STRUCTURAL ANALYSIS OF A 50 KB REGION ON 17q23

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

H. BEGÜM AKMAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOLOGY

AUGUST 2007

Approval of the Thesis

INVESTIGATION OF MICRORNAS ON GENOMIC INSTABILITY REGIONS IN BREAST CANCER

Submitted by **H. BEGÜM AKMAN** in partial fulfilment of the requirements for the degree of **Master of Science in Biology Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Zeki Kaya _____
Head of the Department, **Biology**

Assist. Prof. Dr. A. Elif Erson _____
Supervisor, **Biology Dept., METU**

Examining Committee Members

Assoc. Prof. Dr. Sertaç ÖNDE _____
Dept. of Biological Sciences, METU

Assist. Prof. Dr. Ayşe Elif ERSON _____
Dept. of Biological Sciences, METU

Assist. Prof. Dr. Ayşel GÖZEN _____
Dept. of Biological Sciences, METU

Assist. Pro. Dr. Özlen KONU _____
Depr. of Molecular Biology and Genetic, Bilkent University

Dr. Sreeparna BANERJEE _____
Dept. of Biological Sciences, METU

Date: 28/08/2007

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : H. Begüm Akman

Signature :

ABSTRACT

STRUCTURAL ANALYSIS OF A 50 KB REGION ON 17q23

Akman, H. Begüm
M.Sc., Department of Biology
Supervisor: Assist. Prof. Dr. Ayşe Elif Erson

August 2007, 107 pages

17q23 amplicon is one of the many chromosomal regions that undergo amplification in breast tumors. Such amplicons harbor proto-oncogenes that may be overexpressed due to gene amplifications. Copy number analysis in breast cancer cell lines and breast tumors identified several independently amplified regions within the 17q23 amplicon, suggesting that a number of genes are selected for amplification as they may independently contribute to tumor formation and progression.

To characterize distinct amplicons on 17q23 and localize putative oncogenes, various studies are done on this region. In order to better understand the role of 17q23 amplification in breast cancer, characterizing unidentified genes or ESTs (*E*xpressed *S*equence *T*ags) on the 17q23 amplicon is crucial. In this study, *in silico* analysis of human (*H.sapiens*), chimpanzee (*P.troglodytes*), and

mouse (*M.musculus*) genomes were performed to examine sequence homology between these 3 species for the purpose of identifying novel genes.

The objective of this study was to analyze a 50 kb region between *TBX2* and *TBX4* genes and characterize the EST T02811 located on that region. Analysis and comparisons of these three genomes were established based on genomic sequences and availability of ESTs with gene prediction programs (BLAST, Pipmaker, Vista, GENSCAN etc.). Based on our results, we prepared a homology map between these 3 species, including positions of ESTs that may indicate novel genes. In this 50 kb region, we found *in silico* and experimantal evidence for the presence of an unidentified gene. We managed to extend the 313 bp EST T02811 size to 1423 bp which we think is as of yet incomplete.

Further studies are needed to characterize this novel gene as a potential oncogene candidate. Characterizing the roles of such candidate oncogenes in amplicons will provide a better understanding of genomic amplicon regions and their roles in tumorigenesis.

Keywords: Breast cancer, 17q23, EST, T02811, Human, Chimpanzee, Mouse

ÖZ

17q23 ÜZERİNDE BULUNAN 50 KB'LİK BÖLGENİN YAPISAL ANALİZİ

Akman, H. Begüm
Yüksek Lisans, Biyoloji Bölümü
Tez yöneticisi: Yrd. Doç. Dr. Ayşe Elif Erson

Ağustos 2007, 107 sayfa

17q23 amplikon bölgesi meme kanserinde amplifiye olan birçok bölgeden bir tanesidir. Bu tür amplikonlar gen amplifikasyonlarından dolayı aşırı ekspres edilen proto-onkogenleri barındırırlar. Meme kanseri hücre hatlarında ve meme tümörlerinde yapılmış olan kopya sayısı analizleri amplikon içerisinde birçok birbirinden bağımsız bölgenin amplifiye olduğunu belirlemiştir, bu da bazı genlerin çoğalmak üzere seçilerek tümör oluşumu ve gelişimine katkıda bulunduğunu göstermektedir.

17q23 üzerindeki amplikonları karakterize etmek ve olası onkogenleri belirlemek için çeşitli çalışmalar yapılmıştır. 17q23 amplikonunun çoğalmasının meme kanserinde rolünü anlamak amacıyla, bu amplikon üzerindeki belirlenmemiş genleri veya EST'leri (*E*xpressed *S*equence *T*ags-İfade edilen Dizi Etiketleri) karakterize etmek çok önemlidir. Bu çalışmada, insan (*H.sapiens*), şempanze (*P.troglodytes*) ve fare (*M.musculus*) genomlarının *in siliko* analizleri

diziler arsındaki homolijiyi araştırmak ve yeni genleri ortaya çıkarmak amacıyla yapılmıştır.

Bu çalışmanın amacı, *TBX2* ve *TBX4* genleri arasındaki 50 kb'lik bölgenin analizi edilmesi ve bu bölgede yeralan T02811 EST'sinin karakterizasyonudur. Bu üç memelinin genomlarının analizleri ve karşılaştırılması, 50 kb'lik genomik DNA dizisi ve bu bölgede yeralan EST'ler kullanılarak çeşitli biyoinformatik araçlar (BLAST, Pipmaker, Vista, GENSCAN vs.) ile yapılmıştır. Elde edilen sonuçlar göstermektedir ki, 3 tür arasında hazırlanan homoloji haritasına ve EST'lerin pozisyonlarına göre bu bölgede henüz tanımlanmamış çeşitli büyüklüklerde yeni genlerin olması mümkündür. Bu 50 kb'lik bölgede, henüz tanımlanmamış bir genin varlığına dair *in siliko* ve deneysel ipuçları belirledik. T02811 transkriptini 313 bp'den 1423 bp'ye uzatmayı başardık.

Bu yeni gen potensiyel bir onkogen adayı olup olmadığını bulmak için daha fazla karakterizasyon çalışmaları yapılmalıdır. Amplikonlarda bulunan onkogen adaylarının karakterize edilmeleri, genomik amplikon bölgelerinin ve bu bölgelerin tümör oluşumundaki rollerinin anlaşılması açısından önemlidir.

Anahtar Kelimeler: Meme kanseri, 17q23, EST, T02811, İnsan, Şempanze, Fare

To my grandmother and grandfather

ACKNOWLEDGMENTS

First of all, I would like to express my greatest thanks to my supervisor Assist. Prof. Dr. Ayşe Elif Erson for her guidance, valuable advices, support and insight throughout the research, I feel really lucky for having such a role model like her. I have always been grateful for the opportunity to study in her lab.

I would also like to thank the members of thesis examining committee; Assoc. Prof. Dr. Sertaç Önde, Assist. Prof. Dr. Ayşegül Gözen, Assist. Prof. Dr. Özlen Konu, and Dr. Sreeparna Banerjee for their suggestions and comments.

I would like to mention the members of Erson Lab, Ayşegül Sapmaz, Duygu Selçuklu, Shiva Akhavan, Emre Öktem, and Serkan Tuna, and thank them all for their support during the times we spent together. There is no way I can appreciate the overall support of my grasshopper Kevser Gençalp, she's been a great partner for me throughout the experiments. I should also mention Bilsev Koyuncu which is a lovely and caring friend; she's been with me during the hard times as well as the cheerful times. I would like to thank to the METU Plant Biotechnology Lab members for their help and sharing throughout the experiments.

Last but certainly not least, I would like to acknowledge my family. My mother Gülgün Barın had always supported me and be there for me throughout my life. I couldn't express my thankfulness enough to her. My father Cumhur Akman was the one who encouraged me to be the person whom I am now. My brother Berk Akman has been the one who cheer me up all the tough times.

Ultimately, my greatest and deepest gratitude go to my love Taner for his incredible and unbelievable support during the study. He was the one who made me believe to chase my dreams. I feel extremely lucky for sharing my life with him, his presence and everything.

This work is supported by the research fund: METU BAP-2006-07-02-00-01.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ATP	Adenosine Triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Base Pairs
cDNA	Complementary Deoxyribonucleic Acid
CGH	Comparative Genomic Hybridization
CIP	Calf Intestinal Phosphatase
CNS	Conserved Noncoding Sequence
DEPC	Diethyl pyrocarbonate
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DNase I	Deoxyribonuclease I
dNTP	Deoxyribonucleotide Triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediamine Tetra Acetic Acid
ePCR	Electronic Polymerase Chain Reaction
EST	Expressed Sequence Tag
FASTA	Fast Aye
FISH	Fluorescence in situ Hybridization
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GC	Guanine Cytosine
GI	GeneInfo Identifier
HBSS	Hank's Balanced Salt Solution
HMG	Human Mammary Gland
LB	Lysogeny Broth
MEM	Minimum Essential Medium
NCBI	National Center for Biotechnology Information
LINE	Long Interspersed Nuclear Element

LOH	Loss of Heterozygosity
OD	Optical Density
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PDF	Portable Document Format
pfu	*Pyrococcus furiosus*
PIP	Percent Identity Plot
RACE	Rapid Amplification of cDNA ends
RNA	Ribonucleic Acid
RNase	Ribonuclease
RR-cDNA	RACE Ready cDNA
RT-PCR	Reverse Transcription Polymerase Chain Reaction
rpm	Revolution per Minute
Taq	*Thermus aquaticus*
TAP	Tobacco Acid Pyrophosphatase
TBE	Tris-Boric acid-EDTA
TE	Tris-EDTA
TNE	Tris-NaCl-EDTA
SINE	Short Interspersed Nuclear Element
SOC	Super Optimal Catabolite Repressed Broth
STS	Sequence Tag Site
UniSTS	STS database
UCSC	University of California, Santa Cruz
UTR	Untranslated Region

CHAPTER I

INTRODUCTION

1.1. Breast Cancer and Genomic Instability

The odds of developing breast cancer at some time during a woman's life is about 1 in 8 in the developed world [1]. It is the second leading cause of cancer death in women worldwide, exceeded only by lung cancer. Unfortunately, 1 in 33 (3%) of women's death is reported to be caused by breast cancer [1]. Moreover, global mortality statistics suggest that breast cancer remains as one of the leading causes of cancer deaths among women [2].

Breast cancer is a genetic disease and results from a series of complex genetic abnormalities. Amplifications and/or deletions of certain chromosomal areas, translocations and gene mutations are somatic changes in the genome of breast cancer cells. Therefore, a complex and heterogeneous set of genetic alterations is implicated in the etiology of breast cancer [3] [4]. Amplification of oncogenes (*MYC*, *ERBB2* and *CCND1*), mutations of the tumor suppressor genes (*TP53* and *CDH1*) and loss of heterozygosity (LOH) at chromosome 1, 3p, 6q, 7q, 8p, 10q, 13q, 16q, 17, 18q, 22q, and X are only some of the complex genetic abnormalities found in breast tumors [5].

Chromosome 17 is one of the smallest and gene-rich human chromosomes [6]. In human cancers, chromosome 17 is frequently rearranged and several rearrangement breakpoints are shown to map to either the short arm or the long arm of this chromosome [7]. Besides, chromosome 17 is reported to be harboring

manifold regions of gains or losses in a diversity of other tumors [8]. According to the data coming from CGH (*C*omparative *G*enomic *H*ybridization), LOH, and molecular genetics studies; rearrangement of chromosome 17 is taking place in 30% of breast tumors [9] [10]. CGH studies done on 30 breast cancer cell lines and 22 primary tumors have also reported that chromosome 17p undergoes only losses, while 17q suffers more complicated gains and losses or combination of them [11].

Chromosome arm 17q is frequently rearranged. 17q11.2-q21.1 harboring *HER-2/neu* oncogene is amplified in around 25-30% of breast cancer cases [12] [13]. In addition, LOH has been reported at 17q at various loci [14] [15]. The tumor suppressor gene *BRCA1* which is located on 17q21 was also reported to show LOH [16]. *FRA17B* at 17q23.1 and *RNU2* at 17q21-q22 are the fragile sites which have been found so far [17] [18]. 12 different segments were reported as; 17q12 and five different segments from 17q22-25 showing mainly gains, 17q11.2, 17q21, and 17q24 segments showing mainly losses, 17q21.3, 17q22, and 17q25 segments involving in both gains and losses [11].

Multiple regions of chromosomal amplifications are very often in breast carcinomas [19] [20]. Up to now, over 20 different amplicons have been identified in breast cancer by using CGH technique, with each amplicon varying in size, location, gene density, and frequency of occurrence [21].

In breast cancer, at least two major regions are showing frequent amplification at chromosome 17. One of those major regions is 17q23 and the other is 17q12-21 [22]. According to CGH analysis the 17q23 region shows gain approximately in 20% of primary breast carcinomas, as well as in several breast cancer cell lines [23] [24] [25] [26]. Also, 17q23 amplicon appears to be separate from 17q12-21 amplicon. The presence of at least two separate and largely independent amplicons on 17q which was assessed by expression analysis was further confirmed at the DNA amplification level by FISH (Fluorescent in situ

Hybridization) [22]. Among these regions, 17q23 amplification seems to be a common phenomenon in breast tumors.

Table 1.1 Chromosomal regions involved in DNA copy number increase in 15 breast cancer cell lines [21] (Figure was taken from Kallioniemi et al.; 1994).

Cell line	Regional copy-number increases	Whole-arm gains
BT-20	4q32–q34, 6q21–q22	5p, 7p, 10p, 16q, 18p, 20
BT-474	17q12, 17q22–q24, **20q13**	
BT-483	12q24	1q, 8q, 19q, 20q
MCF-7	1cen–q32, **3p14**, 8q21–qter, 15q21–qter, 16q23–q24, **17q22–q24, 20q13**	5p, 12q, 14q
MDA-157	1q32–qter, 5q32–qter, **13q31–qter**, 14q24–qter, 17q22–qter, **19q13.1**, 19q13.4, **20q13**	2p, 7q, 8q, Xp
MDA-175	11q13	1q, 8q, 20q
MDA-231		4p, 6p, 11q, 19q
MDA-330	3q26–qter, 5q31–qter, 7q22–q32, 8q21–q23, 11p15, 11q13, **17q12**, 17q21–qter, 20q13.2–qter	5p, 10p, 14q
MDA-361	6cen–q21, **12q21.3–q23**, 12q24, **17q22–qter**, 19q13.3–q13.4, 20q13	5p, 8q, 12p, 16
MDA-435	6q12–q13	3p14–qter, 8q, 20q
MDA-436	3p22–pter, 5q31–qter, **8q22–qter**, 14q31–qter, 17q22–qter, 20q13	1q, 5p, 16q, 21, 22
MDA-453	1q31–qter, 3q26–qter, 17q22–q24	8q, 14q, 20, 22
SK-BR-3	3p22–pter, **8q21, 8q23–q24.1**, 10cen–q21, 13q22–qter, 14q31, 17q12, 17q24–qter, 20q	1q, 7pter–q31, 16p
ZR-75-1	11q13, 12q14–q15, 17q22–qter	1q, 7p, 12p, 16p, 20q, 22
ZR-75-30	**8q23–qter, 17cen–q24**	1q, 5p, 20p

1.2. 17q23 Amplicon

Amplification of the 17q23 chromosomal region was first discovered in breast cancer [21] even though 17q23 region gain and amplification is not only specific to breast cancer. While amplification of this chromosomal region is exclusively high in breast cancer, it was indicated that increased 17q23 copy number is also common in brain, lung, ovary, urinary bladder and soft tissue tumors [27]. CGH and FISH studies have revealed gain and amplification of part or all of the region in cancers of the lung [28], ovary [29], pancreas [30], bladder [31 32], liver [33], esophagus [34], stomach [35], and uterus [29 36 37]. Besides, gain and amplification have also been detected in neuroblastomas [38] and meningiomas [39 40]. A wide-ranging study of 17q23 copy number in 3520 tumors from 166 categories, has reported that gain of 17q23 was detected in many tumors, but high level amplification was limited only to breast tumors[27]. Still, it remains possible that 17q23 region may be selected for amplification in other tumor types. So, further studies of the amplicon in other tumors are necessary.

4429 tumor samples representing 166 different tumor categories and 359 normal tissue samples from 40 different tissue categories were examined by FISH on tissue microarrays [27]. It was reported that most frequently effected tissues are lung, mammary gland, and soft tissues. 35% of lung, 21% of the soft tissue, 30% of mammary tumor specimens showed increased copy number. 17q23 was found to be increased in copy number at 11% of primary tumors and 33% of metastatic cases. High-level amplification of 17q23 was reported in 2% of specimens including mammary gland, lung, adrenal gland, ovary, skin, soft tissue, stomach, thyroid gland, urinary bladder, and uterus [27]. In Figure 1.1 the frequencies of increased 17q23 copy number in different tumor categories was shown.

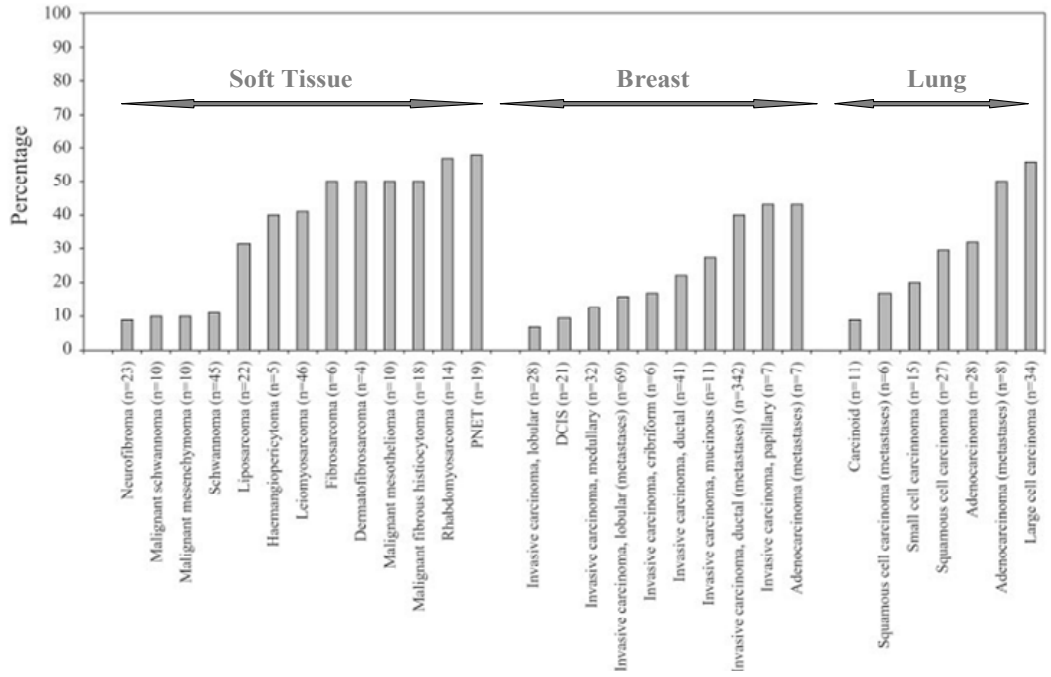

Figure 1.1 The frequency of increased 17q23 copy number in different tumor categories originating from lung, mammary gland, and soft tissue [27]. (Figure was taken from Andersen *et al.;* 2002)

Boundaries of the 17q23 amplicon has been defined previously by several studies [41] [42] [23] [26] [43]. The first study of this type involving FISH analysis of breast cancer cell lines detected three independent peaks of amplification on chromosome 17q in the MCF7 cell line [42]. Levels of amplification in the peaks ranged from 2- to 15- fold. This study offered the first evidence suggesting that 17q22-24 region harbored a complex amplicon. Later on, known STS (*S*equence *T*agged *S*ite) probes were used from the region for Southern blot analysis of breast cancer cell lines [23]. Meanwhile completion of the Human Genome Project [44] [45] provided a higher resolution map of the region, which helped researchers to delineate the size and boundaries of the 17q23 amplicon. Finally 17q23 amplicon was reported to be discontinuous [41] and that it was a very gene-rich region containing nearly 50 genes, 29 of which with known functions [46].

In order to confirm the complex structure of the 17q23 amplicon, physical map of the region was created [47]. In this physical map of the 17q23 region nearly all the gaps were filled and the positions of the known and predicted genes were shown. A combination of STS mapping and sequence analyses were used to assemble this physical map. The physical and transcription map of 17q23 is shown in Figure 1.2.

This map differs from the more updated physical map of the region available in NCBI MapViewer [48] and UCSC Genome Browser [49] databases. There are some recently identified genes which were not available in that map. MapViewer and Genome Browser physical maps heavily depends on sequence analysis, therefore it is possible that the inconsistencies resulted from the upgraded versions of human genome sequences and the new research data.

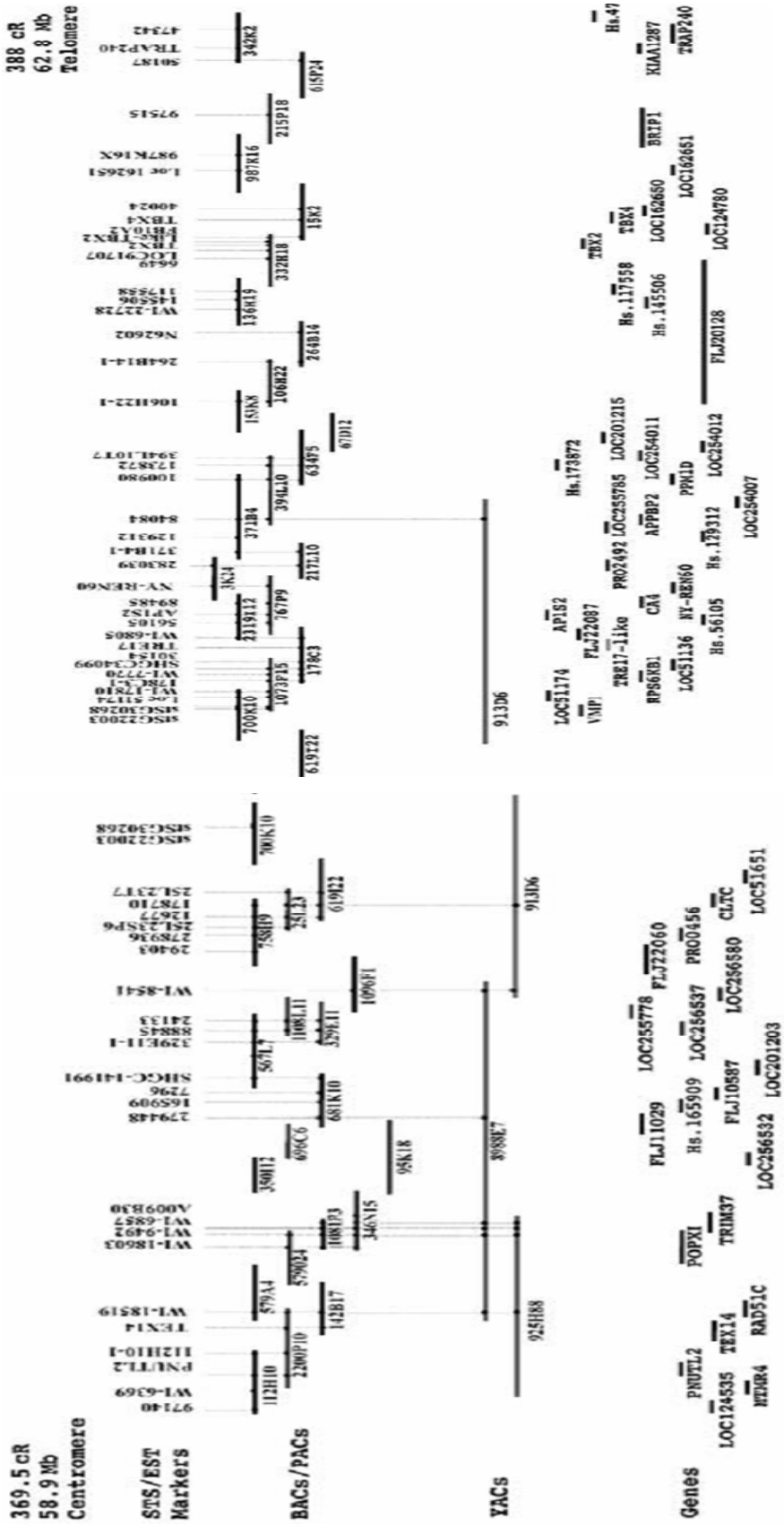

Figure 1.2 Physical and transcription map of the 17q23 amplicon (Figure taken from Sinclair *et al*; 2003).

Finally, an accurate and complete physical map of the 17q23 amplicon is reported recently [46]. The map covers a 5.5 Mb region between 53.5 Mb- 59.00 Mb. The known genes in that region are symbolized with horizontal lines and orientations of the genes are indicated with arrowheads. The accurate physical map of the 17q23 was shown in Figure 1.3.

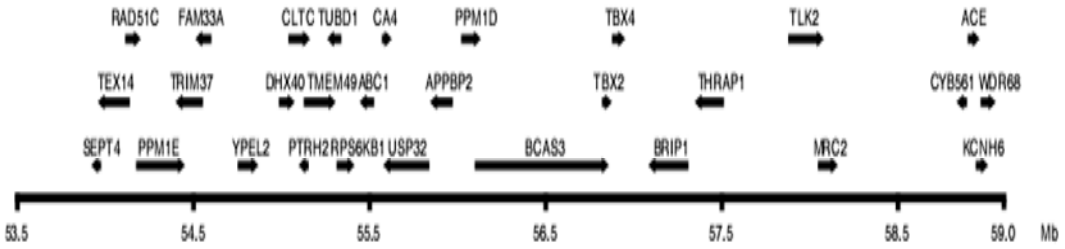

Figure 1.3 Physical map of the 17q23 amplicon (Figure taken from Parssinen *et al.*; 2007).

1.2.1. 17q23 and Breast Cancer

Several studies have attempted to correlate breast cancer progression with amplification of 17q23 chromosomal region. By looking at the correlation between amplification and high-level mRNA expression, these studies aimed to uncover the possible target genes for the amplification of 17q23 chromosomal region [23 50 41]. The initial observation of regional gain of 17q23 in human tumors came from a CGH analysis that detected amplification of 17q23 in 18% of primary breast tumors [21]. Afterwards, a number of studies analyzed gain of 17q and amplification of 17q23 in various groups of breast tumors. In an early study, 33 formalin-fixed, paraffin-embedded breast tumors were showed gain of 17q23 in 1 of 10 diploid tumors and 4 of 10 aneuploid tumors [51]. Besides, CGH of primary tumors and metastatic lesions identified gain of 17q23 in 9 of 29 primary

tumors (31%) and 12 of 29 matched metastases (41%) [52]. In another study, gain and amplification of 17q22-q24 was detected in more than 50% of the *BRCA1* associated tumors and 87% of the *BRCA2* associated tumors compared to 18% in the sporadic controls [53].

Two studies focusing on 17q23 amplicon identified a possible association between clinical outcome of breast cancer patients and genetic aberrations. In one of these studies, 48 primary breast tumors from node-negative patients were screened by CGH [54]. The tumors are classified into those with disease recurrence within 5 years and those disease-free after 5 years. Gain of the 17q22-q24 region was 10% and the total number of chromosomal aberrations was higher in the group with disease recurrence. In a similar study, node-negative breast tumors found to be gaining 17q in tumors with poor prognostic features [55]. According to these studies, it is suggested that gain of 17q along with other gains and losses present in the tumor cells is related with tumor progression and aggressive clinical behaviour.

1.2.2. Amplified and Overexpressed Genes From the 17q23 Amplicon

A number of putative oncogene candidates have been proposed for the 17q23 amplicon; *RPS6KB1*, *PAT1* (*APPBP2*), *RAD51C*, *TBX2*, *TRIM37* (*MUL*), *THRAP1* (*TRAP240*), *PPMID*, *BRIP1*, *FAM33A*, *DHX40*, *CLTC*, *PTRH2*, *TMEM49*, *TUBD1*, *ABC1* and *USP32* as possible oncogenes [41 23 24 25 26 56 57 46].

Unfortunately, most of the studies were carried out at a time when the Human Genome Project was not completed, so genome sequence of this region was incomplete, thus some of the genes within this region had not been identified. Genomic contigs were not fully known and orientation of some contigs weren't defined. After the Human Genome Project was finished, further studies by using FISH and quantitative real-time RT-PCR analyzed 29 genes from that region in detail [46].

RPS6KB1 is a mediator involved in G1-S progression of the cell cycle [58]. This gene encodes a member of the RSK (ribosomal S6 kinase) family of serine/threonine kinases. This kinase contains 2 non-identical kinase catalytic domains and phosphorylates several residues of the S6 ribosomal protein. The kinase activity of this protein leads to an increase in protein synthesis and cell proliferation [59]. *RPS6KB1* is regulating a wide array of cellular processes involved in mitogenic response such as protein synthesis, translation of specific mRNA species, and cell cycle progression G1 to S phase [60]. *RPS6KB1* has been reported to be amplified and overexpressed more than 2.5 times in the MCF7 breast cancer cell line [41]. Also cDNA microarray analysis based studies further confirmed this gene as a putative target for the 17q23 amplification [26].

PAT1 (*APPBP2*) is a cytoplasmic protein involving in cellular trafficking of amyloid precursor protein [61]. It is a microtubule-interacting protein which is recognizing the basolateral sorting signal of amyloid precursor protein. *PAT1* is also confirmed by cDNA microarray analysis as a putative oncogene [26].

TBX2 acts as a potent immortalizing gene by downregulating $p19^{ARF}$ as reported previously [62]. The murine tumor suppressor $p19^{ARF}$ is reported to fulfill an important protective role in preventing primary cells from oncogenic transformation via its action in the p53 pathway. *TBX2* was shown to be amplified and overexpressed in breast tumors [41].

MUL (*TRIM37*) encodes a RING-B-box-Coiled-coil (RBCC) tripartite motif-containing peroxisomal protein which is a member of the zinc finger protein family [63] [64]. This gene has been reported to be amplified and overexpressed in breast cancer cell lines [41] [26]. Members of this zinc finger protein family have a variety of functions such as regulation of development and cell proliferation [65] [66]. *MDM2* and *BRCA1* were found to have roles in cancer and they are members of the same gene family. These make *MUL* a very interesting putative target gene for the 17q23 amplification [26].

TRAP240 (*THRAP1*) is a member of a large multisubunit complex of thyroid hormone receptor-associated proteins [67]. It acts like a ligand, interacting with thyroid receptor and activate transcription of genes involved in different biological processes.

In another study, as can be seen from Figure 1.4 amplification and overexpression of genes and ESTs (*E*xpressed *S*equence *T*ag) found on 17q23 amplicon have been detected by using semi-quantitative duplex PCR and semi-quantitative duplex RT-PCR by using 24 breast cancer cell lines and a primary breast tumor sample [41].

Recently, an amplicon core is proposed according to the expression levels determined by quantitative RT-PCR [46]. According to this study, expression screening of 24 known genes from the 17q23 amplicon were done by qRT-PCR by using RNA samples from 26 primary breast tumors and a normal human mammary gland (HMG). This study revealed 11 genes showing significantly higher expression levels in primary breast tumors with high level 17q23 amplification compared to tumors without amplification. These 11 genes (*FAM33A*, *DHX40*, *CLTC*, *PTRH2*, *TMEM49*, *TUBD1*, *RPS6B1*, *ABC1*, *USP32*, *APPBP2*, *PPM1D*) together were called the amplicon core as can be seen in the Figure 1.5.

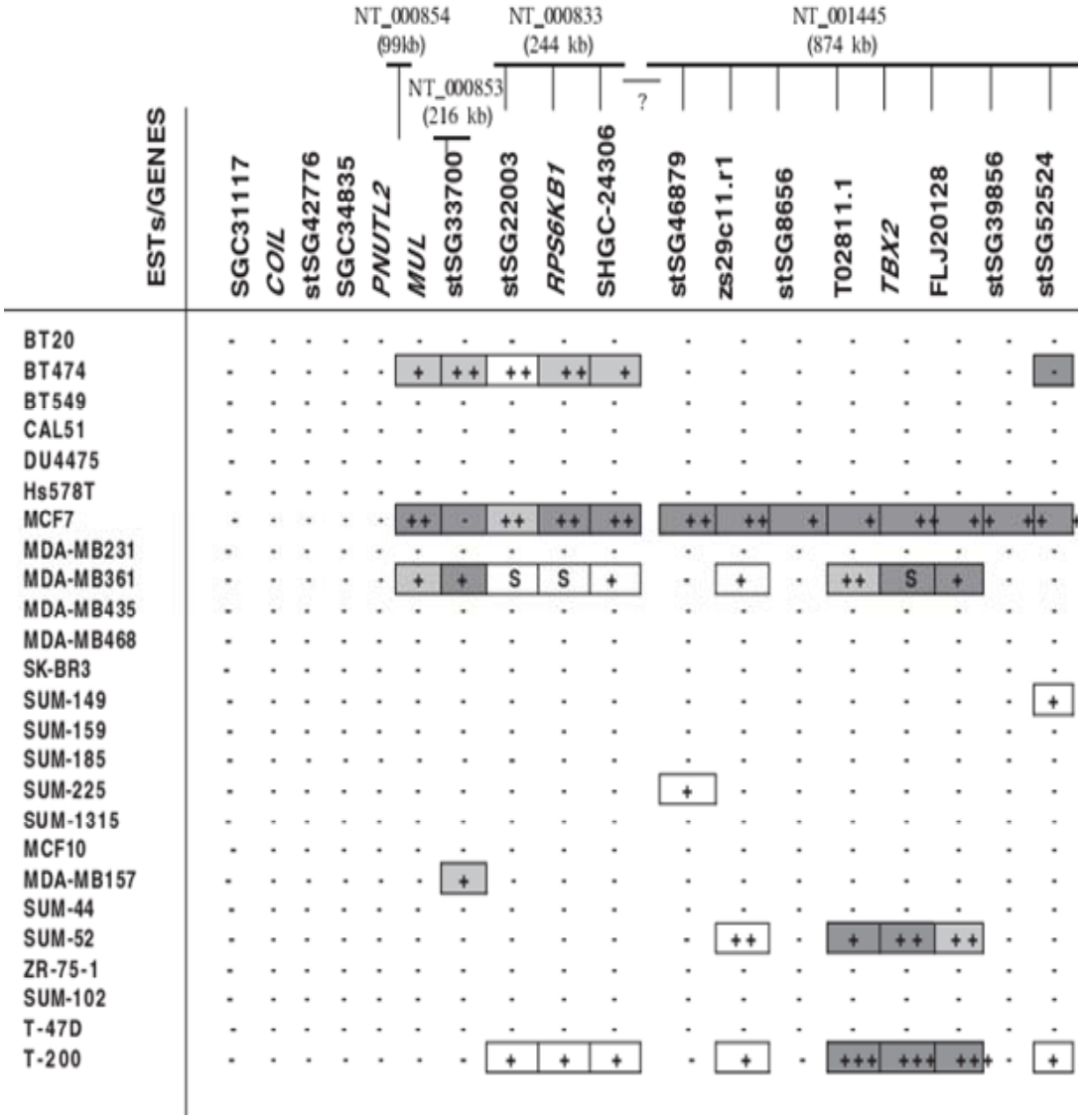

Figure 1.4 Amplified and overexpressed genes and ESTs from 17q23 amplicon [41] (Figure taken from Erson *et al.*; 2001).

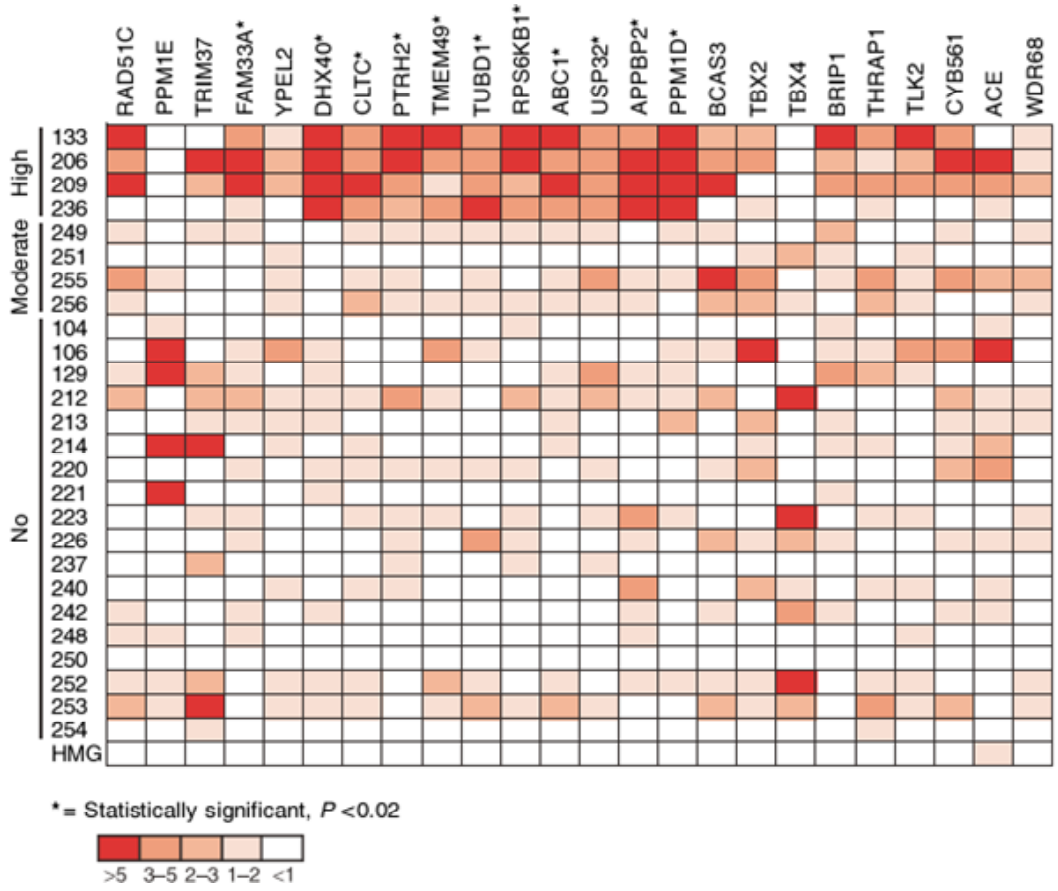

Figure 1.5 Expression levels of 24 genes from 17q23 amplicon in 26 primary breast tumors and normal HMG [46] (Figure taken from Parssinen *et al.*; 2007).

According to the prior studies, multiple genes from 17q23 amplicon were believed to have important roles for cancer initiation and progression; thus, making them promising targets for diagnostic, prognostic and therapeutic approaches.

ESTs are short, approximately 200-800 nucleotide long, unedited, randomly selected single-pass sequence reads derived from cDNA libraries [68]. They can be generated either from 5' or 3' end of cDNA clones. ESTs have multiple applications, they were first used to construct maps of the human genome [69], then they preferred to be used in gene structure predictions [70] [71], to explore alternative

splicing [72] [73] [74] [75], to distinguish genes exhibiting tissue or disease-specific expression [76], and for the discovery and characterization of genes and candidate SNPs [77]. dbEST is the EST database which was specifically created to meet the requirements to reach EST data and submission of EST sequences [78].

Discovery of genes involving in cancer initiation and progression is a crucial work. Exploiting ESTs for identifying potential oncogenes on 17q23 by using molecular genetic tools is very critical because this amplicon harbors genes which seem to be responsible for tumor initiation and progression. When the copy number of 17q23 amplicon increases, multiple adjacent genes are overexpressed [46]. However, there is no strong evidence that one of these genes from 17q23 is possessing oncogenic properties solely, it is believed that these genes are responsible from breast cancer development collectively. Although 29 genes were identified from that region, there remain other unidentified genes which are waiting to be discovered on 17q23.

1.2.3. T02811

T02811 is an EST which was generated by single pass sequencing [79]. The submitted sequence for this EST is 313 bp long for now and it was derived by oligo dT priming from 3' of human fetal brain cDNAs. T02811 is also known as FB10A2, as an STS (*S*equence Tag *S*ite) covering the same sequence.

1.2.3.1. Expression Profile of T02811

Via performing semiquantitative duplex PCR and Southern Blot analysis it was found to be amplified and overexpressed in a primary breast tumor sample and also in MCF7, MDA-MB361 and SUM-52 breast cancer cell lines [41]. ESTs, as indicators of novel genes, are potential oncogene candidates when amplification and overexpression occurs. T02811 was revealed to be amplified more than 20 times in a primary breast tumor sample, 2.5 times in MCF7 and

SUM-52 cell lines, and 3.5 times in MDA-MB361 cell line [41]. Expression of T02811 in primary breast tumor sample and SUM-52 cell line is 5.5-fold, it is 2.9-fold in MCF7 cell line and 1.2-fold in MDA-MB361 cell line [41] (Figure 1.4).

Expression profile of T02811 was also evaluated in another study by using FISH on tissue microarrays [26]. According to the outcomes of that study in MCF7 cell line, copy number of T02811 is more than 20-fold and it is nearly 20-fold in BT-474 breast cancer cell line. In another study considering structural analysis of 17q23, Southern blot analysis revealed that the copy number of T02811 is approximately 23-fold in MCF7 and 5-fold in BT474 breast cancer cell lines [43]. All these studies point out that T02811 is amplified and overexpressed in breast cancer. Gene amplification is one of the major mechanisms that allow cancer cells to promote expression of genes that are involved in tumor development and progression, therefore; characterization of T02811 which is amplified and overexpressed may be important in terms of gene identification purposes as well as understanding its potential contribution to breast tumorigenesis.

T02811 is also found in the genome of *Pan troglodytes* (chimpanzee) genome with respect to UniSTS database of NCBI [48]. Orthologous genes in human and chimpanzee are extremely similar according to the comparison of human and chimpanzee genomes [80]. This provides information that the possible gene encoded by this EST is conserved in primates.

Although there is no evidence that T02811 found in *Mus musculus* (mouse) genome, there exists a probability that the gene encoded by that EST located on genome of the mouse as well. Since chromosome 17 of human is syntenic to the chromosome 11 of mouse, there are orthologous genes present [6]. Conserved synteny between human chromosome 17 and mouse chromosome 11 is shown in Figure 1.6. In the figure, approximate alignments of ideograms of two chromosomes are shown in the center, with red lines indicating the orthologous

genes. Gene density, G+C content and densities of LINEs (red) and SINEs (blue) are specified from top to bottom for each organism.

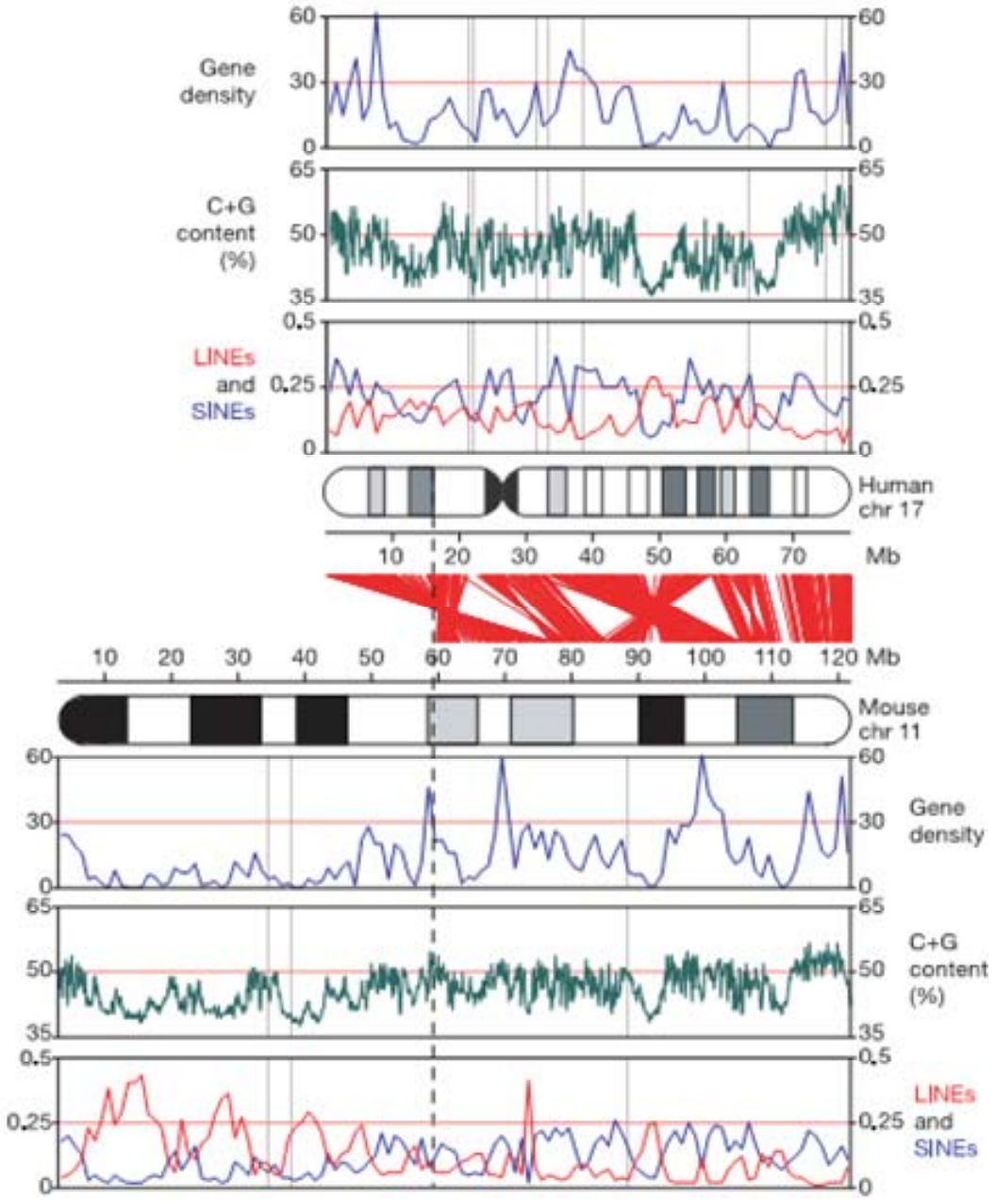

Figure 1.6 Landscapes of human chromosome 17 and mouse chromosome 11 [6]. (Figure taken from Zody *et al.*; 2006).

1.3. Aim of the study

As a gene rich region, 17q23 is an important chromosome band in breast cancer. Overexpressed genes mapping to 17q23 analyzed so far indicate some roles during breast tumorigenesis.

Our aim in this study was to further investigate a significantly amplified segment of this amplicon via *in silico* analyses and to search for existing ESTs that may indicate presence of, novel unidentified genes in the region.

CHAPTER II

MATERIALS AND METHODS

2.1. Materials

2.1.1. MCF7 Breast Cancer Cell Line and Media

MCF7 cell line was used to isolate genomic DNA and total RNA. The cell line was kindly donated by Asst. Prof. Dr. Uygar Tazebay, from Bilkent University, Ankara, Turkey.

Minimum Essential Medium (MEM) with Earl's salts of Biochrom AG (Cat# FG0325) was used as growth media. The composition of the media is presented in Appendix A. 10% Foetal Calf Serum Gold, PAA (Cat# A11-649) and 5% Penicilline / Streptomycine (10 000 IU/ 10 000 µg/mL), Biochrom AG (Cat# A2212) were added into MEM by filtrating through Millipore 0,45 µm HV Durapore membrane (Cat# SCHVU05RE).

Hank's Balanced Salt Solution of Biochrome AG (Cat# L 2055) was used in tissue culture studies to wash monolayer cells from metabolic wastes and dead cells. 1X Trypsin from PAA Laboratories (Cat# L11-022) was used to detach monolayer cells from flasks' surface.

Doubling time of MCF7 cell line is 29 hours, so medium refreshing was done 3 times a week and subculturing was done every two days when the cells

were ~80% confluent. DNA isolation from cells was done at ~90% confluency and total RNA isolation was done at ~70% confluency.

Freezing was done by using growth medium supplemented with 5% DMSO (dimethyl sulfoxide added as the cryoprotectant agent). Cells were stored at vapor phase of the liquid nitrogen which is approximately -150°C. Thawing was done at 37°C in water bath.

All the reagents and chemicals used in cell culture studies were cell culture grade.

2.1.2. Mammalian Cell Culture Conditions

MCF7 cell line was incubated at 37°C with 95% air and 5% CO_2 in a hepa filtered Heraeus Hera Cell 150 incubator. Cell line was handled in a Bilser Class II laminar flow cabinet by using appropriate cell culture techniques.

2.1.3. Bacterial Culture Media and Culture Conditions

LB (Luria Broth) media was used to grow *E.coli* TOP10 and DH5α strains. Also the medium was supplemented with 100 mg/mL ampicillin for bacterial selection. Ingredients of the media are listed in Appendix B. The entire medium was dissolved in distilled water, and then pH of the medium was adjusted to 7.4 with NaOH. Finally the medium was sterilized by autoclaving at 121°C for 15 minutes. Ampicillin was added freshly to the sterile medium.

All the bacterial cultures were grown with a 180-200 rpm (revolution per minute) in a Sanyo Gallenkamp shaker at 37°C. When needed 1.5 % agar was added to solidify the media. A liquid aliquot of each bacterial strain taken from mid-log phase was kept at –80°C in 20% glycerol.

SOC (Super Optimal Catabolite repressed broth) growth media was used during transformation protocols, The composition of bacterial culture medium is provided in Appendix B. SOC medium was prepared by dissolving the ingredients in distilled water, and then pH of the medium was adjusted to 7.0 with NaOH. Finally the medium was sterilized by autoclaving at 121°C for 15 minutes. Filter-sterilized 2M $MgCl_2$ was added freshly to the medium.

2.1.4. Other Chemicals and Materials

The chemicals used in this study were purchased from Sigma Chemical Company (N.Y., USA), and Applichem (Darmstadt, Germany). Other chemicals, kits and enzymes used in molecular biology studies were from MBI Fermentas (Ontario, Canada), Applichem (Darmstadt, Germany), Roche (Mannheim, Germany), Invitrogen (Carlsberg, CA, USA), Biochrom AG (Berlin, Germany), PAA Laboratories (Linz, Austria), Promega (Madison, USA), Genemark (Taiwan), Molecular BioProducts Inc. (San Diego, CA) and Favorgen (Taiwan). All of the media and solutions were prepared by using sterile distilled water.

Autoclavable centrifuge (1,5 mL) and PCR (0,2 and 0,5 mL) tubes, and tips were from Greiner (Frankfurt, Germany). Autoclavable centrifuge wares were from Nalgene (USA). Phase lock gels used in RNA and DNA isolation were purchased from Eppendorf (Hamburg, Germany). 5 mm glass beads used for bacterial spreading were from Merck (Darmstadt, Germany). T25 and T75 sterile, filtered cap tissue culture flasks, sterile serological pipettes and RNase-free filter tips used in cell culture were from Greiner (Frankfurt, Germany). 0.22 μm and 0.45 μm minipore filters were from Macherey-Nagel (Düren, Germany). Sterile syringes used for filtration were from Tyco Healthcare (UK).

All the Gene Specific Primers (GSPs) and other primers were ordered from İontek (İstanbul, Turkey) and Integrated DNA Technologies (Coralville, USA).

2.2. Methods

2.2.1. Bioinformatic Analysis of a 50 kb Region on 17q23

Prior to analysis, boundaries of the region were determined according to the 3' of the *TBX2* and 5' of the *TBX4* genes. Distance between these two genes is 46,980 bp. So, 50,000 bp sequence covering the gap between *TBX2* and *TBX4* genes along with a partial sequence from these genes was selected in order to investigate this region in a variety of bioinformatics programs.

Pan troglodytes (chimpanzee) and *Mus musculus* (mouse) were the two mammalian species chosen to be compared with *Homo sapiens* (human). 50 kb region which was determined for structural analysis is similar in chimpanzee genome according to the locations of *TBX2* and *TBX4* genes found in chimpanzee. On the other hand, chromosome 17 of human is syntenic to the chromosome 11 of mouse [6]. So, while comparing human and mouse genomes, 50 kb region from chromosome 11 syntenic to the human was also determined.

2.2.1.1. Idenfication of ESTs on 50 kb Region

From NCBI Mapviewer [48] and UCSC Genome Browser [49] databases ESTs from human, mouse and chimpanzee aligned to 50 kb region were determined. Locations, sizes, and orientations of these ESTs were analyzed.

2.2.1.2. Gene Prediction Analysis

Exon predictions were done from the 50 kb long genomic DNA sequences with GENSCAN (http:// genes.mit.edu/ GENSCAN. html) [81], and GRAIL (http://comp bio.ornl.gov/Grail-1.3/). While using GENSCAN, the sequence was submitted in FASTA format. The outcome was retrieved both in PDF and text

format. In GRAIL, the outcome was only in text format. Positions of possible exons were shown manually on genomic sequence.

2.2.1.3. Percent Identity Plot (PIP) Analysis

Vista [82] and Multi PipMaker [83] were the percent identity programs which were used to analyze alignment of sequences from human, mouse, and chimpanzee genomes.

50 kb sequences from three species were used in FASTA format according to the directions of Multi PipMaker program. Genomic DNA sequences were used after masking the repeats by RepeatMasker program (http://www.repeatmasker.org). RepeatMasker program was used to screen DNA sequences for interspersed repeats and low complexity DNA sequences. Alternatively in Vista, sequences were derived from UCSC directly by the program itself. While aligning the sequences, program acquired the positions of genes from UCSC database as well.

2.2.1.4. Locating Affymetrix Probes on 50 kb Region

Affymetrix Human Exon 1.0 ST probe sets are designed depending on the existing gene exons or sites where the possible exons could be present. Therefore, Affymetrix probes were determined from UCSC Genome Browser which were aligned to 50 kb region of human genome. Probe set data was obtained from Affymetrix Netaffx Exon/Gene Array website (https://www.affymetrix.com/analysis/netaffx/xmlquery_ex.affx).

2.2.2. Rapid Amplification of cDNA ends (RACE)

In order to synthesize RACE-ready cDNA (RR-cDNA) total RNA isolation was performed from MCF7 breast cancer cell line. Isolated RNA was treated with DNase before cDNA synthesis.

2.2.2.1. Total RNA Isolation from MCF7 Breast Cancer Cell Line

Prior to RNA isolation, eppendorf tubes, and other equipments were washed with DEPC-treated water and placed under a hood for overnight to evaporate the excess DEPC. All the solutions used in the RNA isolation were prepared with DEPC-treated water and autoclaved before use. For 1 L of water 800 µL of DEPC was used.

After cleaning the work area by RNase AWAY from Molecular BioProducts (Cat# 7000) and DNA AWAY from Molecular BioProducts (Cat# 7010) T75 flasks were taken from incubator, 4 flasks of cell culture were enough to obtain a decent RNA concentration. MCF7 cells were not more than 70% confluent when isolating RNA to ensure they were actively dividing. Confluency was checked under a Olympus CKX41 inverted microscope. Growth media was sucked off from flasks, and 8 mL Trizol Reagent (Guanidium thiocyanate) from Invitrogen (Cat# 15596-018) was added to each flask for homogenization. After adding the Trizol, cells were kept at room temperature for 5 minutes to allow complete dissociation of nucleoprotein complexes.

Phase separation was the following step, cells were transferred to a 15 mL sterile blue cap tubes. 1.6 mL chloroform was added to each tube. After capping the tubes securely, tubes were shaked vigorously for 15 seconds by hand. Then tubes were allowed to incubate at room temperature for 3 minutes. After incubation, samples were centrifuged at 4,700 g for 20 minutes at 8°C. After

phases were separated, RNA remained in the aqueous phase and the volume of the aqueous phase was about 60% of the original Trizol volume, approximately 5 mL.

After transferring the aqueous phase to a fresh tube, 4 mL isopropyl alcohol was added to each tube (0.5 mL isopropyl alcohol per 1 mL Trizol used) for RNA precipitation. The samples were incubated at RT for 15 minutes and centrifuged at 4,700 g for 20 minutes at 4°C. RNA precipitate formed a gel-like pellet on the bottom of the tube after centrifugation.

Subsequent to removing supernatant, RNA pellet was washed with 8 mL of 75% ethanol (prepared by using DEPC-treated water). After mixing the sample by gentle vortexing, it was centrifuged at 4,700 g for 7 minutes at 4°C. Following the centrifugation, supernatant was taken without disturbing the RNA pellet and allowed to dry for 10 minutes at RT. RNA pellet was resuspended in 25 µL RNase-free distilled water. After incubating the samples at 55°C for 10 minutes, 4 tubes were collected in one tube for quantification. RNA was either immediately used after determination of its quantity and quality or kept at -80°C for further use.

2.2.2.1.1. Determination of RNA quantity and quality

The RNA concentration is determined by measuring absorbance at 260 nm on a spectrophotometer in distilled water (one absorbance unit = 40 µg/mL RNA). The A260/A280 ratio should be approximately 2.0, but figures between 1.8 and 2.1 are considered acceptable. The RNA concentration should be greater than 1.1 µg/µl to accept it as appropriate isolation. After measuring the optical densities the concentration of RNA can be calculated as follows:

$$\text{RNA } (\mu g/mL) = 40 \times \text{Dilution Factor} \times OD_{260}$$

The integrity of the RNA is tested by agarose gel electrophoresis. For this RNase-free 1% agarose gels in TBE were prepared with DEPC-water. Approximately 2 µg RNA was loaded from each sample in RNase-free DNA loading buffer and run for 30-60 min at 80-100 V. The rRNA bands were observed without any obvious smearing patterns**.**

2.2.2.1.2. DNase Treatment

In order to ensure DNA-free RNA, isolated RNA was treated with Deoxyribonuclease I from Fermentas (DNase I) (Cat# EN0521). 10 µg RNA was treated with RNase-free DNase I. Ingredients of the reaction are listed in Table 2.1, reaction mixture was set up on ice.

Table 2.1 DNase I reaction mixture.

RNA (2.4 µg/µL)	4.5 µL
10X Reaction Buffer	10 µL
DNase (1u/µL)	10 µL
Ribonuclease Inhibitor (40u/µL)	5 µL
DEPC-dH_2O	70.5 µL
Total Volume	**100 µL**

Incubation at 37°C for 40 minutes was done in Memmert water bath. Reaction was stopped by adding equal amount of; 100 µL Phenol: Chloroform: Isoamyl alcohol (25: 24: 1) and tube was vortexed for 30 seconds, gently. Than the sample was kept on ice for 10 minutes and was transferred into a 1.5 mL Phase Lock Gel heavy from Eppendorf (Cat# 955154151). Centrifugation at 14,000 g for 20 minutes was done at 4°C. Afterwards upper phase including RNA was taken into a DEPC-dH_2O treated clean tube. Upper phase was approximately

180 µL, so 540 µL ice cold 100% ethanol (3V) and 18 µL 3M NaAc (1/10 V) was added. Then the sample was incubated at -20°C overnight. Next day, the sample was centrifuged 30 minutes at 14,000 g at 4°C in a pre-cooled centrifuge. After discarding the supernatant, pellet was washed with 70% cold ethanol (prepared with DEPC treated water) and centrifuged 10 minutes at 14000 g at 4°C. Finally, pellet was dissolved in 20 µL RNase-free water. Determination of RNA quantity and quality was done as previously described.

2.2.2.2. RACE-ready cDNA synthesis

GeneRacer™ Kit of Invitrogen (Cat# L1500-01) was used to obtain RACE-ready cDNAs. Total RNA isolated from MCF7 breast cancer cell line was used to synthesize RACE-ready cDNA according to the manual of the kit. Besides, HeLa total RNA provided with the kit was used as a control. The GeneRacer™ Kit provides a procedure to get full-length 5' and 3' ends of cDNA by using known cDNA sequences form ESTs.

Dephosphorylation step is done to remove the 5' phosphates from total RNA with calf intestinal phosphatase (CIP). By this way, truncated mRNA and non-mRNAs were eliminated from ligation with GeneRacer RNA Oligo. Dephosphorylation reaction was set up on ice in a 1.5 mL sterile microcentrifuge tube using the reagents provided by the kit. Reaction mixture was set up on ice as mentioned in Table 2.2.

After reaction mixture was pipetted and vortexed briefly, it was incubated at 50°C for 1 hour. When the incubation was over, tubes were placed on ice and RNA was precipitated. To precipitate RNA, 90 µL DEPC water and 100 µL phenol: chloroform was added to the mixture and vortexed vigorously for 30 seconds. Then the sample was centrifuged at 14,000 g in a microfuge for 5 minutes at room temperature. Aqueous phase was taken to a new microcentrifuge tube. Then, 2 µL 10 mg/mL mussel glycogen, 10 µL 3M sodium acetate (pH 5.2)

were added and mixed. Additionally 220 µL 95% ethanol was added and vortexed briefly. At that point the sample was left overnight at -20°C.

Table 2.2 CIP treatment reaction mixture.

Reagent	**MCF7 RNA**	**HeLA RNA**
RNA	1 µL (1µg)	2 µL (1µg)
10X CIP Buffer	1 µL	1 µL
RNase Out (40U/µL)	1 µL	1 µL
CIP (10U/µL)	1 µL	1 µL
DEPC water	6 µL	5 µL
Total Volume	**10 µL**	**10 µL**

Next day, the sample was centrifuged at 14,000 g for 20 minutes at 4°C in order to pellet RNA. Supernatant was taken by pipet without disturbing the pellet and the pellet was washed with 500 µL 70% ethanol by inverting several times. After centrifugation at 14000 g for 2 minutes at 4°C, ethanol was carefully removed by a pipette. To collect remaining ethanol, tube was spun again and the remaining ethanol was removed. The RNA pellet was left to air-dry for 2 minutes. Finally, the pellet was resuspended in 7 µL DEPC water. At the end of the dephosphorylation, truncated mRNA and non-mRNAs were eliminated. On the other hand, full-length mRNAs were not affected by CIP treatment as they got 5' cap structures.

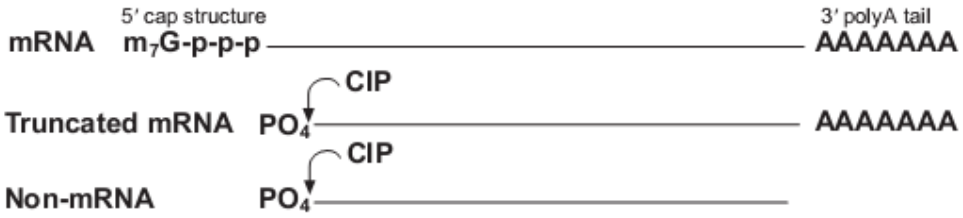

Figure 2.1 CIP treated total RNA. (Figure taken from User Manual of GeneRacer™ Kit of Invitrogen)

Following step was to remove the 5' cap structure from the full-length, intact mRNAs. This was done by using tobacco acid pyrophosphatase (TAP) treatment. After this treatment mRNAs were left with a 5' phosphate which was necessary for GeneRacer RNA Oligo ligation. So, decapping reaction was set up on ice. Reagents listed in Table 2.3 were added on to the RNA collected from the dephosphorylation reaction.

Table 2.3 TAP treatment reaction mixture.

Dephosohorylated RNA	7 µL
10X TAP Buffer	1 µL
RNaseOut (40 U/µL)	1 µL
TAP (0.5 U/µL)	1 µL
Total Volume	**10 µL**

After mixing the sample, it was vortexed briefly. To collect the fluid it was centrifuged briefly and incubated at 37°C for 1 hour. Tube was placed on ice when the incubation was over. RNA precipitation was done as told before, 7 µL

RNA was obtained after precipitation. At the end of the decapping reaction and RNA precipitation, full-length mRNA's were decapped as shown below.

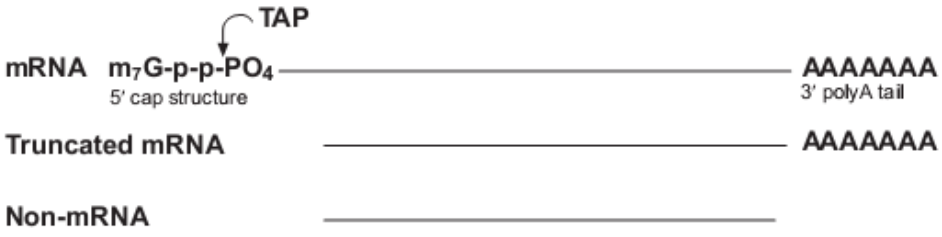

Figure 2.2 TAP treated total RNA. (Figure taken from User Manual of GeneRacer™ Kit of Invitrogen)

After removing the 5' cap, GeneRacer RNA Oligo was ligated to mRNA's 5' phosphate by using T4 RNA ligase. In order to do that, 7 µL dephosphorylated, decapped RNA was added on to the lyophilized GeneRacer™ RNA Oligo (0.25 µg). By pipetting up and down RNA oligo and RNA sample was mixed. Then, the mixture was incubated at 65°C for 5 minutes to relax the RNA secondary structure. When the incubation was over, the tube was chilled on ice for 3 minutes and spun to collect the fluid. Reagents listed in Table 2.4 were added to the tube, mixed gently by pipetting, and spun to collect.

The mixture was incubated at 37°C for 1 hour. Then it was centrifuged briefly and placed on ice. RNA precipitation was repeated as explained before. Resuspensin was done with 10 µL DEPC water. At the end of oligo ligation, RNA was ready for reverse transcription.

Table 2.4 GeneRacer™ RNA Oligo ligation reaction mixture.

10X Ligase Buffer	1 µL
10 mM ATP	1 µL
RNaseOut (40 U/µL)	1 µL
T4 RNA Ligase (5U/µL)	1 µL
Total Volume:	**10 µL**

Figure 2.3 Ligation of GeneRacer™ RNA Oligo to the mRNA. (Figure taken from User Manual of GeneRacer™ Kit of Invitrogen)

Reverse transcription was done by using SuperScript III Reverse Transcriptase and GeneRacer Oligo dT Primer to create RACE-ready first-strand cDNA. First GeneRacer™ Oligo dT Primer, dNTP mix and sterile distilled water were added to the 10 µL RNA sample. This mixture was incubated at 65°C for 5 minutes to remove any RNA secondary structures. After chilling on ice and spinning briefly, 5X First Strand buffer, 0.1 M DTT, RNaseOut and SuperScript III Reverse transcriptase were added and mixed by pipetting.

The mixture was incubated at 50°C for 60 minutes and at the end reverse transciptase was inactivated at 70°C for 15 minutes. After the enzyme inactivation, the sample was chilled on ice and spun to collect fluids. 1 µL RNase H (2 Units) was added to the reaction mix, and incubated at 37°C for 20 minutes.

RACE-ready cDNA was stored at -20°C. This RACE-ready cDNA contained priming sites at the 5' and 3' ends.

Table 2.5 Reverse Transcription with GeneRacer™ Oligo dT Primer reaction mixture.

GeneRacer™ Oligo dT Primer	1 µL
dNTP Mix	1 µL
Sterile distilled water	1 µL
5X First Strand Buffer	4 µL
0.1 M DTT	1 µL
RNaseOut (40 U/µL)	1 µL
SuperScript III RT (200 U/µL)	1 µL
Total Volume:	**20 µL**

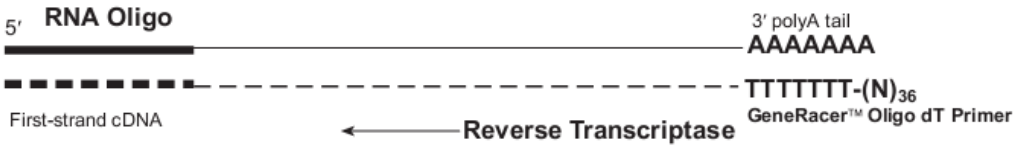

Figure 2.4 Reverse Transcription by GeneRacer™ Oligo dT Primer (Figure taken from User Manual of GeneRacer™ Kit of Invitrogen).

2.2.2.3. Amplifying cDNA Ends by PCR

RACE PCRs were done by using gene specific primer and a second primer from anchored oligos. Generally, nested RACE PCRs were done to ensure the specificity of amplified fragments.

2.2.2.3.1. Gene Specific Primer (GSP) Design

GSPs were the PCR primers designed from T02811 sequence information. GSPs were designed according to the following; first, all GSPs were designed approximately 24-28 nucleotides in length to increase specificity of binding. Second, high annealing temperatures (>72°C) with GC contents ranging between 50-70% were preferable. Moreover, GSPs were aimed to have low GC content at 3' ends to minimize extension by DNA polymerase at non-target sites. Another criteria was to control the GSP sequences for both self-complementarity and complementarity to the primers supplied in the kit. The most important criteria for GSPs were the specificity to the Human genome, so all GSPs were BLASTed to human genome to avoid non-specific annealing during RACE PCRs. A complete list of 3'RACE and 5'RACE GSPs, and primers supplied in the kit were listed in Appendix C.

2.2.2.3.2. Verification of cDNA Quality by Using β-actin Primers

Before amplifying MCF7 and control HeLa RACE-ready cDNAs with GSPs, control PCR with β-actin primers were done to confirm the quality of RACE-ready cDNAs. Control primers were provided with the kit. Two different PCRs were set up to amplify both 3' and 5' ends of β-actin gene. Same PCR conditions were used as described in Table 2.9. As can be seen from Figure 2.5 expected sizes for 3' and 5' RACE were 1800 bp and 900 bp, respectively. After analyzing the RACE-ready cDNAs, 5' RACE and 3' RACE PCRs were set up separately.

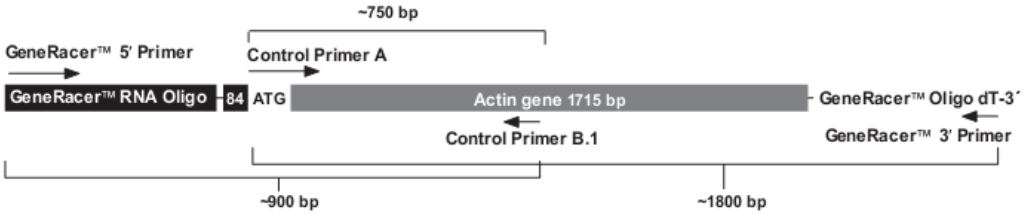

Figure 2.5 The first-strand cDNA having GeneRacer™ Oligo at 5' and GeneRacer™ Oligo dT Primer at 3' (Figure taken from User Manual of GeneRacer™ Kit of Invitrogen).

Table 2.6 Optimized PCR conditions to amplify β-actin gene from 5' end.

Reagent	Volume
dH_2O	36.5 µl
10X Rxn Buffer with $(NH_4)_2SO_4$	5 µl
dNTP Solution (10 mM each)	1 µl
GeneRacer 5' Primer (10 µM)	4.5 µl
Control Primer B.1 (10 µM)	1.5 µl
pfu polymerase	0,5 µl
RACE-ready cDNA Template	1 µl
Total volume	**50 µl**

Table 2.7 Optimized PCR conditions to amplify β-actin gene from 3' end.

Reagent	Volume
dH_2O	36.5 µl
10X Rxn Buf. w/ 20 mM $MgSO_4$	5 µl
dNTP Solution (10 mM each)	1 µl
Control Primer A (10 µM)	1.5 µl
GeneRacer 3' Primer (10 µM)	4.5 µl
pfu polymerase	0,5 µl
RACE-ready cDNA Template	1 µl
Total volume	**50 µl**

Table 2.8 PCR cycling conditions to amplify β-actin gene.

Initial denaturation	95°C 2 min	
Denaturation	95°C 1 min	35 cycles
Annealing	65°C 30 s	
Extension	72°C 4 min	
Final extension	72°C 5 min	

2.2.2.3.3. 3' RACE PCR

3' RACE PCRs were done by using forward GSPs. Nested PCRs were done by using forward nested GSPs as shown in Figure 2.6.

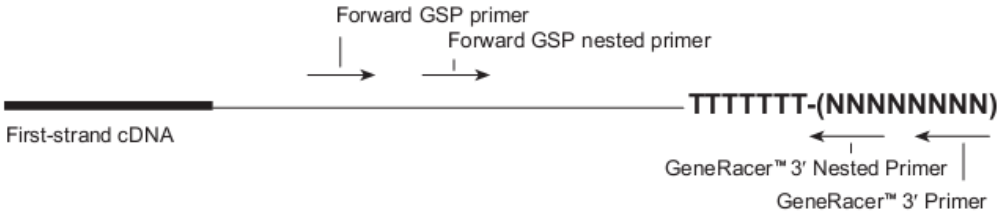

Figure 2.6 3' RACE PCR amplification by using forward GSP and GeneRacer™ 3' Primer (Figure taken from User Manual of GeneRacer™ Kit of Invitrogen).

2.2.2.3.4. 5' RACE PCR

5' RACE PCRs were done by using various GSPs. Nested PCRs were done by using nested GSPs. PCR mixture and PCR cycling conditions for 5' RACE were provided in Table 2.9 and Table 2.10, respectively.

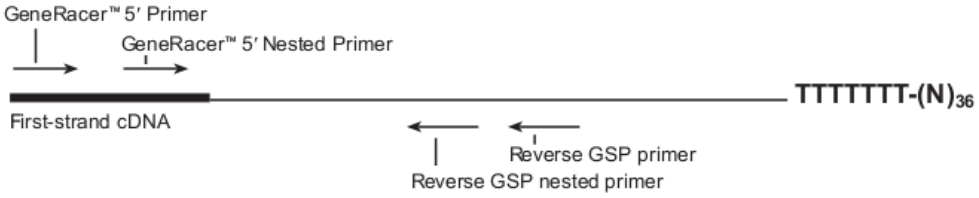

Figure 2.7 5' RACE PCR amplification by using reverse GSP and GeneRacer™ 5' Primer.

Table 2.9 Optimized conditions for 5' RACE PCR.

Reagent	Volume
dH_2O	36.5 µl
10X Rxn Buf. w/ 20 mM $MgSO_4$	5 µl
dNTP Solution (10 mM each)	1 µl
GSP (10 µM)	1.5 µl
GeneRacer 5' Primer (10 µM)	4.5 µl
pfu polymerase	0,5 µl
RACE-ready cDNA Template	1 µl
Total volume	**50 µl**

PCR products obtained throughout this study were visualized on agarose gels. Depending on the size of DNA to be detected 1%, 2% or 3% agarose gels were used. For this purpose 0.5 g, 1 g or 1.5 g of agarose was dissolved in 50 mL of 0.5 X TBE (Appendix D) in microwave oven. When it completely dissolved it was cooled down to approximately 50°C and 2.5 µL of ethidium bromide from 10 mg/mL stock (final concentration was 5 µg/mL) was added to gel solution. The comb was placed for well formation and without making bubbles the solution was poured into the electrophoresis tray. When the gel completely solidified, electrophoresis buffer was poured to the tank till it covered the gel for approximately 5 mm and the comb was removed carefully.

Table 2.10 PCR cycling conditions for 5' RACE Touchdown PCR.

Initial denaturation	95°C 2 min	
Denaturation	95°C 30 s	5 cycles
Annealing	70°C 30 s	
Extension	72°C 4 min	
Denaturation	95°C 30 s	5 cycles
Annealing	68°C 30 s	
Extension	72°C 4 min	
Denaturation	95°C 30 s	25 cycles
Annealing	66°C 30 s	
Extension	72°C 4 min	
Final extension	72°C 10 min	

The sample solution to be loaded was prepared by mixing samples with 6X loading buffer to a final concentration of 1X and loaded into the wells. A suitable molecular size marker, DNA ladder, was also loaded into a separate well. Then the tank was connected to a power supply and electrophoresis was performed under constant voltage of 100 V for 45 minutes - 1 hour. The gel was observed under UV light and photographed by using Vilber Lourmat Gel Imaging System.

2.2.2.4. DNA Extraction From Agarose Gels

2 different DNA extraction kits were used throughout this study. One of them is DNA Clean/ Extraction Kit of GeneMark (Cat. No. DP034-150). Second kit was Roche's Agarose Gel DNA Extraction Kit (Cat. No. 11 696 505 001). Both kits were used according to the instructions of manufacturer.

2.2.2.5. A-Tailing for Blunt-ended PCR Products

In order to accomplish TOPO TA cloning, first it was necessary to add A overhangs to the PCR product. So, the reaction mixture was set up on ice as follows: 1 µL Taq DNA polymerase 10X reaction buffer with $MgCl_2$ was added on 7 µL purified PCR fragment. 1 µL 2 mM dATP was added to a final concentration of 0.2 mM. Finally, 1 µL Taq DNA polymerase (5 units/ µL) was added and mixed by pipetting briefly. The reaction mixture was incubated at 70°C for 30 minutes. A-tailed PCR fragment was ready for ligation after incubation. It was used fresh in order to increase the efficiency of TA cloning.

2.2.2.6. Cloning and Sequencing of PCR Products

TOPO TA Cloning® Kit for Sequencing (Cat# K-4475-J10) of Invitrogen was used for cloning and sequencing PCR products. Prior to the cloning chemically competent *E.coli* cells were prepared.

2.2.2.6.1. Chemically Competent *E.coli* Preparation

Prior to competent *E.coli* preparation, all the necessary autoclavable equipment were sterilized by autoclaving at 121°C for 20 minutes. Aseptic techniques were used while handling the bacterial cultures and growth media.

TOP10 and DH5α *E.coli* strains were inoculated into 20 mL LB and incubated at 37°C and 200 rpm overnight. Next day, 300 µL from growth culture was transferred into 50 mL LB. They were incubated at 37°C and 200 rpm till OD600 reach to 0.6, which was approximately 2 hours. Then 50 mL culture was divided into 2 prechilled centrifuge tubes and kept on ice for 10 minutes. Then, centrifuge was done at 4000 rpm for 10 minutes at 4°C. After the supernatant was removed, pellet was resuspended in 5 mL 10 mM $CaCl_2$ (Appendix D) by vortexing. Second centrifuge was done at 3000 rpm for 10 minutes at 4°C. The

supernatant was discarded and 1 mL 75 mM $CaCl_2$ (Appendix D) was added on to pellet. Also, 200 µL ice cold glycerol was added on to the cells. Glycerol was the cryoprotectant agent to protect cell damage during freezing and long term storage at -80°C. Chemically competent cells were divided into sterile centrifuge tubes and frozen in liquid nitrogen. They were stored at -80°C.

2.2.2.6.2. TOPO Cloning Into pCR®4-TOPO® Plasmid

TOPO cloning reaction was set up on ice with the reagents provided with kit. (Table 2.11)

Table 2.11 TOPO cloning reaction mixture.

A-tailed PCR Product	2 µL
Salt Solution	1 µL
Sterile water	2 µL
pCR®4-TOPO® vector	1 µL
Final Volume	**6 µL**

They were mixed gently by pipetting and incubated at RT for 5 minutes. Then, the mixture was placed on ice and proceeded to transformation. Also, TOPO cloning reaction might be stored at -20°C overnight, but the efficiency of the cloning may be reduced.

2.2.2.6.3. Transforming TOP10 and DH5α Competent Cells

Prior to transformation of competent cells, LB plates with 50 µg/ mL ampicillin or 50 µg/ mL kanamycin were prepared.

100 µL frozen *E.coli* cells were taken from -80°C and they were thawed on ice. 2-5 µL from ligation reaction or intact plasmids were added and mixed gently, without pipetting up and down. Cells were incubated on ice for 30 minutes followed by 45 seconds heat-shock at 42°C. Cells were chilled on ice for 2 minutes and 900 µL SOC medium (Appendix B) was added. They were incubated at 37°C, with 200 rpm shaking for 1 hour. After incubation, 100-250 µL bacterial cells were spread on LB plates. Depending on the transformation efficiency 1:10 dilution was done with SOC medium while spreading. After the plates are dried at RT, they were incubated overnight at 37°C.

2.2.2.6.4. Plasmid Isolation from *E.coli* Cells

Throughout this study, plasmid isolation from bacterial cells were performed by using GeneMark's Plasmid Miniprep Purification Kit (Cat# DP01-100). Isolation of plasmids were performed according to the manufacturer's instructions.

After isolation, quantity of the plasmids was determined by using Hoefer^{TM} DQ300 Fluorometer. For this purpose, calibration of the fluorometer was done by using Assay Solution and 100 ng/ µL Calf Thymus DNA purchased from Sigma (Cat# D4764). Assay Solution was prepared freshly by using 10 µL Hoechst 33258 stock solution (Appendix D), 10 mL 10X TNE buffer (Appendix D) and 90 mL distilled filtered water. Quantification of the samples was done by using 2 mL Assay solution and adding 2µL from the sample.

2.2.2.7. Verification of Inserts by Restriction Enzyme Digestion

In order to verify if the amplified fragment was inserted into the vector, restriction enzyme digestion with *Eco*RI was performed.

2.2.2.8. Sequence Analysis of PCR Products

T3 and T7 sequencing primers (Appendix C) from the pCR®4-TOPO® vector were used during sequencing of the PCR products. For this purpose, the service of Sequence Analysis Unit of İontek, İstanbul was used.

The results of sequencing analysis were investigated using Chromas (v2.23) software program. After analyzing the chromatogram, FASTA format of the sequences were used in BLAST [48] program of NCBI in order to find similar sequences in the human genome.

2.2.3. Extending T02811 by PCR

By using the data obtained from bioinformatics tools, a number of PCR primers were designed and used to extend the size of the T02811. Nucleic acid source was MCF7 breast cancer cell line. Total RNA was isolated as described before.

2.2.3.1. Genomic DNA Isolation from MCF7 Breast Cancer Cell Line

The genomic DNA was isolated from MCF7 breast cancer cell line by using 3 T75 flasks of cell culture. Cells are 90% confluent when the isolation was performed. First, cells were washed with 5 mL Hank's Balanced Salt Solution (HBSS) to remove dead cells and metabolic wastes. After removing the HBSS from each flask with sterile glass pipettes, 2 mL trypsin was added to each flask. Flasks were incubated at 37°C incubator for 3-5 minutes, till the cells were

detached from the surface. Flasks were tapped gently to dislodge cells and observed under microscope. After ensuring all the cells were freely floating, 5 mL growth media was added into the flasks. Adding growth media stopped the action of trypsin, otherwise excess treatment with trypsin could harm cells. Cell clumps were disassociated by resuspending the cell suspension by the help of a serological pipette up and down several times. When all the clumps were disassociated, cell suspensions were taken into 15 mL sterile centrifuge tubes. They were centrifuged at 1,000 rpm for 5 minutes. Supernatants were removed and pellets were completely resuspended with Hank's solution. The suspension was centrifuged at 1,000 rpm for 5 minutes again. After removing the supernatant, 5 mL TNE buffer (Appendix D) was added to each tube and vortexed gently. Then, 50 µL Proteinase K (20 mg/mL) from Applichem (Cat# 4392) was added to each tube giving a final concentration of 200 µg/ mL. In addition, 250 µL from 20% SDS (Appendix D) was added to each tube giving a final concentration of 1% SDS. Tubes were mixed by inverting up and down several times. Then tubes were left at 55°C incubator overnight. Next day, 6 mL Phenol: Chloroform: Isoamyl alcohol (25:24:1) was added to each tube. An emulsion was formed by gently mixing. Suspension was spun at 2000 rpm for 7 minutes at 4°C. Top layer of the solution which is approximately 5 mL was removed carefully into a clean centrifuge tube without disturbing the protein layer and the bottom phase. 3M NaAc (pH 5.2) was added into the solution as 1/10X of total volume, which was 500 µL. 10 mL pre-cooled 100% ethanol was added into each tube. Tubes were shaked back and forth gently, until a clump of DNA formed. A glass pipette was bent into a hook by the help of Bunsen burner, and the DNA clump was taken by wrapping it around the hook. DNA was soaked into 70% ethanol to get rid of the salt. Then, it was left to air-dry for about 1 minute. It shouldn't be too dry otherwise DNA can be damaged. DNA was dissolved in 250 µL 1X TE buffer (Appendix D) and 3 tubes were pooled together. DNA was left in 65^{o}C water bath for 1 hour, and kept at room temperature overnight. Quantification was done by using Hoefer^{TM} DQ300 Fluorometer as described before. The purity and

concentration of DNA was checked by agarose gel electrophoresis. The prepared DNA was stored at –20°C for further usage.

2.2.3.2. cDNA synthesis from Total RNA Samples

Total RNA isolation from MCF7 breast cancer cell line was done as described previously. After DNase I treatment and quantification of RNA sample cDNA synthesis was done. Revert Aid First Strand cDNA Synthesis Kit of Fermentas (Cat# K1622) was used.

Besides, cDNA synthesis from Ambion's Human Breast Total RNA (Cat# 7952) was done with same kit as well. cDNAs were synthesized according to the manufacturer's instructions.

2.2.3.3. Primer Design

PCR primers were designed according to the data coming from bioinformatics analysis of the region. Gene prediction, percent identity plot analyses, positions of ESTs and Affymetrix probes were combined to design primers from the candidate coding sequence. While designing primers, certain parameters were considered. Primer candidates were selected manually from sequences retrieved from NCBI, and they were analyzed by Oligo Analyzer 3.0 of IDT (http://www. idtdna.com / analyzer/ Applications/ OligoAnalyzer/). Primer candidates were analyzed in terms of self-dimerization, hetero-dimer formation, GC content, melting temperature, and specificity. Specificity of all primers to the human genome were checked by using BLAST tool of NCBI [84]. List of the primers are given in Appendix C.

2.2.3.4. Cloning and Sequencing of PCR Products

Amplified fragments were cloned into pCR®4-TOPO® vector as described previously and sent to sequencing.

2.2.3.5. Sequence Analysis

Sequencing results were compared with human genomic DNA sequence by using BLAST tools of NCBI [48].

2.2.3.6. ORF Analysis of Extended Transcript

ORF (Open Reading Frame) Finder of NCBI [48] was used for identifying possible open reading frames.

2.2.4. Duplex RT-PCR to Compare Expression Levels of T02811

Reverse Transcription PCR (RT-PCR) was used to detect the overexpression in the MCF7 breast cancer cell line. MDA-MB231 breast cancer cell line RNA and normal breast RNA were used for cDNA synthesis to observe the difference between MCF7 cDNA.

Duplex Reverse Transcription PCR (D-RT-PCR) was done by using Glyceraldehyde 3-Phosphate Dehydrogenase (GAPDH) and T02811BKout primers. GAPDH primer couple was used as internal control, since GAPDH is a housekeeping gene and can be used to normalize PCR conditions.

PCR and cycling conditions were shown in Table.2.12 and Table 2.13, respectively. GAPDH concentration was optimized with normal breast cDNA by testing at 10% and 30% at the beginning. Then MCF7, MDA-MB231, and breast cDNA were used with 10% GAPDH concentration. GAPDH primer couple was

designed to amplify 115 bp size product. T02811BKout primer couple was amplifying 494 bp.

Table 2.12 Optimized conditions for D-RT-PCR.

Reagent	**Volume**
dH_2O	10.7 µl
10X Rxn Buf. w/ 20 mM $MgCl_2$	3 µl
dNTP Solution (2 mM each)	3 µl
T02811BKoutF Primer (5 µM)	3 µl
T02811BKoutR Primer (5 µM)	3 µl
GAPDH Forward Primer (10%)	3 µl
GAPDH Reverse Primer (10%)	3 µl
Taq polymerase	0,3 µl
Template	1 µl
Total volume	**30 µl**

Table 2.13 PCR cycling conditions for D-RT-PCR.

Initial denaturation	94°C 3 min	
Denaturation	94°C 30 s	34 cycles
Annealing	55°C 30 s	
Extension	72°C 30 s	
Final extension	72°C 5 min	

CHAPTER III

RESULTS AND DISCUSSION

3.1. Bioinformatic Analysis of a 50 kb Region on 17q23

Sequence of the 50 kb region between human *TBX2* and *TBX4* genes was taken from UCSC Genome Browser [49] database in FASTA format. The boundaries of the selected region are shown in Figure 3.1.

Similarly corresponding sequences from chimpanzee (*Pan troglodytes*) and mouse (*Mus musculus*) were also determined for further comparison and analysis. Boundaries of the chimpanzee and mouse sequences were given in Figure 3.2 and Figure 3.3, respectively. Positions and boundaries of the 3 sequences are listed in the Table 3.1. All the bioinformatics analysis through human, chimpanzee, and mouse were performed by using these 50 kb long sequences.

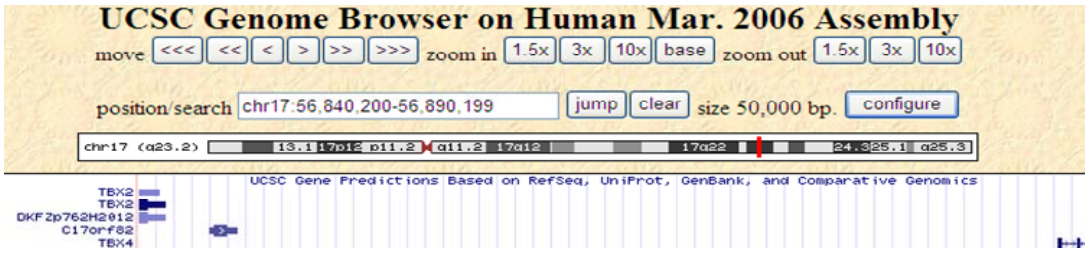

Figure 3.1 Physical map of the 50 kb region on 17q23 belonging to the human from UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

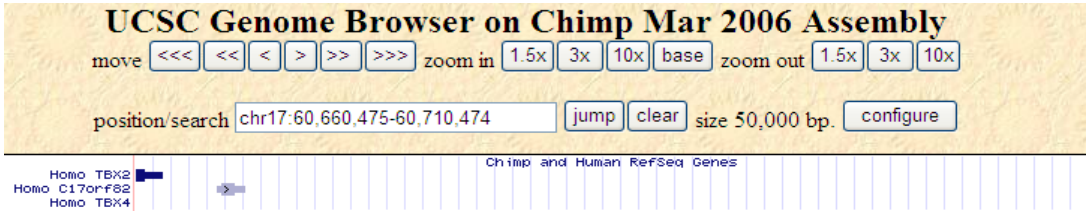

Figure 3.2 Physical map of the 50 kb region belonging to the chimpanzee from UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

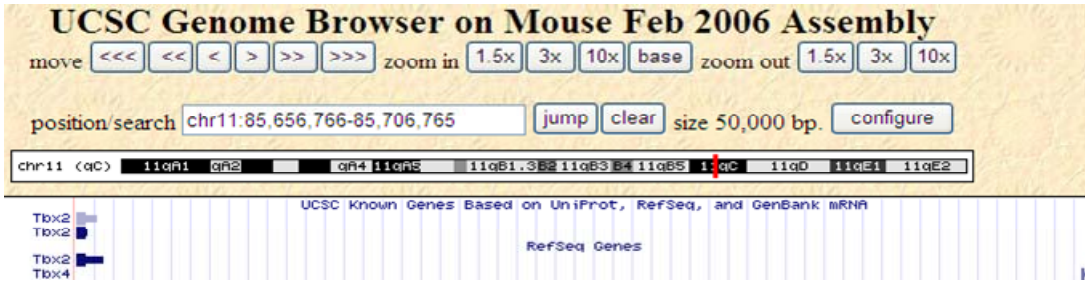

Figure 3.3 Physical map of the 50 kb region belonging to the mouse from UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

Like human, in the chimpanzee, the same region was found on chromosome 17 [80]. Chromosome 17 of human chromosome is syntenic to chromosome 11 of mouse [6]. As a result, 50 kb region which is orthologous to human 50 kb region was determined from chromosome 11 of mouse. In Figure 3.3 the boundaries of the selected 50 kb region from mouse are shown.

Table 3.1 Chromosomal positions of 50 kb region on human, chimpanzee, and mouse genomes from UCSC Genome Browser [49].

Coordinates	**Human**	**Chimpanzee**	**Mouse**
Chromosome	17	17	11
Start	56,840,200	60,660,475	85,656,766
End	56,890,199	60,710,474	85,706765

3.1.1. ESTs Positioned on 50 kb Region

To analyze this region, first existing ESTs were determined by the help of UCSC Genome Browser [49] and NCBI Mapviewer [48]. Physical map of the human, chimpanzee, and mouse sequences with ESTs are shown in Figure 3.4, Figure 3.5, and Figure 3.6, respectively.

On human sequence, there are two ESTs which are T02811 and AA745188. Others positioned on the region are corresponding to the alternatively-spliced forms of *TBX2* and *TBX4* genes.

On chimpanzee sequence, there are no ESTs derived from chimpanzee transcripts. There are human ESTs which are matching to the chimpanzee sequence as the 50 kb region is 98% identical to the human sequence. So, T02811 and AA745188 were also positioned on this sequence.

On mouse sequence there are 10 ESTs derived from mouse transcripts. Some of these ESTs are overlapping or spliced. There are also ESTs corresponding to *TBX2* and *TBX4* gene transcripts.

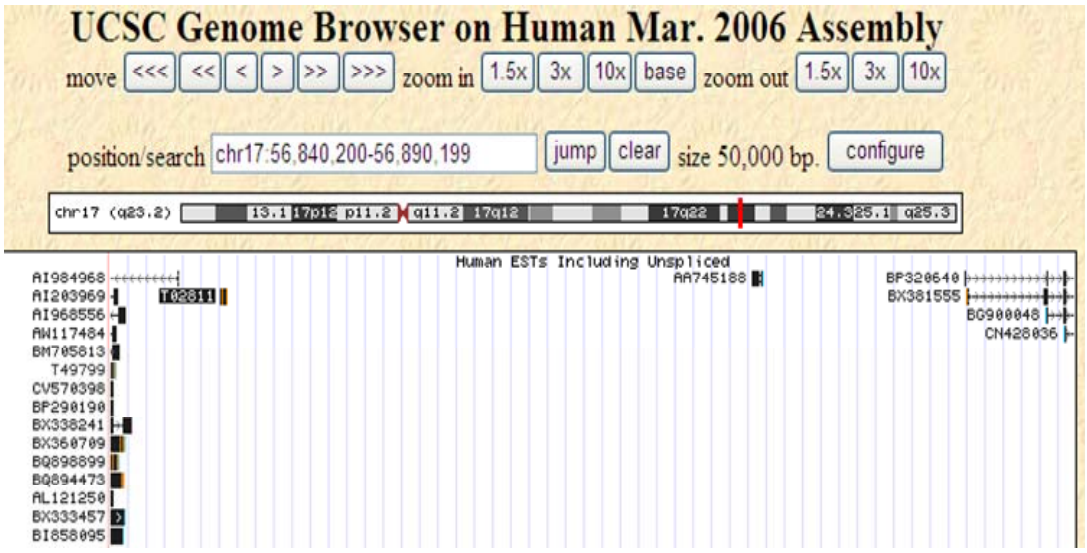

Figure 3.4 Physical map of the 50 kb region showing the ESTs belonging to the human from UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

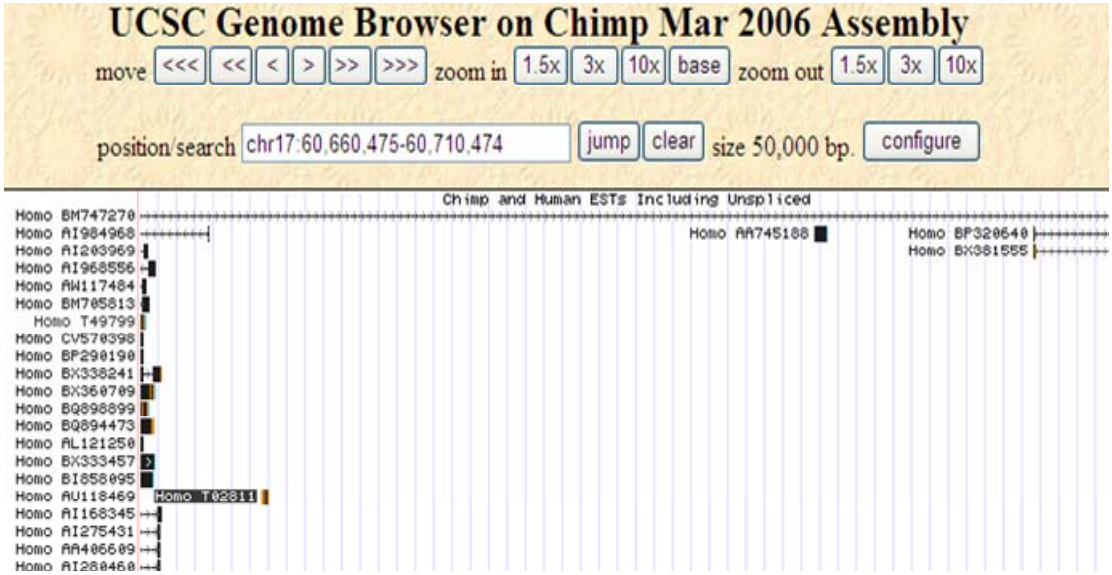

Figure 3.5 Physical map of the 50 kb region showing the ESTs belonging to the chimp from UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

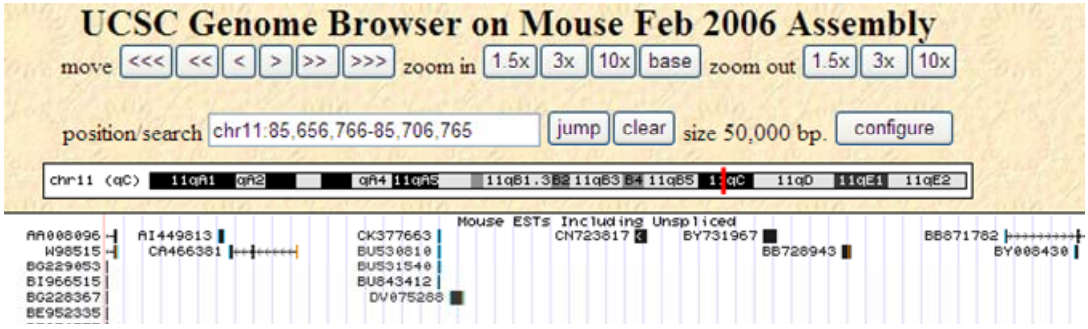

Figure 3.6 Physical map of the 50 kb region showing the ESTs belonging to the mouse from UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

3.1.2. Gene Predictions by GENSCAN

Via using the FASTA sequences obtained from human, chimpanzee, and mouse genomes, gene prediction analysis were done. GENSCAN (http://genes.mit.edu/GENSCAN.html) was used online from MIT server. Outcomes of the gene predictions of human, chimpanzee, and mouse are shown in Figure 3.7, Figure 3.8, and Figure 3.9, respectively.

According to the human exon prediction of the GENSCAN first exon predicted from the 50 kb region corresponds to the 3' of the *TBX2* gene, so it was neglected. Other four exons, illustrated as 9-12, corresponded to the 5' of the *TBX4* gene in human genome, so they were neglected as well. Finally, remaining seven exons, designated as 2-8, showed significant predictions. This predicted gene was approximately 38 kb in size (between 4.0-42.0 kb) and size of the transcript was nearly 1.5 kb. Based on that prediction, the gene was possibly transcribed from the minus strand.

On the other hand, from the GENSCAN prediction of chimpanzee, first predicted exon corresponded to the 3' of the *TBX2*, so exon number 1 was neglected. From the prediction there was a gene having 3 exons, (between 2.0-38.0 kb) approximately 36 kb in size. This gene candidate is expressed from the positive strand. Finally, there was another gene predicted with 4 exons, illustrated as 5-8, but 3' of this gene corresponded to the 5' of the *TBX4* in chimpanzee genome, so this prediction was neglected.

Also, from the prediction of GENSCAN of mouse, first two exons which were shown as 1 and 2 were also neglected as they corresponded to 3' of the *TBX2* gene. Following next six exons, designated as 3-8 pointed out a gene which was 12 kb long (between 4.0-16.0 kb) and expressed from the negative strand.

When the exon predictions were compared to each other, similarity between human and mouse was remarkable. T02811 is known to be expressed from negative strand in the human genome. As the predictions matching to the position of T02811 also pointed out a gene expressed from the negative strand, possibility of the presence of a gene expressed from the negative strand from this 50 kb region was strengthened.

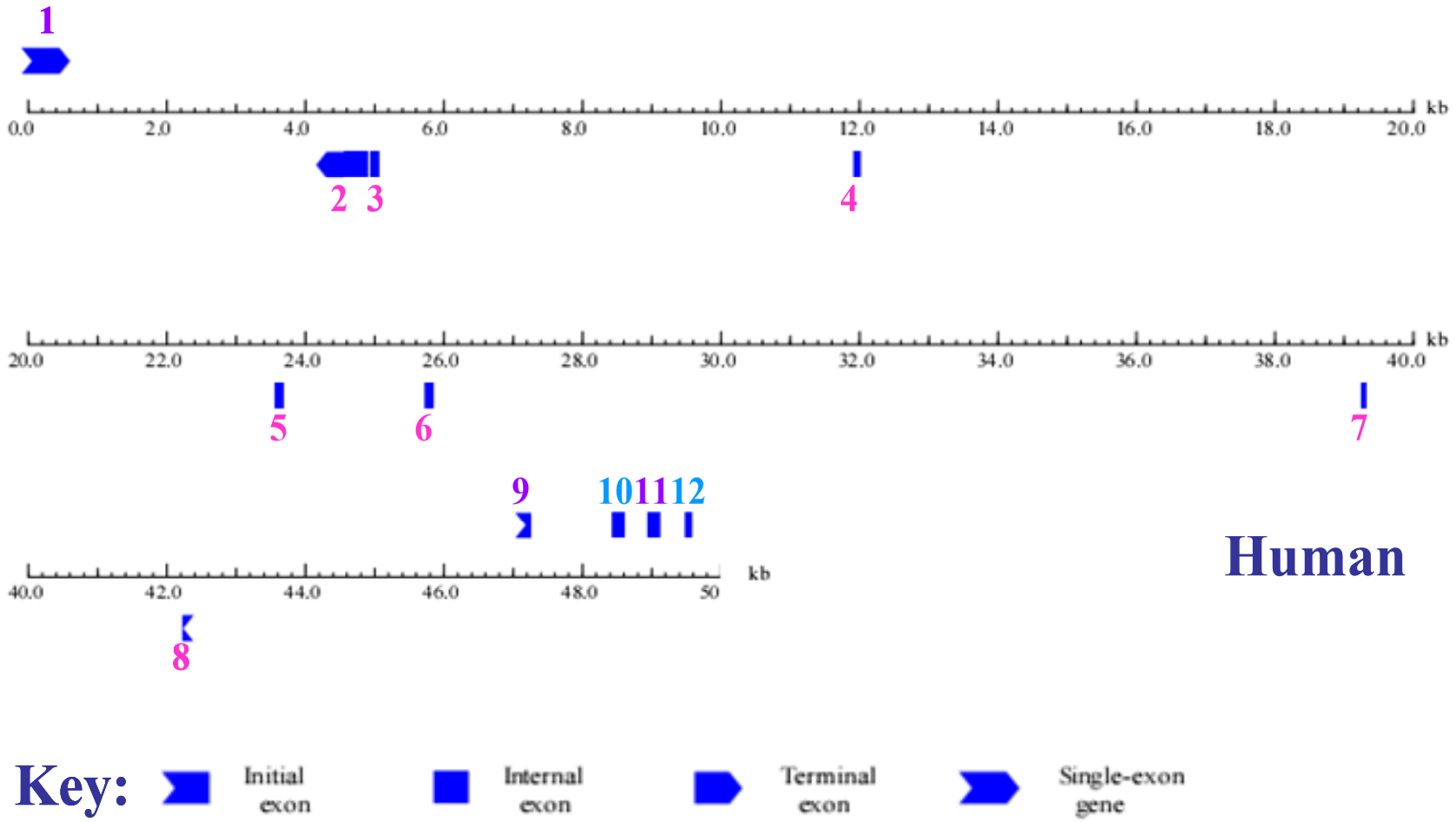

Figure .3.7 Gene prediction analysis by GENSCAN from 50 kb genomic sequence of human.

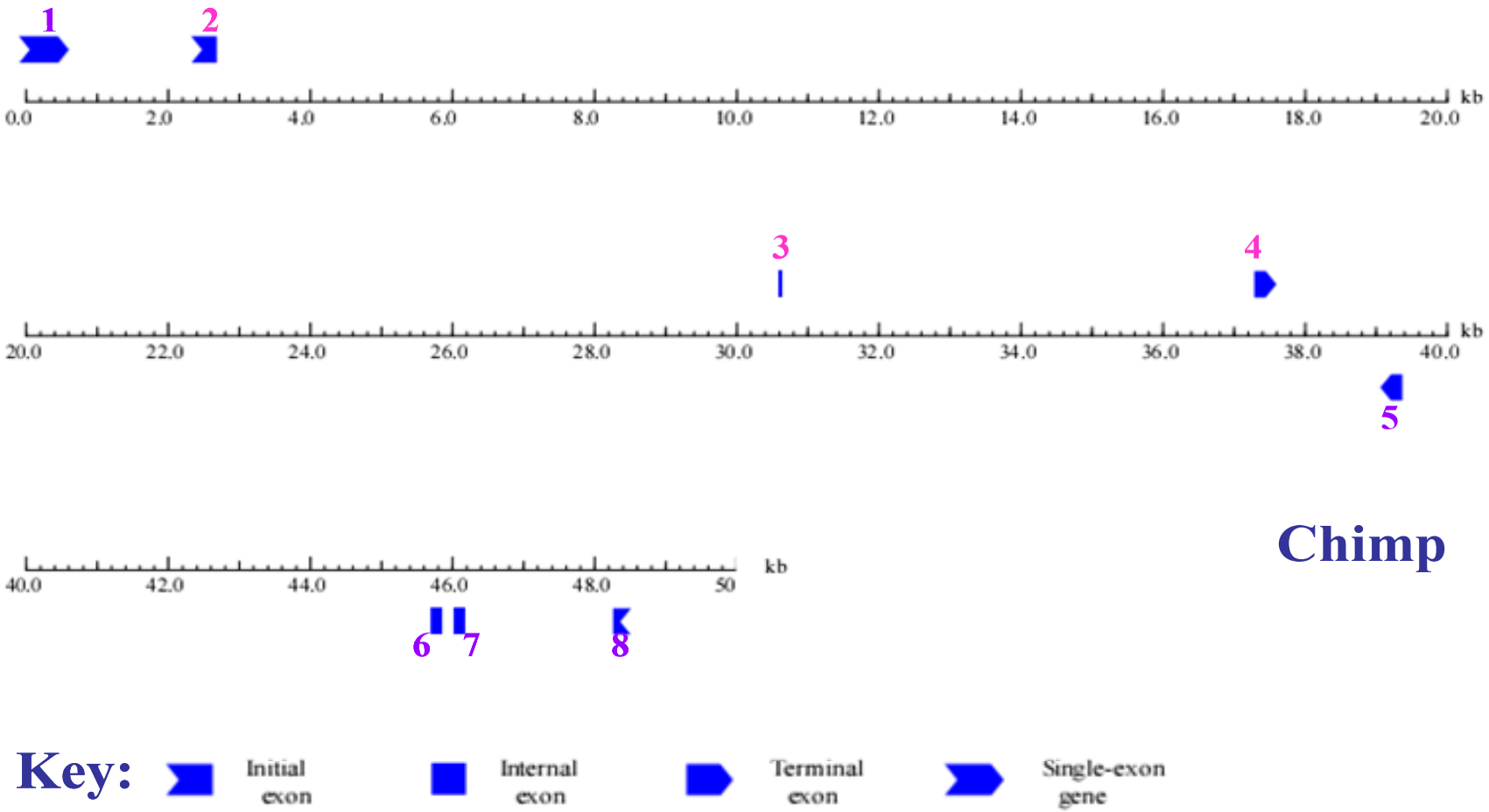

Figure 3.8 Gene prediction analysis by GENSCAN from 50 kb genomic sequence of chimpanzee.

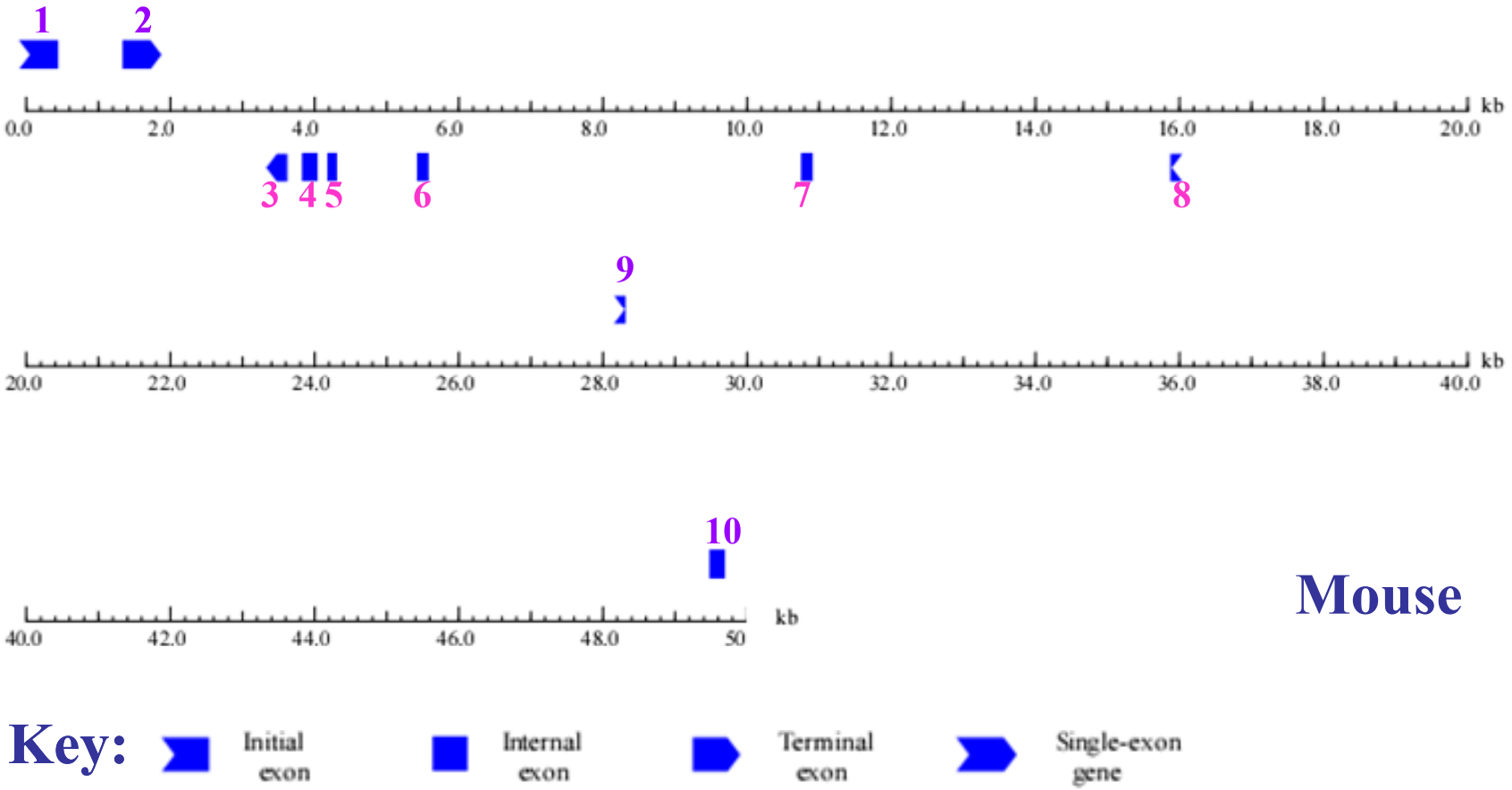

Figure 3.9 Gene prediction analysis by GENSCAN from 50 kb genomic sequence of mouse.

3.1.3. Percent Identity Plot (PIP) Results

PIP analysis was done by using Vista [85] and Multi PipMaker [83] sequence alignment programs. While using Vista, sequences were derived online from database by the program itself (http://pipeline.lbl.gov/cgi-bin/gateway2). Also, existing genes available on the database were also noted on the alignment result as shown in the Figure 3.10 and Figure 3.11 for chimpanzee and mouse, respectively. Besides existing genes, UTR (Untranslated Regions), CNS (Conserved Noncoding Sequences), LINEs (Long Interspersed Nuclear Elements), and SINEs (Short Interspersed Nuclear Elements) were also reported on the PIP alignment.

Percent identity alignments differ from gene prediction programs, as they are based on existing information not predictions. So, when two sequences were aligned in Vista, similarities of the sequences were shown by peaks in different colors. The *x*-axis represents the base sequence; *y*-axis represents the percent identity. As the Vista acquired the annotation files including genes and exons from NCBI databases itself, these were marked at the final plot. The direction of genes was indicated by an arrow, while the coding exons and UTRs were marked with rectangles of different colors. Conserved regions were highlighted under the curve, with red indicating a conserved non-coding region and blue indicating a conserved exon. Conserved UTRs were colored turquoise. A conserved region was defined with percentage and length cutoffs. Conserved segments with percent identity *x* and length *y* were defined to be regions in which every contiguous sub segment of length *y* was at least *x*% identical to its paired sequence.

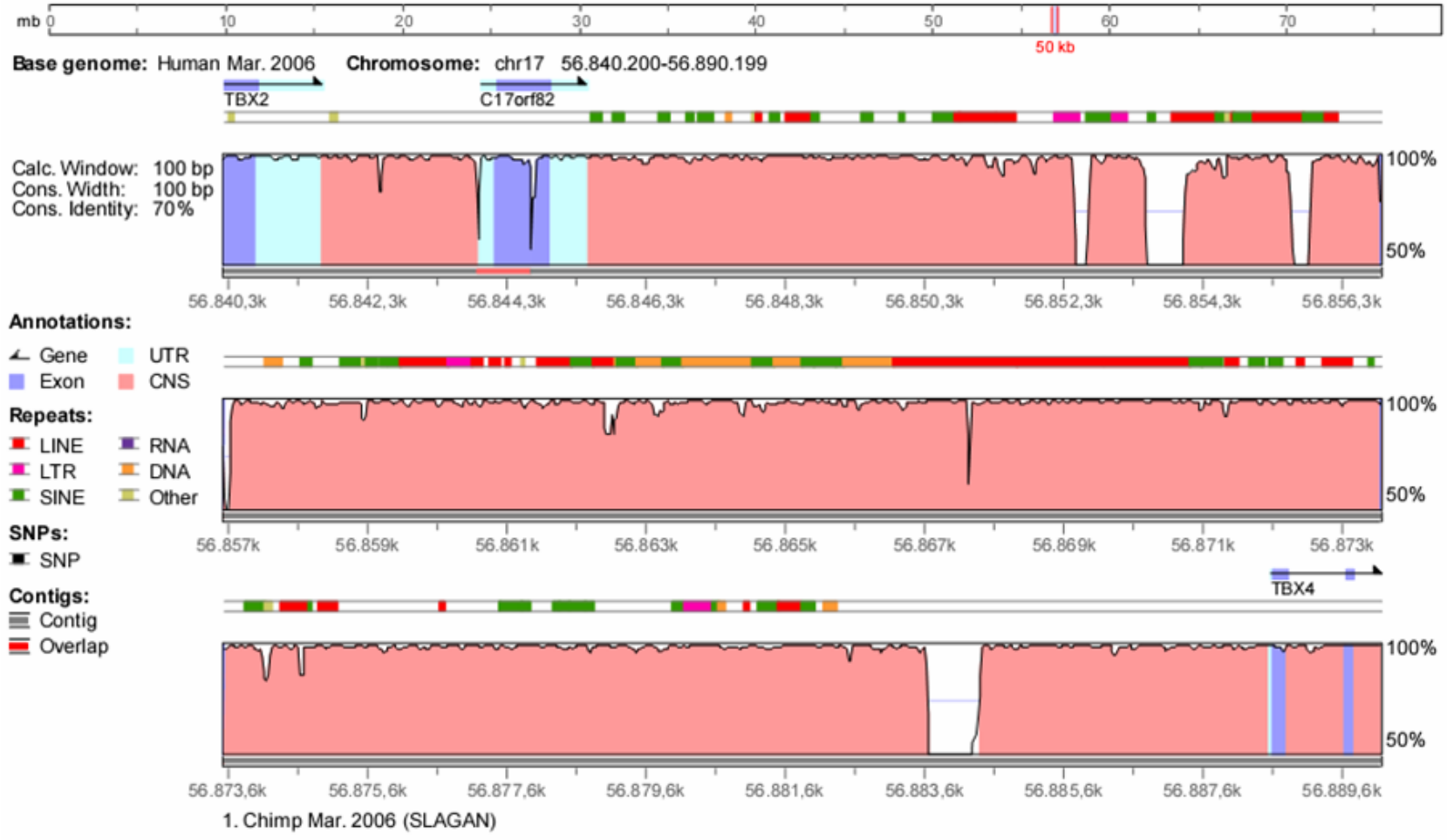

Figure 3.10 Alignment of human and chimpanzee sequences by using Vista.

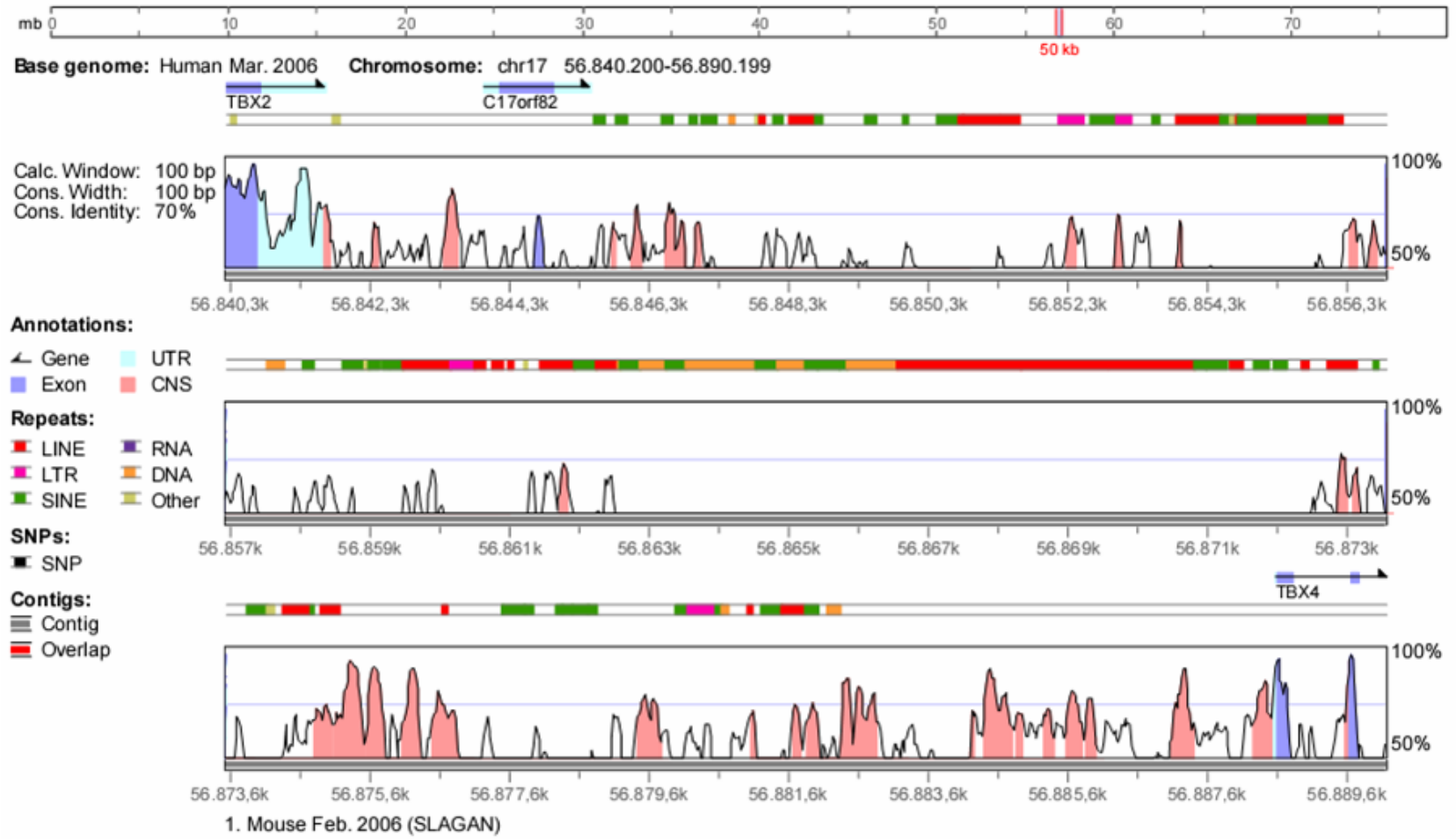

Figure 3.11 Alignment of human and mouse sequences by using Vista.

On the other hand, Multi Pipmaker program was also used for aligning 3 sequences. While using Multi PipMaker, DNA sequences were extracted from UCSC Genome Browser and turned into FASTA format, then human sequence was submitted as the main sequence and chimpanzee and mouse sequences were submitted for alignment to human sequence. RepeatMasker program was also used to mask interspersed repeats.

As can be seen from Figure 3.12, there were conserved regions especially between human and chimpanzee sequences. Also, black dots represented possible exonic regions which were found in both species. When the density of black dots increased, it indicated overlapping coding sequences.

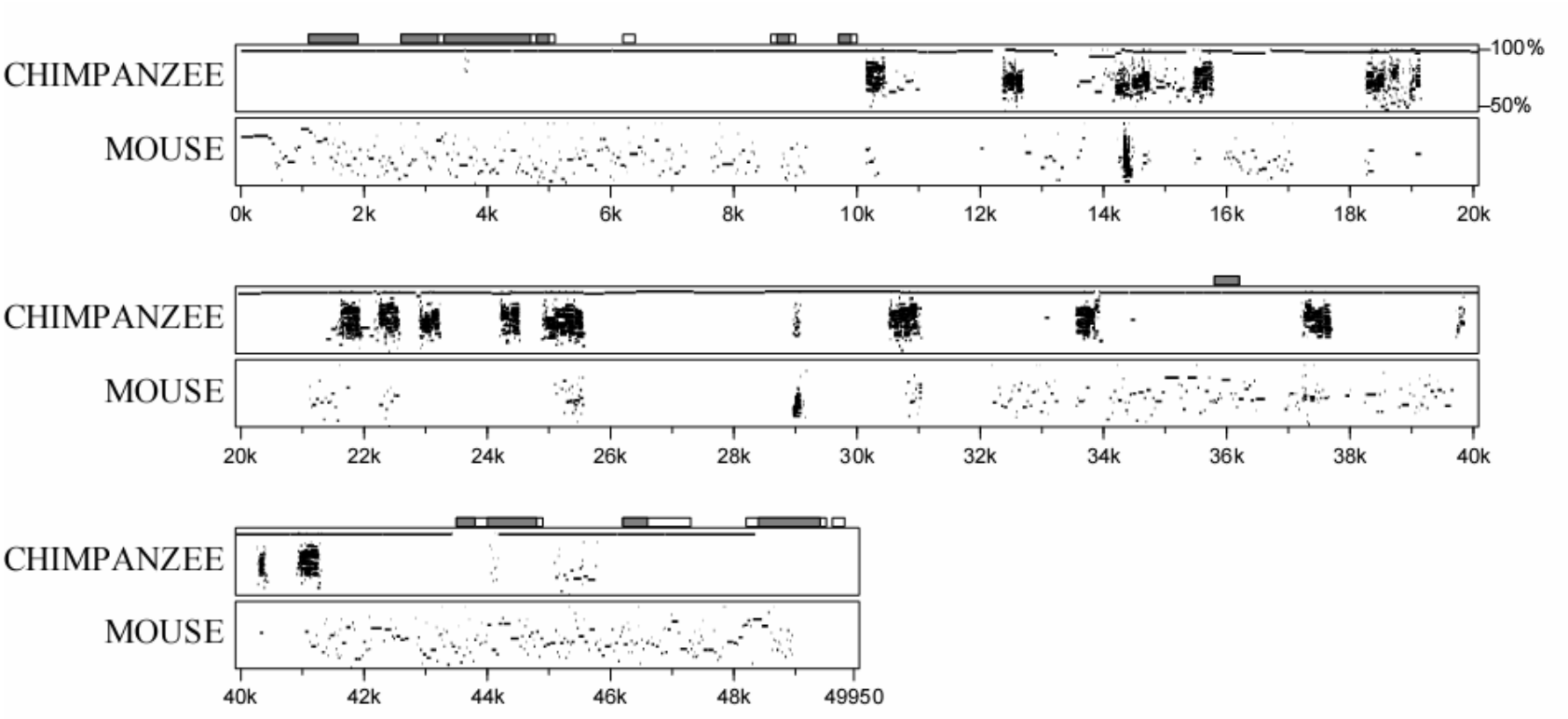

Figure 3.12 Alignment of human, chimpanzee, and mouse sequences by using MultiPipmaker.

3.1.4. Affymetrix Probes Located on 50 kb Region

Affymetrix probes which were designed from existing gene data and predicted exons (https://www.affymetrix.com/analysis/netaffx/xmlquery_ex.affx) in the 50 kb region from human sequence were determined by using UCSC Genome Browser [49]. Locations of the probes are shown in the Figure 3.13.

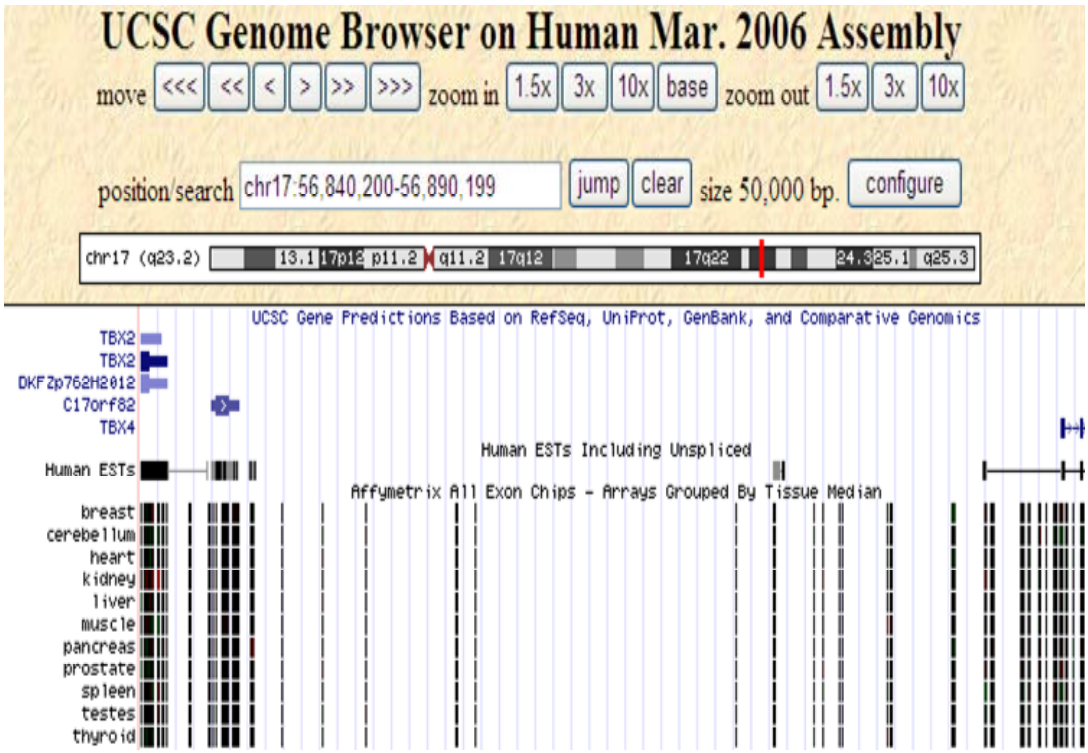

Figure 3.13 Affymetrix probe alignments to the 50 kb region in the human genome. Figure was taken from UCSC Genome Browser [49].

3.1.5. Combination of Bioinformatics Data

All the data obtained by bioinformatics combined together as shown in Figure 3.14. Blue arrows indicated the predicted exons by GENSCAN, upward

pink arrows indicated Affymetrix probes, *TBX2* and *TBX4* genes were indicated as blue arrows, hypothetical protein C17orf82 was indicated with a green arrow, human ESTs; T02811 and AA745188 were designated by purple arrows, mouse ESTS; AI449813 and CA466381 were indicated as pink arrows, and finally green bubbles showed conserved regions with chimpanzee and mouse obtained by percent alignments. This combination of data eased to comment on the 50 kb region on 17q23. As the 17q23 amplicon is a gene-rich region, presence of a gene in that 50 kb region is highly possible. *In silico* analysis also pointed that, a gene with multiple exons on that region may exists. Overlapping data coming from gene prediction analysis, percent identity alignments and existing ESTs were suggesting a gene expressed from the negative strand.

It is also possible that there may be another gene in the region because there is a human EST, namely AA745188 (positioned at 33.0 kb) which is expressed from the positive strand. Moreover, gene prediction from chimpanzee genome showed that there can be a gene expressed from positive strand. So, these information may indicate the presence of more than one gene in the region, one being expressed from the negative strand and the other from the positive strand. These two genes may also be overlapping for some part at this region.

Presence of various Affymetrix probes designed specifically for that region is another sign that there are possible exons positioned on that area as Affymetrix probe sets are designed particularly for existing genes or predicted exons.

Based on bioinformatic based evidence we collected, we switched to molecular biology tools as an attempt to clone this gene that harbors the EST T02811. RACE was the chosen technique to clone the full-length cDNA from the known EST sequence.

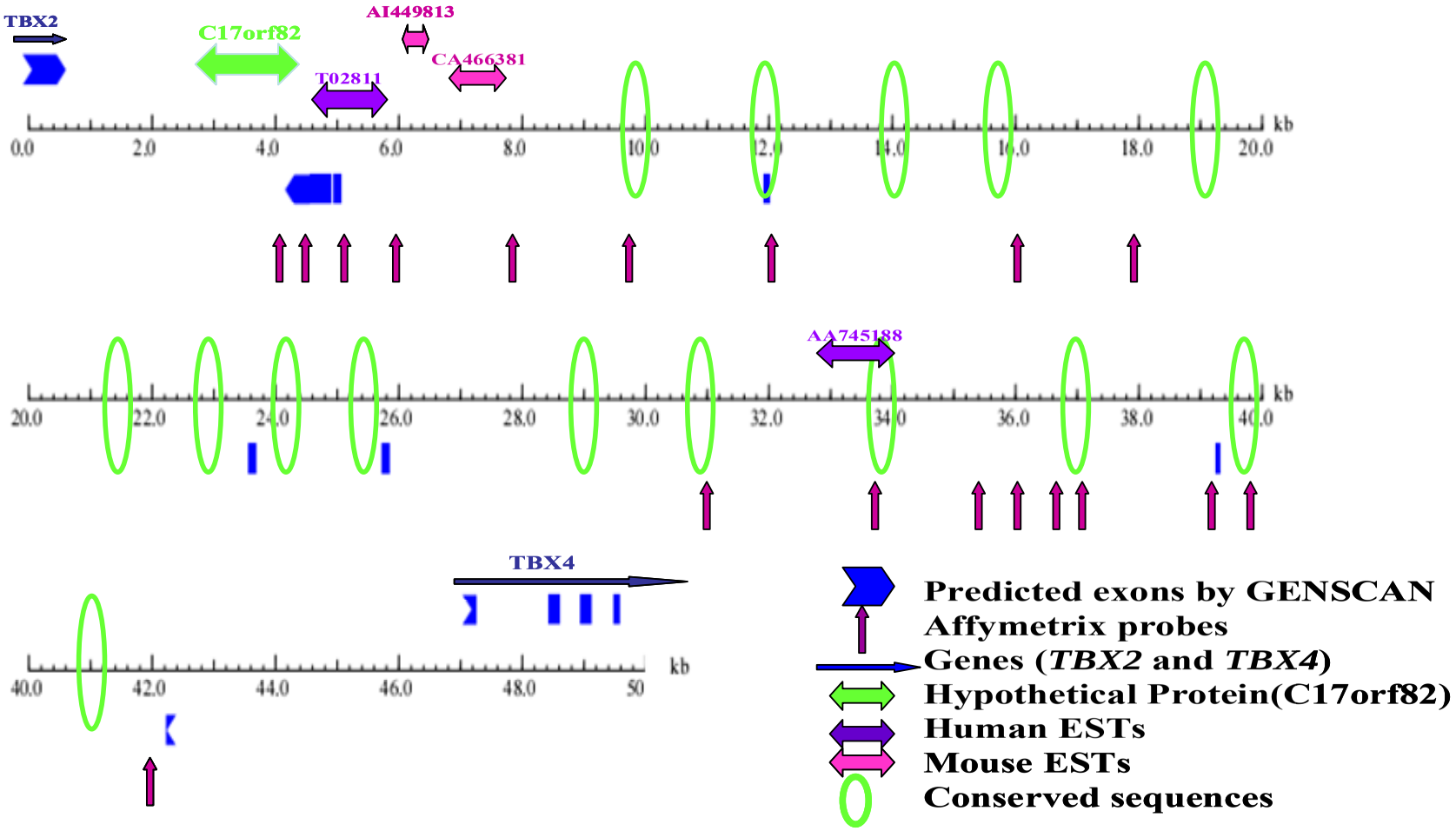

Figure 3.14 Combination of bioinformatics data.

3.2. RACE PCR Results

3.2.1. RNA Quantification and Qualification Results

After DNase treatment MCF7 RNA was quantified by spectrophotometer. The A260/A280 ratio was 1.9 which is an acceptable value for nucleic acid ratio. Quantity of the RNA sample was calculated as 2.4 µg/µL from the equation. Also, agarose gel electrophoresis was done to visualize the RNA bands as seen in Figure 3.15. RACE-ready cDNAs and other cDNAs were synthesized from that sample in further studies.

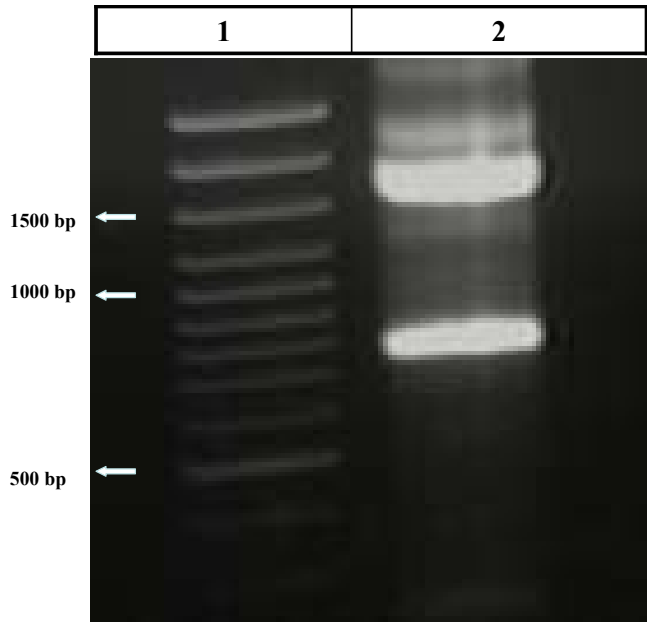

Figure 3.15 The agarose gel (1%) photograph of DNase-treated MCF7 total RNA sample. DNA marker was Gene Ruler 100 bp DNA Ladder Plus from Fermentas. The samples loaded on the gel were:

Lane 1. DNA marker **Lane 2.** DNase-treated MCF7 total RNA.

3.2.2. Verification of RACE-ready cDNA

PCR amplification was used to confirm the quality of RACE-ready cDNAs (RR-cDNA). For this purpose 2 different PCRs were performed by using *β-actin* primers and for 5' and 3' RACE. 2 samples synthesized by using MCF7 RNA (MCF7 I and MCF7 II) and control reaction which was synthesized by using HeLa RNA were used in PCR analysis. Also a blank control was used during PCR which contained all the elements of a typical PCR but not any cDNA sample. The PCR results with *β-actin* specific primers are shown in Figure 3.16.

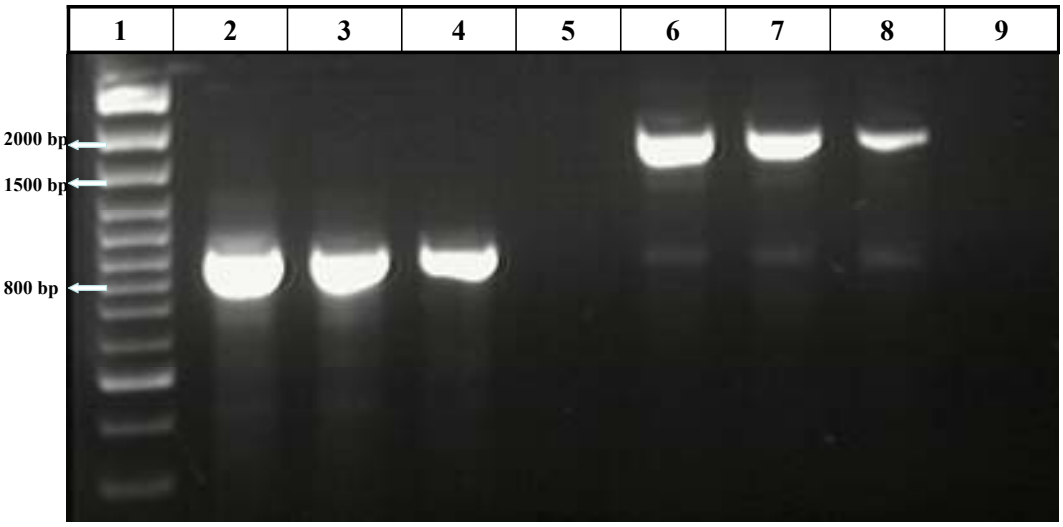

Figure 3.16 The agarose gel (1%) photograph of PCR analysis for *β-actin* gene. DNA marker was Gene Ruler 100 bp DNA Ladder Plus from Fermentas (Cat# SM0321). The samples used in PCR were:

Lane 1. DNA marker **Lane 2.** MCF7 I RR-cDNA **Lane 3.** MCF7 II RR-cDNA **Lane 4.** HeLa RR-cDNA **Lane 5.** Blank control **Lane 6.** MCF7 I RR-cDNA **Lane 7.** MCF7 II RR-cDNA **Lane 8.** HeLa RR-cDNA **Lane 9.** Blank control

It can be seen that in 3' RACE PCR, 872 bp fragment was amplified as expected which is confirmed that the 3' GeneRacer™ Oligo dT Primer was anchored properly. Similarly, 1800 bp fragment which was amplified by 5' RACE PCR verified that the 5' GeneRacer™ RNA Oligo was anchored correctly.

3.2.3. 3' RACE Results

A number of 3'RACE PCRs and nested PCRs were done in different conditions by using GSP1 and nGSP1, but there wasn't any significant PCR amplification. Positions of GSP1 and nGSP1 were shown in the Figure 3.17.

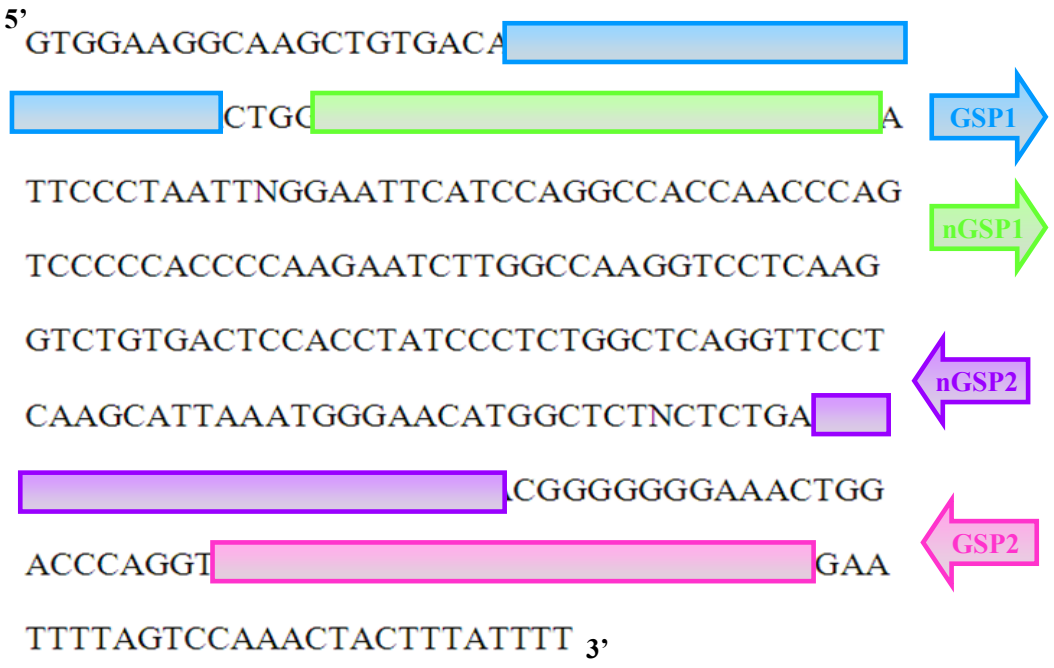

Figure 3.17 Positions and orientations of GSPs on the nucleotide sequence of T02811.

3.2.4. 5' RACE Results

While optimizing 5' RACE PCRs, first amplified product was approximately 900 bp long as can be seen from Figure 3.18. PCR was performed by using GSP2 and GeneRacer™ 5' Primer by using MCF7 I and II RR-cDNAs as template and HeLa RR-cDNA as control. A blank control was also included. This product was not obtained by touchdown PCR, so further trials were continued to remove non-specific bands on the gel. But this product was also cloned for sequencing

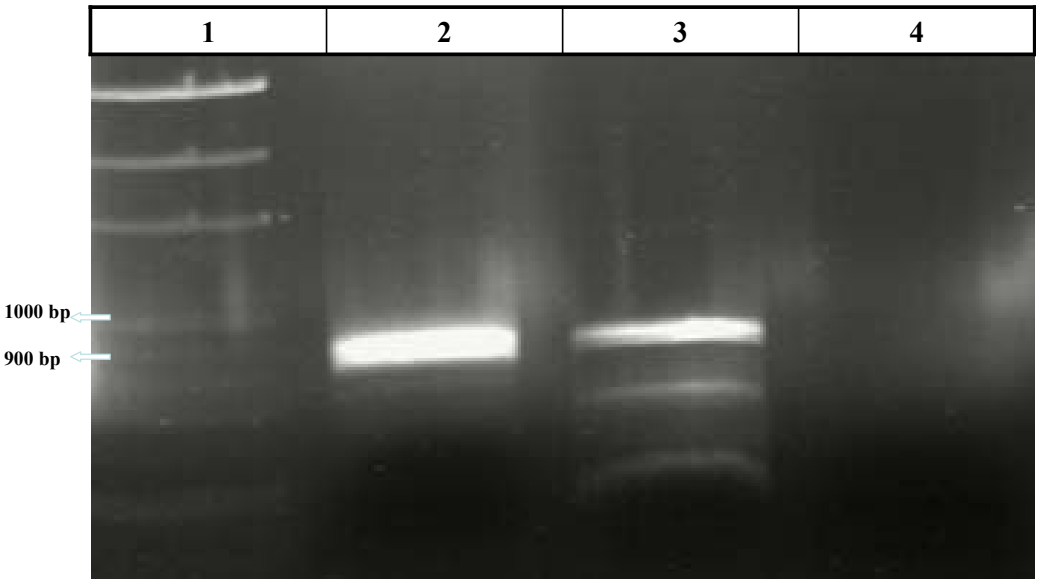

Figure 3.18 The agarose gel (1%) photograph of PCR analysis for 5' RACE. DNA marker was Gene Ruler 100 bp DNA Ladder Plus from Fermentas. The samples used in PCR were:

Lane 1. DNA marker **Lane 2.** MCF7 I RR-cDNA **Lane 3.** HeLa RR-cDNA **Lane 4.** Blank control

After optimizing the 5' RACE PCR conditions, touchdown PCR was done. GSP2 and GeneRacer™ 5' Primer were used together with MCF7 I and II RR-cDNAs as template and HeLa RR-cDNA as control. A Blank control was also included. The results of PCR are shown in Figure 3.19.

Amplified fragment was approximately 900 bp long. As the size of the transcript wasn't known, the amplified fragments were further checked by performing PCR with different GSPs and nested PCRs.

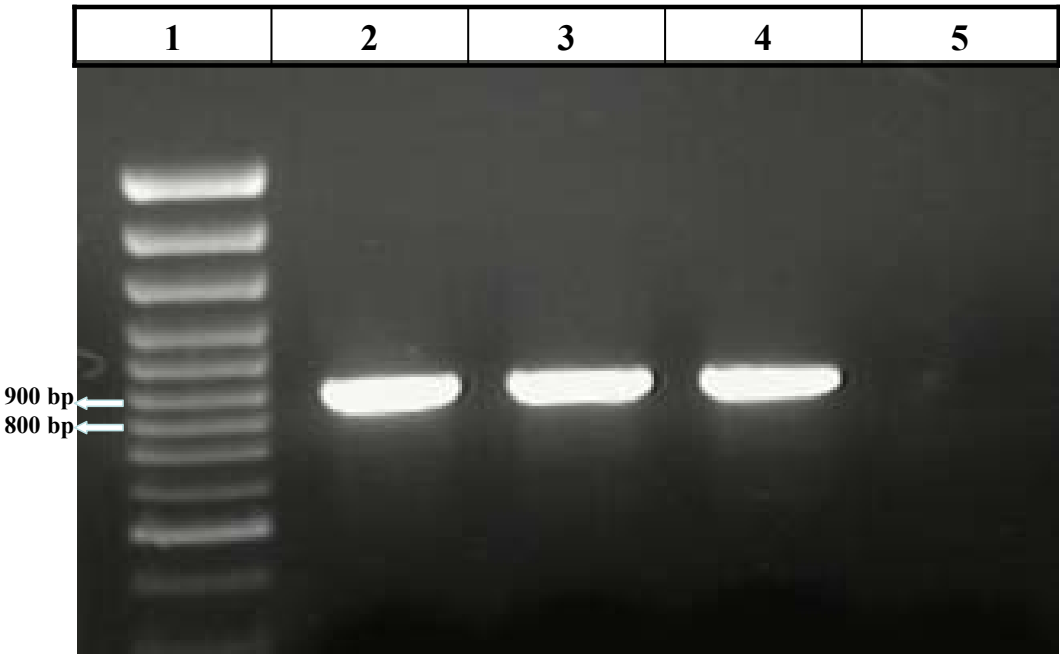

Figure 3.19 The agarose gel (1%) photograph of PCR analysis for 5' RACE. DNA marker was Gene Ruler 100 bp DNA Ladder Plus from Fermentas. The samples used in PCR were:

Lane 1. DNA marker **Lane 2.** MCF7 I RR-cDNA **Lane 3.** MCF7 II RR-cDNA **Lane 4.** HeLa RR-cDNA **Lane 5.** Blank control

For this purpose, another 5' RACE PCR was done with nGSP2 and GeneRacer™ 5' Primer. Positions of GSP2 and nGSP2 were indicated in the Figure 3.17. MCF7 II RR-cDNA was used as template. HeLa RR-cDNA was used as control sample. A blank control was also included. The results of the PCR are shown in Figure 3.20.

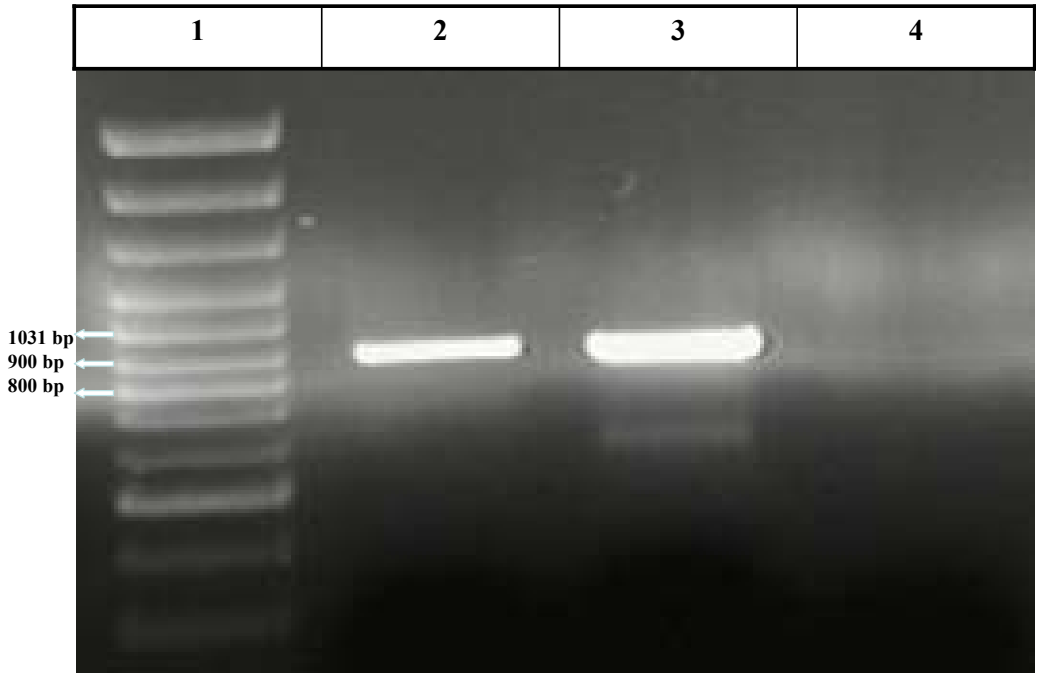

Figure 3.20 The agarose gel (1%) photograph of PCR analysis for 5' RACE. DNA marker was Gene Ruler 100 bp DNA Ladder Plus from Fermentas. The samples used in PCR were:

Lane 1. DNA marker **Lane 2.** MCF7 II RR-cDNA **Lane 3.** HeLa RR-cDNA **Lane 4.** Blank control

The amplified fragment was approximately 900 bp in both samples. This result was consistent with the previous PCR. So, additional confirmation was done by performing nested PCRs.

3.2.5. Nested 5' RACE Results

Nested PCRs were done by using PCR products as template and using nested primers designed from inner sequence to check if the amplified fragment was really the expected product. Positions of nested primers; GSP2-3 and GSP2-4 together with GSP2 and nGSP2 were shown in the Figure 3.21.

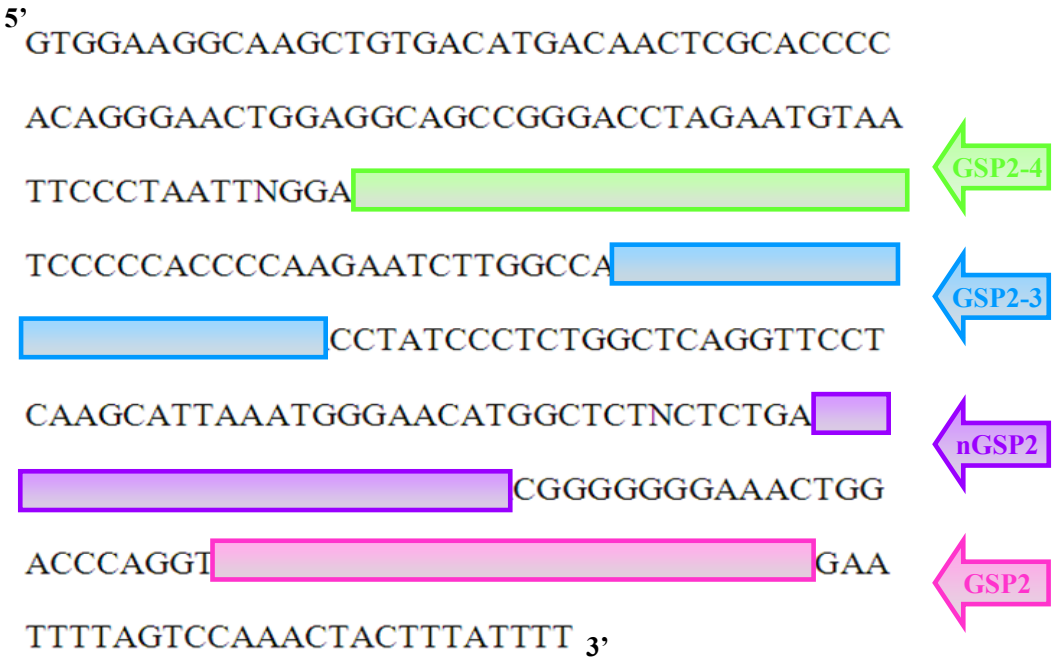

Figure 3.21 Positions and orientations of reverse GSPs and nested GSPs on the nucleotide sequence of T02811.

For this purpose, 2 different master mixes were prepared with different nested primers; in the first master mix GSP2-3 and the GeneRacer™ 5' Nested Primer were used. In the second master mix, GSP2-4 and GeneRacer™ 5' Nested Primer were used. MCF7 II and HeLa PCR products which were obtained by GSP2 and GeneRacer™ 5' Primer were the templates for both master mixes. A blank control was included as well. The PCR results are shown in Figure 3.22.

According to the results of nested PCRs, amplified fragments were around 900 bp, again. Faint bands which were visualized as 500 bp probably resulted from non-specific binding of nested primers.

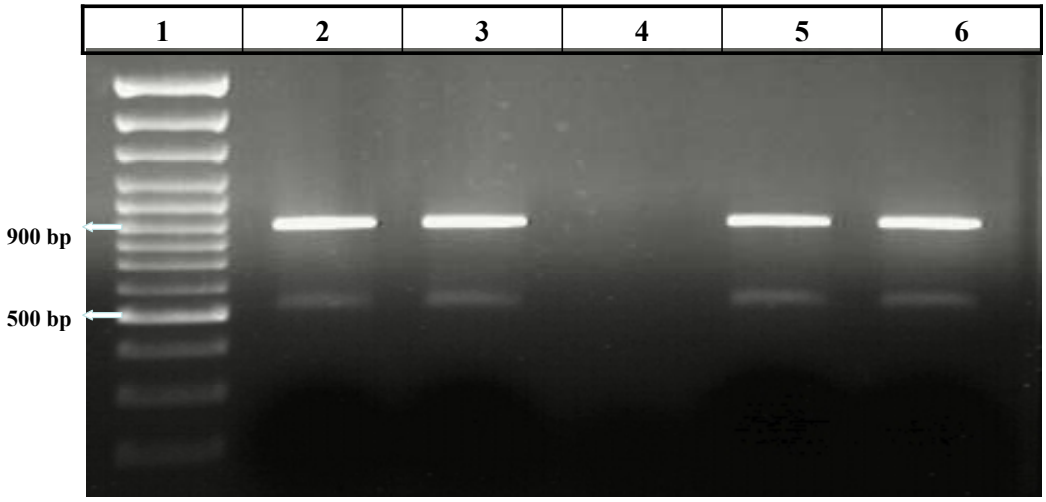

Figure 3.22 The agarose gel (1%) photograph of PCR analysis for nested 5' RACE. DNA marker was Gene Ruler 100 bp DNA Ladder Plus from Fermentas. The samples used in PCR were:

Lane 1. DNA marker **Lane 2.** MCF7 II PCR Product **Lane 3.** HeLa PCR Product **Lane 4.** Blank control **Lane 5.** MCF7 II PCR Product **Lane 6.** HeLa PCR Product

These primers; GSP2, nGSP2, GSP2-3, and GSP2-4 (Appendix C) as can be seen in the Figure 3.21 were designed close only 20 bp apart from each other. Together, two 5' RACE PCR results and two nested 5' RACE PCR results were consistent with each other, all the amplified fragments pointing out a product of approximately 900 bp. So, amplified fragments from first round PCRs were isolated from agarose gel and cloned. Three different PCR products obtained by GSP2 and nGSP2 primers were cloned by TOPO TA cloning into pCR®4-TOPO® vector.

3.2.6. Cloning and Sequencing Results of RACE Products

After cloning the amplified fragments into pCR®4-TOPO®, transformation into TOP10 cells were performed. After isolating plasmids from *E.coli* cells, these plasmids were digested with *Eco*RI restriction enzyme to confirm the size of inserts. Digested plasmids were loaded on agarose gel as seen in Figure 3.23. As the sequence of the transcript wasn't known, restriction enzyme sites were not known either. So, adding the sizes of bands was the only way to determine if the insert was 900 bp or not.

From the agarose gel, it was clear that all the inserts were correct. 3 of these plasmids; #2 (PCR product of GSP2, Figure 3.19), #5 (PCR product of nGSP2, Figure 3.20) and #8 (PCR product of GSP2, Figure 3.18) which were designated with the stars were sent to sequencing.

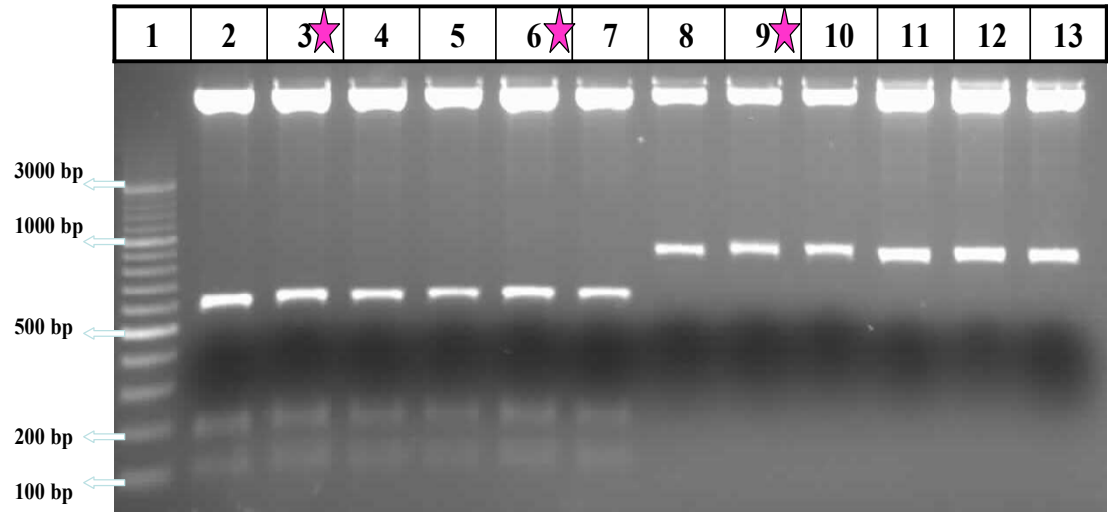

Figure 3.23 The agarose gel (2%) photograph of *Eco*RI digestion. DNA marker was Mass Ruler DNA Ladder Plus from Fermentas. The samples loaded on agarose gel were:

Lane 1. DNA marker **Lane 2-13.** Digested plasmids (#1-12).

3.2.6.1. Sequence Analysis of 5' RACE PCR products

T3 and T7 sequencing primers (Appendix C) from the pCR®4-TOPO® vector were used during sequencing of the PCR products. The results of sequencing analysis were investigated using Chromas (v2.23) software program. After analyzing the chromatogram, FASTA format of the sequences were used in BLAST [48] program of NCBI in order to find similar sequences in human genome. Sequence analysis was carried out in order to obtain the 5' sequence of the transcript. The analysis was done with the products of PCRs which were carried out with MCF7 II RR-cDNA and the gene specific primers GSP2 (two different PCR product were cloned, #2 and #5) and nGSP2 (#8). Alignment with human genome showed that amplified fragments didn't give hit to chromosome 17.

Instead, they matched to other genomic sequences from various parts of human genome.

#2 and #5 gave hit to chromosome 3 (GI # 28974991). Two results coming from #2 and #5 were 100% identical to each other. This identity was not surprising, as the nested PCRs done by PCR products worked similarly for this two fragment at the first place. But it is interesting that 4 GSPs (GSP2, nGSP2, GSP2-3, and GSP2-4) designed arbitrarily from the sequence of T02811 were annealed to another region of human genome and amplified that region. This can be explained by existence of similar sequences resembling T02811 on other chromosomes. So, GSPs annealed to these sequences.

On the other hand, #8 gave hit to only chromosome 1, specifically to the transmembrane protein 125 (*TMEM45)*. This result showed that, even if the GSP was same in a PCR, cycling conditions could affect the result. Because only difference between #2 and # 8 was the cycling conditions. #2 was performed with touchdown PCR which improves the specificity of PCR. On the other hand, #8 was performed at 65 ^{o}C Tm.

These results showed that, GSPs were not specific enough to amplify T02811. Although amplified fragments were very clear on gel and nested PCRs worked as well, an unintended cDNA was amplified.

Besides the unspecific binding of GSPs, another explanation is that the transcript (T02881) belongs to a gene family, so that there could be conserved motifs in the sequence of the transcript. This leads to amplification of other genes from other chromosomes sharing similar motifs. As the known sequence of the EST is 313 bp, this sequence was very limited to design new GSPs for RACE. So, it was necessary to know more sequence from T02811 transcript.

3.3. Results of PCR Analysis to Extend the Size of Transcript

Various primers were designed in order to amplify fragments including T02811. These PCR products were cloned and sent to sequencing individually.

3.3.1. PCR Results by Using T02811BKout Primers

In Figure 3.24 positions of T02811BKout primers are shown with respect to T02811. As can be seen from Figure 3.25 T02811BKoutF and T02811BKoutR primers amplified a fragment both from MCF7 genomic DNA and MCF7 cDNA. Also, DNase-treated RNA which was the source of cDNA was used as negative contol and confirmed that cDNA was not contaminated with DNA.

This result showed that the transcript was larger then 313 bp and there was no intronic site in this 494 bp part of the transcript as the size of genomic DNA and cDNA was equal.

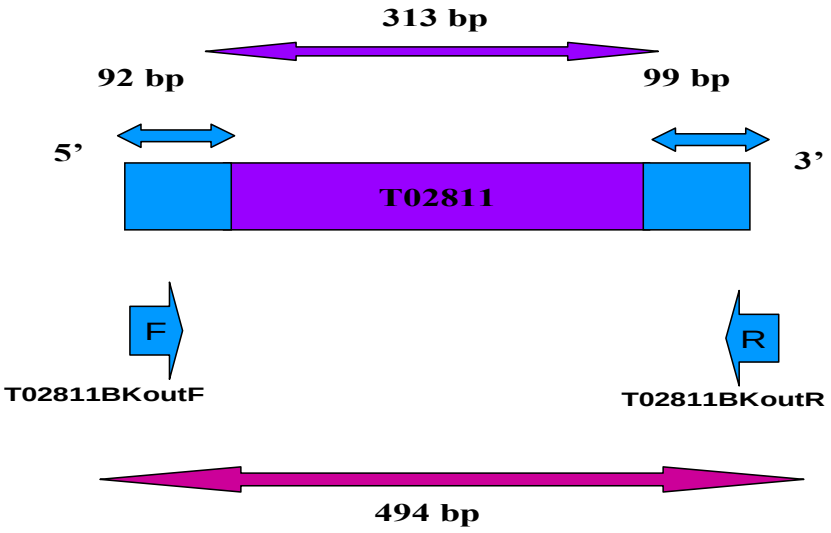

Figure 3.24 T02811BKout Primers for extending T02811 transcript to 494 bp.

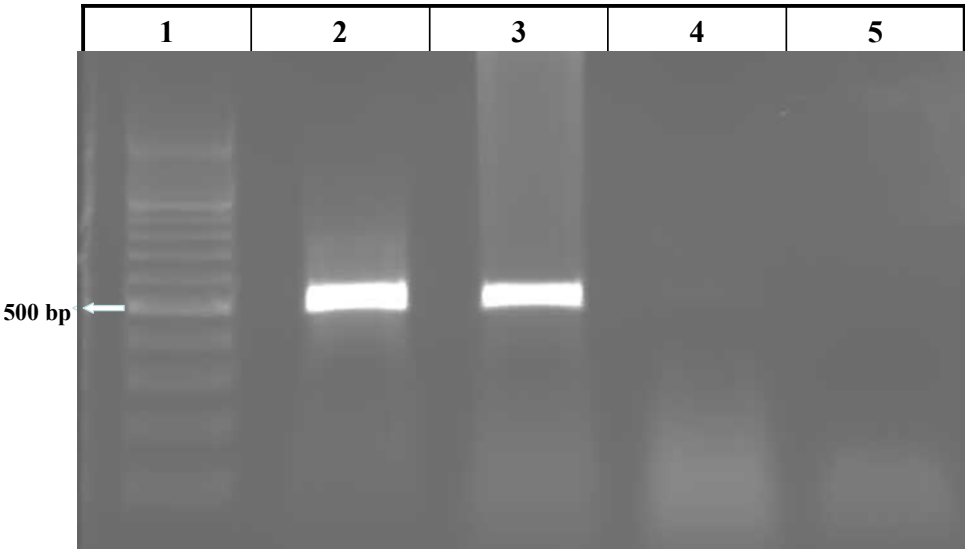

Figure 3.25 The agarose gel (1%) photograph of PCR analysis for extending T02811. DNA marker was Gene Ruler 100 bp DNA Ladder Plus from Fermentas. The samples used in PCR were:

Lane 1. DNA marker **Lane 2.** MCF7 DNA **Lane 3.** MCF7 cDNA **Lane 4.** MCF7 RNA **Lane 5.** Blank control

3.3.2. GSP Design from Extended Sequence and 5'RACE

In order to perform new RACE PCRs, it was necessary to design new GSPs from the extended sequence. So by the help of the BLAST tool, blastn algorithm was used for aligning 494 bp sequence with the human genome. This provided to skip regions which were giving hit to other chromosomes. As can be seen from Figure 3.26, red line is the 494 bp sequence aligning to chromosome 17. On the other hand, green, blue and black lines are representing other sequences from the human genome which are giving hit to the 494 bp partially. So, new GSPs were designed from the regions which are designated with stars to avoid non-specific annealing to these sequences in the genome.

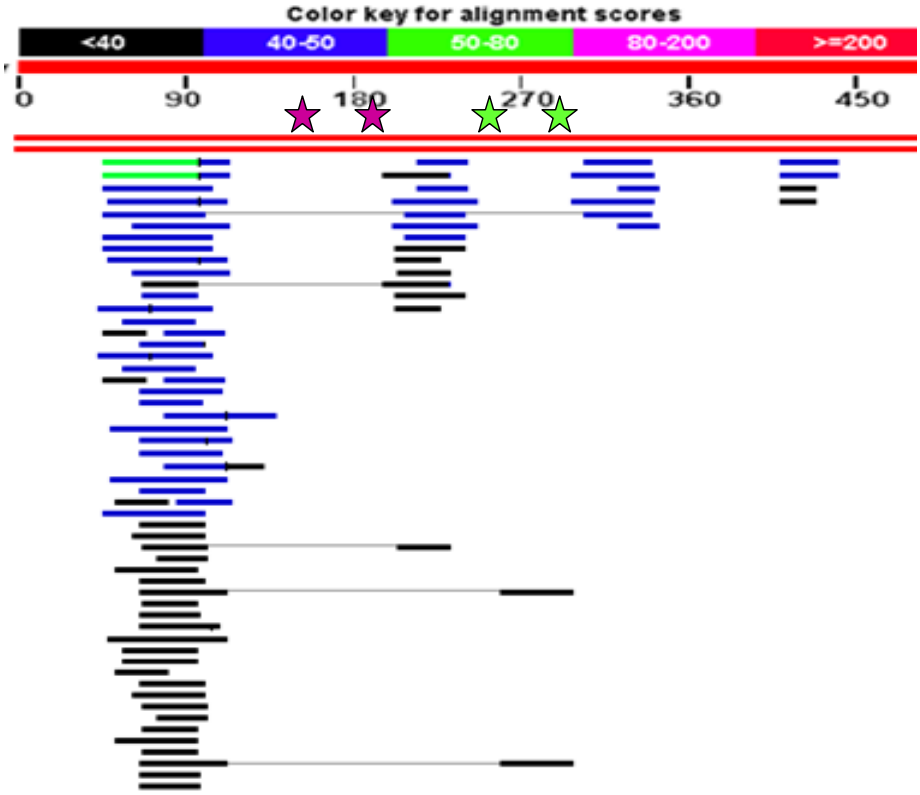

Figure 3.26 GSPs designed from 494 bp sequence specified on BLAST figure. Pink and green stars are indicators of reverse and forward GSPs, respectively.

After the primers arrived, new RACE PCRs were set up. Unfortunately, neither 3' RACE nor 5' RACE PCRs worked.

3.3.3. Extending T02811 from 5'

In order to design new GSPs, it was required to know more sequence belonging to the T02811 transcript. So, by using the data obtained by *in silico* tools, PCR primers were designed from the 5' of T02811. These primers were designed especially from predicted exons to guarantee they could amplify cDNA as well as DNA.

There exists a hypothetical protein, which is called C17orf82 or LOC388407 at the 5' of the T02811 [86]. C17orf82 was identified as an open reading frame through bioinformatics; it was available in the databases. It is 251 amino acid long and suggesting the function of collagen alpha 1 gene family according to the UniGene database of NCBI. mRNA sequence is 1546 bp long. Since, C17orf82 is 500 bp away from T02811, primers were designed from the sequence of C17orf82 as Fx, Fy, and Fz. PCR optimizations were done by using DNA and only Fy worked from genomic DNA, so PCR was performed with DNA; cDNA, and RNA samples. Positions of primers used in the PCR and expected sizes can be seen in Figure 3.27. PCR result with Fy and T02811BKoutR primer is shown in Figure 3.28.

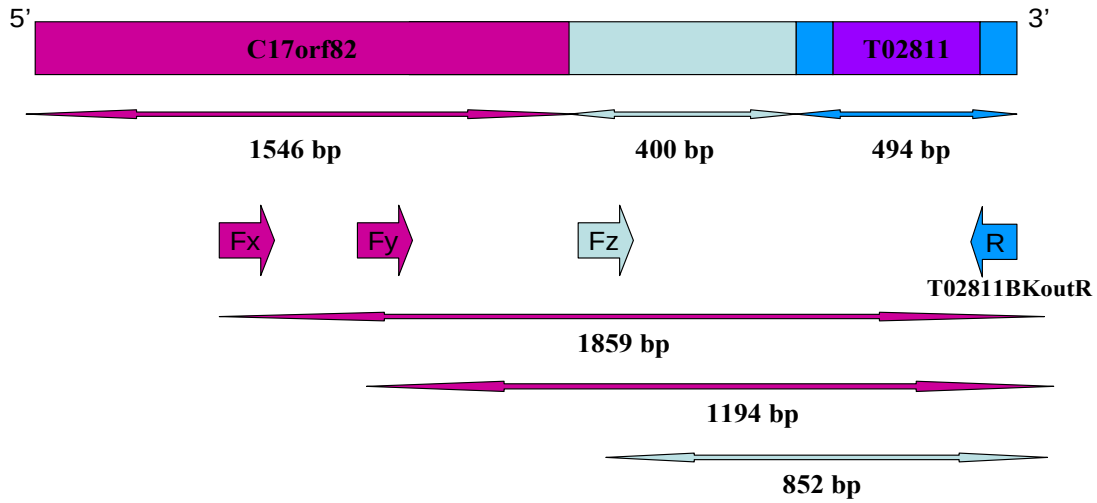

Figure 3.27 Primers for extending T02811 transcript from 5'.

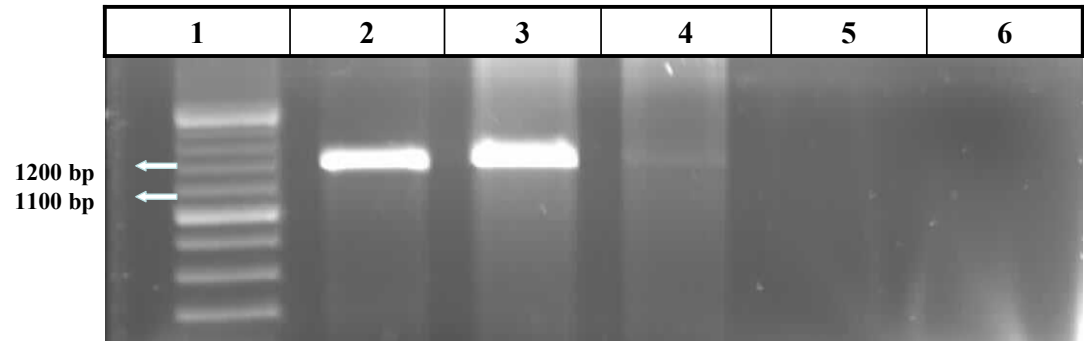

Figure 3.28 The agarose gel (1%) photograph of PCR analysis for extending T02811 from 5'. DNA marker was Gene Ruler 100 bp DNA Ladder Plus from Fermentas. The samples used in PCR were:

Lane 1. DNA marker **Lane 2.**MCF7 DNA **Lane 3.** MCF7 cDNA **Lane 4.** Breast cDNA **Lane 5.** MCF7 RNA **Lane 6.** Blank control

This 1194 bp cDNA PCR product was cloned into pCR®4-TOPO® and sent to sequencing. Sequence results showed that cDNA sequence was the same with the Human genomic sequence meaning there is not any intron, which is consistent with PCR result. Interestingly, this sequence is covering 300 bp from C17orf82, which is a hypothetical protein expressed from positive strand. As T02811 is a negative-strand transcript, this result can be explained as there can be two overlapping genes on that region, one of them expressed from positive and other is expressed from negative strand.

3.3.4. GSP Design from Extended Sequence and 5' RACE PCR

To order new GSPs, 1194 bp sequence was BLASTed and overlapping sequences shown with green and blue lines were skipped. New GSPs were designed from the regions indicated with stars on the Figure 3.29.

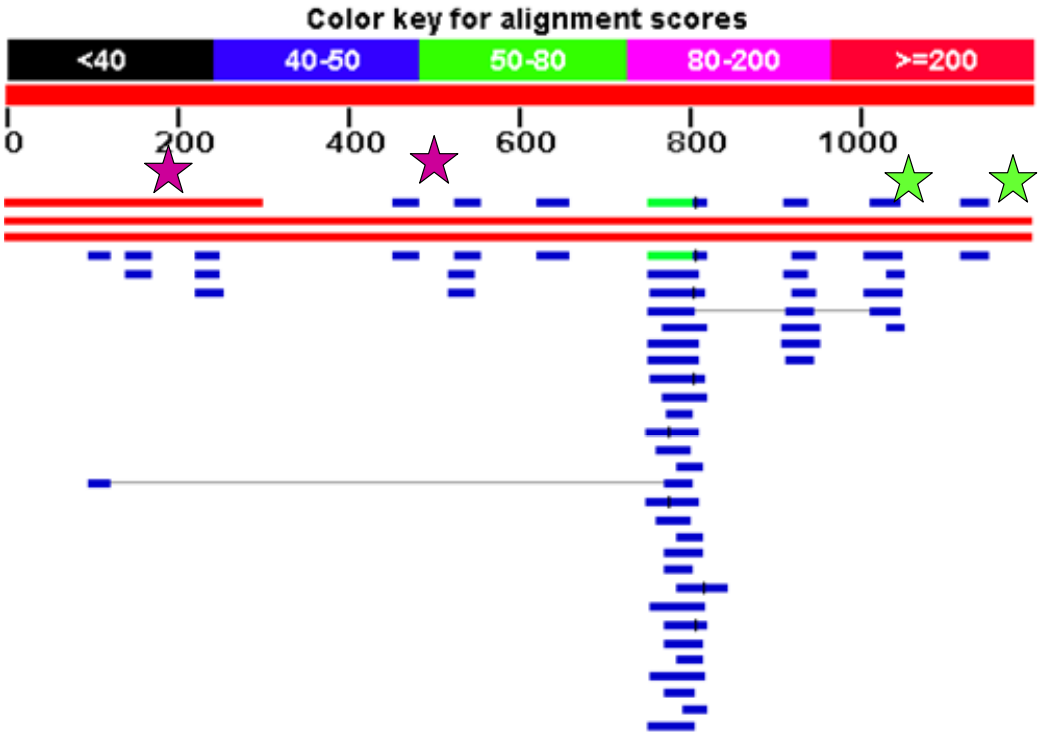

Figure 3.29 GSPs designed from 1194 bp sequence specified on BLAST figure. Pink and green stars are indicators of reverse and forward GSPs, respectively.

Following task was performing RACE with new GSPs. This time 5' RACE PCR was worked with one of the new GSPs (GSP_170_195) as can be seen from Figure 3.30. Amplified fragment was approximately 850 bp and it was cloned and sent to sequencing.

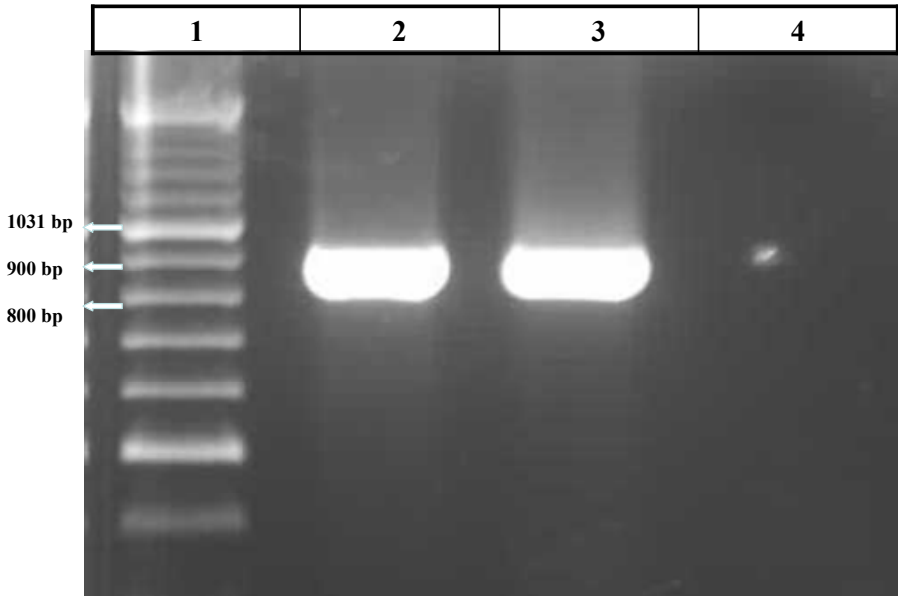

Figure 3.30 The agarose gel (1%) photograph of PCR analysis for 5' RACE. DNA marker was Gene Ruler 100 bp DNA Ladder Plus from Fermentas. The samples used in PCR were:

Lane 1. DNA marker **Lane 2.**MCF7 RR-cDNA **Lane 3.** HeLa RR-cDNA **Lane 4.** Blank control

When the sequence result arrived, it came up that the amplified fragment didn't correspond to T02811 for the second time. Interestingly, BLAST results showed that sequence belonged chromosome 3 and, also this sequence is 100% identical with the first 5' RACE results which is aligning to chromosome 3. Although the GSPs used in PCR were different, sequence results came up identical for three times. An explanation for this, the sequence of T02811 transcript is similar to that part of chromosome 3, so that primers annealed to that region. When T02811 and this region from chromosome 3 aligned via bl2seq algorithm of BLAST tool [48] , significant similarity was not found. Although there is not any sequence alignment, existence of secondary cDNA structures may

cause this unwanted annealing more than one time. Another reason for this result can be that this transcript belongs to a gene family. As a result of conserved motifs between these genes, GSPs were bound to the gene located on chromosome 3. Still, another explanation; there could be pseudogenes sharing the similar sequences which provided the annealing of GSPs to different parts in the genome.

3.3.5. Extending T02811 from 3'

Similar to the 5' extension, 3' extension was done by using bioinformatics data. Possible exon sites were considered during primer design. Besides human ESTs, there are several mouse ESTs that exist in the 50 kb region, so primers were designed from the genomic region where 2 of these ESTs, namely AI449813 and CA466381. As T02811 is expressed from negative strand, mouse ESTs which are expressed from negative strand were chosen. Both of these ESTs are located on 3' of T02811, so reverse primers were designed after BLASTing the mouse sequences to human genome.

AI449813 is aligning to the human genome 100% for 27 bp, and CA466381 is aligning to the human genome 100% for 50 bp. So, reverse primers were designed from these hits as shown in Figure 3.31. While optimizing the PCR conditions by using DNA, M_EST_50 didn't work, so only M_EST_27 was used in PCR analysis.

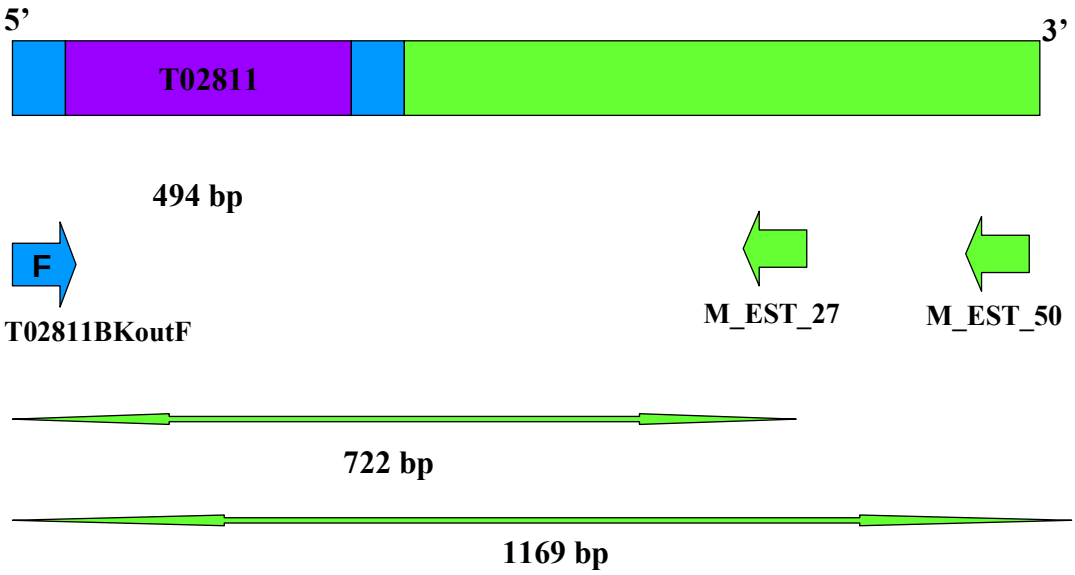

Figure 3.31 Primers for extending T02811 transcript from 3'.

As can be seen from Figure 3.32, PCR was done by using DNA, cDNA and RNA samples extracted from MCF7. Amplified fragments were the same length from DNA and cDNA, so that result pointed out that there still was no intron at the 3' yet. cDNA PCR product was cloned in to pCR®4-TOPO® and sent to sequencing to confirm the sequence. Finally, sequencing results showed that the transcript was extending to the 3'.

Moreover, as the reverse primer which was used in the PCR amplification was designed from a mouse EST (AI449813) which aligned to negative strand on mouse genome, T02811 and AI449813 belong to the same gene which is conserved in two species.

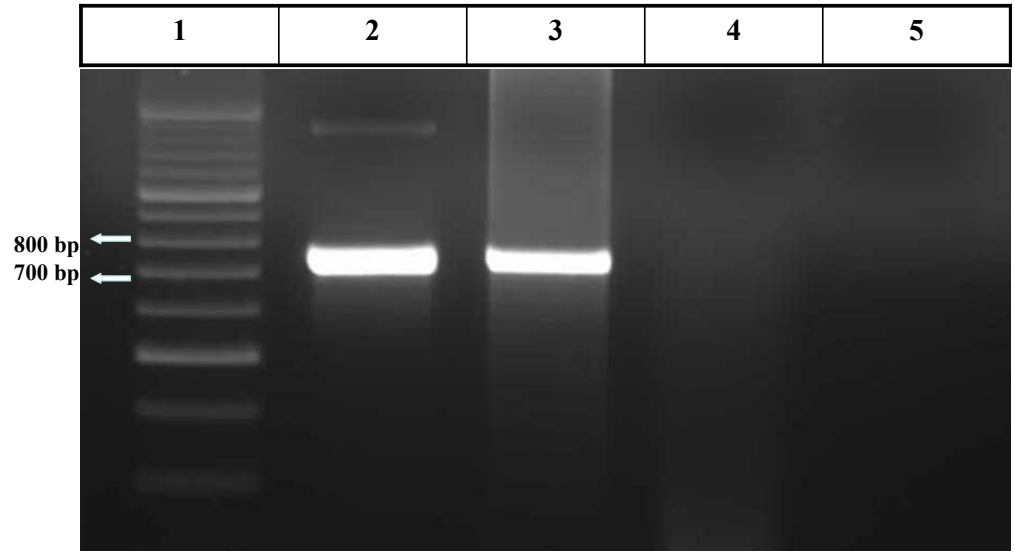

Figure 3.32 The agarose gel (1%) photograph of PCR analysis for extending T02811 from 3'. DNA marker was Gene Ruler 100 bp DNA Ladder Plus from Fermentas. The samples used in PCR were:

Lane 1. DNA marker **Lane 2.**MCF7 DNA **Lane 3.** MCF7 cDNA **Lane 4.** MCF7 RNA **Lane 5.** Blank control

3.3.6. Extended Size of the T02811 Transcript

Via 5' and 3' PCRs, size of the T02811 was extended to 1423 bp as can be seen from Figure 3.33. As confirmed by sequence analysis, there was still no intron in that 1423 bp region.

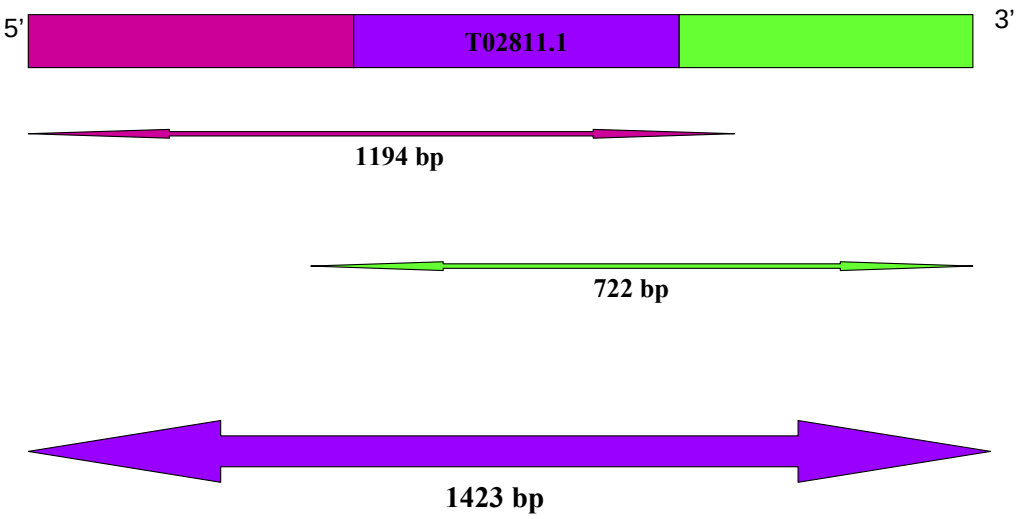

Figure 3.33 Extended size of the transcript.

3.3.7. ORF Analysis of the Extended Transcript

ORF Finder of NCBI [48] was used to identify possible ORFs from 1423 bp sequence. 6 different frames acquired by ORF Finder program are shown in Figure 3.36. The existence of an ORF, especially a long one, is usually a good indication of the presence of a gene. According to this result, we didn't detect a continuous reading frame. As the full-length transcript is still not known, it will be useful to resubmit the full-length cDNA sequence to the ORF Finder program after further cloning of the RACE products.

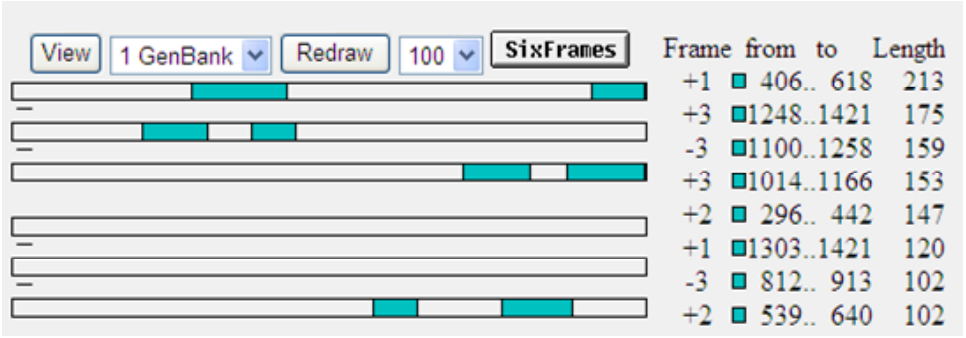

Figure 3.34 Six possible ORFs of 1423 bp transcript.

3.3.8. GSP design from 1423 bp and 3' RACE PCR

As the size of transcript extended to 1423 bp, new GSPs were designed from that sequence. Positions of GSPs are indicated in Figure 3.34.

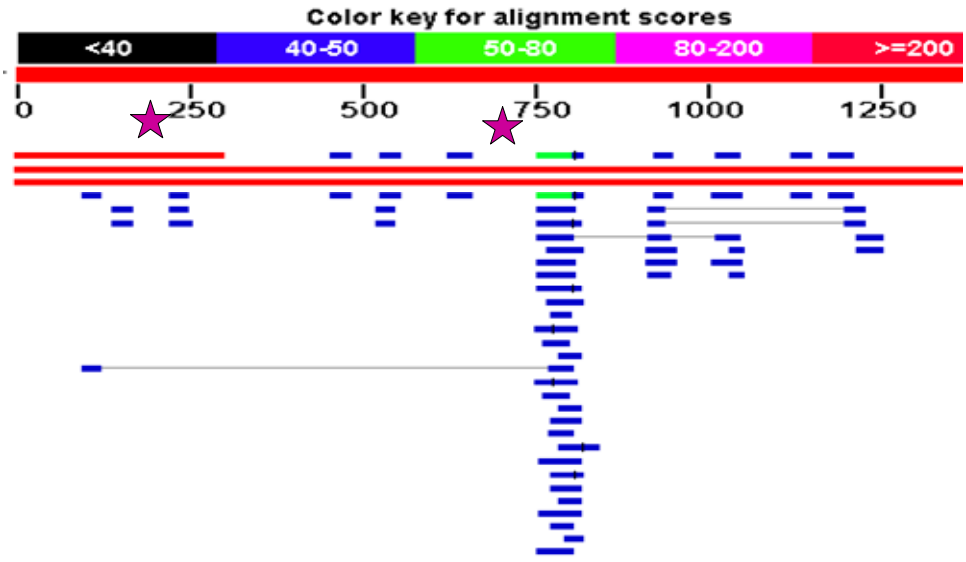

Figure 3.35 GSPs designed from 1423 bp sequence specified on BLAST figure. Pink stars are indicators of overlapping reverse and forward GSPs.

When the GSPs arrived, 3' RACE PCR was done by using MCF7 cDNAs synthesized by Roche's 3'/5' RACE Kit, 2nd Generation (Cat# 03353621001). Also Roche's Expand Long Template PCR System (Cat# 11681834001) was used as DNA polymerase mix in PCR.

From that PCR, as can be seen in Figure 3.35 two fragments were amplified which were approximately 800 bp and 550 bp in size. Although PCR conditions were changed to eliminate lower-sized fragment, it did not disappear. So, both of the products were cloned to plasmid for sequencing as they can be alternatively spliced forms of the same gene.

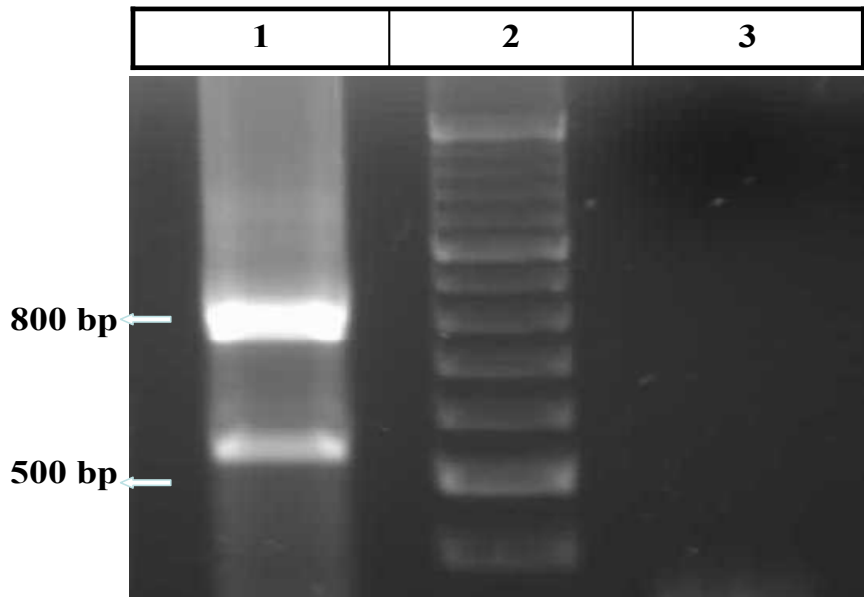

Figure 3.36 The agarose gel (1%) photograph of 3' RACE PCR analysis. DNA marker was Gene Ruler 100 bp DNA Ladder Plus from Fermentas. The samples used in PCR were:

Lane 1. MCF7 RR-cDNA **Lane 2.** DNA marker **Lane 3.** Blank control

3.4. Duplex RT-PCR (D-RT-PCR) Results

Overexpression of T02811 was checked by D-RT-PCR by using MCF7, MDA-MB231 and normal breast cDNA. GAPDH primers were used as an internal control, as the GAPDH is a housekeeping gene and housekeeping genes are expressed stable in different tissues, they can be used to normalize duplex PCR conditions.

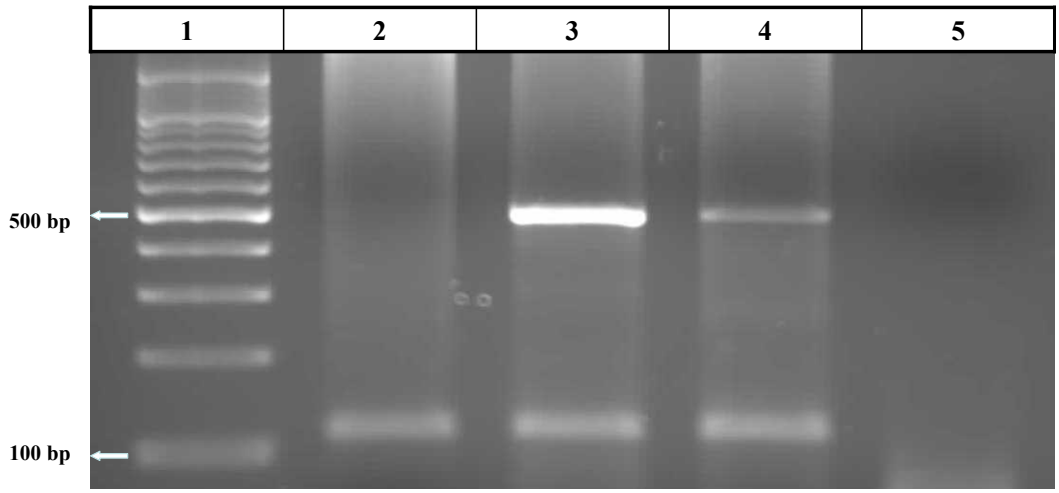

Figure 3.37 The agarose gel (1%) photograph of D-RT-PCR analysis of T02811. DNA marker was Gene Ruler 100 bp DNA Ladder Plus from Fermentas. The samples used in PCR were:

Lane 1. DNA marker **Lane 2.**MDA-MB231 cDNA **Lane 3.** MCF7 cDNA **Lane 4.** Breast cDNA **Lane 5.** Blank control

As can be seen from the Figure 3.36 D-RT-PCR result we detected overexpression in MCF7 cell line. GAPDH intensities were equal in all three of samples which provide comparison between T02811 expression levels. As expected, MCF7 was showing high expression compared to normal breast sample.

On the other hand, MDA-MB231 breast cancer cell line did not show any T02811 expression at all. This can be explained by this cell line was established from the pleural effusion which may represent a different cell population compared to MCF7 or the normal breast tissue.

CHAPTER IV

CONCLUSION

The objective of the present study was to analyze the 50 kb region between *TBX2* and *TBX4* genes positioned on 17q23 of the human genome.

To achieve this, first of all, bioinformatics tools were used to determine the boundaries. Then in addition to human sequence, corresponding sequences from chimpanzee and mouse were used for comparisons. EST data was used to identify possible genes positioned on this region. While analyzing the region, an EST; T02811 was identified which has a noticeably high expression profile in breast cancer cell line MCF7. So, our study focused on this EST, to identify and characterize it by using RACE. Unfortunately, both RACE trials from 3' and 5' were unsuccessful. As the T02811 is a short EST of 313 bp, designing specific primers was very challenging. Moreover, sequencing results revealed that this transcript could belong to a gene family, so presence of conserved motifs lead to non-specific amplification during RACE PCRs. So, extending the transcript's size by PCR was the following task.

By using the bioinformatics data coming from human, chimpanzee and mouse, PCR primers were designed both from 3' and 5' end of T02811. Positions of mouse ESTs, exon prediction analysis, Affymetrix's probe positions and PIP results assisted to design primers from predicted exon sites. In order to extend the size of transcripts, PCRs were set up by using genomic DNA and cDNA, in order to understand if there are intron sites present or not. PCR analysis from 3' extended the transcript from 313 bp to 1194 bp. Similarly, 5' extension led to 722

bp product. Finally, after extracting overlapping sequences, the as of yet incomplete size of the transcript was extended to 1423 bp.

Future work would include identifying full-length cDNA of T02811. After cloning the full-length transcript, further studies will be held to characterize the role of the gene in breast cancer as an oncogene candidate mapping to the 17q23 amplicon. This gene is a potential oncogene candidate as it is overexpressed and amplified in breast cancer cell lines. Thus, characterizing this gene and identifying its cellular function will be important.

REFERENCES

1. Jemal, A. et al. Cancer statistics, 2007. *CA Cancer J Clin* **57**, 43-66 (2007).

2. Mathers, C.D. & Loncar, D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* **3**, e442 (2006).

3. Kerangueven, F. et al. Genome-wide search for loss of heterozygosity shows extensive genetic diversity of human breast carcinomas. *Cancer Res* **57**, 5469-74 (1997).

4. Beckmann, M.W., Niederacher, D., Schnurch, H.G., Gusterson, B.A. & Bender, H.G. Multistep carcinogenesis of breast cancer and tumour heterogeneity. *J Mol Med* **75**, 429-39 (1997).

5. Ingvarsson, S. Molecular genetics of breast cancer progression. *Semin Cancer Biol* **9**, 277-88 (1999).

6. Zody, M.C. et al. DNA sequence of human chromosome 17 and analysis of rearrangement in the human lineage. *Nature* **440**, 1045-9 (2006).

7. Mitelman, F., Mertens, F. & Johansson, B. A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nat Genet* **15 Spec No**, 417-74 (1997).

8. Struski, S., Doco-Fenzy, M. & Cornillet-Lefebvre, P. Compilation of published comparative genomic hybridization studies. *Cancer Genet Cytogenet* **135**, 63-90 (2002).

9. Courjal, F. & Theillet, C. Comparative genomic hybridization analysis of breast tumors with predetermined profiles of DNA amplification. *Cancer Res* **57**, 4368-77 (1997).

10. Forozan, F. et al. Comparative genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary DNA microarray data. *Cancer Res* **60**, 4519-25 (2000).

11. Orsetti, B. et al. Genomic and expression profiling of chromosome 17 in breast cancer reveals complex patterns of alterations and novel candidate genes. *Cancer Res* **64**, 6453-60 (2004).

12. Slamon, D.J. et al. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177-82 (1987).

13. Revillion, F., Bonneterre, J. & Peyrat, J.P. ERBB2 oncogene in human breast cancer and its clinical significance. *Eur J Cancer* **34**, 791-808 (1998).

14. Cropp, C.S., Champeme, M.H., Lidereau, R. & Callahan, R. Identification of three regions on chromosome 17q in primary human breast carcinomas which are frequently deleted. *Cancer Res* **53**, 5617-9 (1993).

15. Niederacher, D. et al. Patterns of allelic loss on chromosome 17 in sporadic breast carcinomas detected by fluorescent-labeled microsatellite analysis. *Genes Chromosomes Cancer* **18**, 181-92 (1997).

16. Casey, G. The BRCA1 and BRCA2 breast cancer genes. *Curr Opin Oncol* **9**, 88-93 (1997).

17. Coquelle, A., Pipiras, E., Toledo, F., Buttin, G. & Debatisse, M. Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. *Cell* **89**, 215-25 (1997).

18. Sutherland, G.R., Baker, E. & Richards, R.I. Fragile sites still breaking. *Trends Genet* **14**, 501-6 (1998).

19. Gebhart, E. et al. Cytogenetic studies on human breast carcinomas. *Breast Cancer Res Treat* **8**, 125-38 (1986).

20. Dutrillaux, B., Gerbault-Seureau, M. & Zafrani, B. Characterization of chromosomal anomalies in human breast cancer. A comparison of 30 paradiploid cases with few chromosome changes. *Cancer Genet Cytogenet* **49**, 203-17 (1990).

21. Kallioniemi, A. et al. Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc Natl Acad Sci U S A* **91**, 2156-60 (1994).

22. Willis, S. et al. Detailed gene copy number and RNA expression analysis of the 17q12-23 region in primary breast cancers. *Genes Chromosomes Cancer* **36**, 382-92 (2003).

23. Couch, F.J. et al. Localization of PS6K to chromosomal region 17q23 and determination of its amplification in breast cancer. *Cancer Res* **59**, 1408-11 (1999).

24. Barlund, M. et al. Multiple genes at 17q23 undergo amplification and overexpression in breast cancer. *Cancer Res* **60**, 5340-4 (2000).

25. Wu, G.J. et al. 17q23 amplifications in breast cancer involve the PAT1, RAD51C, PS6K, and SIGma1B genes. *Cancer Res* **60**, 5371-5 (2000).

26. Monni, O. et al. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc Natl Acad Sci U S A* **98**, 5711-6 (2001).

27. Andersen, C.L. et al. High-throughput copy number analysis of 17q23 in 3520 tissue specimens by fluorescence in situ hybridization to tissue microarrays. *Am J Pathol* **161**, 73-9 (2002).

28. Ried, T. et al. Mapping of multiple DNA gains and losses in primary small cell lung carcinomas by comparative genomic hybridization. *Cancer Res* **54**, 1801-6 (1994).

29. Sonoda, G. et al. Comparative genomic hybridization detects frequent overrepresentation of chromosomal material from 3q26, 8q24, and 20q13 in human ovarian carcinomas. *Genes Chromosomes Cancer* **20**, 320-8 (1997).

30. Solinas-Toldo, S. et al. Mapping of chromosomal imbalances in pancreatic carcinoma by comparative genomic hybridization. *Cancer Res* **56**, 3803-7 (1996).

31. Kallioniemi, A. et al. Identification of gains and losses of DNA sequences in primary bladder cancer by comparative genomic hybridization. *Genes Chromosomes Cancer* **12**, 213-9 (1995).

32. Voorter, C. et al. Detection of chromosomal imbalances in transitional cell carcinoma of the bladder by comparative genomic hybridization. *Am J Pathol* **146**, 1341-54 (1995).

33. Wong, N. et al. Assessment of genetic changes in hepatocellular carcinoma by comparative genomic hybridization analysis: relationship to disease stage, tumor size, and cirrhosis. *Am J Pathol* **154**, 37-43 (1999).

34. Pack, S.D. et al. Molecular cytogenetic fingerprinting of esophageal squamous cell carcinoma by comparative genomic hybridization reveals a consistent pattern of chromosomal alterations. *Genes Chromosomes Cancer* **25**, 160-8 (1999).

35. Koo, S.H., Kwon, K.C., Park, J.W., Lee, Y.E. & Kim, J.W. Characterization of chromosomal breakpoints in an ALL patient using cross-species color banding. *Cancer Genet Cytogenet* **119**, 118-20 (2000).

36. Suehiro, Y. et al. Genetic aberrations detected by comparative genomic hybridization predict outcome in patients with endometrioid carcinoma. *Genes Chromosomes Cancer* **29**, 75-82 (2000).

37. Kiechle, M. et al. Genetic imbalances in precursor lesions of endometrial cancer detected by comparative genomic hybridization. *Am J Pathol* **156**, 1827-33 (2000).

38. Plantaz, D. et al. Gain of chromosome 17 is the most frequent abnormality detected in neuroblastoma by comparative genomic hybridization. *Am J Pathol* **150**, 81-9 (1997).

39. Khan, J., Parsa, N.Z., Harada, T., Meltzer, P.S. & Carter, N.P. Detection of gains and losses in 18 meningiomas by comparative genomic hybridization. *Cancer Genet Cytogenet* **103**, 95-100 (1998).

40. Cai, D.X., James, C.D., Scheithauer, B.W., Couch, F.J. & Perry, A. PS6K amplification characterizes a small subset of anaplastic meningiomas. *Am J Clin Pathol* **115**, 213-8 (2001).

41. Erson, A.E. et al. Overexpressed genes/ESTs and characterization of distinct amplicons on 17q23 in breast cancer cells. *Neoplasia* **3**, 521-6 (2001).

42. Barlund, M. et al. Increased copy number at 17q22-q24 by CGH in breast cancer is due to high-level amplification of two separate regions. *Genes Chromosomes Cancer* **20**, 372-6 (1997).

43. Wu, G. et al. Structural analysis of the 17q22-23 amplicon identifies several independent targets of amplification in breast cancer cell lines and tumors. *Cancer Res* **61**, 4951-5 (2001).

44. Venter, J.C. et al. The sequence of the human genome. *Science* **291**, 1304-51 (2001).

45. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

46. Parssinen, J., Kuukasjarvi, T., Karhu, R. & Kallioniemi, A. High-level amplification at 17q23 leads to coordinated overexpression of multiple adjacent genes in breast cancer. *Br J Cancer* **96**, 1258-64 (2007).

47. Sinclair, C.S., Rowley, M., Naderi, A. & Couch, F.J. The 17q23 amplicon and breast cancer. *Breast Cancer Res Treat* **78**, 313-22 (2003).

48. Wheeler, D.L. et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **35**, D5-12 (2007).

49. Kuhn, R.M. et al. The UCSC genome browser database: update 2007. *Nucleic Acids Res* **35**, D668-73 (2007).

50. Barlund, M. et al. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. *J Natl Cancer Inst* **92**, 1252-9 (2000).

51. Ried, T. et al. Comparative genomic hybridization of formalin-fixed, paraffin-embedded breast tumors reveals different patterns of chromosomal gains and losses in fibroadenomas and diploid and aneuploid carcinomas. *Cancer Res* **55**, 5415-23 (1995).

52. Kuukasjarvi, T. et al. Genetic heterogeneity and clonal evolution underlying development of asynchronous metastasis in human breast cancer. *Cancer Res* **57**, 1597-604 (1997).

53. Tirkkonen, M. et al. Distinct somatic genetic changes associated with tumor progression in carriers of BRCA1 and BRCA2 germ-line mutations. *Cancer Res* **57**, 1222-7 (1997).

54. Isola, J.J. et al. Genetic aberrations detected by comparative genomic hybridization predict outcome in node-negative breast cancer. *Am J Pathol* **147**, 905-11 (1995).

55. Hermsen, M.A. et al. Genetic analysis of 53 lymph node-negative breast carcinomas by CGH and relation to clinical, pathological, morphometric, and DNA cytometric prognostic factors. *J Pathol* **186**, 356-62 (1998).

56. Bulavin, D.V. et al. Amplification of PPM1D in human tumors abrogates p53 tumor-suppressor activity. *Nat Genet* **31**, 210-5 (2002).

57. Li, J. et al. Oncogenic properties of PPM1D located within a breast cancer amplification epicenter at 17q23. *Nat Genet* **31**, 133-4 (2002).

58. Chou, M.M. & Blenis, J. The 70 kDa S6 kinase: regulation of a kinase with multiple roles in mitogenic signalling. *Curr Opin Cell Biol* **7**, 806-14 (1995).

59. Grove, J.R. et al. Cloning and expression of two human p70 S6 kinase polypeptides differing only at their amino termini. *Mol Cell Biol* **11**, 5541-50 (1991).

60. Seufferlein, T. & Rozengurt, E. Rapamycin inhibits constitutive p70s6k phosphorylation, cell proliferation, and colony formation in small cell lung cancer cells. *Cancer Res* **56**, 3895-7 (1996).

61. Zheng, P., Eastman, J., Vande Pol, S. & Pimplikar, S.W. PAT1, a microtubule-interacting protein, recognizes the basolateral sorting signal of amyloid precursor protein. *Proc Natl Acad Sci U S A* **95**, 14745-50 (1998).

62. Jacobs, J.J. et al. Senescence bypass screen identifies TBX2, which represses Cdkn2a (p19(ARF)) and is amplified in a subset of human breast cancers. *Nat Genet* **26**, 291-9 (2000).

63. Kallijarvi, J., Avela, K., Lipsanen-Nyman, M., Ulmanen, I. & Lehesjoki, A.E. The TRIM37 gene encodes a peroxisomal RING-B-box-coiled-coil protein: classification of mulibrey nanism as a new peroxisomal disorder. *Am J Hum Genet* **70**, 1215-28 (2002).

64. Avela, K. et al. Gene encoding a new RING-B-box-Coiled-coil protein is mutated in mulibrey nanism. *Nat Genet* **25**, 298-301 (2000).

65. Saurin, A.J., Borden, K.L., Boddy, M.N. & Freemont, P.S. Does this have a familiar RING? *Trends Biochem Sci* **21**, 208-14 (1996).

66. Slack, F.J. & Ruvkun, G. A novel repeat domain that is often associated with RING finger and B-box motifs. *Trends Biochem Sci* **23**, 474-5 (1998).

67. Mangelsdorf, D.J. & Evans, R.M. The RXR heterodimers and orphan receptors. *Cell* **83**, 841-50 (1995).

68. Nagaraj, S.H., Gasser, R.B. & Ranganathan, S. A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief Bioinform* **8**, 6-21 (2007).

69. Wilcox, A.S., Khan, A.S., Hopkins, J.A. & Sikela, J.M. Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome. *Nucleic Acids Res* **19**, 1837-43 (1991).

70. Kan, Z., Rouchka, E.C., Gish, W.R. & States, D.J. Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res* **11**, 889-900 (2001).

71. Jiang, J. & Jacob, H.J. EbEST: an automated tool using expressed sequence tags to delineate gene structure. *Genome Res* **8**, 268-75 (1998).

72. Brett, D. et al. EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. *FEBS Lett* **474**, 83-6 (2000).

73. Mironov, A.A. & Pevzner, P.A. SST versus EST in gene recognition. *Microb Comp Genomics* **4**, 167-72 (1999).

74. Modrek, B., Resch, A., Grasso, C. & Lee, C. Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res* **29**, 2850-9 (2001).

75. Kan, Z., Castle, J., Johnson, J.M. & Tsinoremas, N.F. Detection of novel splice forms in human and mouse using cross-species approach. *Pac Symp Biocomput*, 42-53 (2004).

76. Schmitt, A.O. et al. Exhaustive mining of EST libraries for genes differentially expressed in normal and tumour tissues. *Nucleic Acids Res* **27**, 4251-60 (1999).

77. Buetow, K.H., Edmonson, M.N. & Cassidy, A.B. Reliable identification of large numbers of candidate SNPs from public EST data. *Nat Genet* **21**, 323-5 (1999).

78. Boguski, M.S., Lowe, T.M. & Tolstoshev, C.M. dbEST--database for "expressed sequence tags". *Nat Genet* **4**, 332-3 (1993).

79. Khan, A.S. et al. Single pass sequencing and physical and genetic mapping of human brain cDNAs. *Nat Genet* **2**, 180-5 (1992).

80. Mikkelsen, T.S. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69-87 (2005).

81. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* **268**, 78-94 (1997).

82. Mayor, C. et al. VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16**, 1046-7 (2000).

83. Schwartz, S. et al. PipMaker--a web server for aligning two genomic DNA sequences. *Genome Res* **10**, 577-86 (2000).

84. Ye, J., McGinnis, S. & Madden, T.L. BLAST: improvements for better sequence analysis. *Nucleic Acids Res* **34**, W6-9 (2006).

85. Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M. & Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res* **32**, W273-9 (2004).

86. Strausberg, R.L. et al. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Natl Acad Sci U S A* **99**, 16899-903 (2002).

APPENDIX A

MAMMALIAN CELL CULTURE MEDIUM

Table A.1 Compositions of MEM with Earle's salts

Component	**mg/L**
L-arginine.HCl	126
L-cystine	24
L-glutamine	292
L-histidine.HCl.H_2O	42
L-isoleucine	52
L-leucine	52
L-lysine.HCl	73
L-methionine	15
L-phenylalanine	32
L-threonine	48
L-tyrptophane	10
L-tyrosine	36
L-valine	46
Folic acid	1
Nicotinamide	1
D-Ca-pantoyhenate	1
Pyridoxal.HCl	1
Thiamine.HCl	1
Riboflavin	0.1
Myo-inositol	2
NaCl	6800
KCl	400
$NaH_2PO_4.H_2O$	140
$MgSO_4.7H_2O$	200
$CaCl_2$	200
D-glucose	1000
Phenol red	10
$NaHCO_3$	2200

APPENDIX B

BACTERIAL CULTURE MEDIUM

LB

For 1 liter,

Yeast Extract	5 g
Tryptone	10 g
NaCl	10 g
1N NaOH	1 mL
Agar	15 g

pH: 7.4

SOC

For 1 liter,

Yeast Extract	5 g
Tryptone	20 g
NaCl	0.5 g
250 mM KCl	10 mL
5N NaOH	0.2 mL
2M $MgCl_2$	5 mL (freshly added)

pH: 7

APPENDIX C

GENE SPECIFIC PRIMERS & OTHER PRIMERS

Table C.1 Forward Gene Specific Primers for 3' RACE PCR.

Gene Specific Primer	Sequence	Bases	Tm *	GCcontent
GSP1-T02811	TGACAACTCGCACCCCACAGGGAA	24	76 °C	%58
nestedGSP1-T02811	AGGCAGCCGGGACCTAGAATGTA	23	72 °C	%57
GSP1_3	ATTCATCCAGGACACCAACCCAGT	24	72°C	%50
GSP1_4	TCCTCAAGGTCTGTGACTCCACCTA	25	76°C	%52
3'RACE-GSP	TCCAGGCCACCAACCCAGTCCCCCACCC	28	96°C	%71
3'RACE-GSP-NSTD	GGTCTGTGACTCCACCTATCCCTCTGGC	28	90°C	%61
GSP-1081-1106R	GAGGTAGAAGGCCAAGCTGTGACTGA	26	80°C	%54
GSP-1271-1242R	TCGCTTTCTCTCTTTGGGGTCTCTCT	26	78°C	%50
3RACE_GSP_TEM	CTAGAGGCAGAGTTCGACCTTGTACCT	27	82°C	
3RACE_NSTD_GSP	CGCTGGAATAACGCAGCCCAAAGCC	25	80°C	%60

Table C.2 Reverse Gene Specific Primers for 5' RACE PCR.

Gene Specific Primer	Sequence	Bases	Tm *	GCcontent
T02811-GSP2	CAGTTGATCTGCAATTGGCCAGTTG	25	74°C	%48
T02811-nestedGSP2	TCCCGCCCCCAGCCTCAACTGCT	23	78°C	%70
RACE_GSP2_3	ACTGGGTTGGTGGCCTGGATGAAT	24	74°C	%54
RACE_GSP2_4	TGGAGTCACAGACCTTGAGGACCT	24	74°C	%54
5'RACE-GSP	CCCGCCCCCAGCCTCAACTGCTTCAGAG	28	94°C	%68
5'RACE-GSP-NSTD	AGAGCCATGTTCCCATTTAATGCTTGAG	28	80°C	%43
GSP-453-478F	GTTTCCTAAGGGCACACGGCTAGAAA	26	78°C	%50
GSP-170-195F	TGCTCACCTAAGTTCCTGCCTCCATA	26	78°C	%50
5RACE_GSP_TEM	GGCTTTGGGCTGCGTTATTCCAGCG	25	80°C	%60
5RACE_NSTD_GSP	AGGTACAAGGTCGAACTCTGCCTCTAG	27	82°C	%62

*Annealing temperatures are calculated according to the 4 x (G+C) + 2 x (A+T) formula

Table C.3 GeneRacer™ Kit Primers.

Name	Bases	Homology	Tm*
GeneRacer™ 5' Primer	23	Position 1-23 of GeneRacer™ RNA oligo	74 °C
GeneRacer™ 5' Nested Primer	26	Position 15-40 of GeneRacer™ RNA oligo	78 °C
GeneRacer™ 3' Primer	25	Position 1-25 of GeneRacer™ Oligo dT Primer	76 °C
GeneRacer™ 3' Nested Primer	23	Position 14-36 of GeneRacer™ Oligo dT Primer	72 °C
Control A Primer	24	Position 67-90 of human β-actin(NM 001101.2)	72 °C
Control B.1 Primer	22	Position 793-814 of human β-actin(NM 001101.2)	76 °C

Table C.4 Primers for extending T02811.

Primer Name	Sequence	Bases	Tm *
T02811BKout100R	CAACAAATTGTAGTGCTGGG	20	58°C
T02811BKout100F	GATAGATAGGTACAAGGTCG	20	58°C
Fx	CTTGGACTTCCTTTTGGGCTCT	22	66°C
Fy	CCACTCGACGGTGAGGACTTAC	22	70°C
Fz	TGGCTGGGGAATGCCAGAGTTG	22	70°C
M_EST_27_R	TTCTCTGAACCCCAATGTCAT	21	60°C
M_EST_50_R	ATGGAGCCAGAGGACATCCT	20	62°C

Table C.5 GAPDH Primers for D-RT-PCR and T3 and T7 primers used for sequencing.

	Sequence	Bases	Tm *
GAPDH Forward	TATGACAACGAATTTGGCTAC	21	58°C
GAPDH Reverse	TCTCTCTTCCTCTTGTGCTCT	21	58°C
T3	ATTAACCCTCACTAAAGGGA	20	56°C
T7	TAATACGACTCACTATAGGG	20	56°C

APPENDIX D

BUFERS & SOLUTIONS

10X TBE (Tris Borate) Buffer (1L)

Tris-base	108 g
Boric acid	55 g
0.5M EDTA (pH: 8.0)	40 mL

All the components were mixed and the volume was completed to 1 L with dH_2O.

10X TNE (Tris-NaCl-EDTA) Buffer (1 L)

Tris	12.11 g
EDTA $Na_2.2H_2O$	3.72 g
NaCl	116.89 g

All the components were mixed, pH was adjusted to 7.4 with concentrated HCl, and the volume is completed to 1 L with dH_2O.

1X TE (Tris-EDTA) Buffer (1 L)

Tris.HCl 10 mM

EDTA 1 mM

All the components were mixed and the volume is adjusted to 1 L with dH_2O.

Hoechst 33258 Stock Solution

Hoechst 33258 dye from Hoefer Inc (Cat# DQ201)	10 mg
Distilled water	10 mL

10 mM $CaCl_2$ (40 mL)

$CaCl_2$	0.045 g
Sterile distilled water	40 mL

Filter sterilization was done with 0.22 μm filter.

75 mM $CaCl_2$ (20 mL)

$CaCl_2$	0.1665 g
Sterile distilled water	20 mL

Filter sterilization was done with 0.22 μm filter.

20% SDS Solution (50 mL)

SDS	10 g
Distilled water	50 mL

3M NaAc Solution (100 mL)

$C_2H_3NaO_2$	24.61 g
Distilled water	100 mL

pH was adjusted to 5.2 with concentrated glacial acetic acid.

APPENDIX E

PLASMID MAP

Figure E.1 Map of the pCR®4-TOPO plasmid

APPENDIX F

GENBANK GI NUMBERS

Table F.1. GenBank GI (GeneInfo Identifier) Numbers

	GenBank GI #
AI449813	4292629
C17orf82	28278251
CA466381	24922733
LOC388407	44662807
T02811	314052
TBX2	44921604
TBX4	18129689