GAAGTGTCACGTCACTTAACCTTCTGAACTGACGTTTATGTTTGTGGGTCCTAGGAAAGAACTAACCAAG 5601 to 5670

TTCCAAACAAATGAGGCATCAGTCTGAAAGCCAGGGAGTTGGTCTGAGTGACAAGGAATTGGTTTCAGAT Base pairs
AAGGTTTGTTTACTCCGTAGTCAGACTTTCGGTCCCTCAACCAGACTCACTGTTCCTTAACCAAAGTCTA 5671 to 5740

GATGAAGAAAGAGGAACGGGCTTGGAAGAAAATAATCAAGAAGAGCAAAGCATGGATTCAAACTTAGGTG Base pairs
CTACTTCTTTCTCCTTGCCCGAACCTTCTTTTATTAGTTCTTCTCGTTTCGTACCTAAGTTTGAATCCAC 5741 to 5810

AAGCAGCATCTGGGTGTGAGAGTGAAACAAGCGTCTCTGAAGACTGCTCAGGGCTATCCTCTCAGAGTGA Base pairs
TTCGTCGTAGACCCACACTCTCACTTTGTTCGCAGAGACTTCTGACGAGTCCCGATAGGAGAGTCTCACT 5811 to 5880

CATTTTAACCACTCAGCAGAGGGATACCATGCAACATAACCTGATAAAGCTCCAGCAGGAAATGGCTGAA Base pairs
GTAAAATTGGTGAGTCGTCTCCCTATGGTACGTTGTATTGGACTATTTCGAGGTCGTCCTTTACCGACTT 5881 to 5950

CTAGAAGCTGTGTTAGAACAGCATGGGAGCCAGCCTTCTAACAGCTACCCTTCCATCATAAGTGACTCTT Base pairs
GATCTTCGACACAATCTTGTCGTACCCTCGGTCGGAAGATTGTCGATGGGAAGGTAGTATTCACTGAGAA 5951 to 6020

CTGCCCTTGAGGACCTGCGAAATCCAGAACAAAGCACATCAGAAAAAGCAGTATTAACTTCACAGAAAAG Base pairs
GACGGGAACTCCTGGACGCTTTAGGTCTTGTTTCGTGTAGTCTTTTTCGTCATAATTGAAGTGTCTTTTC 6021 to 6090

TAGTGAATACCCTATAAGCCAGAATCCAGAAGGCCTTTCTGCTGACAAGTTTGAGGTGTCTGCAGATAGT Base pairs
ATCACTTATGGGATATTCGGTCTTAGGTCTTCCGGAAAGACGACTGTTCAAACTCCACAGACGTCTATCA 6091 to 6160

TCTACCAGTAAAAATAAAGAACCAGGAGTGGAAAGGTCATCCCCTTCTAAATGCCCATCATTAGATGATA Base pairs
AGATGGTCATTTTTATTTCTTGGTCCTCACCTTTCCAGTAGGGGAAGATTTACGGGTAGTAATCTACTAT 6161 to 6230

GGTGGTACATGCACAGTTGCTCTGGGAGTCTTCAGAATAGAAACTACCCATCTCAAGAGGAGCTCATTAA Base pairs
CCACCATGTACGTGTCAACGAGACCCTCAGAAGTCTTATCTTTGATGGGTAGAGTTCTCCTCGAGTAATT 6231 to 6300

GGTTGTTGATGTGGAGGAGCAACAGCTGGAAGAGTCTGGGCCACACGATTTGACGGAAACATCTTACTTG Base pairs
CCAACAACTACACCTCCTCGTTGTCGACCTTCTCAGACCCGGTGTGCTAAACTGCCTTTGTAGAATGAAC 6301 to 6370

CCAAGGCAAGATCTAGAGGGAACCCCTTACCTGGAATCTGGAATCAGCCTCTTCTCTGATGACCCTGAAT Base pairs
GGTTCCGTTCTAGATCTCCCTTGGGGAATGGACCTTAGACCTTAGTCGGAGAAGAGACTACTGGGACTTA 6371 to 6440

CTGATCCTTCTGAAGACAGAGCCCCAGAGTCAGCTCGTGTTGGCAACATACCATCTTCAACCTCTGCATT Base pairs
GACTAGGAAGACTTCTGTCTCGGGGTCTCAGTCGAGCACAACCGTTGTATGGTAGAAGTTGGAGACGTAA 6441 to 6510

GAAAGTTCCCCAATTGAAAGTTGCAGAATCTGCCCAGAGTCCAGCTGCTGCTCATACTACTGATACTGCT Base pairs
CTTTCAAGGGGTTAACTTTCAACGTCTTAGACGGGTCTCAGGTCGACGACGAGTATGATGACTATGACGA 6511 to 6580

GGGTATAATGCAATGGAAGAAAGTGTGAGCAGGGAGAAGCCAGAATTGACAGCTTCAACAGAAAGGGTCA Base pairs
CCCATATTACGTTACCTTCTTTCACACTCGTCCCTCTTCGGTCTTAACTGTCGAAGTTGTCTTTCCCAGT 6581 to 6650

ACAAAAGAATGTCCATGGTGGTGTCTGGCCTGACCCCAGAAGAATTTATGCTCGTGTACAAGTTTGCCAG Base pairs
TGTTTTCTTACAGGTACCACCACAGACCGGACTGGGGTCTTCTTAAATACGAGCACATGTTCAAACGGTC 6651 to 6720

AAAACACCACATCACTTTAACTAATCTAATTACTGAAGAGACTACTCATGTTGTTATGAAAACAGATGCT Base pairs
TTTTGTGGTGTAGTGAAATTGATTAGATTAATGACTTCTCTGATGAGTACAACAATACTTTTGTCTACGA 6721 to 6790

GAGTTTGTGTGTGAACGGACACTGAAATATTTTCTAGGAATTGCGGGAGGAAAATGGGTAGTTAGCTATT Base pairs
CTCAAACACACACTTGCCTGTGACTTTATAAAAGATCCTTAACGCCCTCCTTTTACCCATCAATCGATAA 6791 to 6860

TCTGGGTGACCCAGTCTATTAAAGAAAGAAAAATGCTGAATGAGCATGATTTTGAAGTCAGAGGAGATGT Base pairs
AGACCCACTGGGTCAGATAATTTCTTTCTTTTTACGACTTACTCGTACTAAAACTTCAGTCTCCTCTACA

GGTCAATGGAAGAAACCACCAAGGTCCAAAGCGAGCAAGAGAATCCCAGGACAGAAAGATCTTCAGGGGG Base pairs
CCAGTTACCTTCTTTGGTGGTTCCAGGTTTCGCTCGTTCTCTTAGGGTCCTGTCTTTCTAGAAGTCCCCC 6931 to 7000

CTAGAAATCTGTTGCTATGGGCCCTTCACCAACATGCCCACAGATCAACTGGAATGGATGGTACAGCTGT Base pairs
GATCTTTAGACAACGATACCCGGGAAGTGGTTGTACGGGTGTCTAGTTGACCTTACCTACCATGTCGACA 7001 to 7070

GTGGTGCTTCTGTGGTGAAGGAGCTTTCATCATTCACCCTTGGCACAGGTGTCCACCCAATTGTGGTTGT Base pairs
CACCACGAAGACACCACTTCCTCGAAAGTAGTAAGTGGGAACCGTGTCCACAGGTGGGTTAACACCAACA 7071 to 7140

GCAGCCAGATGCCTGGACAGAGGACAATGGCTTCCATGCAATTGGGCAGATGTGTGAGGCACCTGTGGTG Base pairs
CGTCGGTCTACGGACCTGTCTCCTGTTACCGAAGGTACGTTAACCCGTCTACACACTCCGTGGACACCAC 7141 to 7210

ACCCGAGAGTGGGTGTTGGACAGTGTAGCACTCTACCAGTGCCAGGAGCTGGACACCTACCTGATACCCC Base pairs
TGGGCTCTCACCCACAACCTGTCACATCGTGAGATGGTCACGGTCCTCGACCTGTGGATGGACTATGGGG 7211 to 7280

AGATCCCCCACAGCCACTACTGATGTTAACGGGGATCCTCTAGAGTCGAGGATCCGGCTGTGGAATGTGT Base pairs
TCTAGGGGGTGTCGGTGATGACTACAATTGCCCCTAGGAGATCTCAGCTCCTAGGCCGACACCTTACACA 7281 to 7350

GTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTA Base pairs
CAGTCAATCCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAAT 7351 to 7420

GTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAAT Base pairs
CAGTCGTTGGTCCACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTA 7421 to 7490

TAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATT Base pairs
ATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAA 7491 to 7560

CTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATT Base pairs
GAGGCGGGGTACCGACTGATTAAAAAAAATAAATACGTCTCCGGCTCCGGCGGAGCCGGAGACTCGATAA 7561 to 7630

CCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTTGGGCTGCAGGTCGAGGCG Base pairs
GGTCTTCATCACTCCTCCGAAAAAACCTCCGGATCCGAAAACGTTTTTCGAACCCGACGTCCAGCTCCGC 7631 to 7700

GATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCC Base pairs
CTAGACTAGTTCTCTGTCCTACTCCTAGCAAAGCGTACTAACTTGTTCTACCTAACGTGCGTCCAAGAGG 7701 to 7770

GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCC Base pairs
CCGGCGAACCCACCTCTCCGATAAGCCGATACTGACCCGTGTTGTCTGTTAGCCGACGAGACTACGGCGG 7771 to 7840

GTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATG Base pairs
CACAAGGCCGACAGTCGCGTCCCCGCGGGCCAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTAC 7841 to 7910

AACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGA Base pairs
TTGACGTCCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCT 7911 to 7980

CGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCT Base pairs
GCAACAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTCACGGCCCCGTCCTAGAGGACAGTAGA 7981 to 8050

CACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGG Base pairs
GTGGAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTTACGCCGCCGACGTATGCGAACTAGGCC 8051 to 8120

CTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCT Base pairs
GATGGACGGGTAAGCTGGTGGTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCGGCCAGA 8121 to 8190

TGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAG Base pairs
ACAGCTAGTCCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGCGGTCGGCTTGACAAGCGGTCCGAGTTC 8191 to 8260

GCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGG Base pairs
CGCGCGTACGGGCTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACC 8261 to 8330

AAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGC Base pairs
TTTTACCGGCGAAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCCTGGCGATAGTCCTGTATCG 8331 to 8400

GTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGT Base pairs
CAACCGATGGGCACTATAACGACTTCTCGAACCGCCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCA 8401 to 8470

ATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCT Base pairs
TAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGA 8471 to 8540

GGGGTTCGATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGGAATGAAAGACCCCACCTGTAG Base pairs
CCCCAAGCTATTTTATTTTCTAAAATAAATCAGAGGTCTTTTTCCCCCCTTACTTTCTGGGGTGGACATC 8541 to 8610

GTTTGGCAAGCTAGCTTAAGTAACGCCATTTTGCAAGGCATGGAAAAATACATAACTGAGAATAGAGAAG Base pairs
CAAACCGTTCGATCGAATTCATTGCGGTAAAACGTTCCGTACCTTTTTATGTATTGACTCTTATCTCTTC 8611 to 8680

TTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCAAACAGGATATCTGTGGTAAGCAGTT Base pairs
AAGTCTAGTTCCAGTCCTTGTCTACCTTGTCGACTTATACCCGGTTTGTCCTATAGACACCATTCGTCAA 8681 to 8750

CCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACAGCTGAATATGGGCCAAACAGGATATCTGTGGTAAG Base pairs
GGACGGGGCCGAGTCCCGGTTCTTGTCTACCTTGTCGACTTATACCCGGTTTGTCCTATAGACACCATTC 8751 to 8820

CAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTCCCCAGATGCGGTCCAGCCCTCAGCAGTTTCTA Base pairs
GTCAAGGACGGGGCCGAGTCCCGGTTCTTGTCTACCAGGGGTCTACGCCAGGTCGGGAGTCGTCAAAGAT 8821 to 8890

GAGAACCATCAGATGTTTCCAGGGTGCCCCAAGGACCTGAAATGACCCTGTGCCTTATTTGAACTAACCA Base pairs
CTCTTGGTAGTCTACAAAGGTCCCACGGGGTTCCTGGACTTTACTGGGACACGGAATAAACTTGATTGGT 8891 to 8960

ATCAGTTCGCTTCTCGCTTCTGTTCGCGCGCTTCTGCTCCCCGAGCTCAATAAAAGAGCCCACAACCCCT Base pairs
TAGTCAAGCGAAGAGCGAAGACAAGCGCGCGAAGACGAGGGGCTCGAGTTATTTTCTCGGGTGTTGGGGA 8961 to 9030

CACTCGGGGCGCCAGTCCTCCGATTGACTGAGTCGCCCGGGTACCCGTGTATCCAATAAACCCTCTTGCA Base pairs
GTGAGCCCCGCGGTCAGGAGGCTAACTGACTCAGCGGGCCCATGGGCACATAGGTTATTTGGGAGAACGT 9031 to 9100

GTTGCATCCGACTTGTGGTCTCGCTGTTCCTTGGGAGGGTCTCCTCTGAGTGATTGACTACCCGTCAGCG Base pairs
CAACGTAGGCTGAACACCAGAGCGACAAGGAACCCTCCCAGAGGAGACTCACTAACTGATGGGCAGTCGC 9101 to 9170

GGGGTCTTTCATTTGGGGGCTCGTCCGGGATCGGGAGACCCCTGCCCAGGGACCACCGACCCACCACCGG Base pairs
CCCCAGAAAGTAAACCCCCGAGCAGGCCCTAGCCCTCTGGGGACGGGTCCCTGGTGGCTGGGTGGTGGCC 9171 to 9240

GAGGTAAGCTGGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAG Base pairs
CTCCATTCGACCGACGGAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCTC 9241 to 9310

ACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTG Base pairs
TGCCAGTGTCGAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAAC 9311 to 9380

GCGGGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGCG Base pairs
CGCCCACAGCCCCGCGTCGGTACTGGGTCAGTGCATCGCTATCGCCTCACATATGACCGAATTGATACGC 9381 to 9450

GCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAA Base pairs
CGTAGTCTCGTCTAACATGACTCTCACGTGGTATACGCCACACTTTATGGCGTGTCTACGCATTCCTCTT 9451 to 9520

AATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCG Base pairs
TTATGGCGTAGTCCGCGAGAAGGCGAAGGAGCGAGTGACTGAGCGACGCGAGCCAGCAAGCCGACGCCGC 9521 to 9590

AGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC Base pairs
TCGCCATAGTCGAGTGAGTTTCCGCCATTATGCCAATAGGTGTCTTAGTCCCCTATTGCGTCCTTTCTTG 9591 to 9660

ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGC Base pairs
TACACTCGTTTTCCGGTCGTTTTCCGGTCCTTGGCATTTTTCCGGCGCAACGACCGCAAAAAGGTATCCG 9661 to 9730

TCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATA Base pairs
AGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTCCTGATAT 9731 to 9800

AAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA Base pairs
TTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCT 9801 to 9870

TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTT Base pairs
ATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTATCGAGTGCGACATCCATAGAGTCAA 9871 to 9940

CGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT Base pairs
GCCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAA 9941 to 10010

ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGT Base pairs
TAGGCCATTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCA 10011 to 10080

AACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCT Base pairs
TTGTCCTAATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGA 10081 to 10150

ACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAG Base pairs
TGTGATCTTCCTGTCATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATC 10151 to 10220

CTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGC Base pairs
GAGAACTAGGCCGTTTGTTTGGTGGCGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCG 10221 to 10290

AGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACT Base pairs
TCTTTTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGA 10291 to 10360

CACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATG Base pairs
GTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTCCTAGAAGTGGATCTAGGAAAATTTAATTTTTAC 10361 to 10430

AAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAG Base pairs
TTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGACTGTCAATGGTTACGAATTAGTCACTC 10431 to 10500

GCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTA Base pairs
CGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATCTATTGAT 10501 to 10570

CGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCC Base pairs
GCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGGTGCGAGTGGCCGAGG 10571 to 10640

AGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCC Base pairs
TCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGAAATAGGCGG 10641 to 10710

TCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACG Base pairs
AGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGTTGC 10711 to 10780

TTGTTGCCATTGCTGCAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTC Base pairs
AACAACGGTAACGACGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAG 10781 to 10850

CCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCG Base pairs
GGTTGCTAGTTCCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGC 10851 to 10920

ATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTA Base pairs
TAGCAACAGTCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAAT 10921 to 10990

CTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTG Base pairs
GACAGTACGGTAGGCATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCAC 10991 to 11060

TATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTA Base pairs
ATACGCCGCTGGCTCAACGAGAACGGGCCGCAGTTGTGCCCTATTATGGCGCGGTGTATCGTCTTGAAAT 11061 to 11130

AAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA Base pairs
TTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACTCTAGGT 11131 to 11200

GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTG Base pairs
CAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCAC 11201 to 11270

AGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATA Base pairs
TCGTTTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTATGAGTAT 11271 to 11340

CTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAAT Base pairs
GAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACAGAGTACTCGCCTATGTATAAACTTA 11341 to 11410

GTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGA Base pairs
CATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAGATTCT 11411 to 11480

AACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAA Base pairs
TTGGTAATAATAGTACTGTAATTGGATATTTTTATCCGCATAGTGCTCCGGGAAAGCAGAAGTT 11481 to 11544

The enzymes cutting at least 1 times, and at most 6 times are listed below:

Enzyme	Position(s) in Recognition Site	Recognition Sequence
AatI	3243, 6123, 7660	agg/cct
AatII	985, 2240, 11474	gacgt/c
Acc113I	4861, 11032	agt/act
Acc16I	7966, 10774	tgc/gca
Acc65I	650, 4024, 9070	g/gtacc
AccB7I	2092, 7258	ccannnn/ntgg
AccBSI	579, 1285, 8480, 8534, 9595, 11396	gagcgg
AccI 2244	4648, 5432, 9431	gt/mkac
AclNI	900, 3583, 5198	a/ctagt
AfeI	1163	agc/gct
AflII	208, 1246, 3130, 8625	c/ttaag
AflIII	2690, 9659	a/crygt
AhdI	1260, 9115, 9161, 10552	gacnnn/nngtc
Alw44I	9475, 9973, 11219	g/tgcac
Ama87I	415, 646, 7212, 9000, 9033, 9066	c/ycgrg
AocI 1021	1258, 3252, 4983	cc/tnagg
Aor51HI	1163	agc/gct
ApaI	7023	gggcc/c
ApaLI	9475, 9973, 11219	g/tgcac
AscI	617	gg/cgcgcc
AseI	1755, 4842, 4884, 10724	at/taat
AsnI	1755, 4842, 4884, 10724	at/taat
Asp718I	650, 4024, 9070	yg/gtacc
AspEI	1260, 9115, 9161, 10552	gacnnn/nngtc
AvaI	415, 646, 7212, 9000, 9033, 9066	c/ycgrg
AviII	7966, 10774	tgc/gca
AvrII	7661	c/ctagg
BalI	832, 1351, 3008, 7946	tgg/cca
BamHI	1667, 5650, 7313, 7330	g/gatcc
BbeI	1641, 7867, 9042	ggcgc/c
BbuI	1855, 7409, 7481, 8269	gcatg/c
BcgI	11079	cgannnnnntgc
BclI	1800, 7705	t/gatca
BcoI	415, 646, 7212, 9000, 9033, 9066	c/ycgrg
BfrI	208, 1246, 3130, 8625	c/ttaag
BglI	3691, 7614, 10672	gccnnnn/nggc
BglII	2731, 3955, 3995, 6379, 6987	a/gatct
BlnI	7661	c/ctagg
BlpI	2647, 5267	gc/tnagc
BpmI	1798, 3743, 4881, 5935, 8578, 10642	ctggag
Bpu1102I	264, 5267	gc/tnagc
BsaAI	8168, 9413	yac/gtr
BsaBI	7723	gatnn/nnatc
BsaMI	2742, 2849, 4956	gaatgc
BsaOI	967, 7773, 9575, 9999, 10922, 11071	cgry/cg
BsaWI	138, 7895, 9865, 10012, 10843	w/ccggw
Bse118I	8183, 8364, 10632	r/ccggy
Bse21I	1021, 1258, 3252, 4983	cc/tnagg
Bse8I	7723	gatnn/nnatc
BsePI	566, 617, 8261, 8986	g/cgcgc
BsgI	2471, 4130, 5613, 7144	gtgcag
Bsh1285I	967, 7773, 9575, 9999, 10922, 11071	cgry/cg

Enzyme	Positions of Recognition Sites	Recognition Sequence
Bsh1365I	7723	gatnn/nnatc
BsiEI	967, 7773, 9575, 9999, 10922, 11071	cgry/cg
BsiI	6479, 6706, 8461, 9837, 11221, 11528	ctcgtg
BsmI	2742, 2849, 4956	gaatgc
BsoBI	415, 646, 7212, 9000, 9033, 9066	c/ycgrg
Bsp120I	7019	g/ggccc
Bsp1407I	1522, 4468 6705	t/gtaca
Bsp1720I	2647, 5267	gc/tnagc
Bsp19I	6663, 7568, 8296	c/catgg
BspCI	10922	cgat/cg
BspHI	3019, 3367, 10379, 11387, 11492	t/catga
BspLU11I	2690, 9659	a/catgt
BspMI	3735, 6038, 7693, 7765, 8128	acctgc
BspTI 208	1246, 3130, 8625	c/ttaag
BsrBI	579, 1285, 8480, 8534, 9595, 11396	gagcgg
BsrBRI	7723	gatnn/nnatc
BsrDI	6595, 8095, 10611, 10793	gcaatg
BsrFI	8183, 8364, 10632	r/ccggy
BsrGI	1522, 4468, 6705	t/gtaca
BssAI	8183, 8364, 10632	r/ccggy
BssHII	566, 617, 8261, 8986	g/cgcgc
BssSI	6479, 6706, 8461, 9837, 11221, 11528	ctcgtg
Bst1107I	4649, 9432	gta/tac
Bst98I	208, 1246, 3130, 8625	c/ttaag
BstD102I	579, 1285, 8480, 8534, 9595, 11396	gagcgg
BstDSI	151, 934, 6663, 7568, 8296	c/crygg
BstEII	1327, 5312, 6865, 7207	g/gtnacc
BstI	1667, 5650, 7313, 7330	g/gatcc
BstMCI 967	7773, 9575, 9999, 10922, 11071	cgry/cg
BstPI	1327, 5312, 6865, 7207	g/gtnacc
BstXI	1792, 5708, 7152, 7181	ccannnnn/ntgg
BstZI	964, 7770	c/ggccg
Bsu36I	1021, 1258, 3252, 4983	cc/tnagg
CelII	2647, 5267	gc/tnagc
Cfr10I	8183, 8364, 10632	r/ccggy
Cfr42I	154	ccgc/gg
Cfr9I	646, 9066	c/ccggg
CpoI	8380	cg/gwccg
CspI	8380	cg/gwccg
CvnI	1021, 1258, 3252, 4983	cc/tnagg
DraI	4819, 10418, 10437, 11129	ttt/aaa
DraIII	1824	cacnnn/gtg
DrdI	7891, 9354, 9767	gacnnnn/nngtc
DsaI	151, 934, 6663, 7568, 8296	c/crygg
EagI	964, 7770	c/ggccg
Eam1105I	1260, 9115, 9161, 10552	gacnnn/nngtc
Ecl136II	585, 6292, 9005	gag/ctc
EclHKI	1260 9115 9161 10552	gacnnn/nngtc
EclXI	964, 7770	c/ggccg
Eco147I	3243, 6123, 7660	agg/cct
Eco255I	4861, 11032	agt/act
Eco32I	311, 387, 8733, 8808	gat/atc
Eco47III	1163	agc/gct
Eco52I	964, 7770	c/ggccg
Eco81I	1021, 1258, 3252, 4983	cc/tnagg
Eco88I	415, 646, 7212, 9000, 9033, 9066	c/ycgrg
Eco91I	1327, 5312, 6865, 7207	g/gtnacc

Enzyme	Positions of Recognition Sites	Recognition Sequence
EcoICRI	585, 6292, 9005	gag/ctc
EcoNI	1024, 1632, 3237, 5567	cctnn/nnnagg
EcoO65I	1327, 5312, 6865, 7207	g/gtnacc
EcoRI	1644, 2618	g/aattc
EcoRV	311, 387, 8733, 8808	gat/atc
EcoT22I	5137, 7411, 7483	atgca/t
EheI	1639, 7865, 9040	ggc/gcc
Esp1396I	2092, 7258	ccannnn/ntgg
FauNDI	9482	ca/tatg
FbaI	1800, 7705	t/gatca
FspI	7966, 10774	tgc/gca
GsuI	1798, 3743, 4881, 5935, 8578, 10642	ctggag
HindII	1952, 3816, 4222, 6649, 7307, 11093	gty/rac
HindIII	7677	a/agctt
HpaI	3816, 4222, 7307	gtt/aac
KasI	1637, 7863, 9038	g/gcgcc
KpnI	654, 4028, 9074	ggtac/c
Ksp22I	1800, 7705	t/gatca
KspI	154	ccgc/gg
MamI	7723	gatnn/nnatc
MfeI	11, 21, 6521, 7128, 7179	c/aattg
MluNI	832, 1351, 3008, 7946	tgg/cca
MphI	5137, 741,1 7483	atgca/t
MroNI	8364	g/ccggc
MscI	832, 1351, 3008, 7946	tgg/cca
MspCI	208, 1246, 3130, 8625	c/ttaag
MunI	11, 21, 6521, 7128, 7179	c/aattg
Mva1269I	2742, 2849, 4956	gaatgc
NaeI	8366	gcc/ggc
NarI	1638, 7864, 9039	gg/cgcc
NcoI	6663, 7568, 8296	c/catgg
NdeI	9482	ca/tatg
NgoAIV	8364	g/ccggc
NgoMI	8364	g/ccggc
NheI	6, 16, 26, 203, 5588, 8620	g/ctagc
NsiI	5137, 7411, 7483	atgca/t
PaeI	1855, 7409, 7481, 8269	gcatg/c
PflMI	2092, 7258	ccannnn/ntgg
PleI9I	10922	cgat/cg
Pme55I	3243, 6123, 7660	agg/cct
Ppu10I	5133, 7407, 7479	a/tgcat
PpuMI	502, 1459, 4380, 6031, 8923	rg/gwccy
PshAI	751, 2240	gacnn/nngtc
PshBI	1755, 4842, 4884, 10724	at/taat
Psp124BI	587, 6294, 9007	gagct/c
Psp1406I	10778, 11151	aa/cgtt
Psp5II	502, 1459, 4380, 6031, 8923	rg/gwccy
PspAI	646, 9066	c/ccggg
PspALI	648, 9068	ccc/ggg
PspEI	1327, 5312, 6865, 7207	g/gtnacc
PspOMI	7019	g/ggccc
PstNHI	6, 16, 26, 203, 5588, 8620	g/ctagc
PvuI	10922	cgat/cg
RcaI	3019, 3367, 10379, 11387, 11492	t/catga
RsrII	8380	cg/gwccg
SacI	587, 6294, 9007	gagct/c
SacII	154	ccgc/gg

Enzyme	Positions of Recognition Sites	Recognition Sequence
ScaI	4861, 11032	agt/act
SexAI	1455, 3826, 4027, 5526, 7428	a/ccwggt
SfiI	7614	ggccnnnn/nggcc
Sfr303I	154	ccgc/gg
SmaI	648, 9068	ccc/ggg
SpeI	900, 3583, 5198	a/ctagt
SphI	1855, 7409, 7481, 8269	gcatg/c
SseBI	3243, 6123, 7660	agg/cct
SspBI	1522, 4468, 6705	t/gtaca
SstI	587, 6294, 9007	gagct/c
SstII	154	ccgc/gg
StuI	3243, 6123, 7660	agg/cct
Van91I	2092, 7258	ccannnn/ntgg
Vha464I	208, 1246, 3130, 8625	c/ttaag
VneI	9475, 9973, 11219	g/tgcac
VspI	1755, 4842, 4884, 10724	at/taat
XbaI	1661, 6382, 7319, 8887	t/ctaga
XcmI	7045	ccannnnn/nnnntgg
XmaI	646, 9066	c/ccggg
XmaIII	964, 7770	c/ggccg
Zsp2I	51, 37, 7411, 7483	atgca/t

APPENDIX 4.

The partial sequence of the pEGFP-BRCA1$_{(83-2436)}$ construct analyzed by Sequence Navigator computer program.