Identification of Integration Sites of Mouse Mammary Tumor Virus in Human Breast Cancer Tissues

BY

Talha Bin Rahat

NUST201260362MASAB91012F

Atta-ur-Rahman School of Applied Biosciences,

National University of Sciences and Technology,

Islamabad, Pakistan.

2014

Identification of Integration Sites of Mouse Mammary Tumor Virus in Human Breast Cancer Tissues

A thesis submitted as MS project in partial fulfillment of the requirement for the degree of Master of Science
In
Healthcare Biotechnology

Talha Bin Rahat

NUST201260362MASAB91012F

Supervised by: Dr. Hajra Sadia

Atta-ur-Rahman School of Applied Biosciences,

National University of Sciences and Technology,

Islamabad, Pakistan.

2014

رَبَّنَا مَا خَلَقْتَ هَٰذَا بَاطِلًا سُبْحَانَكَ

فَقِنَا عَذَابَ النَّارِ

(Al-Quran, 3:191)

"Our Lord, You did not create this aimlessly; exalted are You [above such a thing]; then protect us from the punishment of the Fire".

ACKNOWLEDGEMENTS

الْحَمْدُ لِلَّهِ رَبِّ الْعَلَمِين

This thesis is an effort of 1 year of research work and 2 years of carving the idea for this research. All this is nothing but a manifestation of Allah's blessings upon me. He provided me with so many great friends and mentors who helped me to shine this idea and this work. And all Praise is to Him, for blessing me with a family that always encouraged me to explore my abilities.

I want to express my deep gratitude to my supervisor, Dr. Hajra Sadia, who has been a great supervisor and great mentor throughout my time at ASAB. I want to apologize for my lazy and careless behavior which she always tolerated and helped me to overcome my incapacities. I want to thank Dr. Attya Bhatti, HoD healthcare biotechnology and Dr. Peter John, Principal, ASAB, for maintaining a research-pro environment at ASAB. I would like to acknowledge NUST for provision of funds, Dr. Rubina Tabassum, GEC members, the faculty at ASAB, my batch mates, and my lab associates for timely guidance and help throughout this research. I am especially thankful to Ms. Miriam Kathleen Gomez for always being a great help in times of need, Ms. Wasifa Naushad for discussing her views and ideas on this topic, Mr. Imran Raza for being a great friend and mentor and Mrs. Fozia Parveen. I am also grateful to all my classmates and friends especially Dr. Fahd, Hakeem Usman, Umer, Talha, Feroz, Hussain, Ibrahim, Irfan, Hamza, Tabassum and Umair for discussing their ideas and polishing my efforts on this topic. It would not have been possible without all of them.

And finally my utmost respect and love for my parents and family. I don't have words to thank them for their support.

Talha Bin Rahat

TABLE OF CONTENTS

ACRONYMS

BLAST	Basic Local Alignment Tool
BLAT	BLAST-Like Alignment Tool
bp	Base Pairs
CA	Capsid
CiS	Carcinoma in Situ
CIS	Common Integration Site
dATP	deoxyadenosine Triphosphate
dCTP	deoxycytosine Triphosphate
dGTP	deoxyguanosine Triphosphate
dH_2O	distilled H_2O
dNRE	Distal NRE
dNTP	deoxynucleotide Triphosphate
dsDNA	Double Stranded DNA
dTTP	deoxythymine Triphosphate
dUMP	Deoxyuridine Monophosphate
dUTP	Deoxyuridine Triphosphate

EBV	Epstein Barr Virus
EDTA	Ethylenediaminetetraacetic acid
EPC	Enhancer of Polycomb
FFPE	Formalin Fixed Paraffin Embeded
fgf	Fibroblast Growth Factor
GR	Glucocorticoide Receptor
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HERV	Human Endogenous Retrovirus
HPV	Human Papilloma Virus
HRE	Hormone Responsive Elements
HTLV	Human T-cell Lymphoma Virus
IC	Invasive ductal Carcinoma
IL-2	Interleukine 2
IL-4	Interleukine 4
IN	Integrase
IPA	Isopropyl Alcohol
ITAM	Immunoreceptor Tyrosine-based Activation Motif
jNRE	Junctional NRE

LB	Luria-Bertani
LM-PCR	Ligation Mediated Polymerase Chain Reaction
LTR	Long Terminal Repeats
MA	Matrix Protein
MGE	Mammary Gland Enhancer
MMTV	Mouse Mammary Tumor Virus
mRNA	messenger RNA
NBP	NRE Binding Protein
NC	Nucleocapsid
NF	Nuclease Free
NRE	Negative Regulatory Elements
PBS	Primer Binding Site
PCR	Polymerase Chain Reaction
PNK	Polynucleotide Kinase
pNRE	Proximal NRE
rpm	Revolutions per minute
RT	Reverse Transcriptase
RTC	Reverse Transcription Complex
SAg	Super Antigen

SDS	Sodium Dodecyl Sulfate
ssRNA	Single Stranded RNA
TAE	Tris Acetate EDTA
TBLV	Type B Leukemogenic Virus
tfr1	Transferrin 1
TGFβ	Tumor Growth Factor β
TNF	Tumor Necrosis Factor
TRAIL	TNF-Related Apoptosis-Inducing Ligand.
tRNA	Transfer RNA
UBP	Ubiquitous Binding Protein
UV	Ultra Violet

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Mouse mammary tumor virus (MMTV) is an established cause of mammary tumors in mice but its role in human breast cancer is still controversial. MMTV has been detected in human breast cancer samples in several parts of the world, but the establishment of a causal role is under debate. Several groups have reported presence of MMTV-like sequences in human breast cancer. We have also identified MMTV-like sequences in 26% of breast cancer tissues from Pakistani population. Here we propose that integration site analysis of MMTV might help in establishing its role in breast cancer development. We have used LM-PCR to identify integration sites of MMTV. Human DNA from Breast cancer samples was digested and ligated to linkers. An MMTV specific and a linker specific primer were used to amplify integration site. Amplified DNA was isolated by shotgun cloning in bacteria. Identification of thirteen distinct integration sites supports the idea of a viral etiology of breast cancer. Furthermore, our data also indicates anomalous behavior of MMTV-like sequences. Previous data suggests a totally random integration of MMTV in both mouse and human genome in cell lines. However we have identified 5 identical integration sites in eight different breast cancer tissue samples. One particular site at locus 3q26.2 was identified from 7 samples. This data also supports sequence specific integration of MMTV. This study is first of its kind in MMTV research. Although, the link is still not clear, but further studies on these integration sites might reveal a link between MMTV-like sequences and human breast cancer.

CHAPTER 1

INTRODUCTION

Cancer is a multifactorial disorder of cellular proliferation. It is the second leading cause of death in developing countries with Breast cancer being the second most common cancer overall and most common in female population (Grayson, 2012). Breast cancer is the leading cause of death in women with approximately 69% of them in the developing countries contrary to other cancers which are prevalent in developing countries (World Cancer Report, 2008)

Pakistan in particular has the highest incident of breast cancer only behind the Jewish population. It has a significantly higher breast cancer incidence as compared to the neighboring countries India and Iran (Liede *et al.*, 2002). One in every nine women in Pakistan develops breast cancer at some stage in her life. A single center study reports the incidence of breast cancer among female cancer patients to be about 45.9% (Badar *et al.*, 2011). Major risk factors associated with breast cancer can be divided into three groups. At first are genetic and chromosomal basis. These include genetic mutations in genes like p53, BRCA1 and BRCA2, dysregulation of certain genes and chromosomal aberrations. Secondly, there are hormone associated reasons. These include exposure to hormones at unnatural times like early menarche, late menopause, endogenous estrogen regulation, late first child birth etc. and thirdly there are factors that are based on the lifestyle of the individual. Unhealthy lifestyle including alcohol consumption, obesity and physical inactivity etc. (World Cancer Report, 2008).

Another factor contributing to cancers are viruses. Several viruses have been declared as oncogenic due to their established role as a causative agent of cancer. Such viruses of human include HPV (Human Papilloma Virus), HTLV (Human T-cell Lymphoma Virus) and EBV (Epstein Barr Virus). There are other viruses which are strongly associated with cancers with a suggested role in cancer pathways like HBV (Hepatitis B Virus) and HCV (Hepatitis C Virus). In the same regard several studies have associated breast cancer with viruses as well. Commonly associate viruses with human breast cancer include EBV, HPV and MMTV (Mouse Mammary Tumor Virus) (Joshi *and* Buehring, 2012).

MMTV was first discovered as a milk-transmitted filterable agent that caused mammary tumors in mice (Bittner, 1943). Later studies confirmed it as a virus. This virus has since been extensively studied. It is a complex beta-retrovirus that alongside signature proteins of retroviruses (Gag, Pol and Env) also produces accessory proteins (SAg, Rem and Dut) (Buchen-Osmund, 2006). The genome is flanked by relatively longer Long Terminal Repeats (LTRs). Open Reading Frame (ORF) for Super Antigen (SAg) is also present in these LTRs (Choi *et al.,* 1991). In MMTV these repeats contain at least 3 regulatory elements that control transcription regulation of both MMTV and genes downstream of MMTV integration sites (Bramblett *et al.,* 1995)). One of these regulatory elements is Hormone Responsive Element (HRE) that regulates transcription of MMTV in response to steroid hormones (Majors *and* Varmus, 1983).

MMTV belongs to genus *Betaretrovirus*, subfamily *Orthoretrovirinae*, family *Retroviridea.* It carries two copies of single stranded RNA (ssRNA) genome which is not segmented and like other retroviruses it encodes a Reverse Transcriptase enzyme

to replicate its genome (Büchen-Osmond, 2006). The viral envelope is composed of a lipid bilayer derived from cell membrane. The envelope bilayer contains spikes of viral envelope glycoproteins, SU (gp52) and TM (gp36), linked together by disulfide bonds. The virion is composed of a nucleoprotein core containing two RNA molecules (two full-length genomes) in association with the nucleocapsid (NC) protein. A shell of capsid proteins (CA) surrounds the viral core. It constitutes the icosahedral capsid. The region between capsid and envelope bilayer is localized by Matrix protein (MA). Two essential enzymes, polymerase (reverse transcriptase, RT) and integrase (IN) are found within the core where cellular tRNAs are also found that function as primers (Buchschacher Jr, 2001). The detailed structure of the virus is given in the figure 1.1 (Maitra U, *et al.*, 2006).

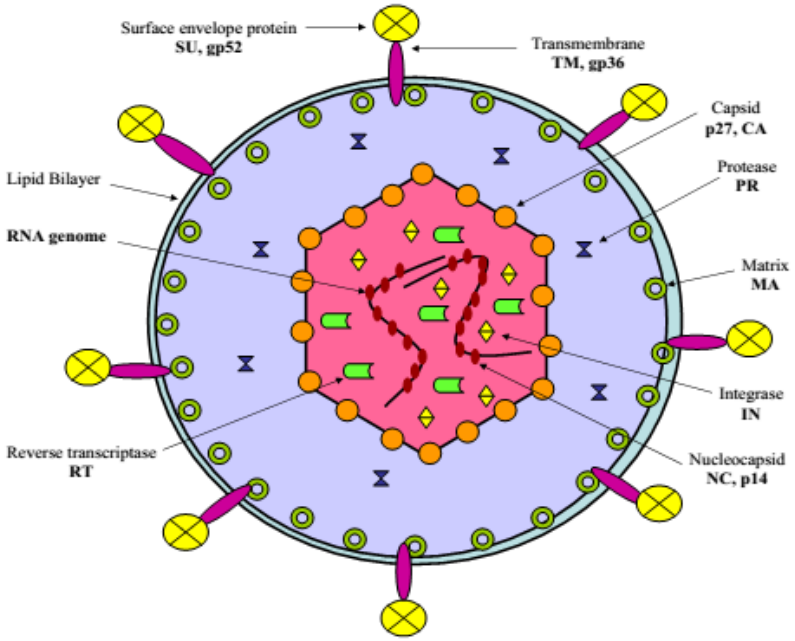


Figure 1.1 MMTV Structure.

Structure of MMTV showing a mature virus and location of proteins. Adapted from Maitra *et al.* (2006).

The genome encodes for 7 major proteins gag (group specific antigen), Pro (protease), Pol (polymerase and integrase), Env (envelope), Rem (regulator of export of MMTV mRNAs), DU (dUTPase) and SAg (superantigen). Because of the discovery of Rem Protein along with two accessory proteins SAg and DU (dUTPase), MMTV is now considered a complex betaretrovirus (Mertz *et al.*, 2005).

Oncogenic potential of MMTV is explained under three different theories. The first and most acceptable is the upregulation of downstream genes. MMTV integrations near *wnt* and *fgf*-family proteins trigger upregulation of these genes. In fact several of the mouse oncogenes have been discovered because of MMTV integration (Ross, 2010). Due to this property Theodorou *et al.* (2007) have suggested the use of this virus to locate potential oncogenes. Second property of MMTV that describes its oncogenic potential is SAg protein (Mukhopadhyay *et al.*, 1995). And third one is an ITAM domain in its envelope protein (Katz *et al.*, 2005).

Due to some similarities between mammary tumors in mice and breast cancer in humans and hormone responsiveness of the onset of breast cancer in humans triggered a search for human homolog of MMTV. Initial studies base on immunochemistry did yield some evidence but were later hampered by the discovery of HERVs (Human Endogenous Retroviruses). In fact studies on MMTV led to the discovery of first HERV (Ross, 2010). In 1990s PCR based detection of MMTV-like viruses triggered a renewed interest in MMTV. Several groups amplified MMTV like sequences from human breast cancer samples. Initially Env protein was focused, then Wang *et al.* (2001) detected LTR sequences by PCR amplification. Even a full-length proviral genome was identified by (Liu *et al.*, 2001).

However, results were not consistent everywhere, some groups like Zangen *et al.* reported zero prevalence of MMTV-like sequences (Zangen *et al.*, 2002). Mant *et al.,* (2004) amplified a few sequences from breast cancer samples, reported those sequences to be of human origin upon sequencing (Mant *et al.*, 2004). Previous studies in Pakistani population reported LTR and Env sequences from 26% and 20% of the samples respectively (Naushad *et al.*, 2014).

These association studies however are neither consistent nor efficient enough to establish any causal role of this virus. Since MMTV integration sites are not known from cancerous cells, we decided to identify MMTV integration sites from Breast cancer samples positive for MMTV-LTR sequences and search for any potential genetic changes that might result in Breast Cancer.

We have used LM-PCR (Ligation Mediated PCR) based detection of insertion sites. The method was used for the detection of retroviral integration sites by Wu *et al.*, 2003. Genomic DNA is digested into smaller fragments and ligated to linkers. One primer from linker and other from end sequence of the insert, are used to amplify the insertion site. These sites are then shotgun cloned into a plasmid vector followed by multiplication of plasmids in bacteria. Plasmids from random colonies are picked from each sample and are screened for integrations. Positive clones are sequenced and aligned against human genome using alignment tools like BLAT (BLAST-Like Alignment Tool) and BLAST (Basic Local Alignment Search Tool).

CHAPTER 2

LITERATURE REVIEW

Mouse Mammary Tumor Virus (MMTV) is an oncogenic betaretrovirus. Over the course of evolution it has stably integrated in its natural host's (i.e. mouse) genome as an endogenous retrovirus (Mtv). Functional endogenous retrovirus is transmitted vertically to the next generation while the exogenous virus is mainly transmitted to lactating pups via infected female. It causes palpable mammary tumors in mice by integrating in certain genes and gene pathways. Its LTRs have Hormone Responsive Elements (HREs) which in response to steroid hormones enhance the downstream gene expression. MMTV like sequences including complete proviral genome have been isolated from several human breast cancers as well (Luther *and* Acha-Orbea, 1997). Although several hypotheses have been presented about molecular events that might lead an MMTV infection to cancer in humans, the causal role of MMTV in human breast cancer is still debatable (Salmons and Gunzburg, 2013).

2.1 MMTV GENOME AND PROTEINS

Viral genomic RNA of MMTV contains short direct repeats R at their ends. These are followed by Unique regions, 120 bases long U5 (at 5' end) and 1200 bases in length, U3 (at 3' end). Just downstream of U5 there is an 18 bases long primer binding site (PBS). This site is complementary to the 3' end of tRNA-lys. It functions as the primer for the synthesis of minus strand DNA (Peters *and* Glover, 1980). Following U5 region there are overlapping reading frames that encode the viral genes gag, pro, pol, env, rem and sag. The Genome structure is that of mRNA i.e.

methylguanosine cap at 5' end and poly-A tail at 3' end (Mertz *et al.*, 2005; Schlom *et al.*, 1973). The genomic RNA also contains a packaging signal ψ. The ends containing U3, R and U5 regions constitute long terminal repeats (LTR). Viral promoters and enhancers are present in the U3 region of the 5' LTR. The initiation of transcription occurs at 5' U3/R region (Buchschacher Jr, 2001). It terminates at R/U5 junction within the 3'LTR. There are few other promoters described for MMTV as well (Mink *et al.*, 1990). Three polyprotein precursors are encoded by MMTV. Two "-1" ribosomal frameshift events during genomic RNA translation lead to the production of these three polyprotein precursors Pr77Gag, Pr110Gag-Pro and Pr160Gag-Pro-Pol (Jacks *et al.*, 1987).

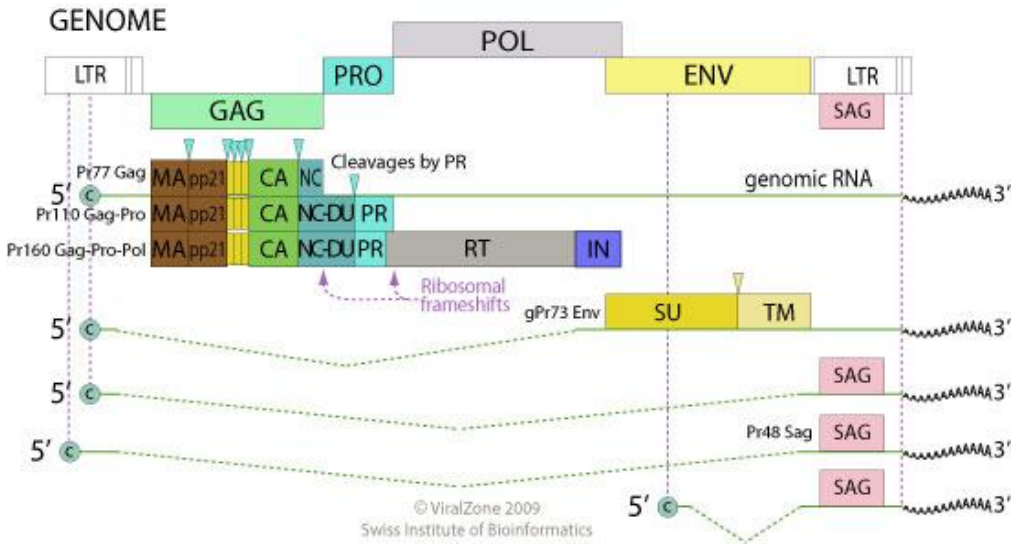


Figure 2.1 MMTV genome.

Monopartite, linear, dimeric, ssRNA(+) genome of about 8-10 kb, with a 5'-cap and a 3'poly-A tail. There are two long terminal repeats (LTRs) of about 600nt long at the 5' and 3' ends. The LTRs contain the U3, R, and U5 regions. There is also a primer binding site (PBS) at the 5'end and a polypurine tract (PPT) at the 3'end (viralzone, 2009).

2.1.1 Gag proteins

Pr77 precursor encodes nonglycosylated structural proteins of the viral core. It is subsequently cleaved by the viral Protease (PR) to generate mature proteins (Hizi *et al.*, 1987). The ribonucleoprotein complex is surrounded by the CA protein (capsid, p27), the major structural virion protein (Clive Dickson *and* Skehel, 1974)). The NC protein (nucleocapsid P14) is associated closely with the genomic RNA because of its highly basic nature. It is required for budding and packaging of the virus (Wills *and* Craven, 1991). There are other proteins generated by Protease activity that have potential structural role (Sarkar *et al.*, 1978).

2.1.2 Pro proteins (protease and dUTPase)

The Pr110 precursor encodes viral protease PR and dUTPase (DU). A proteolytic process results in a 30kDa protein DU from Gag-6 Pro (Hizi *et al.*, 1987). It degrades deoxyuridine triphosphate (dUTP) into deoxyuridine monophosphate (dUMP) and pyrophosphate to prevent its incorporation into viral DNA during replication (Threadgill *et al.*, 1993).

2.1.3 Pol proteins

RT an IN are produced by Pr160. They are involved in replication and integration of viral genome. RT is a reverse transcriptase that reverse transcribes the genomic RNA into dsDNA. IN (integrase) is involved in the integration of this dsDNA.

2.1.4 Env Proteins

'env' gene encodes envelope glycoprotein (Env). Env is further processed into transmembrane unit, TM (gp36) and surface unit, SU (gp52). The process of synthesis

of Env precursor takes place at rough endoplasmic reticulum and is followed by cleavage of signal peptide (Dickson *and* Atterwill, 1980). This polyprotein then moves to Golgi apparatus for glycosylation. After glycosylation SU and TM are incorporated into host cell plasma membrane. SU serves as the receptor to bind host cell during infection. The function of TM is to mediate the envelope fusion with the host membrane. During budding the envelope proteins become the part of the MMTV envelope.

2.1.5 SAg

MMTV is unique in a sense that unlike most retroviruses there is an open reading frame within the U3 region. Either of the LTR or env promotor can drive the expression of sag gene (Choi *et al.,* 1991). The sag gene encodes 36kDa type ii transmembrane protein. This protein is essential for viral expression in B and T cells and plays an important role in milk-borne transmission of MMTV to mammary gland cells.

2.1.6 Rem protein

Retroviruses need to export unspliced mRNA to cytoplasm to be used as genome for packaging. Betaretroviruses use two mechanisms, a cis-actin, RNA stem-loop structure or a trans-acting protein. For the export of unspliced mRNA from nucleus to cytoplasm MMTV employs an essential trans-acting element Rem. Due to the use of this transport element MMTV is classified as a complex betaretrovirus (Indik *et al.*, 2005b), (Mertz *et al.*, 2005).

2.2 MMTV REPLICATION

MMTV entry initiates through the interaction of the envelope SU glycoprotein with a cellular receptor *tfr1* (transferrin 1). The viral DNA synthesis takes place within the cytoplasm in a large ribonucleoprotein structure called the reverse transcription complex (RTC), which is believed to result from structural rearrangements of the virion core (Mougel *et al.*, 2009). Details of retroviral replication are illustrated in the figure 2.2 (Principles of Virology, 2009).

Viral genome is uncoated and viral RT reverse transcribes it into dsDNA. A cellular tRNA molecule is used as a primer. The ribonuclease H activity of RT degrades the majority of the virion RNA in the RNA-DNA hybrid. The resulting double-stranded DNA (provirus) has identical LTRs on either end composed of U3, R and U5. The provirus remains associated with nucleoprotein complexes, known as the pre-integration complex, which is then integrated into the host chromosome by the viral integrase protein. MMTV is believed to infect actively dividing cells for their replication to gain access to the nucleus when the nuclear membrane is disassembled during mitosis. Once the provirus has integrated, cellular RNA polymerase II synthesizes at least 5 different length mRNAs which include a full length genomic RNA. It uses cellular machinery for the production of proteins from mRNA sequences. The viral proteins are assembled in the cytoplasm into immature virions containing a spherical nucleocapsid core of 70 nm in diameter. The viral envelope proteins are transported to the plasma membrane via host secretory pathway (Morita *and* Sundquist, 2004). The immature virion particle associates with the cell membrane prior to budding. PR-mediated processing of viral proteins is required to generate mature infectious virions.

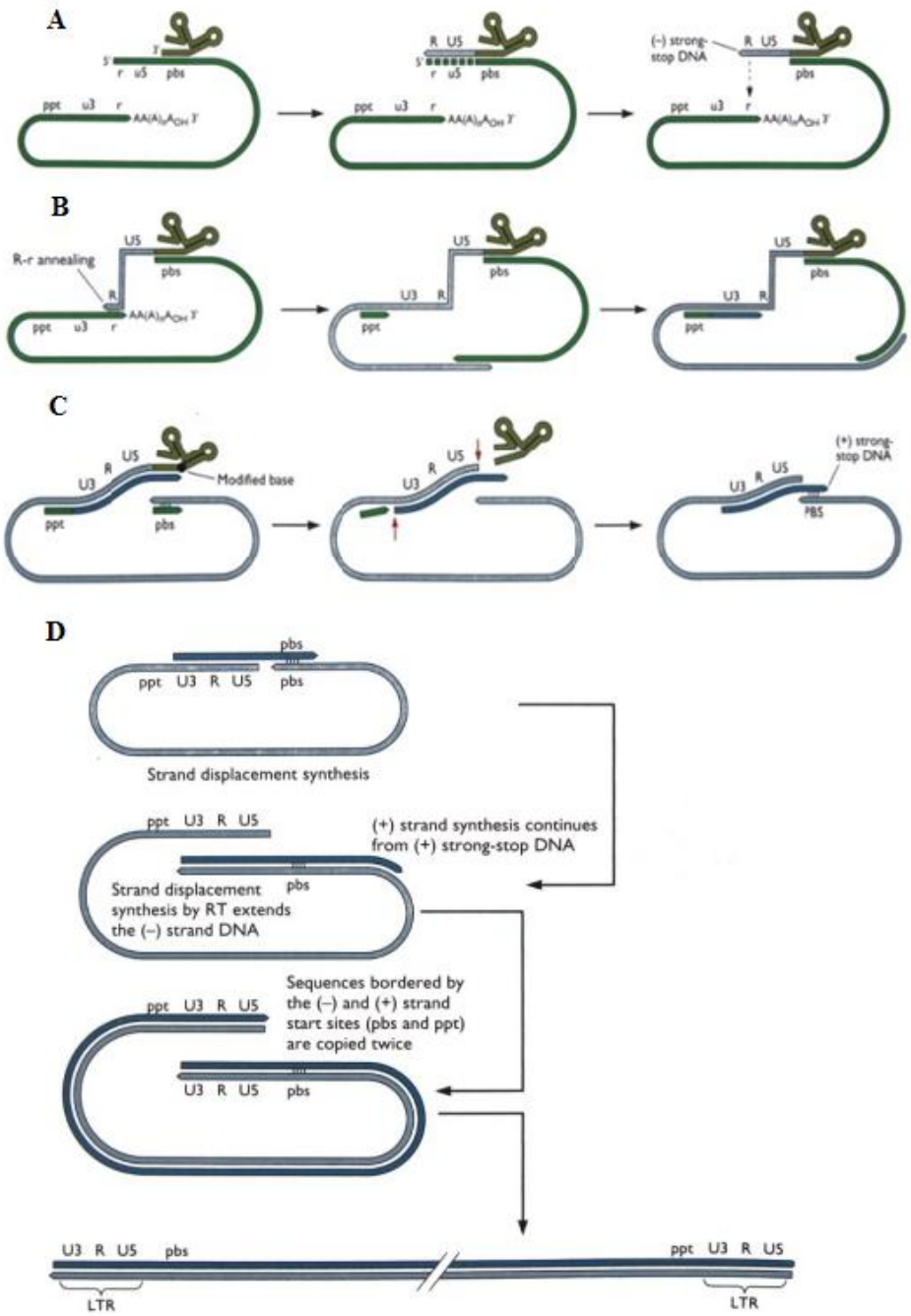


Figure 2.2 MMTV Replication.

A. cellular rRNA binds on the PBS and reverse transcription quickly proceeds until it reaches r(repeat sequences). This product acts as primer on the other end of genome. B. The first template exchange and the initiation of +strand DNA proceeds. C. One early digestion product of the genomic RNA is a fragment comprising the polypurine tract (ppt). These few nucleotides serve as primer for the initiation of the (+) strand DNA synthesis, which starts prior to the completion of the (-) strand DNA. Again, the RT is forced to arrest after several nucleotides as it

encounters a modified base which cannot be copied. D. synthesis of (+) strand of DNA. (adapted from principles of virology, volume 2. Flint *et al.*, 2009)

2.3 MMTV TRANSMISSION

As stated previously, MMTV infection is caused by both endogenous and exogenous virus. However there are differences in both infections. Exogenous transmission of MMTV is associated with a relatively higher incidence of mammary tumors i.e. 90% and they tend to occur by 6 months of age in mice. As for endogenous infection the incidence rate of MMTV is 40% and they tend to occur by 20 months of age (Bentvelzen, 1974; Nandi *and* McGrath, 1973). Figure 2.3 shows the diagrammatic representation of infectious cycle of exogenous MMTV in mice (Johal, 2009).

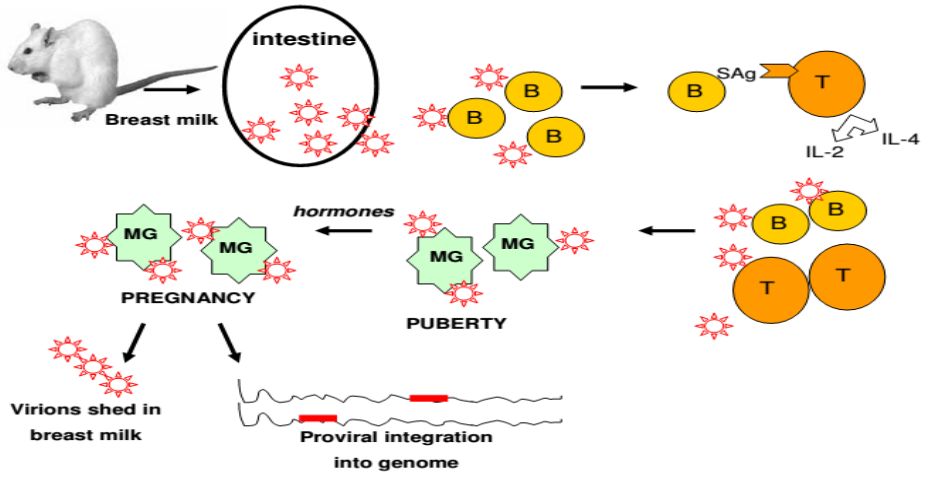


Figure 2.3 MMTV infectious cycle.

MMTV gains entry into B cells in the gut via TfR1. B cells process the virus and present the SAg to T cells. The T cells release cytokines that cause bystander cell division and viral amplification. The virus then travels to the mammary gland, where it infects these cells. During pregnancy, steroid hormones upregulate the transcription of virus. The virus is then transmitted exogenously via breast milk or endogenously via germ line integration. (Johal, 2009)

After the ingestion of MMTV-infected breast milk, the virus makes its way to B cells in Peyer's patches in the alimentary canal via Toll-like receptor 4 (Rassa *et al.*, 2002). It utilizes the body's immune system for its own spread. B cells, after processing the virus present MMTV's superantigen (SAg) to T cells. T cells are activated and they release IL-2 and IL-4. This causes bystander cell division and number of lymphoid cells containing the integrated virus increases thus increasing the viral particles in blood (Held *et al.*, 1993). Viral particles leave the gut through mesenteric lymph nodes and travel to the mammary gland. In the glands they gain entry into the mammary cells via interaction transferrin receptor 1 (TfR1) (Karapetian *et al.*, 1994) (Zhang *et al.*, 2003). The envelope protein undergoes a conformational change upon interaction with TfR1 and infects by trafficking to a low pH compartment of the cell in mice. However, the case is different with humans, it doesn't use human TfR1 (E. Wang *et al.*, 2008). Influence of steroid hormones during pregnancy and lactation causes a dramatic upregulation of MMTV expression (Cato *et al.*, 1987).

In the absence of B lymphocytes, mice mammary epithelial cells exposed to MMTV are not infected and do not form tumors (Golovkina *et al.*, 1998). MMTV does not infect non-dividing cells because the nuclear envelope must be disintegrated to allow the viral dsDNA pre-integration intermediate into the nucleus. This environment is only given dividing cells. This is further strengthened by the fact that most adult mice are resistant to oral MMTV infection as viral receptors in the intestinal tract may be downregulated in adult mice (Velin *et al.*, 1996). Thus the lymphoid tissue is the first site of propagation of virus. Velin *et al.* infected mice via nasal route and showed that nasal associated lymphoid tissue was the site of primary

infection, the infection then spreads to circulation lymphocytes from where it is carried further to different organs (Velin *et al.*, 1997)

2.4 MMTV TRANSCRIPTION REGULATION

Transcription regulation is a very important topic for role of MMTV in cancer, since MMTV is generally known to cause cancers by inducing the expression of downstream gene. U3 region of the MMTV LTR contains at least 3 different regulatory elements i.e. a hormone response element (HRE) (Majors *and* Varmus, 1983), a mammary gland enhancer (MGE) (E. Mok *et al.*, 1992) and negative regulatory elements (NREs) (Bramblett *et al.*, 1995).

2.4.1 HRE (hormone responsive element)

At least six glucocorticoid receptor (GR) binding sites have been identified within MMTV LTR. They are responsible for the hormonally regulated transcription of MMTV (Fletcher *et al.*, 2000). Glucocorticoids are steroid hormones that bind to receptors localized in the cytoplasm (Rao, 1981). A hexanucleotide consensus sequence has been identified (TGTTCT) that serves as the receptor-binding site (Kühnel *et al.*, 1986). GR entry is followed by recruitment of other transcription factors, e.g., nuclear factor 1 (NF-1) and octamer transcription factor (OCT-1) and the basal transcriptional machinery to the promoter leading to increased MMTV transcription (Beato, 1996). The actual mechanism involves chromosomal remodeling. The binding sites for different transcription factors are otherwise inaccessible. Binding of hormones, results in the recruitment of chromosome remodeling machinery (Collingwood 1999). Transcription regulation based on nucleosomal rearrangement is well characterized in MMTV (Fletcher *et al.*, 2000).

2.4.2 Mammary gland enhancer (MGE)

MMTV LTR contains a bipartite mammary gland enhancer, which is divided into a BanII element (located between -1075 to -978 upstream of the transcription start site) and the mammary-specific enhancer of MMTV (MEM) element (-938 to -862 upstream of the transcription start site) (Grimm *and* Nordeen, 1999). The enhancer is functional in both lactating and non-lactating mammary glands. Deletion of the MGE region decreases both glucocorticoid-induced and basal level transcription (Haraguchi *et al.*, 1997). In addition, six cis-acting elements have also been identified within the enhancer region that harbored multiple binding sites for nuclear factors, such as mp4, mp5/AP-2 and NF1 (Mellentin-Michelotti *et al.*, 1994).

2.4.3 Negative Regulatory elements (NRE)

TBLV (type B leukemogenic virus) is an MMTV T-cell variant that causes thymic lymphomas in mice. Only the substitution of U3 region resulted in conversion of a mammary gland specific virus to a lymphomagenic virus (Yanagawa *et al.*, 1993), (Hsu *et al.*, 1988). Deletions of NRE resulted in MMTV expression in other tissues like brain, skeletal muscles, heart and liver (Henrard *and* Ross, 1988). These studies demonstrate that only HRE and MGE are not necessary for its tissue tropism. Several negative regulatory elements (NREs) have been reported in MMTV LTR as an answer to its tropism (Lee *et al.*, 1991). They are divided into distal-NRE (dNRE), junctional-NRE (jNRE) and proximal-NRE (pNRE) with reference to promoter (Zhu *and* Dudley, 2002). Different cellular factors bind NREs to regulate virus expression. Transcription enhancer factor-1 (TEF-1) binds to pNRE and blocks basal and glucocorticoid mediated viral transcription (Maeda *et al.*, 2002). Similarly, transcription factor Ku recruits DNA-dependent protein kinase to MMTV NRE and

modulates the transcription activity (Giffin *et al.*, 1996). Two more factors have been identified to bind NREs i.e. NRE-binding protein, NBP (identified as special AT-rich sequence-binding protein 1 (SATB1)) and Ubiquitous binding protein, UBP (identified as a homeodomain protein, CDP) (Zhu *and* Dudley, 2002), (J. Liu *et al.*, 1997).

2.5 MMTV AND CANCER

Salmons and Gunzburg have reviewed the various mechanisms by which MMTV contributes to cancer. There are two major hypotheses for the role of MMTV in cancer. The first is Insertion Mutations in and near genes disrupting their expression and function and the second is oncoproteins coded by MMTV. Another hypothesis has also been presented that MMTV might integrate near another potential oncovirus like HPV (Human Papilloma Virus) and EBV (Epstein Barr Virus) (Salmons *and* Gunzburg, 2013).

2.5.1 Insertional mutagenesis

MMTV causes tumors mainly by integrating into a region near proto-oncogenes. LTR contain Hormone Responsive Elements (HREs) and Mammary gland enhancers and that in response to steroid hormones drive the upregulation of downstream genes (Callahan *and* Smith, 2008). Theodoru *et al.* has utilized this ability of MMTV to identify new cancer genes. They used high throughput retroviral insertion screening for mammary tumors and identified at least 17 new common insertion sites (CIS) and new genes involved in mammary tumors (Theodorou *et al.*, 2007). In their study, Theodorou *et al.* found Multiple insertions within the Arf3(2), Nsf(3) and Fbxw4(13) genes, which are located adjacent to the Wnt1/Wnt10b,

Wnt3and Fgf8genes respectively. Expression levels of the former genes were not significantly affected, the expression levels of the latter were increased in the respective tumors, indicating enhancer-mediated activation of Wnt1, Wnt3 and Fgf8over long distances up to 157 kb (Theodorou *et al.*, 2007).

Most of the common integration sites (CIS) in mice belong to Wnt and Fgf family of genes (Ross, 2010). Table 2.1 summarizes known CIS in mice. There is a hypothesis that activation of multiple oncogenes is required for mammary tumorigenesis. This hypothesis is supported by several studies Investigation of a large percentage of mammary tumors derived from MMTV-infected wild type mice have had "hits" at both *Wnt1* and *Fgf3* therefore, it is likely that additional genes are activated in these tumors. Also, double transgenic mice, such as MMTV-*Wnt1*/MMTV-*Fgf3* have been seen to have accelerated mammary tumorigenesis (Kwan *et al.*, 1992), (Shackleford *et al.*, 1993), (MacArthur *et al.*, 1995). Similarly, transgenic mice with a genetic predisposition to mammary tumorigenesis, such as those with express the *wnt1*, c-neu or TGFβ transgene expression in mammary tissue show accelerated induction of tumors on MMTV infection (Shackleford *et al.*, 1993), (Dievart *et al.*, 1999).

2.5.2 MMTV oncoproteins

Like some other oncogenic viruses MMTV has also been speculated to encode oncoproteins. In this regard two proteins SAg and Env have shown some properties of oncogenesis. Mukhopadhyay *et al.* (1995) showed the oncogenic potential of SAg by expressing SAg in immortalized cell lines and transplanting it into mammary fat pads. Only those cell lines transfected with SAg were able to cause tumors in mice (Mukhopadhyay *et al.*, 1995).

Table 2.1 Oncogenes affected in MMTV induced Cancers

Mouse Oncogene Reference (Ross, 2010)	
Gene	Reference
Wild type *Wnt*1/*Wnt*10b	(Nusse *et al.*, 1984)
Wild type *Fgf*3	(C Dickson *et al.*, 1984)
Wild type int-5/aromatase	(Durgam *and* Tekmal, 1994)
Wild type eIF3e-p48	(Marchetti *et al.*, 1995)
Wild type *Notch*4	(MacArthur *et al.*, 1995)
MMTV-*Wnt*1 *Fgf*8	(MacArthur *et al.*, 1995)
MMTV-neu *Notch*1	(Diévart *et al.*, 1999)
WAP-TGFβ *Wnt*1/*Wnt*3	(Schroeder *et al.*, 2000)
Wild type *Fgf*10	(Theodorou *et al.*, 2004)
Wild type *Rspo*2	(Lowther *et al.*, 2005)
Wild type *Rspo*3	(Gattelli *et al.*, 2006) (Theodorou *et al.*, 2007)

Env protein was also shown to be involved in carcinogenesis by immunoreceptor tyrosine based-activation motif (ITAM). ITAM is located in SU subunit of Env. Env-expressing cells showed hallmarks of cell transformation such as sensitivity to apoptosis induced by tumor necrosis factor (TNF)–related apoptosis-inducing ligand (TRAIL) or TNFα, as well as down-regulation of E-cadherin and Keratin-18 (Katz *et al.*, 2005). Feldman and colleagues (2012) demonstrated tumorigenic properties of signal peptide of MMTV Env. The phosphorylation pattern of this protein determines whether it works as oncogenic or anti-oncogenic. Microarray analysis suggested that phosphorylation at serine 18 or at serine 65 is associated with transcriptional regulation of the L5 nucleolar ribosomal protein (a p14 target) and the Erb-B signal transduction pathway.

2.6 MMTV IN HUMANS

There are several similarities in MMTV induced tumors in mice and Breast Cancer in humans. Both are influenced by hormones (Labat 1998). During pregnancy, the expression levels of MMTV increase in mice (Luther and Acha-Orbea, 1997). In humans 62% of gestational breast cancers were found to be MMTV positive as compared 30-38% of sporadic breast cancers (Wang *et al.*, 2003). Different genes associated with human breast cancer are also associated with MMTV insertions in mouse the table 2.2 taken from Johal *et al.* (2007) shows comparison of these genes.

Initial studies based on neutralizing antibodies showed some evidence of MMTV then the discovery of endogenous retroviruses with homology to MMTV sequences and ability to produce a few proteins as well questioned these studies. A renewed interest in MMTV's association with human breast cancer arose when Wang *et al.* (1995) amplified DNA sequences from env region in 39% of the samples processed. The table 2.3 summarizes various studies and percentages of MMTV-like sequences in human breast cancer tissues from different populations and geographic regions. Mostly env sequences were detected in these studies. However, a few studies (Wang *et al.*, 2002, Lawson *et al.*, 2004, Etkind *et al.*, 2004) and Naushad *et al.*, 2014) also detected LTR regions. Zangen *et al.* (2002) and Bindra *et al.* (2007) used qRT PCR for env expression but found no results. Similarly, Mant and Cason (2004) detected MMTV-like env sequences in 16% of the samples but sequencing result showed that these sequences to be of human origin. Liu *et al.* (2001) also identified a complete proviral sequence from 2 human breast cancer samples (B. Liu *et al.*, 2001).

Table 2.2 MMTV affected genes in mouse vs. commonly cited genes in human cancers

Gene	Chromosome		Gene product	Site of expression
	Mouse	Human		
Int1/Wnt1	15	12	Morphogen	Testis
Int2/Fgf3	7	11	Growth factor	None
Int3/Notch4	17	6	Notch related protein	All
Int4/Wnt3	11	17	Wnt related protein	none
Int5	9	15	Aromatase	Breast, ovaries, Testis
Int6	11	8	e-subunit of translation initiation factor eIF3	All
Fgf4/Hst	7	11	Growth Factor	None
Wnt10b	15	12	Signaling Protein	Mammary Glands
Fgf8	19	10	Androgen-induced growth factor	Ovaries, Testis
Rspo2/Int7	15	8	Wnt related protein	All
Rspo3/Thsd2	10	6	Wnt related protein	All

Table 2.3 Prevalence of MMTV sequences in human breast cancer in different populations.

Abbreviations: IC (invasive ductal carcinoma), CIS (Carcinoma in situ).

Investigators	Year	Population	percentage	References
Wang *et al.*	1995	USA	39%	(Yue Wang *et al.*, 1995)
Pogo *et al.*	1999	Italy	38%	(Pogo *et al.*, 1999)
Etkind *et al.*	2000	USA	37%	(P. Etkind *et al.*, 2000)
Wang *et al.*	2001	USA	38%	(Y Wang *et al.*, 2001)
Melana *et al.*	2001	USA	38%	(Melana *et al.*, 2001)
Wang *et al.*	2001	USA	42% (LTR)	(Yue Wang *et al.*, 2001)
Zangen *et al.*	2002	Italy	0%	(Zangen *et al.*, 2002)
Ford *et al.*	2003	Australia	42%	(Ford *et al.*, 2003)
Wang *et al.*	2003	USA	62% in gestational breast cancer	(Y Wang *et al.*, 2003)
Witt *et al.*	2003	Austria	0%	(Witt *et al.*, 2003)
Ford *et al.*	2004	Australia	78%	(Ford *et al.*, 2004)
Lawson *et al.*	2004	Australia	75% in IC and 23% in CIS	(James S Lawson *et al.*, 2004)
Mant *et al.*	2004	UK	16% (sequencing suggests Human DNA)	(Mant *et al.*, 2004)
Faedo *et al.*	2004	Australia	29%	(Faedo *et al.*, 2004)
Levine *et al.*	2004	Tunisia	74%	(Levine *et al.*, 2004)
Etkind *et al.*	2004	USA	50%	(P. R. Etkind *et al.*, 2004)
Ford *et al.*	2004	Australia	32% in IC and 25% in CIS	(Ford *et al.*, 2004)
		Vietnam	62% in IC and 50% in CiS	
Zammarchi *et al.*	2006	Italy	32%	(Zammarchi *et al.*, 2006)
Lawson *et al.*	2006	Australia	40% in IC and 25% in CiS	(James Sutherland Lawson *et al.*, 2006)

Bindra *et al.*	2007	Sweden	0%	(Bindra *et al.*, 2007)
Zapata-Benavides *et al.*	2007	Mexico	4%	(Zapata-Benavides *et al.*, 2007)
Mok *et al.*	2008	Australia	56%	(M. T. Mok *et al.*, 2008)
Fukuoka *et al.*	2008	Japan	0%	(Fukuoka *et al.*, 2008)
Hachana *et al.*	2008	Tunisia	14%	(Hachana *et al.*, 2008)
Park *et al.*	2011	Australia	0%	(Park *et al.*, 2011)
Naushad *et al.*	2014	Pakistan	26% LTR and 20% env	(Naushad *et al.*, 2014)

Some of the sublines of MCF-7 breast cancer cell lines MMTV env and LTR bear MMTV env and LTR sequences and there is evidence of env sequence at RNA level. Expression profiling of one such subline is done by Fernandez-cobo *et al.* (2006) Among the 27 upregulated genes there were six interferon-inducible ones: IFI6, TRIM22, IFITM1, IFITM2+IFITM3, IFI27 and IP-30 and a receptor IFNGR2. In addition, there were five upregulated genes that have a connection with TNF or are involved in its signaling, like LTBR, TRAF3, MMP17, PKN1 and MAPK13. The cytokine TGFβ, and its downstream effector early growth response protein 1 (EGR1), were also upregulated in env+ cells. Twenty genes were down regulated in MMTV env+ cells (Fernandez-Cobo *et al.*, 2006).

MMTV has successfully been cultured in human cell lines (Indik *et al.*, 2005a). However, MMTV has been known to enter the cell using *tfr1* receptor of mouse but not the human transferrin receptor (E. Wang *et al.*, 2008). It is suggested that it might use some other mechanism for entry into the cell (Ross, 2010). Salmons and Gunzburg analyzed the evidence for MMTV-like virus infecting humans to cause

breast cancer on the basis of Fredericks and Relman's revised Koch's postulates (Fredericks *and* Relman, 1996). They came to a conclusion that MMTV fulfills almost all criteria except for the reproducibility which differs because of the different methods of detection. However, the mechanism still remains unknown (Salmons *and* Gunzburg, 2013). Faschinger *et al.*, (2008) have analyzed initial integration sites in mouse and human cell lines. There results show that MMTV do not show any bias towards specific genomic sequences, instead it integrates more randomly than any known retrovirus. Salmons and Gunzburg (2013) have suggested that MMTV integration sites might be sequenced from human breast cancer samples in order to find out whether MMTV-like sequences follow the same mechanism of carcinogenesis in humans as MMTV in mice.

To identify insertion sites different methods are used. There is vast variety of techniques to identify integration sites. There are commercially available kits based on these methods as well Leoni *et al.* (2011) have systematically reviewed all the strategies for identification of such integration sites (Leoni *et al.*, 2011). We have used a PCR based method for the identification of integration sites. Ligation mediated PCR has already been used for the identification of retroviral integration sites (Wu *et al.*, 2003). Faschinger *et al.* (2008) has also used the same method for identification of viral integration sites of MMTV in mouse and human cell lines. In short, human DNA is digested to small fragments using a four-cutter typeII restriction enzyme (mseI). Digested DNA is ligated to linker/adapter sequences and amplified using one integration sequence-specific and other linker sequence-specific primer. It is then cloned in a T/A vector and amplified in bacteria. Plasmids from transformed colonies are then sequenced to identify integration sites. However, in contrast to Faschinger *et*

al. (2008) we have not sequenced all positive clones instead we have randomLy chosen colonies from each sample and screened them for sequences of human origin containing a part of MMTV.

CHAPTER 3

MATERIALS AND METHODS

3.1 SELECTION OF SAMPLES

3.1.1 Sample Collection

Formalin fixed Breast Cancer tissues saved in paraffin blocks were collected from Benazir Bhutto Hospital and Holy Family Hospital Rawalpindi. The samples were analyzed by histopathological analyses and confirmed positive for breast cancer at respective pathology labs under an authorized pathologists. The FFPE (Formalin Fixed Paraffin Embedded) tissue blocks were then stored at room temperature.

3.1.2 DNA Extraction from FFPE Tissue

From each FFPE tissue, 5-10um thick 10-15 sections were cut using microtome and placed in microfuge tubes. Xylene (1mL) was added to each tube and shaken for 30min at room temperature to dissolve paraffin. Dissolved paraffin was removed by centrifugation at 14000 rpm for 10 minute at room temperature followed by decanting the supernatant. The procedure of dissolving and decanting paraffin in xylene was repeated to remove paraffin completely. Sections were washed with 100% ethanol to remove xylene followed by centrifugation at 14000 rpm to decant ethanol. Tissue pellets were air dried. To each tube, 500 µL of tissue digesting buffer was added and incubated at 50°C for overnight. Digestion buffer stock of 100 mL was prepared. Digestion buffer contains 1M Tris-HCl (4 mL), 0.5M EDTA (0.2 mL), 10% SDS solution (10 mL), distilled water (80.8 mL) and proteinase K (5 mL) in a total volume of 100mL. Following day, proteinase K was inactivated by heating at 95 °C for 10 minute. 120 µL of NaCl (6M) was added in each tube and centrifuged at 14000

rpm for 10 minute to pellet down proteins. Supernatant was transferred to another tube and 1 volume 3M sodium acetate was added and shaken vigorously using vortex mixer. DNA was precipitated down by adding 600 µL isopropyl alcohol (IPA) and incubating at -20 °C for overnight.

Next day, DNA suspensions in IPA were centrifuged at 14000 rpm for 20 minute at -4 °C. Supernatant was discarded and the pellet was washed with 70% ethanol. After centrifugation, ethanol was discarded and pellet was air dried. Finally, DNA was dissolved in 50 µL of Nuclease Free water (NF H_2O).

3.1.3 Amplification of housekeeping gene

PCR amplification for 118bp segment of human β-globin gene was used to assess the quality of the DNA. PCR mixture consisted of Taq polymerase buffer (1x), $MgCl_2$ (2 mM), 2 mM of dNTPs (dATP, dGTP, dTTP, dCTP), forward and reverser primers (20 pmole each), thermostable *Taq* polymerase (1 U) and 20 ng of extracted DNA as template. Nuclease free water was added to make the volume up to 50 µL. The primers used to amplify the β-globin gene are given in table 3.1.

Table 3.1 List of Primers used

Serial no.	Primer	sequence 5'—> 3'	Annealing Temperature
1	β-globin Forward	ACACAACTGTGTTCACTAGC	58.4^0C
2	β-globin Reverse	CAACTTCATCCACGTTCACC	60.4^0C
3	ltr Forward	GGTGGCAACCAGGGACTTAT	57.3^0C
4	ltr Reverse	CGTGTGTTTGTGTCTGTTCG	54.5^0C
5	Lin P1	GTAATACGACTCACTATAGGGC	53^0C
6	Lin P2	AGGGCTCCGCTTAAGGGAC	55.4^0C
7	ltr1	CCGTCTCCGCTCGTCACTTAT	56.3^0C
8	ltr2	CGTCACTTATCCTTCACTTTCCA	53.5^0C

The reaction mixture was placed in thermocycler. Thermocycler conditions used to amplify the gene are given in the figure 3.1.

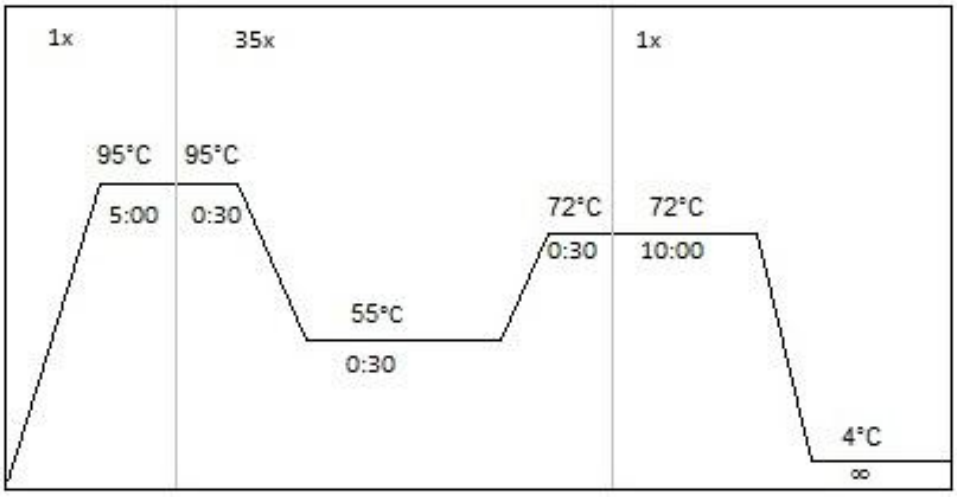


Figure 3.1 Thermocycler profile for B-globin fragment amplification

3.1.4 Gel Electrophoresis

The product was observed on TAE buffer (tris-acetate-EDTA) based 2% agarose gel. TAE (10X) stock solution was made as by adding 48.4 g of Tris base [tris(hydroxymethyl)aminomethane], 11.4 mL of glacial acetic acid (17.4 M), 3.7 g of EDTA dissolved in 800 mL distilled water. After balancing the pH at 8.3, distilled water was added to make final volume 1 liter. Gel was stained with Ethidium bromide and visualized under UV.

3.1.5 Detection of MMTV LTR region

DNA samples with sharp band for B-globin gene were then screened for the presence of MMTV ltr gene. A sequence of 663bp fragment of MMTV ltr gene was amplified using the primers ltr forward and ltr reverse, given in the table 1.1. PCR mixture consisted of Taq polymerase buffer (1x), $MgCl_2$ (2 mM), 2 mM of dNTPs (dATP, dGTP, dTTP, dCTP), forward and reverser primers (20 pmole each),

thermostable *Taq* polymerase (1 U) and 20 ng of extracted DNA as template. Nuclease free water was added to make the volume up to 50 µL. Thermocycler profile for this reaction is given in the figure 3.2. The product was visualized on TAE-based 1% agarose gel. Sample was mixed with loading dye and 5uL was loaded for observation. Gel stained with ethidium bromide was observed under UV.

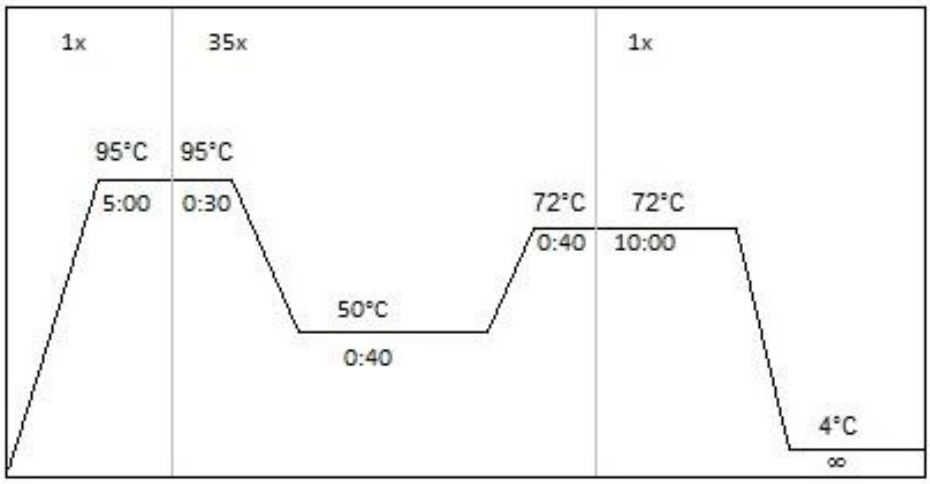


Figure 3.2 Thermocycler profile for LTR fragment amplification

3.2 CLONING OF INTEGRATION SITES

3.2.1 Linker and Nested Primer Design

Linkers containing overhangs for *MseI* restriction enzyme site were designed. Two complementary sequences (43bp long) were designed randomLy with two additional nucleotides as overhangs on the 5' end (table 3.2). Any complementarity with human genome was removed after using BLAST. Two overlapping Primers were chosen against the 5' end of the linker sequence (table 3.1) and two overlapping primers from the 3'end of MMTV (LTR region) were chosen, ltr1 and ltr2 (table 3.1). BLAST was used to remove any similarity of the primers with human genome.

Table 3.2 Linker sequences

S no.	Name	Sequence 5'—> 3'
1.	LinSeq1	TAAGTCCCTTAAGCGGAGCCCTATAGTGAGTCGTATTACCCTTTC
2.	LinSeq2	GAAAGGGTAATACGAATAACTATAGGGCTCCGCTTAAGGGACT

3.2.2 Linker Hybridization

Both linker strands were synthesized separately and purchased from Eurofins Genomics. An amount of 20 pmol of each strand was dissolved in 4x Hybridization buffer in microfuge tubes. Tubes were sealed with paraffin film and heated to 95°C and allowed to cool down from 95°C to room temperature overnight in water bath.

3.2.3 Restriction Digestion of genomic DNA

A four cutter enzyme *MseI* (Thermo Scientific) also known as *TruI* was used to digest human DNA (Restriction site T*TAA). The median restriction length of this enzyme is 70bp in human genome. Genomic DNA was digested using manufacturer's recommended protocol. Genomic DNA was digested in total volume of 20µL reaction mixture containing 2uL human genome, 10X digestion buffer, 10 Units of enzyme and NF water. Reaction mixture was placed at 65°C for 3 hrs. *MseI* enzyme is thermostable and is not deactivated by high temperature. The enzyme was removed using PureLink PCR purification kit (Invitrogen).

3.2.3.1 Purification of digested DNA

PureLink PCR Purification Kit is based on the selective binding of dsDNA to silica-based membrane in the presence of chaotropic salts. Manufacturer's protocol was used to purify the DNA. Digested DNA was mixed with 4 volumes of the binding buffer (B3) and transferred to PureLink Spin column in a collection tube. It was then centrifuged at 13000rpm and flow-through was discarded. DNA was washed by adding 650 µL of wash buffer (containing ethanol), and centrifuge twice at 13,000 rpm for 2 minutes to remove any residual wash buffer. The PureLink Spin column was placed in a fresh sterile PureLink Elution tube. 50 µL of Elution Buffer (10 mM Tris-HCl, pH 8.5) was added and incubated for 2 minutes at room temperature followed by centrifugation at 13000 rpm for 2 minutes. The Spin column was discarded and purified DNA was stored at -20°C.

3.2.4 PNK Treatment of digested DNA

Hybridized linker was treated with T4 Polynucleotide Kinase (PNK) (thermoscientific) in a reaction mixture containing 1mM ATP, 1X reaction buffer, 10Units of PNK enzyme as the final concentration in a volume of 20µL. Reaction was carried out at 37°C for 20 min followed by enzyme denaturation at 75°C for 10 min.

3.2.5 Linker Ligation with digested DNA

Linker was ligated with genomic DNA digest at room temperature for 1 h by adding linkers (final concentration of 10pM), and 5U of T4 ligase to formulate a 30 µL ligation reaction.

3.2.6 Nested PCR of ligated product

Two rounds of PCR were used to amplify the integration sites. First round utilized the primers ltr1 and linker 1 and linker ligated-Digested DNA as template, while the second round was performed using the product of first round as template and ltr2 and linker2 as primers (table 3.1).

Round I PCR reaction mixture (50 µL) consisted of PCR buffer (1X), $MgCl_2$ (2.5 mM), 4 dNTPs (0.2 mM), primers (25 pmole each), Taq polymerase (1U), template DNA (20-50ng) and dH_2O. Round II PCR contained 1µL product of the first round as template. All other conditions were same. PCR thermal cycler profile for both reactions is as given in the figure 3.3.

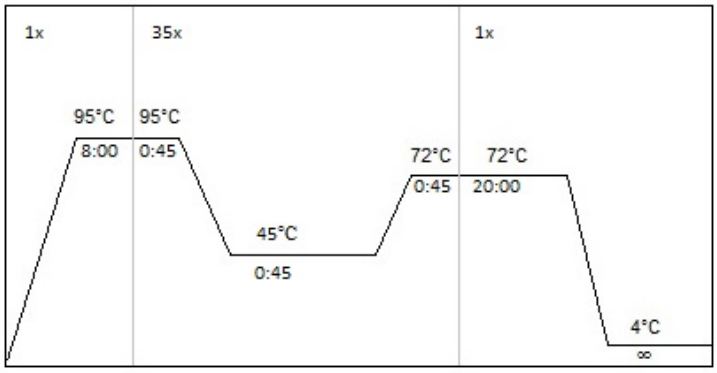


Figure 3.3 Thermocycler profile for LM-PCR

3.2.7 TA Cloning of amplified product

Purified PCR product was ligated in a T/A cloning vector pCR® II (Invitrogen). The vector contains two marker genes and lac promoter which allows bacterial expression of the lacZα fragment for α-complementation (blue-white screening); lacZα fragment contains multiple cloning site and encodes the first 146 amino acids of β-galactosidase, in case of an insert the product is a truncated non-functional protein. It also contains antibiotic resistant genes, ampicillin and

kanamycin, allowing selection and maintenance of *Escherichia coli* (*E.coli*) in antibiotic containing media. The pUC origin maintains high copy number of plasmid in *E. coli*. T7 promoter and priming site allows *in vivo* or *in vitro* transcription of antisense RNA and also provide a site for sequencing. M13 Forward and Reverse priming sites are also available to aid sequencing of the cloned fragment. The map of TA vector is given in Figure 3.4.

In transformation procedure, β-galactosidase cleaves the colorless substrate X-gal (5-bromo-4-chloro-3-indolyl-β -galactopyranoside) into galactose and a blue insoluble product of the cleavage that turns the colony into blue (i.e. blue/white selection). The complete lacZ gene is too large to be carried by a plasmid; cloning vectors contain only the alpha region of the lacZ gene. Expression of the alpha region of lacZ yields an inactive protein. However, the truncated protein encoded by the alpha region can combine into active β-gal when complemented with a protein from lacZ lacking the alpha region. Such "intragenic complementation" can be achieved if a plasmid carrying the alpha region of lacZ is inserted into a bacterial cell containing a mutated alpha region of the lacZ gene. Such cells will turn β-gal positive if the plasmid is present.

DNA cloning into lacZ vectors occurs such that the fragment is likely to be inserted into the alpha lacZ sequence of the plasmid which renders the alpha-region dysfunctional. In this case intragenic complementation no longer occurs. Hence, cells carrying plasmids with transgene will remain β-gal negative. β-gal negative cells can readily be isolated from positive ones by incubation of bacteria on chromogenic culture media containing X-gal and isopropyl β-D-1-thiogalactopyranoside (IPTG) an inducer of lacZ promoter.

3.2.7.1 DNA Purification of PCR product

The PCR product was purified of single stranded DNA, dNTPs, enzymes and buffers by using PureLink PCR purification kit, Invitrogen, using the manufacturer's protocol. The protocol was same as stated in section 3.2.3.1.

3.2.7.2 Ligation of PCR product and Vector

For ligation in T/A vector, 2 µL of vector, 5 µL of purified product, 1 µL of 10X ligase buffer and 1 U of T4 ligase along with 1 µL of NF water was used and made total volume 10µL. The ligation mixture was incubated at 16 °C for overnight. Top 10 strain of *E. coli* were transformed with the ligation mixture.

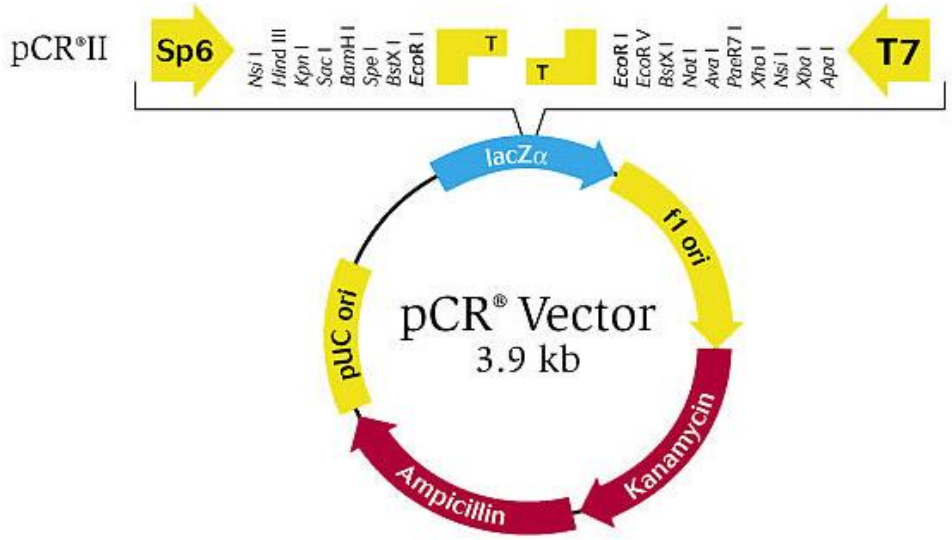


Figure 3.4 Map of TA vector pCR®II.

The figure shows the site of cloning, lacZ-α, ampicillin and kanamycin resistant gene and origin of replication.

3.2.7.3 Preparation of *E. coli* Competent Cells

A single colony of Top-10 cells was grown in Luria-Bertani (LB) medium. The medium was prepared by using 1% trypton, 0.5 % yeast extract and 0.5% NaCl in

1L of distilled water. The pH was adjusted at 7.5 using sodium hydroxide pellets and sterilized by autoclaving. The medium was incubated at 37 °C at 125 rpm with constant shaking for 16-24 hours. Next morning, 1 mL of overnight growth culture was inoculated in 50 mL LB medium. This culture was incubated at 37 °C with constant shaking. The concentration of cells was checked by spectrophotometer after every hour. When the optical density (OD_{600}) of the cells reached 0.6, cells were harvested in a 50 mL centrifuge tube by centrifuging at 4000 rpm at 4 °C. Supernatant was discarded and the pellet was dissolved in 10 mL of ice cold solution A (100mM $CaCl_2$, 50mM $MgCl_2$) and placed on ice for half an hour. The cells were then harvested by centrifuging at 4000 rpm for 10 minutes at 4 °C. The supernatant was discarded and cells were resuspended in 2 mL of Solution B (50mM $MgCl_2$, 25% glycerol). 100 µL of this cell suspension was dispensed in microfuge tubes for single transformation reaction. Competent cells were stored at -80 °C.

3.2.7.4 Transformation of bacteria with recombinant plasmid

Transformation was carried out using Heat Shock Method (Woodcock *et al.*, 1989). A volume of 5µL of ligation mixture was mixed with 100µL of competent cells and the mixture was incubated on ice for half an hour. Heat shock was given to the cells at 42 °C for 2 minutes and snaps cooled on ice for 2-3 minutes. Transformants were allowed to grow in 1 mL LB media, without antibiotic, for 1 hour at 37 °C for the expression of antibiotic resistance gene. Transformants were harvested by centrifugation at 14,000 rpm for 2 minutes and were plated on LB agar plate containing ampicillin (200 µg/ mL). Using spread plate method, IPTG (40 µL of 100mM) and X-Gal (40 µL) were spread on the plates prior to spreading

transformants for the induction of *lacZ* gene and blue/white selection. Plates were incubated at 37 °C overnight.

3.2.7.5 Screening and Selection of Clones

From each plate, 8 different white colonies were selected and each colony was inoculated in 3mL LB media containing ampicillin (200µg/mL) and the cells were allowed to grow at 37 °C with continuous shaking. After 16 hours of growth, plasmids were isolated by Alkaline Lysis Method (Sambrook *and* Russell, 2001). For plasmid preparation, 1.5 mL of overnight grown culture was taken in microfuge tube and spun for 1 minute at 14,000 rpm. Supernatant was aspirated and the cells were resuspended in 100 µL solution I (25 mMTris-HCl, pH 8.0, 10 mM EDTA). To each tube, added 200 µL solution II (0.2 N sodium hydroxide (NaOH), 1.0 % sodiumdodecylsulfate (SDS) and mixed gently by inversion. Added 150 µL solution III (60 mL 3 M potassium acetate (KOAc), pH 4.8, 11.5 mL acetic acid, 28.5 mL H_2O), vigorously mixed briefly to mix, and was spun for 5 minutes at 14,000 rpm. Supernatant was transferred to fresh tube containing 500 µL phenol:chloroform:isoamyl alcohol (25:25:1), vortexed and centrifuged for 5 minutes at 14,000 rpm. Aqueous layer was transferred to fresh tubes containing 1 mL ethanol, mixed by inversion, and spun for 5 minutes at 14,000 rpm. Supernatant was discarded and the pellet was washed with 100 µL of 75 % ethanol and spun for 10 minutes at 14,000 rpm. Ethanol was removed and the pellet was air-dried by leaving on bench with the lid open for 5 minutes. DNA was resuspended in 40 µL dH_2O containing 20 µg/mL RNaseA. The extracted plasmid DNA was resolved on 1 % agarose gel, stained with ethidium bromide and visualized under UV trans-illuminator to check the quality of isolated plasmid.

3.2.8 Colony PCR

Insert was first confirmed by colony PCR. For colony PCR, upto 30 different colonies from each sample were picked. These were dissolved in 15 µL of NF water and vortexed vigorously. Five µL of this solution was used as template for PCR amplification. PCR mixture consisted of Taq buffer (1X), dNTPs (2 mM), $MgCl_2$(2 mM), forward and reverse primers (20 pmole each), and 1 U of Taq polymerase in a final volume of 25 µL. Thermal profile was the same as for the amplification of linker-ligated product (nested PCR; section 1.2.6), except that the initial denaturation time was of 15 minutes. After run over, PCR product was resolved on 1.2 % agarose gel under the constant current of 60 mA and visualized under UV transilluminator.

3.3 SEQUENCE ANALYSIS

3.3.1 Plasmid prep

Colonies confirmed positive for foreign DNA sequence by colony PCR were further processed for sequencing. Firstly the colonies were suspended in 15µL sterile Nuclease Free water and 5µL of this was suspended in 5mL LB media and allowed to grow overnight for plasmid prep. Plasmids were extracted using Thermoscientific GeneJET plasmid miniprep kit according to manufacturer's protocol. After 16hrs of growth, cells were pelleted down by centrifugation at 8000 rpm for 5 minutes; supernatant was discarded. Cells were resuspended in 250µL Resuspension slolution provided with the kit. Lysis solution (250µL) was added and mixed thoroughly by inverting the tube several times to lyse the cells until the solution became slightly viscous and clear. To this lysate 350 µL of neutralization solution was added and mixed again. The mixture was centrifuged at 1300rpm for 5 minutes to pellet down cell debris and chromosomal DNA. Supernatant was carefully transferred to the

supplied GeneJET spin column by a pipette. The mixture was centrifuged again for 1 minute at 13000rpm and flow-through was discarded. Columns were placed back into the same collection tube. Plasmids in the columns were washed by 500µL of the Wash Solution and centrifuged for 1 minute at 13000rpm. This step was repeated twice to avoid residual ethanol in the plasmid preps. The plasmid was eluted by adding 50µL of the elution buffer, incubating 2 minutes at room temperature and centrifuging the column in a sterile microfuge tube for 2 minutes at 13000rpm. Plasmids were then stored at -20°C till further processing.

3.3.2 Sequencing of inserts

Purified plasmids were subjected to automated sequencing services at Macrogen Inc. Korea. Each plasmid was sequenced twice using M13 forward and M13 reverse universal primers.

3.3.3 Sequence alignment

Sequence alignments were performed using the free Biological Sequence Alignment editor 'BioEdit' version 7.1.9. Copyright ©1997-2013, Tom Hall, Ibis Biosciences. One sequence from each plasmid was aligned with reverse complementary sequence of the other sequence of same plasmid. Region between left and right flanks of the TA insertion site was then aligned with ltr 2 and linker2 primer sequences. The flanking plasmid region was omitted and alignment file for each sequence was saved in "*.fas" (fasta) format.

3.3.4 Genomic insertion site identification

The DNA sequence between two primers was aligned against the human genome using an online tool BLAT (BLAST like Alignment Tool). On DNA, Blat

works by keeping an index of an entire genome in memory. Thus, the target database of BLAT is not a set of GenBank sequences, but instead an index derived from the assembly of the entire genome. By default, the index consists of all non-overlapping 11-mers except for those heavily involved in repeats.

Reasons for using BLAT instead of BLAST were

1. Speed
2. Result with only >95% identity are included
3. Provides a direct link UCSC genome browser. This was helpful in identifying the locus and different features on that locus.

However there is a disadvantage in BLAT. It does not use a query length less than 20bp. For Sequences less than 20 bp BLAST was used. For these sequences the inclusion criteria was more stringent. These sequences were considered as integration site only if

1. They showed 100% identity for 100% query coverage at only one locus in the genome.

The sequence identical to human genomic DNA must have started within 3bp adjacent to MMTV provirus end sequence.

For the rest of the sequences >95% identity and location adjacent to MMTV provirus was considered sufficient to call it a true integration event.

CHAPTER 4

RESULTS

4.1 MMTV-LTR

Samples were collected from hospitals of Rawalpindi and Islamabad. A total of 80 FFPE breast tumor biopsy and blood samples diagnosed by histopathology examination by a qualified pathologist were considered for the study. DNA quality assessment was done by amplification of 121bp fragment of β-globin gene by PCR. Samples with sharp band of β-globin were further processed for the presence of LTR region. 663bp fragment LTR region was amplified from 26 samples (previously published data (Naushad *et al.*, 2014)).

4.2 LM-PCR

Out of these 8 breast tumor samples were randomLy selected and subjected to nested LM-PCR (ligation mediated PCR) (figure 4.1).

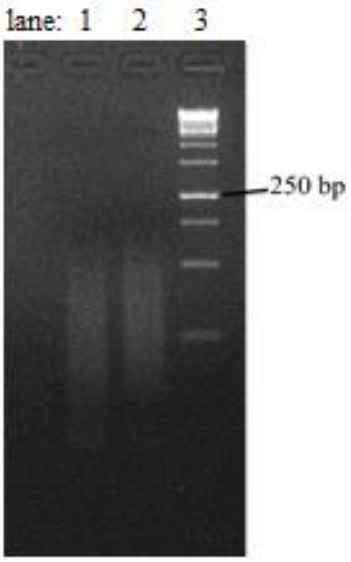


Figure 4.1 LM-PCR.
Product in lane 1 and 2. Marker in lane3. A smear of amplified DNA is visible.

4.3 CLONING OF SEQUENCES

The product of 2nd round was shotgun cloned and up to 30 transformed colonies were randomLy selected from each sample. Each colony was subjected to colony PCR using the same primers to confirm true insertion. Plasmids were extracted from each positive colony.

4.4 SEQUENCING AND ALIGNMENT

All plasmids were sequenced twice using two universal primers flanking the insert sequence. All sequences were aligned against primer sequences (ltr2 and linker2). Based on the criteria defined in chapter no. 3, 59 sequences were considered as true integrations.

4.5 INTEGRATION SITES

We identified 13 unique integration loci. At least 5 loci were cloned from more than one sample. One locus 3q26.2 was cloned from 7 out of 8 samples and 4q21.23 was cloned from 5 out of eight samples. The graph below (fig 4.2) shows the frequency of cloned loci.

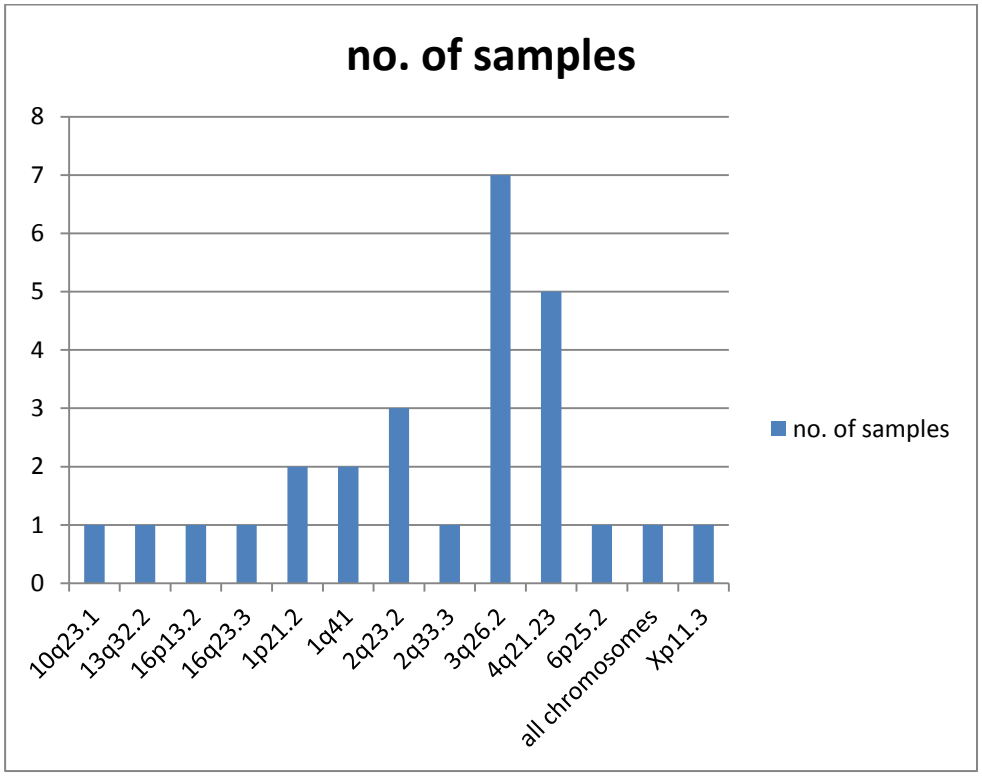
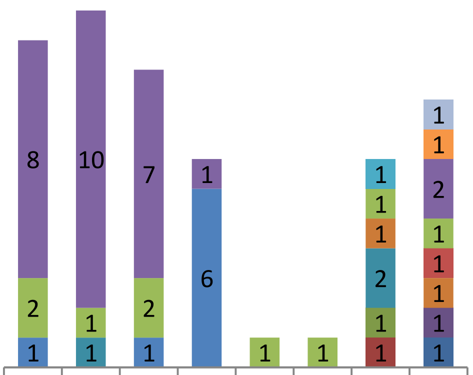


Figure 4.2 Frequency of Identified loci. 3q26.2 was identified in most of the samples (7/8) followed by 4q21.23 (5/8).

Number of clones from same loci in each sample, are given in figure 4.3. The graph shows that most of sites were found in 1 or 2 clones from each sample except for 4q21.23. A detailed account of all sites identified, showing their sequences, nearest genes, and role of these genes is given in the table 4.1.

Number of clones from each locus



	S No.6	S No. 7	S No. 12	S No. 19	S No. 43	S No. 46	S No. 48	S No. 54
all chromosomes								1
Xp11.3								1
6p25.2							1	
4q21.23	8	10	7	1				2
3q26.2	2	1	2		1	1	1	1
2q33.3								1
2q23.2	1		1	6				
1q41							1	1
1p21.2		1					2	
16q23.3								1
16p13.2							1	
13q32.2							1	
10q23.1								1

Figure 4.3 Number of clones for each locus

Table 4.1 Detailed account of all identified sites.

Sample	Clone	Sequence	Repetitive Region	Integration locus	Nearest gene	Distance to nearest gene	Gene description
6	1	CCTTCGAAGCGTAGGCTTAA		3q26.2	GOLIM4	50855	Plays a role in endosome to Golgi protein trafficking; mediates protein transport along the late endosome-bypass pathway from the early endosome to the Golgi. Associated with nasopharynx cancer
	2	AACGA*TGTTCT*CTTAA	LINE (L1PA4)	2q23.2	MBD5	43206	Binds to polycomb repressor complex. Does not interact with either methylated or unmethylated DNA (in vitro)
	3	CAACCAGCAGAATTCCCTTACCAAACTACACTGTTGCTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	9	CACAACCAGCAGAATTCCCTTACCAAACTACACTGTTGCTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	14	CACAACCAGCAGAATTCCCTTACCAAACTACA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes

		CTGTTGCTTAA					
	17	CACAACCAGCAGAATTCCCTTACCAAACTACACTGTTGCTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	20	CACAACCAGCAGAATTCCCTTACCAAACTACACTGTTGCTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	22	CACAACCAGCAGAATTCCCTTACCAAACTACACTGTTGCTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	24	CCTTCGAAGCGTAGGCTTAA		3q26.2	GOLIM4	50855	Plays a role in endosome to Golgi protein trafficking; mediates protein transport along the late endosome-bypass pathway from the early endosome to the Golgi. Associated with nasopharynx cancer
	25	CACAACCAGCAGAATTCCCTTACC	LINE (L2C)	4q21.23	unidentified cDNA AK095285.	same region. 322425 bp homeobox	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number

		AAACTACA CTGTTGCTT AA			Nkx6.1	protein Nkx-6.1	of genes
	27	CACAACCA GCAGAATT CCCTTACC AAACTACA CTGTTGCTT AA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
7	4	CACAACCA GCAGAATT CCCTTACC AAACTACA CTGTTGCTT AA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	6	CACAACCA GCAGAATT CCCTTACC AAACTACA CTGTTGCTT AA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	8	CACAACCA GCAGAATT CCCTTACC AAACTACA CTGTTGTTT AA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	10	CTAATTCT GCTCTGAC		1p21.2	3 unidentified mRNA and	3 unidentified sequences in	This gene encodes a component of the diphthamide synthesis pathway

		TTTACCATTGTTGTTTTAA			cDNA clones. Nearest identified gene is diphtine synthase isoform a	same region. Diphtine synthase: 36477	
	11	CACAACCAGCAGAATTCCCTTACCAAACTACACTGTTGTTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	12	CACAACCAGCAGAATTCCCTTACCAAACTACACTGTTGTTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	14	CACAACCAGCAGAATTCCCTTACCAAACTACACTGTTGTTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	16	CACAACCAGCAGAATTCCCTTACCAAACTACA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes

		CTGTTGTTTAA					
	21	CACAACCAGCAGAATTCCCTTACCAAACTACACTGTTGTTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	23	CCTTCGAAGCGTAGGCTTAA		3q26.2	GOLIM4	50855	Plays a role in endosome to Golgi protein trafficking; mediates protein transport along the late endosome-bypass pathway from the early endosome to the Golgi. Associated with nasopharynx cancer
	24	CACAACCAGCAGAATTCCCTTACCAAACTACACTGTTGTTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	26	CACAACCAGCAGAATTCCCTTACCAAACTACACTGTTGTTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
12	3	CACAACCAGCAGAATTCCCTTACC	LINE (L2C)	4q21.23	unidentified cDNA AK095285.	same region. 322425 bp homeobox	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number

		AAACTACA CTGTTGTTT AA			Nkx6.1	protein Nkx-6.1	of genes
	6	CACAACCA GCAGAATT CCCTTACC AAACTACA CTGTTGTTT AA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	7	CACAACCA GCAGAATT CCCTTACC AAACTACA CTGTTGTTT AA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	9	CCTTCGAA GCGTAGGC TTAA		3q26.2	GOLIM4	50855	Plays a role in endosome to Golgi protein trafficking; mediates protein transport along the late endosome-bypass pathway from the early endosome to the Golgi. Associated with nasopharynx cancer
	11	CACAACCA GCAGAATT CCCTTACC AAACTACA CTGTTGTTT AA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	13	CACAACCA GCAGAATT	LINE (L2C)	4q21.23	unidentified cDNA	same region. 322425 bp	Transcription factor which binds to specific A/T-rich DNA sequences in the

		CCCTTACC AAACTACA CTGTTGTTT AA			AK095285. Nkx6.1	homeobox protein Nkx-6.1	promoter regions of a number of genes
	19	AACGATGT TCTCTTAA	LINE (L1PA4)	2q23.2	MBD5	43206	Binds to heterochromatin. Does not interact with either methylated or unmethylated DNA (in vitro)
	8	CACAACCA GCAGAATT CCCTTACC AAACTACA CTGTTGTTT AA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	20	CCTTCGAA GCGTAGGC TTAA		3q26.2	GOLIM4	50855	Plays a role in endosome to Golgi protein trafficking; mediates protein transport along the late endosome-bypass pathway from the early endosome to the Golgi. Associated with nasopharynx cancer
	26	CACAACCA GCAGAATT CCCTTACC AAACTACA CTGTTGTTT AA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
19	2	CACAACCA GCAGAATT CCCTTACC AAACTACA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes

		CTGTTGCTTAA					
	6	AACGATGTTCTCTTAA	LINE (L1PA4)	2q23.2	MBD5	43206	Binds to heterochromatin. Does not interact with either methylated or unmethylated DNA (in vitro)
	9	CACAACCAGCAGAATTCCCTTACCAAACTACACTGTTGCTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	10	CACAACCAGCAGAATTCCCTTACCAAACTACACTGTTGCTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	11	CACAACCAGCAGAATTCCCTTACCAAACTACACTGTTGCTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	13	CACAACCAGCAGAATTCCCTTACCAAACTACACTGTTGCTTAA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes

	17	CACAACCA GCAGAATT CCCTTACC AAACTACA CTGTTGCTT AA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
43	1	CCTTCGAA GCGTAGGC TTAA		3q26.2	GOLIM4	50855	Plays a role in endosome to Golgi protein trafficking; mediates protein transport along the late endosome-bypass pathway from the early endosome to the Golgi. Associated with nasopharynx cancer
46	3	CCTTCGAA GCGTAGGC TTAA		3q26.2	GOLIM4	50855	Plays a role in endosome to Golgi protein trafficking; mediates protein transport along the late endosome-bypass pathway from the early endosome to the Golgi. Associated with nasopharynx cancer
48	17	CTAATTCT GCTCTGAC TTTACCATT GTTGTCTA A		1p21.2	3 unidentified mRNA and cDNA clones. Nearest identified gene is diphtine synthase isoform a	3 unidentified sequences in same region. Diphtine synthase: 36477	This gene encodes a component of the diphthamide synthesis pathway
	9	CCTTCGAA GCGTAGGC		3q26.2	GOLIM4	50855	Plays a role in endosome to Golgi protein trafficking; mediates protein

		TTAA					transport along the late endosome-bypass pathway from the early endosome to the Golgi. Associated with nasopharynx cancer
	8	ACGATGCT CTCCTAA		16p13.2	UPF0472	65101	Function not yet described
	7	CCAGCGAA TTTTTAGAT GATTATTA TATACTTTA TTTCTTTTT TTTTTC		13q32.2	FARP1	gene disrupted	Functions as guanine nucleotide exchange factor for RAC1.
	4	AACGATGT TCTCCTAA		1q41	GPATCH2	gene disrupted	G-patch domain containing protein. It is involved in negative regulation of phosphatase activity (UNIPROT).
	12	CTAATTCT GCTCTGAC TTTACCATT GTTGTTCTA A		1p21.2	3 unidentified mRNA and cDNA clones. Nearest identified gene is diphtine synthase isoform a	3 unidentified sequences in same region. Diphtine synthase: 36477	This gene encodes a component of the diphthamide synthesis pathway
	14	CATTTCAG CACTCAGT TAA	LINE	6p25.2	PX domain containing protein1	13957	phosphatadyl inositol binding
54	13	ACACATAC AGCTGTAG	LINE(L 1ME2z)	16q23.3	MPP6	gene disrupted	RNA-binding protein that associates with the RNA exosome complex.

		ACTTAA					
	16	TTGCAGGT GTTGTCTGT AAGGACGC TCTCTAA	LTR16A	Xp11.3	Retinitis pigmentosa 2	gene disrupted	GTP binding and GTPase activator activity.
	28	AAAGAAAG AAGGAGGA ACATCTTCT GCCCTAA		2q33.3	PARD3B	gene disrupted	Putative adapter protein involved in asymmetrical cell division and cell polarization processes
	27	CACAACCA GCAGAATT CCCTTACC AAACTACA CTGTTGCTT AA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes
	20	CCTTCGAA GCATAGGC TTAA		3q26.2	GOLIM4	50855	Plays a role in endosome to Golgi protein trafficking; mediates protein transport along the late endosome-bypass pathway from the early endosome to the Golgi. Associated with nasopharynx cancer
	19	AACGATGT TCTCCTA		1q41	GPATCH2	gene disrupted	protein phosphatase1, regulatory subunit 30
	18	CACAACCA GCAGAATT CCCTTACC AAACTACA CTGTTGCTT AA	LINE (L2C)	4q21.23	unidentified cDNA AK095285. Nkx6.1	same region. 322425 bp homeobox protein Nkx-6.1	Transcription factor which binds to specific A/T-rich DNA sequences in the promoter regions of a number of genes

	1	GTTTATTTC CTAA		all chromosomes			
	9	TCCTCATCC ACTTCTTCA CTGCCTGA ATCATTATT TCCTAA		10q23.1	NRG3	476660	Direct ligand for the ERBB4 tyrosine kinase receptor.

CHAPTER 5

DISCUSSION

Available data on the presence of MMTV-like sequences in human breast cancer is not consistent. The variability of results range from 0% prevalence (Fukuoka *et al.*, 2008; Witt *et al.*, 2003; Zangen *et al.*, 2002) to 78% in Tunisian population (Hachana *et al.*, 2008). Similarly, some groups have demonstrated MMTV expression in human breast cancer (James S Lawson *et al.*, 2004) while others have been unsuccessful in showing any expression of MMTV genes in human breast cancers. One group (Mant *et al.*, 2004) amplified MMTV env sequences but later sequencing results identified these sequences to be of human origin. This variation in results can be explained by geographical differences, socio-economic differences and different detection strategies. However, most of the studies do show the presence of MMTV-like sequences in some percentage of breast cancer population. Our previous results were consistent with this notion. In Pakistani population 660 bp env region (identified by Wang *et al.* 1995) was prevalent in 20% of breast cancer samples and 663bp LTR region (identified by Wang *et al.*, 2001) was prevalent in 26% of the samples (Naushad *et al., 2014).* Besides these prevalence studies, Indik *et al.* has demonstrated that MMTV can infect and grow in human breast cancer cell lines (Indik *et al.*, 2005a). They have also identified a few flanking regions to confirm the integration of MMTV in human cell lines. Faschinger and colleagues (2008) infected human cell lines and identified 298 unique initial integration sites in human cell lines. Despite all this data, these association studies cannot identify any mechanism for MMTV-induced cancers in humans. Moreover, MMTV pathogenesis in mouse cannot be

taken as model for carcinogenesis in humans. There are similarities which include hormone responsiveness, upregulation of homologous genes in both species and totally random initial integration in cell lines derived from both species but there are differences and vague scenarios too. Infection in human cell lines, do not utilize the same receptor as in mice. Also there is lack of data on mode of transmission in humans, common integration sites and interaction of viral and cellular proteins. This lack of information makes the pathogenesis model of mouse questionable. Since there is no data on MMTV integration sites cloned from MMTV positive cancer samples, we have identified integration sites from different cancer tissues positive for MMTV-like sequences using LM-PCR.

In our study we processed 8 human breast cancer biopsy samples in which 13 different sites were identified. At least 5 sites were cloned from more than one sample, suggesting a bias towards these sites. Integration at exactly same site in a more than one unrelated samples cannot be by mere chance. One site, 3q26.2, was cloned from 7 samples and another one, 4q21.23, was amplified from 5 different samples. These findings go against the current model of MMTV infection in human cells i.e. MMTV integrates randomLy in human cells (Faschinger *et al.*, 2008). One reason can be that MMTV-like virus of humans utilizes a different mode of infection than MMTV infection in mice. In order to eliminate the chance of another insertion sequence being amplified instead of MMTV, two samples were sequenced for 663bp MMTV-LTR, it showed 95% homology to the available sequence (results from previous study (Naushad *et al.* 2014)).

One sample drew attention as 8 different sites were cloned from this sample (most cloned from one sample); these included 4 of the 5 repeating insertion sites.

Only this particular sample's data is consistent with available data. Firstly, there are more than one integration sites in the genome, secondly, more than one gene was disrupted and thirdly more than one gene lied in the regions under influence of MMTV enhancers.

No specific pattern among repeating sites was observable, however, due to small sample size it is too early for such a claim. Five RefSeq genes were disrupted in by integration events, 3 of these disruptions (PARD3B (2q33.3), MPP6 (16q23.3) and Retinitis pigmentosa 2 (Xp11.3) genes) were unique to the sample with 8 different insertion sites and fourth one (GPATCH2(1q41) gene) was identified in two samples. 5 other integration sites had genes within 66kb region of insertion site, it should be mentioned again that MMTV enhancer elements can regulate gene expression in up to 157kb region (Theodorou *et al.* 2007). These genes were GOLIM4, MBD5, UPF0472, PX domain containing protein and diphthine synthase isoform a. One specific sequence of 13 bp was identified which showed 100% identity at several locations on all chromosomes.

Sites which have been identified from more than two samples are identified as Common Integration Sites (CIS) in our study. Although to mark a site as a common integration site require a much larger sample and at least more than 2 samples should have those sites but due to our small sample size, we have to call those sites which have been amplified from more than one sample as CIS. As expected these sites should match those of mice which mainly include integration in or near *wnt* and *fgf* genes but our data yielded much different results. CIS identified in our samples are discussed below.

Most commonly identified site, 3q26.2, was 50,855 bp upstream of the GOLIM4. According to the available data, this gene must have been overexpressed and its expression should be analyzed in further studies. GOLIM4 encodes Golgi Integral Membrane Protein 4. It is a protein localized in golgi complex and late endosomal compartments but the exact molecular functions of this gene are yet unknown. However, there has been an observed association of gain of copy number at this locus with breast cancer (Weber-Mangal *et al.*, 2003). Similarly, over expression of another gene further downstream at this locus EVI1 (ecotropic viral integration site 1) has been related to breast cancer (Nanjundan *et al.*, 2007). Whether MMTV-like virus is involved in causing chromosomal aberrations or disturbs the expression of genes further downstream than previously reported, needs to be confirmed with further studies.

This site was second most cloned site. Nearest genes at this site are Nkx6.1, COQ2 and HELQ genes. Nkx6.1 is a homeobox protein that binds AT rich DNA and is a potent transcription activator (Iype *et al.*, 2004). COQ2 encodes coenzyme Q2 4-hydroxybenzoate polyprenyltransferase and is involved in oxidative stress mechanisms. Similarly HELQ is a protein with proposed function in DNA repair (Ward *et al.*, 2010). Aberrant expression of all these genes can be associated with progression of cancer.

This site was cloned from 3 different samples. Genes in the vicinity of this site include MBD5, EPC2 *and* ORC4. MBD5 encodes methyl-CpG binding domain protein 5. This protein associates with chromatin but doesn't bind either methylated or unmethylated DNA. However it associates with polycomb repressor complex which is involved in gene silencing and stem cell fate (Baymaz *et al.*, 2014). In addition to the

MBD domain, this protein contains a PWWP domain (Pro-Trp-Trp-Pro motif), which consists of 100-150 amino acids and is found in numerous proteins that are involved in cell division, growth and differentiation (provided by RefSeq, Mar 2010). Similiarly, EPC2 is Enhancer of Polycomb homolog 2. Based on similarity with drosophila EPC, UniProt suggests it to be involved in DNA repair. ORC4 encodes for Origin Recognition Complex subunit 4. This complex is involved in DNA replication. Again all these proteins have a potential role in carcinogenesis.

This site was identified in two samples, but there were no significant genes in its vicinity. However, UCSC genome browser did show 3 unidentified mRNA/cDNA sequences that most probably were expressed in this region and disrupted by this integration. As have been suggested MMTV can identify genes involved in carcinogenesis (Theodorou *et al.*, 2004), detailed study of such unidentified mRNAs may lead to discovery of new genes involved in cancer.

This site was also identified in two samples. It is an insertion in the start of GPATCH2 gene (G Patch Domain containing protein 2) just adjacent to CDS, causing an obvious overexpression. Its overexpression is associated with breast cancer (Lin *et al.*, 2009).

Other sites identified were unique in each sample. Important genes disrupted included FARP1 and PARD3B. FARP1 encodes FERM, RhoGEF And Pleckstrin Domain-Containing Protein. This protein functions as guanine nucleotide exchange factor for RAC1 (UniProt). PARD3B is a gene that codes for Partitioning defective 3 homolog B. This protein is present at tight junctions and may play a role in the formation of tight junctions (UniProt).

Any change in the expression levels of above mentioned proteins or loss of function might contribute to different steps involved in cancer. However, a definite statement cannot be made unless an expression analysis of all these genes is done in samples with integration at these sites.

CONCLUSION AND FUTURE PROSPECTS

We have identified 13 different integration sites of MMTV-like sequences in human breast cancer tissues. There are downstream genes which have not yet been related with breast cancer. Based on MMTV induced cancer studies in mice, it can be stated that these genes might be potential oncogenes which need to be studied further. However, any definitive statement cannot be made until expression analysis of all these genes is done in MMTV-positive tissues. Furthermore role of retroviruses in chromosomal rearrangements should also be studied to identify exact role of MMTV in human breast cancers.

REFERENCES

Baymaz, H. I., Fournier, A., Laget, S., Ji, Z., Jansen, P. W. T. C., Smits, A. H., . . . Vermeulen, M. (2014). MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. *PROTEOMICS*: n/a-n/a.

Beato, M. (1996). Chromatin structure and the regulation of gene expression: remodeling at the MMTV promoter. *Journal of molecular medicine, 74*(12): 711-724.

Bentvelzen, P. (1974). Host-virus interactions in murine mammary carcinogenesis. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 355*(3): 236-259.

Bindra, A., Muradrasoli, S., Kisekka, R., Nordgren, H., Wärnberg, F., *and* Blomberg, J. (2007). Search for DNA of exogenous mouse mammary tumor virus-related virus in human breast cancer samples. *Journal of general virology, 88*(6): 1806-1809.

Bittner, J. J. (1943). Mammary cancer in fostered and unfostered C3H breeding females and their hybrids. *Cancer research, 3*(7): 441-447.

Bramblett, D., Hsu, C., Lozano, M., Earnest, K., Fabritius, C., *and* Dudley, J. (1995). A redundant nuclear protein binding site contributes to negative regulation of the mouse mammary tumor virus long terminal repeat. *Journal of virology, 69*(12): 7868-7876.

Boyle, P., *and* Levin, B. (2008). World cancer report 2008. IARC Press, International Agency for Research on Cancer.

Büchen-Osmond, C. (2006). ICTVdB Management. 00.061.1.01.001. Mouse mammary tumor virus. In: ICTVdB from http://ictvdb.bio-mirror.cn/ICTVdB/00.061.1.01.004.htm

Buchschacher Jr, G. L. (2001). Introduction to retroviruses and retroviral vectors. *Somatic cell and molecular genetics, 26*(1-6): 1-11.

Callahan, R., *and* Smith, G. H. (2008). Common integration sites for MMTV in viral induced mouse mammary tumors. *Journal of mammary gland biology and neoplasia, 13*(3): 309-321.

Cato, A., Henderson, D., *and* Ponta, H. (1987). The hormone response element of the mouse mammary tumour virus DNA mediates the progestin and androgen induction of transcription in the proviral long terminal repeat region. *The EMBO journal, 6*(2): 363.

Choi, Y., Kappler, J. W., *and* Marrack, P. (1991). A superantigen encoded in the open reading frame of the 3′ long terminal repeat of mouse mammary tumour virus.

Dickson, C., *and* Atterwill, M. (1980). Structure and processing of the mouse mammary tumor virus glycoprotein precursor pr73env. *Journal of virology, 35*(2): 349-361.

Dickson, C., *and* Skehel, J. J. (1974). The polypeptide composition of mouse mammary tumor virus. *Virology, 58*(2): 387-395.

Dickson, C., Smith, R., Brookes, S., *and* Peters, G. (1984). Tumorigenesis by mouse mammary tumor virus: Proviral activation of a cellular gene in the common integration region< i> int-2</i>. *Cell, 37*(2): 529-536.

Diévart, A., Beaulieu, N., *and* Jolicoeur, P. (1999). Involvement of Notch1 in the development of mouse mammary tumors. *Oncogene, 18*(44): 5973-5981.

Durgam, V. R., *and* Tekmal, R. R. (1994). The nature and expression of< i> int-5</i>, a novel MMTV integration locus gene in carcinogen-induced mammary tumors. *Cancer letters, 87*(2): 179-186.

Etkind, P., Du, J., Khan, A., Pillitteri, J., *and* Wiernik, P. H. (2000). Mouse mammary tumor virus-like ENV gene sequences in human breast tumors and in a lymphoma of a breast cancer patient. *Clinical cancer research, 6*(4): 1273-1278.

Etkind, P. R., Stewart, A. F., Dorai, T., Purcell, D. J., *and* Wiernik, P. H. (2004). Clonal isolation of different strains of mouse mammary tumor virus-like DNA sequences from both the breast tumors and non-Hodgkin's lymphomas of individual patients diagnosed with both malignancies. *Clinical cancer research, 10*(17): 5656-5664.

Faedo, M., Ford, C. E., Mehta, R., Blazek, K., *and* Rawlinson, W. D. (2004). Mouse mammary tumor-like virus is associated with p53 nuclear accumulation and progesterone receptor positivity but not estrogen positivity in human female breast cancer. *Clinical cancer research, 10*(13): 4417-4419.

Faschinger, A., Rouault, F., Sollner, J., Lukas, A., Salmons, B., Günzburg, W. H., *and* Indik, S. (2008). Mouse mammary tumor virus integration site selection in human and mouse genomes. *Journal of virology, 82*(3): 1360-1367.

Fernandez-Cobo, M., Melana, S., Holland, J., *and* Pogo, B. (2006). Transcription profile of a human breast cancer cell line expressing MMTV-like sequences. *Infectious Agents and Cancer, 1*(1): 7.

Fletcher, T. M., Ryu, B.-W., Baumann, C. T., Warren, B. S., Fragoso, G., John, S., *and* Hager, G. L. (2000). Structure and dynamic properties of a glucocorticoid receptor-induced chromatin transition. *Molecular and cellular biology, 20*(17): 6466-6475.

Flint, S. J., Racaniello, V. R., Enquist, L. W., *and* Skalka, A. M. (2009). Principles of virology, Volume 2: pathogenesis and control (No. Ed. 3). ASM press.

Ford, C. E., Faedo, M., *and* Rawlinson, W. D. (2004). Mouse mammary tumor virus-like RNA transcripts and DNA are found in affected cells of human breast cancer. *Clinical cancer research, 10*(21): 7284-7289.

Ford, C. E., Tran, D., Deng, Y., Rawlinson, W. D., *and* Lawson, J. S. (2003). Mouse mammary tumor virus-like gene sequences in breast tumors of Australian and Vietnamese women. *Clinical cancer research, 9*(3): 1118-1120.

Fredericks, D., *and* Relman, D. A. (1996). Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clinical microbiology reviews, 9*(1): 18-33.

Fukuoka, H., Moriuchi, M., Yano, H., Nagayasu, T., *and* Moriuchi, H. (2008). No association of mouse mammary tumor virus-related retrovirus with Japanese cases of breast cancer. *Journal of medical virology, 80*(8): 1447-1451.

Gattelli, A., Zimberlin, M. N., Meiss, R. P., Castilla, L. H., *and* Kordon, E. C. (2006). Selection of early-occurring mutations dictates hormone-independent

progression in mouse mammary tumor lines. *Journal of virology, 80*(22): 11409-11415.

Giffin, W., Torrance, H., Rodda, D. J., Préfontaine, G. G., Pope, L., *and* Haché, R. J. (1996). Sequence-specific DNA binding by Ku autoantigen and its effects on transcription.

Golovkina, T. V., Dudley, J. P., *and* Ross, S. R. (1998). B and T cells are required for mouse mammary tumor virus spread within the mammary gland. *The Journal of Immunology, 161*(5): 2375-2382.

Grayson, M. (2012). Breast cancer. [10.1038/485S49a]. *Nature, 485*(7400): S49-S49.

Grimm, S. L., *and* Nordeen, S. K. (1999). A composite enhancer element directing tissue-specific expression of mouse mammary tumor virus requires both ubiquitous and tissue-restricted factors. *Journal of Biological Chemistry, 274*(18): 12790-12796.

Hachana, M., Trimeche, M., Ziadi, S., Amara, K., Gaddas, N., Mokni, M., *and* Korbi, S. (2008). Prevalence and characteristics of the MMTV-like associated breast carcinomas in Tunisia. *Cancer letters, 271*(2): 222-230.

Haraguchi, S., Good, R., Engelman, R., Greene, S., *and* Day, N. (1997). Prolactin, epidermal growth factor or transforming growth factor-< i> α</i> activate a mammary cell-specific enhancer in mouse mammary tumor virus-long terminal repeat. *Molecular and cellular endocrinology, 129*(2): 145-155.

Held, W., Waanders, G. A., Shakhov, A. N., Scarpellino, L., Acha-Orbea, H., *and* MacDonald, H. R. (1993). Superantigen-induced immune stimulation

amplifies mouse mammary tumor virus infection and allows virus transmission. *Cell*, *74*(3): 529-540.

Henrard, D., *and* Ross, S. R. (1988). Endogenous mouse mammary tumor virus is expressed in several organs in addition to the lactating mammary gland. *Journal of virology*, *62*(8): 3046-3049.

Hizi, A., Henderson, L., Copeland, T., Sowder, R., Hixson, C., *and* Oroszlan, S. (1987). Characterization of mouse mammary tumor virus gag-pro gene products and the ribosomal frameshift site by protein sequencing. *Proceedings of the National Academy of Sciences*, *84*(20): 7041-7045.

Hsu, C., Fabritius, C., *and* Dudley, J. (1988). Mouse mammary tumor virus proviruses in T-cell lymphomas lack a negative regulatory element in the long terminal repeat. *Journal of virology*, *62*(12): 4644-4652.

Indik, S., Günzburg, W. H., Salmons, B., *and* Rouault, F. (2005a). Mouse mammary tumor virus infects human cells. *Cancer research*, *65*(15): 6651-6659.

Indik, S., Günzburg, W. H., Salmons, B., *and* Rouault, F. (2005b). A novel, mouse mammary tumor virus encoded protein with Rev-like properties. *Virology*, *337*(1): 1-6.

Iype, T., Taylor, D. G., Ziesmann, S. M., Garmey, J. C., Watada, H., *and* Mirmira, R. G. (2004). The transcriptional repressor Nkx6.1 also functions as a deoxyribonucleic acid context-dependent transcriptional activator during pancreatic beta-cell differentiation: evidence for feedback activation of the nkx6.1 gene by Nkx6.1. *Mol Endocrinol*, *18*(6): 1363-1375.

Jacks, T., Townsley, K., Varmus, H. E., *and* Majors, J. (1987). Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus gag-related polyproteins. *Proceedings of the National Academy of Sciences, 84*(12): 4298-4302.

Johal, H. (2009). *The Mouse Mammary Tumour Virus-Like Virus in hormonally influenced human tissues.* University of New South Wales.

Joshi, D., *and* Buehring, G. C. (2012). Are viruses associated with human breast cancer? Scrutinizing the molecular evidence. [Review]. *Breast Cancer Res Treat, 135*(1): 1-15.

Karapetian, O., Shakhov, A. N., Kraehenbuhl, J.-P., *and* Acha-Orbea, H. (1994). Retroviral infection of neonatal Peyer's patch lymphocytes: the mouse mammary tumor virus model. *The Journal of experimental medicine, 180*(4): 1511-1516.

Katz, E., Lareef, M. H., Rassa, J. C., Grande, S. M., King, L. B., Russo, J., . . . Monroe, J. G. (2005). MMTV Env encodes an ITAM responsible for transformation of mammary epithelial cells in three-dimensional culture. *The Journal of experimental medicine, 201*(3): 431-439.

Kühnel, B., Buetti, E., *and* Diggelmann, H. (1986). Functional analysis of the glucocorticoid regulatory elements present in the mouse mammary tumor virus long terminal repeat: a synthetic distal binding site can replace the proximal binding domain. *Journal of molecular biology, 190*(3): 367-378.

Kwan, H., Pecenka, V., Tsukamoto, A., Parslow, T., Guzman, R., Lin, T., . . . Varmus, H. (1992). Transgenes expressing the Wnt-1 and int-2 proto-

oncogenes cooperate during mammary carcinogenesis in doubly transgenic mice. *Molecular and cellular biology, 12*(1): 147-154.

Lawson, J. S., Tran, D. D., Carpenter, E., Ford, C. E., Rawlinson, W., Whitaker, N., *and* Delprado, W. (2006). Presence of mouse mammary tumour-like virus gene sequences may be associated with morphology of specific human breast cancer. *Journal of clinical pathology, 59*(12): 1287-1292.

Lawson, J. S., Tran, D. D., Ford, C., *and* Rawlinson, W. D. (2004). Elevated expression of the tumor suppressing protein p53 is associated with the presence of mouse mammary tumor-like env gene sequences (MMTV-like) in human breast cancer. *Breast Cancer Res Treat, 87*(1): 13-17.

Lee, J., Moffitt, P., Morley, K., *and* Peterson, D. (1991). Multipartite structure of a negative regulatory element associated with a steroid hormone-inducible promoter. *Journal of Biological Chemistry, 266*(35): 24101-24108.

Leoni, C., Volpicella, M., De Leo, F., Gallerani, R., *and* Ceci, L. R. (2011). Genome walking in eukaryotes. *FEBS Journal, 278*(21): 3953-3977.

Levine, P. H., Pogo, B. G. T., Klouj, A., Coronel, S., Woodson, K., Melana, S. M., . . . Holland, J. F. (2004). Increasing evidence for a human breast carcinoma virus with geographic differences. *Cancer, 101*(4): 721-726.

Liede, A., Malik, I. A., Aziz, Z., Rios, P. d. l., Kwan, E., *and* Narod, S. A. (2002). Contribution of< i> BRCA1</i> and< i> BRCA2</i> Mutations to Breast and Ovarian Cancer in Pakistan. *The American Journal of Human Genetics, 71*(3): 595-606.

Lin, M. L., Fukukawa, C., Park, J. H., Naito, K., Kijima, K., Shimo, A., . . . Katagiri, T. (2009). Involvement of G-patch domain containing 2 overexpression in breast carcinogenesis. *Cancer Sci, 100*(8): 1443-1450.

Liu, B., Wang, Y., Melana, S. M., Pelisson, I., Najfeld, V., Holland, J. F., *and* Pogo, B. G. (2001). Identification of a proviral structure in human breast cancer. *Cancer research, 61*(4): 1754-1759.

Liu, J., Bramblett, D., Zhu, Q., Lozano, M., Kobayashi, R., Ross, S. R., *and* Dudley, J. P. (1997). The matrix attachment region-binding protein SATB1 participates in negative regulation of tissue-specific gene expression. *Molecular and cellular biology, 17*(9): 5275-5287.

Lowther, W., Wiley, K., Smith, G. H., *and* Callahan, R. (2005). A new common integration site, Int7, for the mouse mammary tumor virus in mouse mammary tumors identifies a gene whose product has furin-like and thrombospondin-like sequences. *Journal of virology, 79*(15): 10093-10096.

Luther, S. A., *and* Acha-Orbea, H. (1997). Mouse mammary tumor virus: immunological interplays between virus and host. *Advances in immunology, 65*: 139-244.

MacArthur, C. A., Shankar, D. B., *and* Shackleford, G. M. (1995). Fgf-8, activated by proviral insertion, cooperates with the Wnt-1 transgene in murine mammary tumorigenesis. *Journal of virology, 69*(4): 2501-2507.

Maeda, T., Maeda, M., *and* Stewart, A. F. (2002). TEF-1 transcription factors regulate activity of the mouse mammary tumor virus LTR. *Biochemical and biophysical research communications, 296*(5): 1279-1285.

Maitra, U., Johnson, D. G., Krug, R. M., *and* Paull, T. (2006). CDP/Cutl1 controls differentiation-specific MMTV and cellular gene expression in the mammary gland.

Majors, J., *and* Varmus, H. E. (1983). A small region of the mouse mammary tumor virus long terminal repeat confers glucocorticoid hormone regulation on a linked heterologous gene. *Proceedings of the National Academy of Sciences, 80*(19): 5866-5870.

Mant, C., Gillett, C., D'Arrigo, C., *and* Cason, J. (2004). Human murine mammary tumour virus-like agents are genetically distinct from endogenous retroviruses and are not detectable in breast cancer cell lines or biopsies. *Virology, 318*(1): 393-404.

Marchetti, A., Buttitta, F., Miyazaki, S., Gallahan, D., Smith, G. H., *and* Callahan, R. (1995). Int-6, a highly conserved, widely expressed gene, is mutated by mouse mammary tumor virus in mammary preneoplasia. *Journal of virology, 69*(3): 1932-1938.

Melana, S. M., Holland, J. F., *and* Pogo, B. G. (2001). Search for mouse mammary tumor virus-like env sequences in cancer and normal breast from the same individuals. *Clinical cancer research, 7*(2): 283-284.

Mellentin-Michelotti, J., John, S., Pennie, W. D., Williams, T., *and* Hager, G. L. (1994). The 5'enhancer of the mouse mammary tumor virus long terminal repeat contains a functional AP-2 element. *Journal of Biological Chemistry, 269*(50): 31983-31990.

Mertz, J. A., Simper, M. S., Lozano, M. M., Payne, S. M., *and* Dudley, J. P. (2005). Mouse mammary tumor virus encodes a self-regulatory RNA export protein and is a complex retrovirus. *Journal of virology, 79*(23): 14737-14747.

Mink, S., Ponta, H., *and* Cato, A. C. (1990). The long terminal repeat region of the mouse mammary tumour virus contains multiple regulatory elements. *Nucleic acids research, 18*(8): 2017-2023.

Mok, E., Golovkina, T. V., *and* Ross, S. R. (1992). A mouse mammary tumor virus mammary gland enhancer confers tissue-specific but not lactation-dependent expression in transgenic mice. *Journal of virology, 66*(12): 7529-7532.

Mok, M. T., Lawson, J. S., Iacopetta, B. J., *and* Whitaker, N. J. (2008). Mouse mammary tumor virus-like env sequences in human breast cancer. *International Journal of Cancer, 122*(12): 2864-2870.

Morita, E., *and* Sundquist, W. I. (2004). Retrovirus budding. *Annu. Rev. Cell Dev. Biol., 20*: 395-425.

Mougel, M., Houzet, L., *and* Darlix, J.-L. (2009). When is it time for reverse transcription to start and go? *Retrovirology, 6*(1): 24.

Mukhopadhyay, R., Medina, D., *and* Butel, J. S. (1995). Expression of the mouse mammary tumor virus long terminal repeat open reading frame promotes tumorigenic potential of hyperplastic mouse mammary epithelial cells. *Virology, 211*(1): 84-93.

Nandi, S., *and* McGrath, C. M. (1973). Mammary neoplasia in mice. *Advances in Cancer Research, 17*: 353-414.

Nanjundan, M., Nakayama, Y., Cheng, K. W., Lahad, J., Liu, J., Lu, K., . . . Gray, J. W. (2007). Amplification of MDS1/EVI1 and EVI1, located in the 3q26. 2 amplicon, is associated with favorable patient prognosis in ovarian cancer. *Cancer research, 67*(7): 3074-3084.

Naushad, W., bin Rahat, T., Gomez, M. K., Ashiq, M. T., Younas, M., *and* Sadia, H. (2014). Detection and identification of mouse mammary tumor virus-like DNA sequences in blood and breast tissues of breast cancer patients. *Tumor Biology*: 1-10.

Nusse, R., van Ooyen, A., Cox, D., Fung, Y. K. T., *and* Varmus, H. (1984). Mode of proviral activation of a putative mammary oncogene (int-1) on mouse chromosome 15.

Park, D. J., Southey, M. C., Giles, G. G., *and* Hopper, J. L. (2011). No evidence of MMTV-like env sequences in specimens from the Australian Breast Cancer Family Study. *Breast Cancer Res Treat, 125*(1): 229-235.

Peters, G., *and* Glover, C. (1980). tRNA's and priming of RNA-directed DNA synthesis in mouse mammary tumor virus. *Journal of virology, 35*(1): 31-40.

Pogo, B. G., Melana, S. M., Holland, J. F., Mandeli, J. F., Pilotti, S., Casalini, P., *and* Ménard, S. (1999). Sequences homologous to the mouse mammary tumor virus env gene in human breast carcinoma correlate with overexpression of laminin receptor. *Clinical cancer research, 5*(8): 2108-2111.

Rao, G. S. (1981). Mode of entry of steroid and thyroid hormones into cells. *Molecular and cellular endocrinology, 21*(2): 97-108.

Rassa, J. C., Meyers, J. L., Zhang, Y., Kudaravalli, R., *and* Ross, S. R. (2002). Murine retroviruses activate B cells via interaction with toll-like receptor 4. *Proceedings of the National Academy of Sciences, 99*(4): 2281-2286.

Ross, S. R. (2010). Mouse mammary tumor virus molecular biology and oncogenesis. *Viruses, 2*(9): 2000-2012.

Salmons, B., *and* Gunzburg, W. H. (2013). Revisiting a role for a mammary tumor retrovirus in human breast cancer. *International Journal of Cancer, 133*(7): 1530-1535.

Sambrook, J., *and* Russell, D. W. (2001). *Molecular cloning: a laboratory manual (3-volume set)* (Vol. 999): Cold spring harbor laboratory press Cold Spring Harbor, New York:.

Sarkar, N. H., Whittington, E. S., Racevskis, J., *and* Marcus, S. L. (1978). Phosphoproteins of the murine mammary tumor virus. *Virology, 91*(2): 407-422.

Schlom, J., Colcher, D., Spiegelman, S., Gillespie, S., *and* Gillespie, D. (1973). Quantitation of RNA tumor viruses and viruslike particles in human milk by hybridization to polyadenylic acid sequences. *Science, 179*(4074): 696-698.

Schroeder, J. A., Troyer, K. L., *and* Lee, D. C. (2000). Cooperative induction of mammary tumorigenesis by TGFalpha and Wnts. *Oncogene, 19*(28): 3193-3199.

Shackleford, G. M., MacArthur, C. A., Kwan, H. C., *and* Varmus, H. E. (1993). Mouse mammary tumor virus infection accelerates mammary carcinogenesis

in Wnt-1 transgenic mice by insertional activation of int-2/Fgf-3 and hst/Fgf-4. *Proceedings of the National Academy of Sciences, 90*(2): 740-744.

Theodorou, V., Boer, M., Weigelt, B., Jonkers, J., Van Der Valk, M., *and* Hilkens, J. (2004). Fgf10 is an oncogene activated by MMTV insertional mutagenesis in mouse mammary tumors and overexpressed in a subset of human breast carcinomas. *Oncogene, 23*(36): 6047-6055.

Theodorou, V., Kimm, M. A., Boer, M., Wessels, L., Theelen, W., Jonkers, J., *and* Hilkens, J. (2007). MMTV insertional mutagenesis identifies genes, gene families and pathways involved in mammary cancer. *Nature genetics*, *39*(6): 759-769.

Threadgill, D., Steagall, W., Flaherty, M., Fuller, F., Perry, S., Rushlow, K., . . . Payne, S. (1993). Characterization of equine infectious anemia virus dUTPase: growth properties of a dUTPase-deficient mutant. *Journal of virology, 67*(5): 2592-2600.

Velin, D., Acha-Orbea, H., *and* Kraehenbuhl, J.-P. (1996). The neonatal Fc receptor is not required for mucosal infection by mouse mammary tumor virus. *Journal of virology, 70*(10): 7250-7254.

Velin, D., Fotopoulos, G., Luthi, F., *and* Kraehenbuhl, J.-P. (1997). The nasal-associated lymphoid tissue of adult mice acts as an entry site for the mouse mammary tumor retrovirus. *The Journal of experimental medicine, 185*(10): 1871-1876.

Viralzone, S. I. B. Viralzone; Betaretrovirus- MMTV, from http://viralzone.expasy.org/all_by_species/66.htmL#tab6

Wang, E., Obeng-Adjei, N., Ying, Q., Meertens, L., Dragic, T., Davey, R. A., *and* Ross, S. R. (2008). Mouse mammary tumor virus uses mouse but not human transferrin receptor 1 to reach a low pH compartment and infect cells. *Virology, 381*(2): 230-240.

Wang, Y., Holland, J. F., Bleiweiss, I. J., Melana, S., Liu, X., Pelisson, I., . . . Pogo, B. G. (1995). Detection of mammary tumor virus env gene-like sequences in human breast cancer. *Cancer research, 55*(22): 5173-5179.

Wang, Y., Melana, S., Baker, B., Bleiweiss, I., Fernandez-Cobo, M., Mandeli, J., . . . sPogo, B. (2003). High prevalence of MMTV-like env gene sequences in gestational breast cancer. *Medical Oncology, 20*(3): 233-236.

Wang, Y., Pelisson, I., Melana, S., Go, V., Holland, J., *and* Pogo, B.-T. (2001). MMTV-like env gene sequences in human breast cancer. *Archives of virology, 146*(1): 171-180.

Wang, Y., Pelisson, I., Melana, S. M., Holland, J. F., *and* Pogo, B. G.-T. (2001). Detection of MMTV-like LTR and LTR-env gene sequences in human breast cancer. *International journal of oncology, 18*(5): 1041-1044.

Ward, J. D., Muzzini, D. M., Petalcorin, M. I., Martinez-Perez, E., Martin, J. S., Plevani, P., . . . Boulton, S. J. (2010). Overlapping mechanisms promote postsynaptic RAD-51 filament disassembly during meiotic double-strand break repair. *Mol Cell, 37*(2): 259-272.

Weber-Mangal, S., Sinn, H.-P., Popp, S., Klaes, R., Emig, R., Bentz, M., . . . Jauch, A. (2003). Breast cancer in young women (≤35 years): Genomic aberrations

detected by comparative genomic hybridization. *International Journal of Cancer*, *107*(4): 583-592.

Wills, J. W., *and* Craven, R. C. (1991). Form, function, and use of retroviral gag proteins. *Aids*, *5*(6): 639-654.

Witt, A., Hartmann, B., Marton, E., Zeillinger, R., Schreiber, M., *and* Kubista, E. (2003). The mouse mammary tumor virus-like env gene sequence is not detectable in breast cancer tissue of Austrian patients. *Oncology reports*, *10*(4): 1025-1029.

Woodcock, D., Crowther, P., Doherty, J., Jefferson, S., DeCruz, E., Noyer-Weidner, M., . . . Graham, M. (1989). Quantitative evaluation of Escherichia coli host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic acids research*, *17*(9): 3469-3478.

Wu, X., Li, Y., Crise, B., *and* Burgess, S. M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. *Science*, *300*(5626): 1749-1751.

Yanagawa, S., Kakimi, K., Tanaka, H., Murakami, A., Nakagawa, Y., Kubo, Y., . . . Masuda, T. (1993). Mouse mammary tumor virus with rearranged long terminal repeats causes murine lymphomas. *Journal of virology*, *67*(1): 112-118.

Zammarchi, F., Pistello, M., Piersigilli, A., Murr, R., Di Cristofano, C., Naccarato, A., *and* Bevilacqua, G. (2006). MMTV-like sequences in human breast cancer: a fluorescent PCR/laser microdissection approach. *The Journal of pathology*, *209*(4): 436-444.

Zangen, R., Harden, S., Cohen, D., Parrella, P., *and* Sidransky, D. (2002). Mouse mammary tumor-like env gene as a molecular marker for breast cancer? *International Journal of Cancer, 102*(3): 304-307.

Zapata-Benavides, P., Saavedra-Alonso, S., Zamora-Avila, D., Vargas-Rodarte, C., Barrera-Rodriguez, R., Salinas-Silva, J., . . . Trejo-Avila, L. (2007). Mouse mammary tumor virus-like gene sequences in breast cancer samples of Mexican women. *Intervirology, 50*(6): 402-407.

Zhang, Y., Rassa, J. C., Elena deObaldia, M., Albritton, L. M., *and* Ross, S. R. (2003). Identification of the receptor binding domain of the mouse mammary tumor virus envelope protein. *Journal of virology, 77*(19): 10468-10478.

Zhu, Q., *and* Dudley, J. P. (2002). CDP binding to multiple sites in the mouse mammary tumor virus long terminal repeat suppresses basal and glucocorticoid-induced transcription. *Journal of virology, 76*(5): 2168-2179.