STUDIES ON HUMAN LINE-1 RETROTRANSPOSON ENCODED ORF2 PROTEIN

Ph.D. THESIS

by

SOFIA PILLI

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKE
ROORKEE – 247667 (INDIA)
JULY, 2019

STDIES ON HUAMN LINE-1 RETROTRANSPOSON ENCODED ORF2 PROTEIN

A THESIS

Submitted in partial fulfilment of the requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY

in

BIOTECHNOLOGY

by

SOFIA PILLI

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE – 247 667 (INDIA)
JULY, 2019

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **"STUDIES ON HUMAN LINE-1 RETROTRANSPOSON ENCODED ORF2 PROTEIN''** in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy and submitted in the Department of Biotechnology of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from January, 2014 to July, 2019 under the supervision of Dr. Prabhat K Mandal, Assistant Professor, Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other Institution.

(SOFIA PILLI)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(Prabhat K Mandal)
Supervisor

Date:

Abstract

Almost two-third of the human genome is derived from the repetitive DNA sequences which are believed to play significant role in shaping the human genome in the course of evolution. One such repeat called retrotransposons occupied around one-third of the human genome. Retrotransposons are sequences which move from one place of the genome to another place in the same cell using RNA as an intermediate and the process is called retrotransposition. The major consequence of retrotransposition is, it disrupts the sequence where it inserts and thus cause insertion mutation. The Long Interspersed Elements (LINEs or L1s) is the most abundant retrotransposon in the human genome with almost 500,000 copies occupied around 17% of the human genome. Although most of the copies are inactive around 100-150 copies are actively jumping in the recent day human genome. It was first reported in 1988 when Kazazian et al. (Nature 1988; V332; 164-166) found hemophilia A patient resulting from *de-novo* insertion of LINE-1 sequence without any trace of pedigree for the disease. Sequencing of the factor VIII gene from patient showed LINE1 inserted in exon 14 and disrupt the factor VIII gene. Subsequently, the disease causing L1 sequence showed active retrotransposition in cell culture based retrotransposition assay. These observations confirmed for the first time that the transposable element is active in the recent day human genome. An active L1 is 6 kb long with around 906 bp 5'UTR with internal promoter activity, two open reading frames (ORF1 and ORF2) separated by 63 bp spacer sequence and a 3'UTR with around 205 bp long. The element end with Poly A signal sequences at the 3' end and the complete element is flanked by around 20-100 bp target site duplication (TSD) sequences.ORF1p encodes a 40 kDa protein with single stranded nucleic acid binding activity. The ORF2p encoded protein (150 kDa) showed endonuclease (EN) and reverse transcriptase (RT) activities. Although, both proteins are critical in the process of retrotransposition, the exact function in the process of retrotransposition remains unclear. The 5'-UTR of L1 also contains a potent antisense promoter (L1ASP) which transcribes the 5' flanking sequences and thus form chimeric transcripts. L1 is believed to jump by a mechanism called target primed reverse transcription (TPRT) where EN activity of

L1ORF2p introduce a nick at the chromosomal target which creates a 3'OH group. The RT of ORF2p then primes the nick and synthesizes cDNA using L1 mRNA as a template.

Human L1 ORF2p is the key protein in the process of retrotransposition and is extremely difficult to express in-vitro. The expression of this protein is used as an marker to find the active L1 retrotransposition in cells and tissues. Although few antibody against L1-ORF2p is available with some L1 researcher, their quality is not good; many non-specific bands found in immunoblotting.

Prior it was believed that LINE-1 retrotransposons are only active in germ cells (sperm and ovum) and early stage of development. But recent transgenic animal models and high throughput sequencing analysis revealed that L1 is also active in certain parts of normal brain and in few cancers. The studies also showed that cancer of epithelial origin showed more L1 insertion compared to other types of cancer. Although it is known that L1 is highly active in certain cancers, its role towards the development or progression of cancer is completely unknown. Oral cancer a subtype of head and neck is very deadly and highly prevalent in India due to excessive use of tobacco. Very limited study has been performed to see the activity of L1 retrotransposons in oral cancer samples. In this study I have tried to synthesize an human L1 ORF2p (hL1ORF2p) antibody. I have also explored hL1ORF2p expression in oral squamous cell carcinoma (OSCC) samples obtained from Indian patients.

The thesis has been divided into three chapters. **Chapter 1** includes the introduction and detailed literature review about transposable elements specifically about mammalian LINE1 retrotransposons structure, mechanism of retrotransposition and its role in health and disease. The chapter also focuses about LINE1 ORF2 protein activity in retrotransposition and its aberrant expression in different types of cancer along with literature about oral cancer is included.

Chapter 2 comprise the materials and methods used in the research work, Those includes recipes for reagents, solutions, protocols for cloning, expression and purification

of proteins, Site directed mutagenesis, biophysical characterization of purified proteins, protocols for in house antibody generation, tissue processing, immunohistochemistry, animal cell culture, western blotting along with other general techniques.

Chapter 3 embodies details of the results obtained in the study. The main objectives of the study

I) To clone and express fragments of hL1-ORF2p for antibody generation

II) To generate an in house antibody specific to hL1-ORF2p

III) Detection of hL1-ORF2p expression in OSCC samples.

I) To clone and express fragments of hL1-ORF2p for antibody generation

ORF2 protein of human LINE1 contains three domains: - N-terminal endonuclease (EN) domain, central reverse transcriptase (RT) domain and C- terminal CCHC domain. In the present study fragments from CCHC and EN domains were cloned in a bacterial expression vector and its expression was checked in *E.coli* expression cells. Both the domains showed very less expression and the amount was not sufficient to make antibody. Hence, a peptide was designed using bioinformatics tool and a peptide antibody was generated. Although, the peptide antigen showed antibody response, the antibody showed some cross reaction. Next, enhanced green fluorescent protein (EGFP) fused ORF2p sequence was used as antigen for making antibody against ORF2p. The first bleed show good response and the process is ongoing.

II) To generate an in house antibody specific to hL1-ORF2p

Prior generating antibody against hL1-ORF2p, the protocol was standardized using EGFP as an antigen and generated EGFP specific antibody and the results obtained showed specific immune response against EGFP. By following the same procedure the peptide stretch from EN along with carrier protein was injected in to rat for generation of antibody.

III) To detect hORF2p expression in Oral Squamous Cell Carcinoma (OSCC) samples.

OSCC samples were collected from Acharya Tulsi Regional Centre for Cancer and Treatment Bikaner, Rajasthan as all experiments were performed as per institute human ethics committee approval and guidelines. The neoplastic nature of all cancer samples used in this study was confirmed by Hematoxylin and Eosin staining. Next the samples were proceeded to make paraffin block. Slides made from these blocks were then proceeded for Immunohistochemistry with anti human L1 ORF1p antibody [α-hORF1p (RRM)]. The ORF1p positive samples were examined with RT domain specific antibody [α-hORF2p (RT)] which is available in the laboratory. Around 50% samples showed ORF2p positive suggesting human L1 retrotransposon pathway is highly active in OSCC samples in the cancer tissues compared to normal. Data showed very high L1 retrotransposon activity in OSCC which might have some significant role in the onset and progression of this particular type of cancer.

Chapter 4 includes the discussion part of the thesis which concludes the inferences obtained from the results. Further conclusion and future prospective of the work has been discussed.

Acknowledgements

I would like to express my profound gratitude to all the people involved in this journey. First and foremost I would like to take immense pleasure in expressing my deep sense of gratitude to my supervisor **Dr. Prabhat K. Mandal,** who agreed to let me join in his lab five years ago and introduced to the world of L1 biology. I still remember chilly winter mornings often feel warm with his enthusiastic and prepared talks about research and recent findings kindled us to start our research with high hopes. I also want to thank for his constant and untiring support for developing my scientific skills. I am grateful for his constructive criticism and invaluable advice whenever needed while his bounded optimism and courage has always kept my spirits high. Had a chance to learn many lessons along with technical capability that are very fundamental in life. I will sure continue to value his scientific opinions and mentorship.

I would also like to thank the members of my SRC committee for their critiques and insights: Chairman Dr. Bijon Choudhury, Internal expert Dr. Debabrata Sircar and External expert Dr. Kaushik Ghosh. I want to express my gratitude to the current head of the department Prof A.K.Sharma, who has always been helpful and available for having an open door with encouraging attitude.

I take this opportunity to sincerely acknowledge the Ministry of Human Resource Development (MHRD), Government of India for providing me the financial support to complete my course.

Most importantly I am thankful to Dr. Jitendra Nangal (Oncologist, Bikaner, Rajasthan) for providing me with important research material, which was very essential to complete the study. I would like to thank Dr. Vinubalarram (Pathologist, Military hospital, roorkee) for helping me with the suggestions whenever needed and providing her lab facilities for the experimental work.

The members of the RL-8 have earned my gratitude for their help and support in my experiments, Savita, Sur, Mukherjee, Rajkishore and Pooja.

I also want to thank all my colleagues in department for their contribution by allowing to use their lab instruments and offering required chemicals in time of need (Jyothi, Namrata, Timsy and Ananya). In particular, I would like to thank friends for being there for me through thick and thin upholding with their love and warmth.

I am especially thankful for my parents **Sh. Suvarna Raju and Smt Ratna Kumari** who always believed me and supported me in every way they could. I also extend my thanks to my sister Sony younger for their unconditional love and support at all times. I am forever indebted to my family, for their constant love, care, and guidance - in addition to the odd pep talk when I needed a push to keep going. Finally, I owe to almighty for grace and strength throughout this journey.

Sofia Pilli

Publications

1. Extensive LINE-1 retrotransposon promoter Hypomethylation and ORF1p expression in oral Squamous Cell Carcinoma

Savita Budania[1], Debpali Sur[1±], Jitendra Nangal[2±], **Sofia Pilli[1]**, Shilpi Saxena[3], Manash Biswas[3], Ramasare Prasad[1], Vijay Yadav[4], Dustin C. Hancks[5], and Prabhat K. Mandal[1*,]
Head and Neck Wiley journal (Under final revision)

2. Glycine at position 92 is crucial for EGFP fluorescence.
(Manuscript prepared)
Sofia Pilli, Dharma T reddy and Prabhat K Mandal

International Conference

3. Generation of Human LINE1 retrotransposon encoded ORF2p antibody and its detection in Human Brain and Oral Cancer tissues

Sofia Pilli, Koel Mukherjee, Debpali Sur, Savita Budania and Prabhat K Mandal

Indo-US conference on 'Transcription, Chromatin Structure DNA Repair and Genomic Instability.

Organized by: IISc Bangalore

National Conference

4. Generation of Human LINE1 retrotransposon encoded ORF2p antibody and its detection in human tissues

Sofia Pilli, Koel Mukherjee, Debpali Sur, Savita Budhania and Prabhat K Mandal

National Conference on Protein Structure and Dynamics in Health and Agriculture
Organized by: Department of Biosciences and Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, under the aegis of Protein Society, India. Nov 2017

LIST OF ABBREVIATIONS

µg: microgram

Ac: Activator

APEs: apurinic/apyrimidinic type endonucleases

BSA: Bovine Serum Albumin

CC: Coiled Coil

CD: Circular Dichroism

CTD: Carboxy Terminal Domain

CCHC: Cysteine Cysteine Histidine Cysteine

DAB: 3-3´- Diaaminobenzidinetetrahydrochloride (DAB substrate)

DNA : Deoxyribonucleic Acid

Ds: Dissociator

ECL: Enhanced chemiluminescence

EDTA: Ethylene Diamine Tetra-Acetic Acid

ERV: Endogenous Retro Virus

FFPE: Formalin-Fixed Paraffin Embedded

GAPDH: Glyceraldehyde 3-phosphate dehydrogenase

H&E staining: Haematoxylene and Eosin

HNSCC: Head and Neck squamous cell carcinoma

HPV: Human Papilloma Virus

HRP: Horseradish peroxidase

IHC: Immunohistochemistry

IPTG: Isopropyl β-D-1-thiogalactopyranoside

kDa: kilo Dalton

L1-RNPs : LINE-1 ribonucleoparticles

LINE-1: Long INterpersedElement-1

LTR: Long Terminal Repeat

MaLR mammalian apparent long terminal repeat retrotransposon

MCF-7: Michigan Cancer Foundation-7

mm: millimetre

mM: millmolar

Mol Wt: Molecular Weight

NaCl: sodium chloride

NRTIs: Nucleoside Reverse-Transcriptase Inhibitors

°C: degree centigrade

ORF: Open Reading Frame

OSCC: Oral Squamous Cell Carcinoma

PBS: Phosphate Buffered Saline

PBS-T: Phosphate Buffered Saline-Tween

PPs: processed pseudogenes

qPCR: quantitive Polymerase Chain Reaction

RNA: Ribonucleic Acid

RRM: RNA Recognition Motif

RT: Reverse transcriptase

SDM: Site Directed Mutagenesis

SDS-PAGE: Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis

SVA: (SINE-R/VNTR/Alu)

TPRT: Target-Primed Reverse Transcription

TBST: Tris Buffered Saline-Tween

TP53: Tumor Protein 53

TSD: Target Site Duplication

UTR: Untranslated region

V: Volt

x g: times gravity

X-gal: 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

List of Contents

Chapter 1

Introduction

1.1. Overview of Human Genome

The human genome comprises DNA sequences present in nucleus and some organelles like mitochondria. The genome sequence published in 2001 [Lander et al. 2001] *estimated that* the haploid human genome is roughly made with 3 billion base pairs. It contains both protein coding and non coding DNA and estimates suggests that around only 1-2% of the human genome comprises protein coding genes and the remaining is occupied by different types of repeat sequences and those repeats mainly includes microsatellite, satellite and transposable elements [Lander et al. 2001]. The microsatellite and satellite repeats are generally tandem in nature whereas transposable elements are interspersed. The estimates showed that more than one third of the human genome is made with transposon types of repeat. (Figure 1.1 and 1.2)

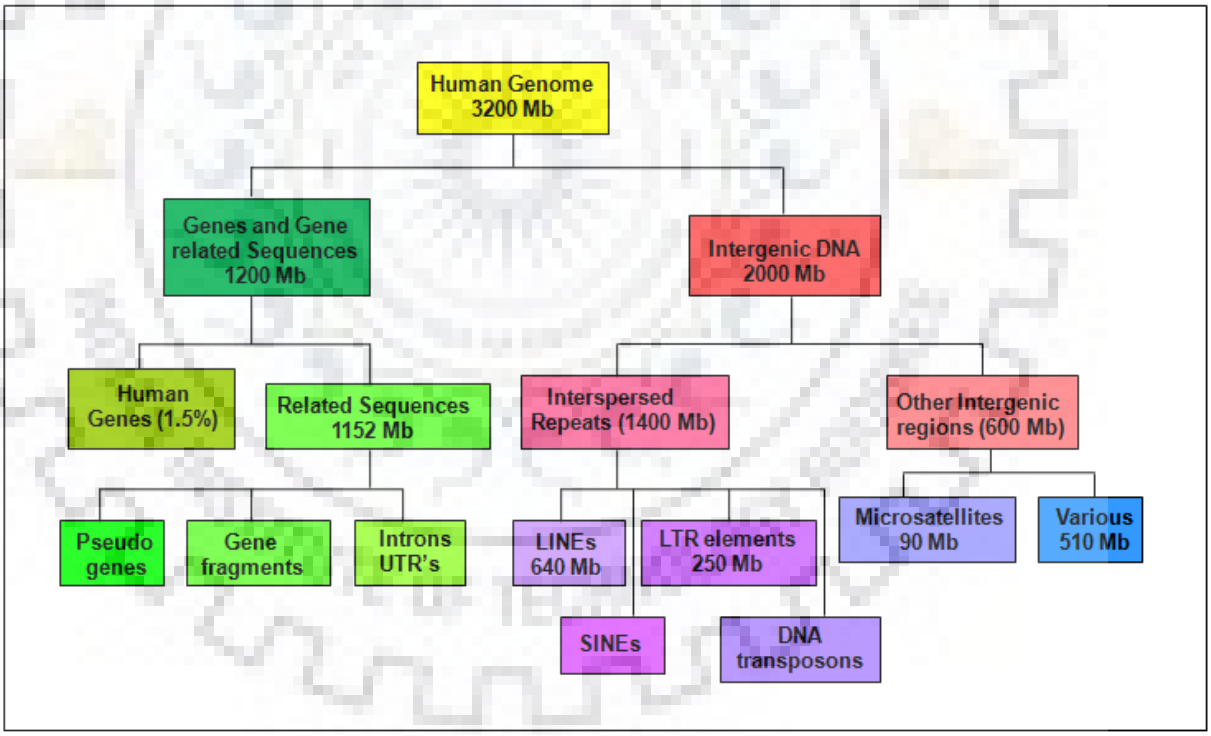


Figure 1.1: Overview of Human genome. Three billion base pairs of human genome divided to two major groups based on function to Genes and Intergenic DNA. Each group further branched into various sub groups depending on their location and mode of action.

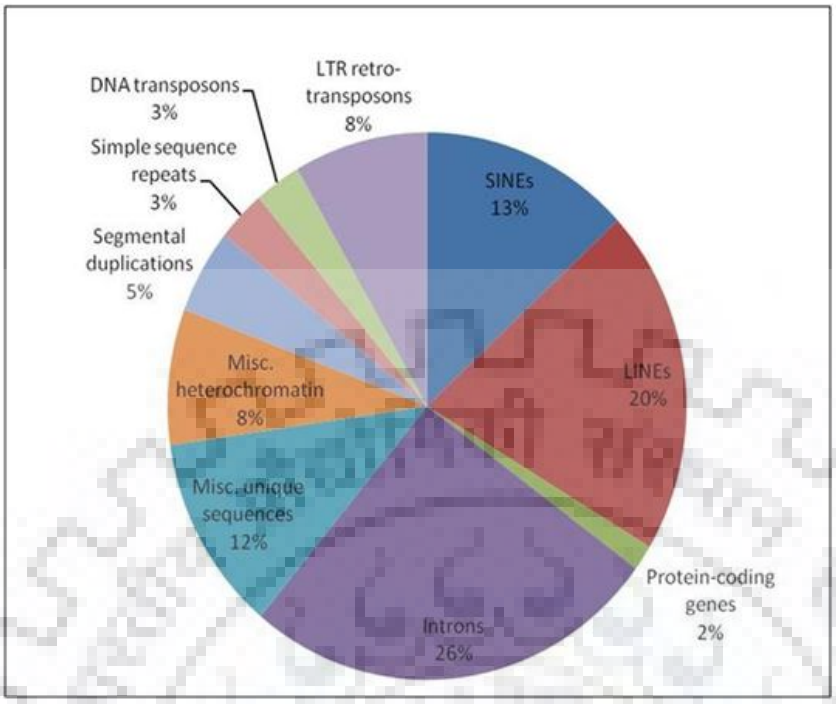


Figure 1.2: Components of human genome: The pie chart showing only around 2% of human genome is responsible for protein coding where as different types of repeat sequences occupied more than 50% of the genome. Among all repeats Long Interspersed Elements (LINEs) occupied the highest proportion which is around 20% of the human genome and then Short Interspersed Element (SINE) which occupied around 13 %. DNA transposons and LTR retrotransposons occupied around 3% and 8% of human genome respectively.

1.2. Transposons

1.2.1. Definition

Transposable elements (TEs) also called as transposons or "jumping genes," are defined as sequences of DNA that have inherent ability to move from one location of the genome to another location. TEs have been identified in both prokaryotes as well as eukaryotes and occupied a substantial fraction of their genomes [Hancks and Kazazian 2016]. For example TEs constitute approximately 12%, 10%, 45% and 37 % and 80% of *C. elega*ns, fish mouse and human and maize genomes respectively [Huang and Boeke et. al., 2012].

1.2.2. Discovery of Transposons

Transposons were first discovered by Barbara McClintock, an American scientist from Cold Spring Harbor Laboratory, New York. Subsequently she won Noble prize for her discovery in 1940 [McClintock B, 1950]. While she was studying maize genetics, she noticed how separate kernels of corn displayed different colors in the same cob. Her research revealed that each corn kernel looked different because it was genetically different. McClintock understood that this could only occur if the genes were not stationary, but mobile, and able to 'jump' from one chromosome to another. Her study concluded that as transposons move from one location to another, their mobility affects the gene expression. Biologists were initially sceptical about McClintock's discovery as it was against principles of Mendelian genetics at that time.

1.2.3. Classification of Transposons

Though there are many ways to categorize transposons, most prominent division is based on the method of transposition involved and type of intermediate product involved. There are two major groups based on the requirement of reverse transcriptase reaction: Class I elements or retrotransposons, and Class II or DNA transposons. Class I elements are a type of TEs that mobilize through reverse transcription [Goodier and Kazazian, 2008; Finnegan DJ, 1997]. They produced RNA transcripts that converted to cDNA by the action of reverse transcriptase before integrated to a new location by copy and paste manner. The Class II TEs are further categorized based on the presence of sequences called Long Terminal Repeats (LTR) on either side. The structure and life cycle of LTR retrotransposons are matched with the life-cycle of retroviruses. Structurally, LTR retroviruses are very similar to retroviruses and differ by lacking envelop gene. In human and mouse genomes, these LTR retrotransposons occupy ~ 8% and ~10% respectively [Lander et al., 2001; Mandal and Kazazian, 2008]. The major class LTR retrotransposon which present in both the organisms genome are belongs to endogenous retroviruses (ERV) and Mammalian apparent long terminal repeat retrotransposon (MaLR) [Ostertag and Kazazian, 2001]. In mouse these elements are active where as in human not a single active element has been found yet [Waterston et al, 2002; Ostertag and Kazazian 2001]. Long INterspersed Elements (LINEs) and Short INterpersed Elements (SINEs)

are belonging to the category of non-LTR retrotransposons. LINEs are autonomous and encode their own protein for retrotransposition whereas SINEs are non-autonomous elements that don't encode any protein and depends on LINEs encoded protein for their mobilisation. More than 20% of the human and mouse genome are occupied by the non-LTR type of retrotransposable elements [Richardson et al., 2014 and Lander et al, 2001].

1.2.4. DNA Transposons

DNA transposons move to new genomic locations via a DNA intermediate. During transposition, transcription of a DNA transposon is initiated from sequences within the transposable element [Craig et al, 2002]. The resultant messenger RNA is transported back to the cytoplasm where translation occurs, leading to the synthesis of a transposase protein. A generic transposase protein, such as that encoded by the eukaryotic Sleeping Beauty DNA transposon [Ivics Z et al, 1997], contains a nuclear localization signal as well as DNA binding and integrase activities. After translation, transposase is imported into the nucleus, where it binds either within or near the transposon inverted terminal repeat sequences to promote transposition by a "cut and paste" or copy and paste mechanism. A given transposase protein can mobilize both protein coding (*i.e.*, autonomous) and non-protein coding sequences (*i.e.*, non-autonomous) to new genome locations. Due to transposase activity, the resultant newly inserted DNA transposon mostly boarded by short target-site duplications (TSD's) on either side of a given class of element.

1.2.5. Long Terminal Repeat (LTR) retrotransposons

LTR retrotransposons or Long Terminal Repeat (LTR) retrotransposons also called are very similar to retroviruses and these elements are boarded by straight repeats of few hundred base pairs long at each ends called LTR. The retrovirus encodes three proteins named: Gag, Pol and Env; all three proteins are required for making new progeny virus in infected cells [Richardson et al, 2015]. Gag is an RNA binding protein where as *pol* encodes a protein with RT activity. The envelope gene encodes a protein named envelop and make the boundary of the virus. To complete the virus life cycle, viral RNA first converted cDNA using pol encoded reverse transcriptase enzyme which

then integrated into the host genome. The integrated virus DNA then transcribed to make new progeny virus when the virus RNA is packed with env protein. Retrotransposons followed the same process like LTR viruses; however, as it does not have a functional env gene, it can't make virus like particle. The unique feature to identify such integrated elements is the presence of direct repeat sequences flanking DNA on either ends and is found throughout genome of *Drosophila melanogaster* genome. Endogenous retroviruses (ERV) are the type of LTR retrotransposon are present in human and mouse genome. Although ERVS are dead in human genome, it is highly active in mice genome. Almost 450,000 copies of ERVs occupied approximately 8% of human genome; whereas in mouse it is around 10% with greater than 600,000 copies [Lander et al., 2001; Waterston et al., 2002 and Mandal and Kazazian, 2008].

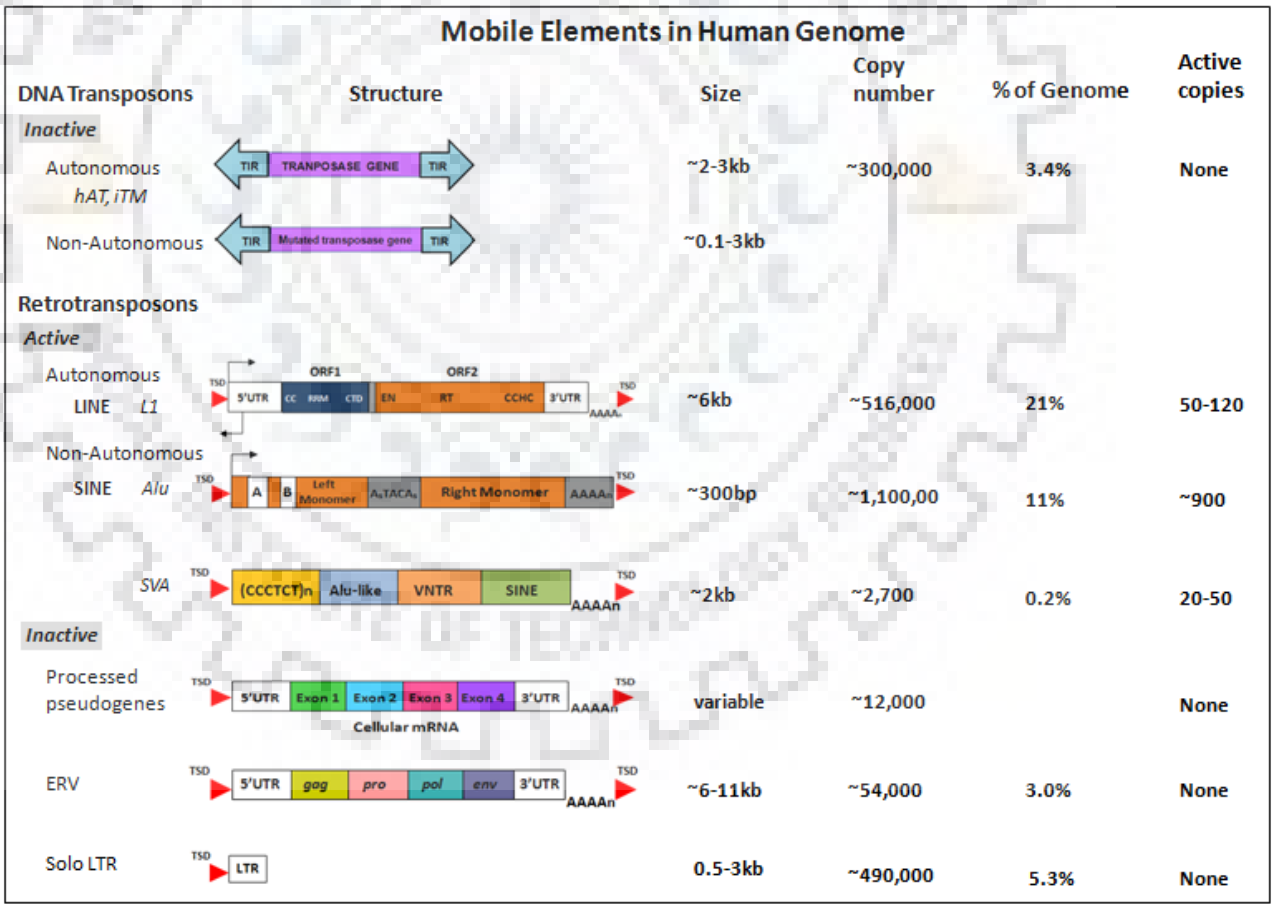


Figure 1.3: Diagrammatic representation of different types of transposable elements. DNA transposons belong to Class II elements and are again divided into two types as autonomous and non-autonomous based on their mode of action. Autonomous can be able to made their own transposes which move the element from one place of the genome to another. Non-

autonomous transposons use transposes encoded by autonomous transposons.TSD: target site duplication, ORF1: Open reading frame 1, ORF2: Open reading frame 2, UTR: untranslated region, A and B: are particular sequence present only in SINE, TIR: terminal inverted repeat, DR: direct repeat. Non-LTR retrotransposons or Class II elements. Autonomous class 1 elements again are divided into two types based on the presence of long terminal repeat (LTR) sequence. The LTR-retrotransposons are like retroviruses with lack of functional envelop gene. The element contains central reverse transcriptase (RT) domain designated as pol. Autonomous non-LTR retrotransposon also called LINEs don't have LTR repeat sequence at the end and contain either two or one open reading frames (ORF2). The ORF2 contains a central RT domain. The non autonomous class element also called SINEs are generally 0.1-0.5 kb in length and depend on autonomous elements machinery for its jumping Details of the structure and abbreviation for LINE-1 UTR: untranslated regions (white box); sense and anti sense internal promoter black arrow. ORF1: denoted by blue box have three domains that comprise of a CCD (coiled coil domain), RNA recognition motif (RRM) and a C- terminal domain (CTD). Intergenic ORF spacer (light blue box between ORF1 and ORF2); ORF2 is yellow box consists of Endonuclease domain (EN), Reverse transcriptase domain (RT) and a Carboxy terminal Cysteine rich domain (CCHC); Poly A tract (long sequences of Adenine downstream of 3'UTR). Human Alu contain left and right monomer with an A-rich linker sequence in between and ends with poly A tail. For human SVA hexameric CCCTCT repeat yellow box; inverted Alu like repeat light blue box; GC rich VNTR brown; SINE-R derived from HERV-K10 (white box). Processed pseudogenes (PP): cellular mRNA spliced with UTRs (white box) and coding ORF where four exons are fused (no intron present between two exons)

1.2.6. Non-LTR retrotransposons

The structure of non-LTR retrotransposons is very different from LTR retrotransposons except the reverse transcriptase domain. Non-LTR retroelements do not have any LTR sequences present at the ends. A regular non-LTR retrotransposon contains two open reading frames (ORF's) ends with a poly(A) at the 3´ end and flanked by target site duplication of around 20 bp (average) length [Richardson et al, 2015]. The mechanism employed by L1 elements differ from LTR retrotransposons. The reason behind this is as there is no sequence present for binding of tRNA primer. The mechanism of retrotransposition of L1 elements initiated by a single strand break (bottom strand) at the target site by element encoded endonuclease [Luan and Eickbush, 1995]. The exposed 3´-OH cats as an primer for reverse transcription reaction which formed a new copy that integrated in to new target location. This process is called target-primed reverse transcription (TPRT) [Luan and Eickbush, 1995; Cost et al., 2002].

1.2.7. Long INterpersed Elements (LINEs) and its discovery

LINEs is a type of non-LTR retrotransposons occupied a significant part of most of the mammalian genome including human. In early 1980's Maxine Frank Singer had just begun an effort to understand the structure and function of repetitive DNA present in many copies of very similar sequence in human and primate genomes. Her lab had found that one kind of repetitive DNA was present in the human genome in roughly 100,000 copies at that time, and called these sequences as *l*ong *in*ter*s*pers*e*d *e*lements (LINEs). In 1986, Alan Scott at Johns Hopkins, been working on L1 and after much work he published consensus L1 sequence in 1987 (Scott et al., 1987) where he showed L1 is around 6 kb in length with two non over-lapping ORFs (ORF1 and ORF2). In 1986, M. Hattori showed that a reverse transcriptase like domain like retroviruses present in the second ORFs of many L1s obtained from various mammals (Hattori et al., 1986). In 1985, Jef Boeke, David Garfinkel, C.A. Styles, and Gerry Fink showed that a yeast retrotransposon named Ty1 use RNA intermediate for its mobility (Boeke et al., 1985) and Jef B for the first time named this kind of sequence as retrotransposons. At the same time Haig H.Kazazian at Johns Hopkins was studying factor VIII mutations in haemophilia A patients. They had collected DNA from 240 patients and analysis of DNA by Southern hybridization had shown abnormal size of factor VIII gene in case of two patients. When it was cloned, this lead to the discovery of human LINE-1, a piece of repetitive DNA suspected of being a transposable element had inserted into the coding region of a gene and likely caused haemophilia A in patient (Johns Hopkins-27) (JH-27) ([Kazazian et al. 1988) . This is the first report showing active retrotransposition can disrupt a gene like factor VIII and causes genetic disease in human.

1.3. Structure of Human LINE-1 element:

Although there are more than 500,000 copies of L1 present in an average human genome, only 100-150 are potentially jumping in recent day human genome. A full length retrotransposition-competent L1(RC-L1) is 6 kb in length with the following features : 1) a ~900 bp CG-rich 5' untranslated region (5' UTR) functioning as an internal promoter, 2) two non-overlapping open-reading frames (ORFs designated ORF1p and ORF2p) separated by a 63 bp spacer sequence, 3) a ~200 bp 3'-UTR and

4) a polyA tail of variable length at the 3'-end of the insertion (Figure 1.4) [Ostertag et al., 2001; Bao et al., 2015].

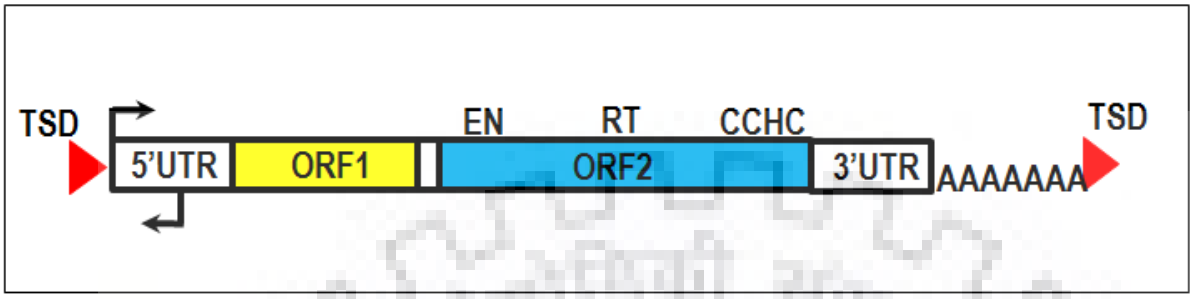


Figure 1.4: Structure of retrotransposition competent human Long Interspersed Nuclear Element (LINE-1). Red triangles represent target site duplication sequences, a unique specific nucleotide sequence flanks at both directions. Both the arrows represent the bi directional promoter followed by 5' untranslated region; ORF1 denoted the first open reading frame encodes 40kDa protein responsible for DNA binding and nucleic acid chaperone activity. Two ORF's are separated by a 63bp Intergenic spacer. ORF2 encodes 150kDa protein with Endonuclease (EN) and reverse transcriptase (RT) activities which were indispensible for process of retrotransposition. CCHC, a cysteine rich domain believed to have RNA binding property that facilitates RT activity in TPRT mechanism.

1.3.1. Human L1 promoter

A full length LINE 1 sequence contains approximately 900bp long 5'UTR which contains bi-directional promoter activities (sense and anti-sense) [Ostertag et al 2001; Speek M 2001; Yang and Kazazian 2006]. The consensus LINE1 sequence contains a polyA signal AATAAA a the 200 bp 3'UTR region, which is required for RNA polymerase II termination suggesting L1 is transcribed by Pol II RNA polymerase. The LINE 1 promoter contains around 40 CpG di-nucleotide at the promoter region. These CpG di-nucleotides are methylated in normal non-dividing cells leading to silencing of L1 transcription [Bao et al 2015; Moutri et al, 2010] However, in many cancers and early stage of development the CpG di-nucleotides show reduced methylation leading to L1 transcription and followed by translation of L1 encoded proteins (L1-ORF1p and L1-ORF2p) [Shukla et al., 2013]. Several transcription factors have been identified which regulate L1 transcription. These includes Ying Yang 1(YY1) [Athanikar et al., 2004], runt related transcription factor 3 (RUNX3) [Yang et al., 2003], testis determining factor gene SRY [Tchénio et al, 2000] and others. The 900 nt of 5′-UTR of

LINE‑1 also contains an antisense promoter (ASP) activity at around 400- 500 nt region which is capable of initiating transcripts in the opposite orientation. Recently an ORF0 protein have been identified and believed to be transcribed by L1 antisense promoter [Denli et al, 2015].

1.3.2. Human L1 ORF1 Protein

Human L1 sequence contains two ORFs in different reading frames. The first ORF designated as ORF1 which encodes a 40 kDa protein (338 amino acid in length) The ORF1p has three distinct domains; Coiled Coil (CC) (amino acids 52-153), RNA Recognition Motif (RRM) (amino acids 157-252), and Carboxy Terminal Domain (CTD) (amino acids 264-323) [Holmes et al.,1992; Khazina et al., 2009, Khazina et al. 2011]. Studies showed that that human ORF1p prefers binding single-stranded (ss) nucleic acid in non-sequence specific manner. The ORF1p also showed nucleic acid chaperone activity [Martin and Bushman, 2001]. Martin et al [Martin and Bushman, 2001 and Khazina et al [Khazina et al., 2011] showed that both human and mouse ORF1p bind to ssRNA and ssDNA with high affinity. Cellular localization showed that the ORF1p is mainly cytoplasmic, occasionally found in nucleus and prefers to bind with own RNA (cis preference) [Wei et al., 2001; Kulpa and Moran 2005]. It is hypothesized that The ORF1p formed L1-RNP with L1RNA and L1-ORF2p and some cellular factors which is the proposed retrotransposition intermediate [Doucet and Mandal et al., 2013; Goodier 2016]. The L1-ORF1p is a unique protein as its amino acids does not match with any known protein present in any organisms available in the database. Structural studies have revealed that the N-terminus CC domain of human ORF1p facilitates trimerization of ORF1p molecules [Martin et al., 2003; Khazina et al., 2009; Khazina et al., 2011]. The central part of ORF1p (RRM domain) with assistance of its C-terminal domain (CTD), is required for ORF1p RNA binding. Alignment of human ORF1 sequence with mouse and rat ORF1 showed that the CTD domain contains three conserved blocks of amino acids ARR at residues 260-262, REKG residues 235-238, and YPAKLS residues 281-287 [Moran et al, 1996]. The ORF1p is extremely sensitive protein as missense mutations of the conserved amino acids present in CTD RRM and CC domains have been reported to abolish or adversely affect retrotransposition of engineered L1 in the cell culture based retrotransposition

assay [Moran et al., 2006; Goodier et al., 2007 and Doucet et al., 2010]. Although ORF1p has shown nucleic acid binding and chaperone activity, the exact function of ORF1p in the process of retrotransposition is completely unknown [Martin 2006 and Martin 2010].

1.3.3. Human L1 ORF2 protein

Human L1-ORF2 encodes a protein with 1267 amino acids in length (MW: ~150kDa) It has three conserved domains: those are i) N-terminal endonuclease domain (EN) (1-239 amino acids) [Feng et al.,1996] ii) central reverse transcriptase domain (RT) (453-880 aa) [Mathias et al.,1991, Malik et al., 1999; Singer and Clements, 1998] and iii) C-terminal CCHC like domain (1096-1275 aa) [Moran et al., 1996; Piskareva et al., 2013]. The L1 ORF2p encoded endonuclease shows similarity to apurinic/apyrimidinic type endonucleases (APEs). I*n vitro* DNA cleavage assay and computational studies revealed that L1ORF2p EN nicks at 5'-TTTT/AA-3' sequence in the bottom strand where the slash ("/") indicates the site of nick [Feng et al.,1996; Mandal et al., 2004; Mandal et al., 2006]. The human L1 RT domain is present downstream of EN domain and have shown sequence similarity to RT domains encoded by other non-LTR retrotransposon, LTR retrotransposon, retroviruses, group II introns and telomerase [Malik et al, 1999]. Biochemical and genetic assay with recombinant ORF2p produced in bacculovirus expression system revealed that full length ORF2p showed RT active on poly rA/oligo dT12 primer template complex [Singer and Clements 1998; Piskareva et al., 2003; Piskareva et al., 2006]. The assay also showed that ORF2p mediated RT activity is metal ion dependent where Mg2+is preferred over Mn2+ [Singer and Clements, 1998]. Additional studies revealed that the human L1 RT exhibited both RNA dependent and DNA dependent polymerase activities, is highly processive (when compared to MMLV RT) and lacks detectable RNaseH activity [Piskareva et al., 2003; Piskareva et al., 2006].

L1 RT activity has also been detected in purified L1-RNP prepared from cells transfected with engineered L1 expression vector [Kulpa and Moran, 2006; Doucet et al., 2010; Mandal et al., 2013]. Investigating types of RNA present in the L1-RNP revealed that apart from retrotransposed RNA (L1, Alu and SVA), the L1-RNPs are highly enriched with cellular mRNA which have PP in the human genome [Mandal et

al., 2013; Pink et al., 2011]. These enriched mRNA are part of L1-RNP as these serve as ORF2p template in a reverse transcriptase assay. Although, ORF2p mediated RT activity and ORF1p were easily detectable in purified L1-RNP, the detection of ORF2p was very difficult. Epitope tagging followed by affinity purification have allowed the detection of ORF2p in purified L1-RNPs from cells transfected with engineered L1 [Goodier et al., 2013; Doucet et al., 2010; Mandal et al., 2013; Taylor et al 2013]. It is believed that ORF2p is a toxic protein for cells because of its intrinsic endonuclease activity. Importantly, the EN mutant ORF2p expressed in significant high amount, have allowed to detect in Coomassie stained denatured SDS-PAGE gel after affinity purification [Mandal et al, 2013].

A third conserved domain of ORF2 is cysteine-rich domain (C-domain) at its C-terminus. The domain has been suggested to function as RNA binding or Zinc knuckle domain [Fanning et al.,1987; Moran et al., 1996; Piskareva et al., 2006]. Preliminary study indicate that recombinant C-domain protein exhibits non-sequence specific RNA binding activity *in vitro* but mutation of the cysteine to another amino acid does not affect the binding property [Piskareva et al., 2006] . Hence, more studies are required to find out the exact function of CCHC domain in L1 retrotransposition.

1.3.4. Human LINE-1 3′-UTR

The human LINE-1 3' UTR is ~206 bp in length and contains a conserved polypurine tract that is predicted to form a G-quadruplex structure [Usdin K and Furano AV 1989]. LINE-1 3' UTRs contain a functional RNA polymerase II polyadenylation signal near their 3' ends. In-vitro studies showed that this LINE-1 poly (A) signal is relatively weak and thus often skipped and utilizes strong poly A signal sequence present in the 3'-flnaking genomic DNA sequences [Moran JV et al.1999]. The use of these 3'- genomic poly(A) signal sequences result in formation of fusion transcript, thus responsible for exon shuffling. Recent cancer genome sequencing showed that such exon shuffling event is very active in many cancers and make cancer genome more complicated. Finally, recent data suggest that the human and mouse 3' UTRs have promoter activity that leads to the generation of alternative LINE-1 transcripts in various tissues [Faulkner GJ et al. 2009].

1.3.5. LINE-1 retrotransposition mechanism

L1 retrotransposition occurs by a process called target primed reverse transcriptase (TPRT). The mechanism was discovered using R2Bm, a site specific non-LTR retrotransposon present in ribosomal DNA locus of silkworm *Bombyxmori* [Luan et al., 1993; Luan and Eickbush, 1995; Cost et al., 2002]. It is hypothesized that following transcription from a chromosomal locus, a full length L1-RNA exported to cytoplasm where both ORFs are translated (Figure 1.5). The newly synthesized ORF1p and ORF2p in the cytoplasm are then interacting with their encoding RNA, a phenomenon known as *cis* preference to form L1-RNPs, the proposed functional intermediate [Wei et a., 2001]. It is also proposed that apart from L1 encoded proteins (ORF1p and ORF2p) some cellular proteins also take part in the formation of L1-RNPs [Pazarro and Cristofari 2016; Goodier 2016]. The L1 RNP then enter the nucleus where a new L1 copy synthesize at the site of insertion via a coupled reverse transcription and integration mechanism termed TPRT [Luan and Eickbush, 1995; Cost et al., 2002].. In this process ORF2p nicks the bottom strand of a target site in a degenerate consensus sequence (5'-TTTT/AA-3') that generate a free 3'-OH that act as a primer for reverse transcription of L1 RNA (Figure 1.5). This results in a new insertion which ends in poly A sequence and flanked by target site duplication (TSD). Several steps in this process are poorly understood which are, import of LINE1-RNP in the nucleus, second strand nick, second strand synthesis and integration of new LINE-1 copy at a new chromosomal location etc. [Richardson et al., 2014].

1.3.6. LINE-1 Retrotransposition Assay

The development of cell culture based retrotransposition assay is one of the important discoveries in the field of retrotransposon. This assay allowed to investigate the activity of L1 elements in real time. In this assay a retrotransposon indicator cassette is inserted into the 3'-UTR of a functional L1 element in opposite orientation of the L1 transcript. The indicator cassette consists with a selectable marker (neo) CMV promoter and SV40 polyA signal sequence. The reporter gene is interrupted by an gamma-globin intron in the opposite orientation with respect to neo marker gene and in the same orientation with respect to L1 transcript. Thus the reporter gene will only expressed upon transcription of the L1, splicing of gamma globin intron followed

reverse transcription of L1mRNA which will insert a new L1 copy along with the reporter cassette. Once integrated, the cells with new L1 insertion are selected or screened for reporter gene expression. Over the years, the neo reporter cassette and a derivative has been used to study the retrotransposition of SVA, Alu, U6SnRNA and cellular mRNA [Richardson Modified version of the retrotransposition cassette has been used to find out the insertion sites which helped to analyze how L1 retrotransposon modulate host genome. [Wei et al., 20001; Moran et al., 1996; Gilbert et al., 2002; Symer et al., 2002].

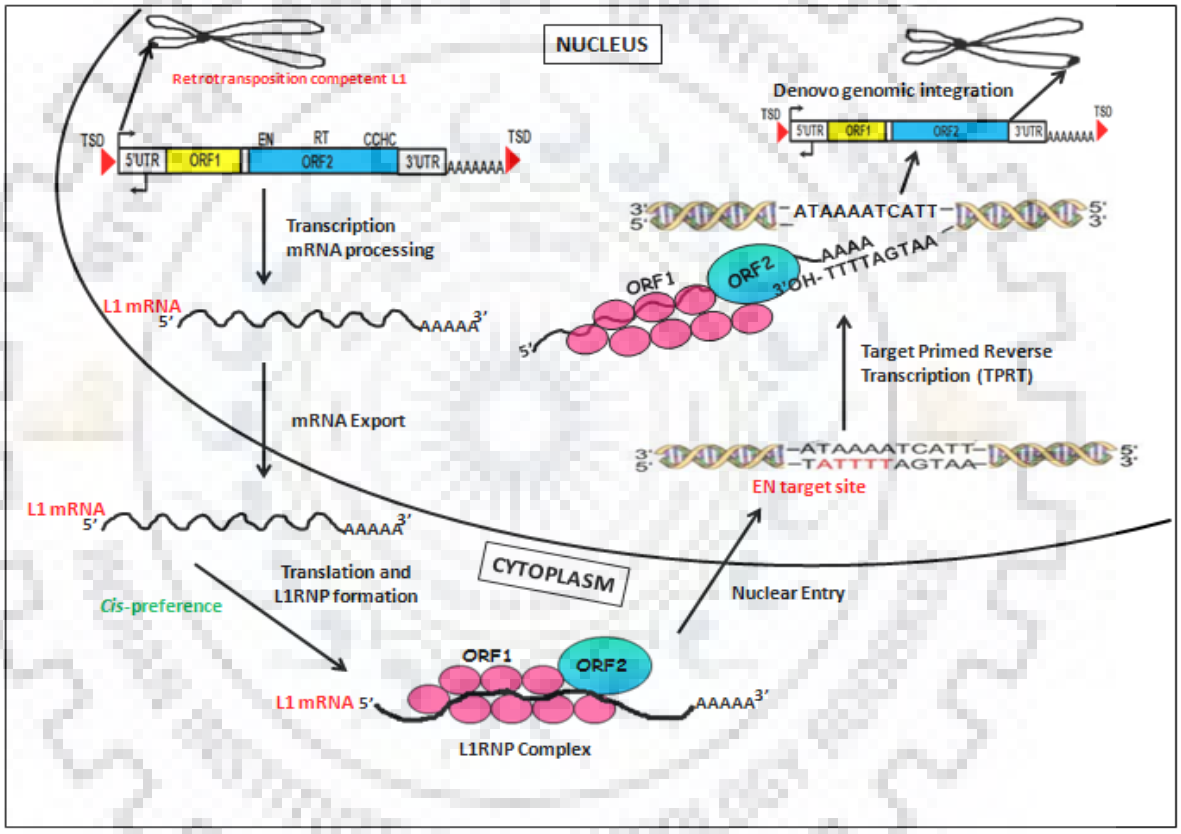


Figure 1.5: Mechanism of L1 retrotransposition- An active L1 present in one chromosomal locus transcribed to mRNA. In cytoplasm L1mRNA undergoes translation of both ORF1 and ORF2 along with some cellular factors bind with L1 RNA and forming L1-RNP complex, the retrotransposition intermediate which enter into nucleus by an Unknown mechanism. The L1RNPs then search the target site where insertion occurs through a mechanism called target-site primed reverse transcription mechanism (TPRT). After export back to the nucleus endonuclease of L1ORF2p cuts at a consensus sequence 5'-TTTT/A-3' (sign "/ "indicates nicking site). The exposed 3'-OH then primed by L1RT resulting in the synthesis of first cDNA using its own mRNA as template. The second strand nick, second strand synthesis and make a new copy to a new chromosomal locus. The new L1 element now formed usually flanked by target site duplication (TSD's) on both sides and a stretch of polyA tail downstream 3'UTR.

1.3.7. Animal models:

To understand the L1 retrotransposition *in vivo,* L1 transgenic mouse was generated where the 3'UTR of an active L1 was tagged with EGFP retrotransposon cassette [Ostertag et al, 2002]. The cassette equipped with an EGFP selectable marker (replacing original neo selection marker), CMV promoter and SV40 poly A signal sequence. The EGFP is disrupted by an intron in opposite transcriptional orientation of EGFP. A successful retrotransposition event will produce green cells. The study showed that L1 elements can retrotranspose in male germ cells [Ostertag et al., 2002 and Athanikar and Moran, 2002]. In another study Park et al showed that L1 elements retrotransposed only in the mature testis of transgenic mice. In 2006, Moutri et al. using another L1 transgenic mouse showed L1 is highly active in certain parts of mice brain like hippocampus, caudate nucleus. All these transgenic mice study showed that L1 retrotransposons is mainly active in two places; in germ tissues and in certain parts of the normal human brain.

1.3.8. L1 Retrotransposition in Brain

In order to ensure its existence, it is believed that LINE-1 retrotransposition must occur in germ cells for its propagation to the next generation [Kazazian and Moran, 2017]. This view was overturned by the unexpected observation that: 1) the differentiation neural stem cells into neurons leads to an increase in LINE-1 transcription and; 2) engineered LINE-1 retrotransposons could retrotranspose in cultured neuronal precursor cells of rat and 3) in the transgenic mice brain [Muotri et al. 2005]. Subsequent studies revealed that engineered L1 can retrotranspose in both fetal and adult human embryonic stem cell-derived neuronal progenitor cells (NPCs) [Coufal et al., 2009]. The sensitive quantitative PCR-based assay detects increased copy number of L1 in human brain tissues compared to matched non-brain tissues (liver or heart samples) [Coufal NG et al., 2009]. The engineered human LINE-1 retrotransposons exhibit enhanced somatic retrotransposition in the mouse models lacking either the methyl-CpG-binding protein 2 (*MeCP2*) or *ATM* genes, [Coufal NG et al., 2009 and Moutri AR et al., 2005]. The use of next generation sequencing also showed the copy number of L1 is higher in certain parts of the brain compare to

matched non brain tissues [Faulkner et al, 2011 and Faulkner, 2104] .Recently our laboratory showed significant L1 ORF1p expression in different anatomic regions of human brain [Sur et al., 2017] compared to non-brain tissues like liver, lung kidney etc. Another interesting observation of the study was increased L1 ORF1p expression in aged brain compare to the younger one. Overall the study showed L1 might have some role in aging process.

1.4. L1 ORF1p antibody:

The first human specific L1 ORF1p (L1HsORF1p) antibody was reported by Leibold et al [Leibold et al., 1990]. Here the ORF1p was expressed as a fusion protein in a bacterial expression vector p3HS1 [Tanese et al., 1985], where L1HsORF1 follows *Escherichia coli* trpE gene in frame. The partially purified 76 kDa trpE-L1Hs ORF1 (trpE- 36.2 kDa, ORF1p-39.8 kDa) fusion protein was used to immunize rabbits. The resulting antiserum was pre-adsorbed using total cell extract made from *E. coli* containing p3HS3. Such pre-adsorption procedure eliminated most of those antibodies which raised to TrpE epitopes or other bacterial proteins that contaminated the antigen preparation. The antibody is very specific and showed single band at around 40 kDa in immunoblotting in total lysate from human cell lines (Ntera2D1 and 2102EP).

The next L1–ORF1p antibody was reported by Goodier et al. [Goodier et al., 2004], where the authors used peptide antigen [20 amino acids stretch (318-338 aa) from ORF1protein]. The antibody was raised in rabbit and affinity-purified. The resulting antibody was not very specific and shows lots of non-specific bands in Western blotting when total cell lysate from human cell lines were used.

Another L1-ORF1 antibody was reported by Harries et al. [Harris et al., 2010]. The authors prepared a GST fused ORF1p, resolved in SDS-PAGE gel, band excised and directly injected into New Zealand rabbits. The anti-sera was purified after passing through CNBr-sepharose 4B coupled ORF1p affinity column. The Western blot performed with lysates from human cell lines known to express endogenous ORF1p showed a single band at around 43 kDa.

Riaz et al. [Riaz et al., 2012] reported a human L1 ORF1p antibody where the author swapped full length ORF1p sequence from plasmid pJM101/L1.3 to pET16B bacterial expression vector. The bacterial expressed His-tagged human ORF1p was

purified using nickel-NTA agarose (Qiagen) followed by removal of His-tagged by factor Xa digestion. The ORF1p without His-tagged was resolved in SDS-PAGE; the band corresponds to 40kDa was excised and purified from gel piece by electrophoretic elution. The purified human ORF1p was then immunized to rabbits to generate polyclonal antibody (termed as Ab#984).

Chen et al. [Chen et al. 2012] used codon optimized synthetic human L1 ORF1p sequence to make Human L1 anti-ORF1p antibody. Here the authors made GST fused ORF1p and then injected the ORF1p in rabbits after removing GST from the fusion protein. The authors selectively purified and enriched antibodies by two steps: first by selective purification of IgG fraction and second ORF1p antigen specific affinity purification. The antibody showed specific detection of ORF1p in cancer samples and human cancer cell lines by immunoblotting, immunocytochemistry and Immunohistochemistry.

Rodic et al. [Rodic et al. 2014] reported a monoclonal anti-human L1 ORF1p antibody where the authors used fifteen amino acid stretch peptide sequence from ORF1p (amino acid position 35-44 from N-terminal of ORF1p and the sequence of the peptide is MENDFDELRE) to immunize BALB/c mice. The sequence used to raise ORF1p antibody is relatively diverged between human and mice LINE-1 ORF1p protein. The antibody is very specific to detect L1 ORF1p by immunoblotting, immunofluorescence and Immunohistochemistry in human cell lines and cancer samples.

Recently our laboratory reported ORF1p antibody [Sur et al. 2017] in rabbit using RRM domain of human ORF1p. The RRM domain was cloned in bacterial pET expression vector; the expressed protein was purified by nickel-NTA agarose chromatography and then injected to rabbits. The antibody showed specific detection of ORF1p in human cell lines and normal human brain tissues by employing immunoblotting, immunofluorescence, Immunohistochemistry and Immunocytochemistry.

1.5. L1 ORF2p antibody:

The first human specific L1 ORF2p (L1HsORF1p) antibody was reported by Cost et al. [Cost et al. 2002]. The 150 kDa ORF2p contains three partially characterized domains;

N- terminus endonuclease (EN), central reverse transcriptase (RT) and C-ter cysteine rich domain of unknown function (CCHC). The authors reported that they have used a mouse antibody directed against the L1 EN domain. However, no detailed description is available in the article (Materials and Methods). Also the authors did not site this antibody in their other publications.

The second human specific L1 ORF2p antibody was reported by Goodier et al. [Goodier et al. 2004]. Two ORF2 peptide sequences, one from N-terminal (aa 154-167 anti-ORF2-N) and other from C-terminal (aa 1259-1275) of ORF2 protein were used to raise polyclonal anti-ORF2 antibodies. The antibodies are not specific and showed many non-specific bands in western blotting.

Ergun et al. [Ergun et al. 2004] raised the antibody against the EN domain of human L1 ORF2 using two peptides (aa 48-63 and 152-166); Numbering corresponds to the amino acid positions in the active L1.2 element). The peptides were chosen based on algorithms that predict antigenic site (Lasergene Software, DNASTAR). The antibody was raised in chicken and purified by affinity chromatography from IgGY fraction. The antibody was referred as anti-EN antibody. The antibody is not specific and showed 70kDa band instead of 150 kDa (MW of L1ORF2p) in western blotting. The authors commercialized the antibody and available in Abcam, USA [anti-LINE-1 ORF2p (ab106004)].

Chen et al. [Chen et al. 2012] used codon optimized synthetic human L1 ORF2p sequence to make Human L1 anti-ORF2p antibody. Here the authors made recombinant 6XHis tagged ORF2p in bacteria, remove the His- tag sequence by thrombin digestion and then injected the ORF2p in rabbit to generate antibody. The authors selectively purified and enriched antibodies by two steps: first by selective purification of IgG fraction and second ORF2p antigen specific affinity purification. The antibody showed specific detection of ORF2p in cancer samples and human cancer cell lines by immunoblotting, immunocytochemistry and Immunohistochemistry. Here, the authors raised antibody against codon optimized human ORF2 sequence and only one literature is available where the antibody is reported.

Sokolowski et al. [Sokolowski et al. 2014] used bacterial expressed His-tagged EN domain and generate monoclonal anti-ORF2 antibody in BALB/C mice. The antibody detects many bands apart from the band of interest (150 kDa) and thus not specific to

human L1ORF2p.They have commercialized the antibody [IG4E11] and it is available with Kerafast, USA [Anti-LINE1 retrotransposable element ORF2 protein (1G4E11)antibody].

Luca et al. [Luca et al. 2016] reported a monoclonal anti human ORF2p antibody where authors tried six separate peptides to generate antibody in BALB/C mice. The author used vector NTI software (Life Technologies) to identify regions with the lowest homology (below 40%) between human and mouse L1 sequences. Only one peptide (#39 aa 119-138: TGARPFIKQVLSDLQRDLDS) showed specific antibody response. The monoclonal antibody referred to as chA1-L1 (isotype IgG2a). They have patented the antibody with patent number US 10214591 and title "Monoclonal antibody to human line-1 ORF2 protein and method for early detection of transforming cells in pre-neoplastic tissue of a human subject" in the year 2016.

1.6. Brief overview on cancer:

Cancer is defined as a abnormally growth of the body cells and occasional spreading of the abnormal cells into neighbouring tissues. It is a deadly disease and for most of the cancers the survival rate is very less. The statistics suggests one in seven deaths is due to the cause of cancer. [International Agency for Research on Cancer. World cancer report 2014]. The way the numbers of cancer patients are increasing world-wide day by day is alarming. As per 2016 WHO's report globally more than 14 million people are affecting due to cancer and more than 8 million death occurs every year (which is about 22,000 cancer deaths per day; ~ 2.9 million in economically developed countries and ~5.3 million economically developing countries). By 2030 it is estimated that the global burden of cancer patients is expected to grow over 22 million.

. Today's society has already realized the threat posed by numerous types of cancer. There are more than 100 types of cancer identified which are named as per the organs or tissues where the cancers form. Among man the five most common site of cancer diagnosed in lung, prostate, colorectal, stomach and liver. Among women those five sites are breast, colorectal, lung, cervix and stomach. All those sites are represented the common cause of cancer death in both men and women. Among men, the lung cancer has highest incident and mortality rate (34.2 and 30.0 per 100000 respectively) than any other cancer. Whereas among women it is breast

cancer which has considerably at higher incident rate (43.3 per 100 000) [International Agency for Research on Cancer. [World cancer report 2014].

1.6.1. Head and Neck Squamous Cell Carcinoma (HNSCC):

Head and neck cancers also called head neck squamous cell carcinoma (HNSCC) begins in the squamous cells that lines the mucosal surfaces inside the head and neck (the throat, the nose and inside the mouth) [Gupta et al., 2016]. The areas which are affected by head and neck cancer are the oral cavity, the pharynx, the larynx, the paranasal sinuses, nasal cavity and sometimes the salivary glands (Figure 1.7). The oral cavity includes lips, tongue, and the gum, the lining inside the cheeks and lips, the floor of the mouth under the tongue, the head palate and the gum behind the wisdom teeth. Cancer inside the squamous cell lining of oral cavity named as oral squamous cell carcinoma (OSCC) represents more than 90% of all forms of head and neck cancer [Agrawal et al., 2011;Stransky et al., 2011]. HNSCC is the sixth most common cancer occurring worldwide affecting more than half a million individuals. HNSCC are frequently very lethal with a five year survival of only ~50%. HNSCC and its treatment mostly resulted in cosmetic deformity and functional impairment of vital functions, including swallowing, hearing, breathing, speech and smell [Agrawal et al., 2011; Stransky et al., 2011].

1.6.2. Genes altered in OSCC:

Similar to all other types of cancers, OSCC is thought to be initiated and progress by a series of genetic alterations [Agrawal et al., 2011; Stranskyet al., 2011; India Project Team of the ICGC, 2013; Majumder 2014]. In addition to point mutations in genes like TP53, DNA rearrangements along with insertions and deletions are known to facilitate transformation. For instance, insertions generated by L1 retrotransposons have inactivated tumor suppressors such as APC [Miki et al., 1992]. Sequencing of the genome revealed some genes such as USP9X, UNC13C, ARID2, TRPM3, MLL4 are specifically altered in OSCC revealed in the report by ICGC project. USP9X is a tumour suppressor coding gene encoding a deubiquitinating enzyme [Agrawal et al., 2012; Pe´rez-Mancera et al., 2012] UNC13C and TRPM3 are related to the release of neurotransmitters [Ariel et al., 2012; Zamudio-Bulcock et al., 2011] and ARID2, MLL4

genes are related to chromatin remodelling [Biankin et al., 2012; Shain et al., 2013; Cho et al., 2007]. Also some genes that are altered in HNSCC like CASP8, TP53, FAT1, NOTCH1. Patients consuming tobacco either by smoking or directly chewing were found to have high C>G transversions mutations [Agrawal et al., 2011; Stranskyet al., 2011; India Project Team of the ICGC, 2013].

1.7. LINE-1 and cancer

Insertion of LINE1 into a gene can disrupt and hence can make it non-functional. Miki *et al.,* in 1992, firstly reported the insertion of a 750bp LINE1 fragment in adenomatous polyposis coli (APC) gene which resulted in colon cancer. Studies have shown that L1 insertions are mostly occurs in epithelial cancers like colorectal, prostate and ovarian rather than in blood or brain cancers [Lee et al., 2012]. There are more than 100 L1 insertions have been identified in one colorectal cancer patient [Lee et al., 2012]. Solyom et al. [2013] screened 16 colorectal cancer patients and identified 69 tumor specific L1 insertions which was absent in normal counterpart. Other cancers where L1 retrotransposition have been studied are oesophageal carcinomas [Doucet-O'Hare et al., 2015] pancreatic ductal adeno carcinoma [Rodic et al., 2015], hepatocellular carcinomas [Shukla et al., 2013], head and neck cancers [Helman et al., 2014], Prostate [Lee et al., 2012] and ovarian cancers [Tang et al., 2017; Lee et al. 2012].

1.8. Objectives of the Study:

Human L1 ORF2p is the key protein in the process of retrotransposition and is extremely difficult to express *in-vitro*. The expression of this protein is used as a marker to find the active L1 retrotransposition in cells and tissues. Although few antibody against L1-ORF2p is available with some L1 researcher, their quality is not good; many non-specific bands found in immunoblotting.

Prior it was believed that LINE-1 retrotransposons are only active in germ cells (sperm and ovum) and early stage of development. But recent transgenic animal models and high throughput sequencing analysis revealed that L1 is also active in certain parts of normal brain and in few cancers. The studies also showed that cancer of epithelial origin showed more L1 insertion compared to other types of cancer. Although it is known that L1 is highly active in certain cancers, its role towards the development or progression of cancer is completely unknown. Oral cancer a subtype

of head and neck is very deadly and highly prevalent in India due to excessive use of tobacco. Very limited study has been performed to see the activity of L1 retrotransposons in oral cancer samples. In this study I have tried to synthesize an human L1 ORF2p (hL1ORF2p) antibody. I have also explored hL1ORF2p expression in oral squamous cell carcinoma (OSCC) samples obtained from Indian patients.

The main objective was answered by performing following sub-objectives.

I) To clone and express fragments of hL1-ORF2p for antibody generation

II) To generate an in house antibody specific to hL1-ORF2p

III) Detection of hL1-ORF2p expression in OSCC samples.

Chapter 2

Materials and Methods

2.1. Material used and its Source:

All *E. coli* bacterial strains used in this study for recombinant DNA work was available in the laboratory. The plasmids used for cloning and expression studies were also available in the laboratory. Reagents used were of molecular grade procured from Roche Biochemicals (Germany), Amersham Pharmacia (USA), Promega (USA), Sigma (USA), SRL(India) and Himedia (India). Enzymes used for restriction digestion in cloning purpose were acquired from New England Biolabs (NEB, UK), TAKARA and Promega (USA). Double distilled water used for preparing buffers and solutions were unless stated otherwise. Autoclaving was done at a pressure of 15 lbs per square inch for 20 min.

2.2. Collection of tissue specimens:

All the paired normal cancer tissues were collected post-operative following proper consent from the patient and their immediate family member from the Acharya Tulsi Regional Cancer Treatment and Research Institute, Bikaner, Rajasthan, India. Following initial collection, samples were stored in RNA later solution (Qiagen) at -20°C. Subsequently, these tissues were used for genomic DNA and protein isolation along with the preparation of formalin fixed paraffin embedded blocks. All investigations were conducted in accordance with ethical principles embodied in the declaration of tissue request and material transfer agreement (IHEC No. BT/IHEC-IITR/2017/6673; Institute Human Ethics Committee (IHEC), Indian Institute of Technology Roorkee, Uttrakhand, India).

2.3. Organisms and growth conditions:

Cells from frozen glycerol stocks were first streaked on an LB plate (containing the appropriate antibiotic wherever necessary) and allowed to grow overnight at 37°C. Liquid cultures in LB medium were initiated from a single colony and were grown with constant shaking at 200rpm at 37°C. The cells were grown overnight, were used for further growth by diluting 100 fold in fresh LB medium and grown with proper aeration at 37°C for 3-4 hrs to obtain log phase cultures.

2.4. Culture media:

2.4.1 Luria Broth (LB):

Composition of per litre

Bactotrypton	10gm
Yeast extract	5gm
NaCl	10gm

The components were dissolved in double distilled water and pH adjusted to 7.0 using 1N NaOH. The medium was sterilised by autoclaving.

2.4.2. SOB medium:

Composition of per litre

Bactotrypton	20gm
Yeast extract	5gm
NaCl	5gm
1M KCl	2.5ml

The components were dissolved in double distilled water and pH adjusted to 7.0 using 1N NaOH. The medium was sterilized by autoclaving.

2.4.3. LB Agar:

LB agar was prepared by adding 1.5% (w/v) of Bacto-agar to LB medium and sterilised by autoclaving. Ampicillin was added to a final concentration of 100µg/ml, kanamycin was added to a final concentration of 50µg/ml and Chloramphenicol to a final concentration of 20µg/ml added after cooling the LB agar to around 55°C and plates poured.

2.5. Preparation of solutions:

2.5.1. Acid salt Buffer:

Composition of per litre

$CaCl_2$ (M.W. 146.9)	14.69gm
$MnCl_2$ (M.W. 197.9)	13.85gm
CH_3COONa (M.W. 136.08)	5.44gm

The components were dissolved in double distilled water and pH adjusted to 7.0 using 1N NaOH. The medium was sterilised by autoclaving

2.5.2. INOUE solution:

Composition of per litre

$CaCl_2$ (M.W. 146.9)	2.20gm
$MnCl_2$ (M.W. 197.9)	10.9gm
KCl (M.W.74.55)	18.7gm
0.5M PIPES solution	20ml

0.5M PIPES solution (10ml): Add 1.51gm PIPES buffer in 8ml water. Adjust pH with KOH and HCl to 6.7 and make volume up to 10ml.

2.5.3. $Cacl_2$ Buffer

$CaCl_2$ (M.W. 146.9)	100mM
$MgCl_2$ (M.W. 197.9)	100mM
$CaCl_2$ (M.W. 146.9)	85Mm
Glycerol (100%)	15% v/v

2.5.4. Lysis buffer A: Composition (in molarity)

Tris HCl (M.W. 154.24)	50mM
NaCl	150mM
EDTA	1mM
PMSF (M.W.)	1mM
Imidazole (M.W.)	25Mm
βME	0.05%

2.6. Plasmid DNA isolation:

2.6.1. Mini prep DNA isolation (Alkaline lysis method)

A single bacterial colony containing the plasmid was inoculated in 10ml of LB broth medium containing appropriate antibiotic and grown for overnight at 37°C with shaking at 200 rpm. The overnight culture was transferred to 1.5ml microcentrifuge tube and centrifuged at maximum speed for 30 seconds at 4°C. The medium was completely drained off and the pellet was resuspended in 200 µl of ice cold alkaline lysis solution I

[50mM Glucose, 25mM Tris-Cl (pH-8.0)] and 10mM EDTA (pH-8.0) by vigorous vortexing. After that, 200 µl of freshly prepared alkaline lysis solution II (0.2N NaOH and 1% SDS) was added to each bacterial suspension. The tubes were closed and mixed the contents by inverting the tubes rapidly (five times). Then 200 µl alkaline lysis solution III (3mM potassium acetate pH- 5.5) was added and dispersed through the viscous bacterial lysate by inverting the tubes several times. The tubes were stored on ice for 10 minutes and centrifuged at maximum speed for 10 minutes at 4°C. The supernatant was transferred to a fresh centrifuge tubes. The DNA was precipitated from the supernatant by adding 0.6 volume of isopropanol and mixed by inverting. The precipitated nucleic acid was collected by centrifugation at maximum speed for 10 minutes at room temperature. The pellet so obtained was washed with 500 µl of 70% (v/v) ethanol, air dried and dissolved in 20 µl of $T_{10}E_1$ (10mM Tris. Cl, pH 8.0, 1mM EDTA, pH 8.0) containing 0.1mg/ml RNAseA. The tube was incubated at 37°C for 30 minutes for RNAseA digestion and 2 µl was loaded on a 0.8% agarose gel to quantify the amount of DNA.

2.6.2. Midi prep DNA isolation (Alkaline lysis method)

Bacterial cells were grown overnight in 50ml of LB broth medium containing appropriate antibiotic and grown for overnight at 37°C with shaking at 200 rpm. The overnight culture was pelleted by centrifugation in 50ml falcon. The cell pellet was resuspended in resuspension buffer [50mM Glucose, 25mM Tris-Cl (pH-8.0)], 10mM EDTA (pH-8.0) , containing 2mg/ml lysozyme and kept on ice for 10 min. To the lysed cells add 3ml of lysis solution (Freshly prepared solution containing 0.2N NaOH and 1% SDS) was added and mixed by inverting. To the tube was added 1.6ml of 3M sodium acetate (pH-4.6). RNA was removed by incubating the supernatant with 12-15 µl of RNAseA (10mg/ml) at 37°C for 45 min. The supernatant was extracted twice with phenol: chloroform: isoamyl alcohol [25:24:1(v/v/v)] and once with chloroform: isoamyl alcohol [24:1(v/v)]. The upper aqueous layer was transferred to an Oakridge tube and the DNA was precipitated by addition of 2.5 volumes of chilled ethanol and left for overnight incubation at -20°C or 45 min at -80°C. The DNA was precipitated by centrifugation at 12,0000rpm for 20 minutes at 4°C. The pellet was resuspended in 0.4ml of nuclease free water and to this was added 120 µl of 4M NaCl and 0.5ml 13% PEG 8000. The tube was incubated in ice for 1hr and centrifuged at RT at 12, 0000

rpm for 15 minutes. The pellet was washed with 200 µl of 70% ethanol, dried in a 37°C incubator and resuspended in 300-500 µl $T_{10}E_1$ solution. .

2.7. Transformation of *E.coli* cells:

2.7.1. Preparation of competent cells by The Inoue Method [Sambrook. J, 2006]

A fresh single bacterial colony of *E.coli* (strain DH5α) was inoculated in 50ml LB medium in a 250ml conical flask, grown for overnight at 37°C with shaking at 200 rpm. One percent of overnight culture was added to 50ml SOB medium and grown at 30°C/220rpm till the O.D reach around at 0.4. Next, the culture was removed from shaker and incubated in ice for 20 minutes with intermittent shaking. After that, cells were pellet down by centrifuging at 3000rpm/10min/4°C. Next the pellet was re-suspended 20ml ice cold Inoue solution by gentle shaking in ice followed by incubation for another 20 minutes on ice. After that, cells were pellet down by centrifuging at 3000rpm/10min/4°C. The pellet was dissolved gently by adding 4ml ice cold Inoue solution + 375 µl 100 % DMSO. Immediately 200 µl aliquots were snap freezed by immersing in liquid nitrogen and stored at -80°C until use.

2.7.2. Transformation of competent cells E.coli cells

Competent cells were thawed on ice and to 100 µl cells, 5-10ng of plasmid DNA was added. The cells were incubated on ice for 45 min. cells were then given a heat shock at 42°C for 90 seconds and incubated on ice for 2 min. 0.8 ml LB was added to the cells and the cells were grown at 37°C for 1hr at 200rpm. Transformants were plated on LB agar plates with appropriate antibiotic and incubated at 37°C for 14-16hr.

2.8. Restriction enzyme digestion:

Restriction enzyme digestions were carried out in usually in 30-50 µl volume. Appropriate amount of DNA was digested in a reaction mixture containing enzyme buffer (as per manufacturer's instructions) and 5-10 units of enzyme at the recommended temperature for 4-16 hours. After incubation the reaction mixtures were loaded with 1X gel loading buffer (GLB) (6.5% sucrose w/v, 1-mM Tris-HCl pH-7.5, 1mM EDTA and 0.03% Bromophenol Blue) into agarose gel and run for appropriate time.

2.8.1. Insert isolation:

The agarose slice was trimmed as much as possible and minced with a sterile blade. The gel slice was then transferred to a 1.5 ml tube and used the Qiagen gel extraction kit for isolating DNA as per manufacturer instruction. (Detailed procedure given in section **2.18.3**)

2.9. Antarctic Phosphatase (AnP) treatment:

The digested vector DNA (0.5µg/µl) was incubated with Antarctic phosphatase and the provided AnP buffer at 37°C for 15 minutes. The enzyme was inactivated by putting it at 80°C for 5 min and the DNA was purified by using QIAquick gel extraction kit as per instruction manual.

2.10. Polymerase chain reaction:

All the primers used for PCR were purchased from GCC biotech (Kolkata, India) and Eurofins (Bangalore) are listed in (**Table 1; Appendix)** PCR amplification was carried out either in 20 or 50 µl reaction volumes with the desired number of cycles. The reaction mixture was prepared as per following table. The PCR was performed in the thermo cycler machine (Applied biosystem veriti 96 well thermal cycler) using the following program. One cycle at 94°C for 30 sec followed by 30 cycles at 94°C for 20 sec, 52°C -62°C for 20 sec and 72°C for 30sec to 1 minute and finally one cycle at 72°C for 2 minutes. The product was checked by Agarose gel electrophoresis.

Component	20 µl Reaction	50 µl Reaction	Final Concentration
Nuclease-free water	to 20 µl	to 50 µl	
5X Reaction buffer	4 µl	10 µl	1X
10 mM dNTPs	0.4 µl	1 µl	200 µM
10 µM Forward Primer	1 µl	2.5 µl	0.5 µM
10 µM Reverse Primer	1 µl	2.5 µl	0.5 µM
Template DNA	variable	variable	< 100 ng
Enzyme	0.2 µl	0.5 µl	1.0 units/50 µl PCR reaction

2.11. Ligation of vector and insert:

The vector and insert DNA was mixed with 1X T4 DNA Ligase Buffer (NEB) and 0.5µl of 2000U/µl T4 DNA Ligase (NEB) in a total volume of 15µl. the mix was incubated at 16°C for 16 hrs and transformed into *E. Coli* competent cells. When the vector and insert DNA are almost similar in size a molar ratio of 1:3 (vector: insert DNA) was used. When vector and insert DNA sizes are not similar a molar ration of 1:1 or 1:2 was taken.

2.12. Agarose gel electrophoresis:

The agarose concentration for DNA gel electrophoresis was determined based on the size of the DNA to be resolved. Agarose was melted in 0.5X or 1X TBE (45mM Tris-borate and 1mM EDTA, pH 8.0) by heating and was cooled to about 50°C before adding 0.5 µg/ml of ethidium bromide. The molten agarose was poured in a tray and allowed to set, After the gel had set, DNA samples were loaded and electrophoresed in 0.5X to 1X TBE in appropriate electric field strength for optimum separation. The DNA was visualized by UV transilluminator at 302 nm wavelength.

2.13. SDS-PAGE Analysis of proteins

SDS-PAGE was carried out in 30 % acrylamide (acrylamide: bisacrylamide 29:1) gels, under reducing conditions according to Laemmli, 1970. After electrophoresis, proteins were detected by staining with Coomassie Brilliant Blue (0.25% CBB R-250 in 45% methanol and 10% acetic acid) solution. The gels were destained by soaking it in the methanol/ acetic acid solution (45% methanol and 10% acetic acid).

2.14. Protein Estimation:

2.14.1. Protein estimation by Bradford's reagent

Protein concentration is estimated using Bio-Rad Bradford's protein assay reagent. The concentrated reagent diluted to 1x prior to use. Protein concentrations of 20µg to 1mg/µl BSA were made for standard curve. Then the test protein dissolved in 1:10 diluted reagent, kept in dark for 5minutes then solutions were measured at 595nm using spectrophotometer.

2.14.2. Protein estimation by BCA assay

The amount of protein in a sample was estimated by the bicinchonninic acid assay using BSA as the standard. The working solution was prepared by mixing bicinchonninic acid (Sigma) and 4% copper sulphate in a ratio of 50:1. 10µl of protein sample was mixed with 200µl of the working solution in a microtitre plate and incubated at 37°C for 30 min. The absorbance was taken at 560nm. BSA of known concentration was used as standard.

2.15. Cloning, expression and purification of human L1CCHC protein:

CCHC fragment (length 536 bp) was cloned in EcoR1-Xho1 restriction sites of vector pET30b. 1F/1R primers (Table1, appendix) were designed to amplify the specific fragment from PBS-L1RP clone available in the laboratory. Clone was transformed into BL21 strain of *E.coli* to check the expression. A single colony of *E. coli* cells containing ORF1 clone inoculated in 10ml of fresh LB medium containing appropriate amount of Kanamycin antibiotic and incubated at 37°C/220 rpm for overnight to get primary culture. Next day, 50ml of LB medium was inoculated with 1% primary culture with appropriate kanamycin antibiotic and grown at 37°C/220 rpm till A_{600} of 0.4.The culture was induced with 0.4mM IPTG (isopropyl- β-D-1-thiogalacto pyranoside) and cells were further incubated for 4 hrs.The cells were then harvested by centrifugation and then resuspended in 5 ml of lysis buffer A. Cells were lysed by sonication on ice and the lysate was centrifuged at 75,000 rpm for 15 minutes at 4°C. The supernatant was incubated with 200µl of pre-equilibrated Ni-NTA agarose beads (Qiagen) for 1 hr at 4°C with gentle mixing. The recombinant protein was eluted with an Imidazole gradient (50-500 mM).

2.16. CCHC Protein purification using Gel Filtration Chromatography:

The protein samples purified from Ni-NTA method were further purified using ÄKTAprime plus size exclusion chromatography with superdex 200pg column (GE Healthcare). This technique separates the macromolecules based on their size and sometimes molecular weight. This technique also helps in estimating molecular weight of unknown proteins. Initially, the column is equilibrated with one column volume of 50mM Tris pH 8.0, 150mM NaCl buffer. After that, protein (1% column volume) was injected into column and allowed to pass through the column with one column volume

of buffer. Protein eluted during this process was collected in fraction tubes. The various fractions were resolved in denaturing PAGE. The fractions containing EGFP or mutant were pooled together and concentrated using 3kDa cut off centricon. This highly pure protein is then used for various biophysical techniques which help in characterizing the protein.

2.17. Cloning, expression and purification of human L1 ORF2p EN domain protein:

EN1 fragment (length 720 bp) and EN2 fragment (length 456 bp) was cloned in Sac1-Xho1 and BamH1 and Xho1 restriction sites of vector pET28c. Primers (2F, 2R, 3F and 3R; Table 1, appendix) were designed to amplify the specific fragment from PBS-L1RP. Clones were transformed into BL21 strain of *E.coli* to check the expression. A single colony of E. coli cells containing pET-hL1EN1 and pET-hL1EN2 was taken from the plate and inoculated in 10ml of fresh LB medium containing desired amount of kanamycin antibiotic and was grown at 37°C for overnight as primary culture. Next day, 0.5ml (1%) culture was taken from the primary culture and inoculated in 50ml of LB medium containing kanamycin. The culture flask was incubated at 37°C/220 rpm till the culture reach A_{600} of 0.4. Next, culture was induced by adding IPTG (final concentration 0.4 mM) followed by growing for 4 hours at 37°C/220 rpm. After analyzing on 12 % SDSPAGE the expressed protein was relatively low, the expression was checked at 37ºC for 3hrs, 30°C for 6hrs, 25ºC for 10hrs and 16°C for 14hrs respectively. The total protein extracts were collected and suspended in lysis buffer and further subjected to sonication for cell disruption. The obtained soluble and insoluble fractions were loaded and analyzed separately using SDS-PAGE. The soluble fraction from 30ºC grown culture was taken and performed affinity chromatography using Ni-NTA beads. The fractions of purified protein along with wash and flow through were also analyzed Along with the protein was also extracted from fraction insoluble fraction and affinity purified and performed dialysis to refold the protein.

2.18. Cloning of Enhanced Green Fluorescence Protein (EGFP) gene into expression vector:

Gene coding for EGFP was taken from commercially available mammalian vector pEGFP-N1. This gene was sub cloned into bacterial expression vector pET 30b (+)

under T7 promoter which is regulated by Lac operon.

2.18.1. pEGFP-N1 and pET 30b (+) plasmid preparation:

pEGFP-N1 and pET 30b (+) plasmids were transformed into *E.coli* DH5α competent cells according to the protocol mentioned in the book: Molecular Cloning- A laboratory Manual 3rd Edition by J Sambrook and D.W. Russell. Single colony were picked and inoculated in 50mL LB Broth along with Kanamycin (30µg/mL) and incubated at 37°C, 250 rpm overnight. Plasmid DNA was isolated using alkaline lysis method with SDS: Midi preparation described in the book: Molecular Cloning- A laboratory Manual 3rd. The plasmid DNA was analyzed on 1% Agarose gel resolved at 1V/cm^2 until gel loading dye front has reached 80% of the gel length. The resolved agarose gel is visualized under UV transilluminator.

2.18.2. Double digestion of EGFP insert and pET30b vector:

2ug of both pEGFP-N1 and pET30b were digested with restriction enzymes BamHI and NotI (New England Biolabs) with 3.1 buffer. The reaction mixture was mixed by tapping gently and centrifuged briefly and incubated at 37°C in circulating water bath for 3 hours.

2.18.3. Gel purification of Insert and Vector:

After the digestion is completed, 4uL of 6X Gel loading dyes is added to 20uL digested sample and resolved in 1% agarose gel at 1V/cm^2. The gel was resolved until the dye has reached 80% of the gel length. DNA band of interest were identified and respective gel piece is cut by visualizing the gel under UV transilluminator. DNA from this gel was extracted by using commercially available kit: Purelink® Quick Gel extraction kit supplied by Invitrogen. The protocol is mentioned in brief. The gel slice was weighed and gel solubilization buffer (L3) was added in 1:3 ratio. The tube containing gel slice and buffer is kept into 50°C water bath for 10-15 minutes by mixing intermittently. Make sure that the gel is completely dissolved. This mixture was loaded onto Quick gel extraction column and centrifuge at 12000xg for 1 minute. Flow through was discarded and empty column was centrifuged with wash buffer containing ethanol was added to the column in order to remove residual ethanol. Now the column was placed in recovery tube and 40uL of elution Gel extraction column and centrifuged at 12000xg for 1 minute. 500uL wash buffer and left for 1 minute at room temperature

and centrifuged at 12000xg for 1 minute. The purified DNA was checked by loading onto 1% agarose gel and resolved at 1V/cm^2 and visualized under UV transilluminator.

2.18.4. Ligation of insert and vector and Transformation of ligated product into *E.coli* DH5α competent cells:

Ligation reaction was setup as according to protocol mentioned in 2.11. The reaction mixture was mixed by tapping gently and centrifuged briefly and incubated in 16ºC water bath for 16 hours. 2uL of this ligated product was transformed into *E.coli* DH5α competent cells using the protocol described in the book: Molecular Cloning- A laboratory Manual 3rd Edition by J Sambrook and D.W. Russell. The protocol is mentioned briefly. 100uL of *E.coli* DH5α competent cells were thawed on ice and to these cells; 2uL of ligated product was added and mixed by tapping gently. The cells with DNA were placed on ice for 45 minutes without disturbing. Heat shock was given to cells by placing the tube in 42ºC water bath for 45 seconds and immediately transferred onto ice and incubated for 5 minutes. Now, 700uL of LB broth was added and the cells were incubated at 37ºC for 1 hour at 250 rpm. After the incubation, the cells were centrifuged at 3000rpm for 5 minutes. The cell pellet was resuspended in 100uL of medium and everything was plated on LB agar plate containing 60µg/mL kanamycin and incubated at 37ºC overnight.

2.18.5. Confirmation of clones by retardation and restriction digestion:

From the single colonies obtained after transformation, 5 single isolated colonies were picked and inoculated in 10mL LB broth medium containing 30µg/mL kanamycin, incubated at 37ºC 250rpm overnight. From these overnight grown cultures, plasmid DNA was isolated using alkaline lysis method which is mentioned earlier. Plasmid DNA from all the 5 colonies along with control was resolved in 1% agarose gel at 1V/cm^2 until the dye has reached 80% of the gel length. Further the plasmid DNA was digested with restriction enzymes to confirm the clones. The reaction mixture was mixed by tapping gently, centrifuged briefly and incubated in 37º C water bath for 3 hours. After this, the samples were resolved in 1% agarose gel at 1V/cm^2 until dye has reached 80% of the gel length.

2.18.6. EGFP expression and optimization:

The confirmed positive clone for EGFP was transformed in to BL-21 and over

expressed using IPTG inducible T7 promoter in pET30b expression vector using *E.coli* BL21 as the expression host. Single colonies were picked after transformation in order to check expression. Primary culture was inoculated from single colonies in 10mL LB broth medium containing 30µg/mL Kanamycin and incubated at 37ºC, 250 rpm overnight. Secondary culture was inoculated by taking 1% inoculum from primary culture and incubated at 37ºC, 250 rpm until OD of the bacterial culture reached 0.4. Secondary culture was induced with 0.5mM IPTG and cells were incubated for 3 hours at 37°C and 250 rpm. For the confirmation of expression, 1ml cell culture was taken in a 1.5ml microcentrifuge tube (Tarsons, India) and pelleted down at 13000 rpm for 2 minutes. For protein expression analysis, sample was prepared by Boiling-Prep method. The expression of EGFP and mutant protein was analyzed on 12% SDS-PAGE (Bio-RAD mini gel assembly).Further expression of EGFP was optimized by varying incubation temperature, IPTG concentration and time period after induction so that both the proteins are in soluble fraction. Finally, the conditions that were obtained after optimization are 20ºC, 0.5mM IPTG induction for 24 hours.

2.18.7 Purification of EGFP and mutant using Ni-NTA affinity chromatography:

EGFP was purified using Nickel Ni-NTA affinity chromatography in batch method at 4ºC. The protocol followed is mentioned in brief. The bacterial cell pellet from 100mL culture after 24 hours induction with 0.5mM IPTG was resuspended in 5 mL lysis buffer (50mM Tris pH 8.0, 150mM NaCl, 1mM EDTA, 1mM PMSF 25mM Imidazole and 0.05% β-mercaptoethanol). Cells were lysed by ultra sonication (5 sec pulse on and 20 sec pulse off for 5 minutes). Lysed cell mixture was centrifuged at 12000rpm for 30 minutes at 4ºC in order to separate soluble fraction from cell debris and insoluble fraction. 1mL Ni-NTA beads (Qiagen) were taken in five 1.5mL micro centrifuge tubes. Beads were washed with wash buffer (50mM Tris pH 8.0, 150mM NaCl, 50mM Imidazole) for 5 times by centrifuging at 4000 rpm for 1 minute every time. Supernatant (soluble fraction) after centrifugation of lysed sample was added to the beads and incubated at 4ºC for 1 hour under constant shaking. Samples along with beads were centrifuged at 4000 rpm for 1 minute and supernatant (flowthrough) was collected and stored. The beads were washed again with wash buffer for 5 times as mentioned earlier. To each tube 500uL of elution buffer (50mM Tris pH 8.0, 150mM NaCl, 500mM Imidazole, and 1mM PMSF) was added and incubated for 30 minutes. Tubes were centrifuged and supernatant was collected (contains purified protein) and

beads were stored in elution buffer at 4°C for further use. Affinity purified EGFP was further purified using gel filtration chromatography using ÄKTAprime plus size exclusion chromatography.

2.19. Immunization protocol:

For immunization, 4-week age Swiss albino mice and Sprague Dawley rat were procured from NIPER (National Institute of Pharmaceutical Education and Research), CAF (Central Animal Facility); Mohali. The animals kept under quarantine for one week in animal facility, Dept. of Biotechnology, IIT Roorkee as per institute animal use guidelines. After passing the quarantine the animals were now can be immunized. The immunization procedure was first standardised by raising antibody against EGFP. Now the fusion protein (pET30b-EGFP-EN) injected in to mice using the standardized protocol tabulated below. Three booster doses of 50µg antigen per mice and 200µg per rat were given followed by primary immunization along with Incomplete Freund's Adjuvant.

PROCEDURE	PROTOCOL	DESCRIPTION
Control serum collection	Day 0	Pre-immune bleed
Primary injection	Day 1	Immunize with 50µg antigen in complete Freud's adjuvant
1st Booster	Day 21	Boost with 25 µg antigen in incomplete Freud's adjuvant, 1st bleed collection
2nd Booster	Day 42	Boost with 25 µg antigen in incomplete Freud's adjuvant, 2nd bleed collection
3rd Booster	Day 62	Boost with 25 µg antigen in incomplete Freud's adjuvant, 3rd bleed collection
Sacrifice	Day 80	Terminal Bleed collection and termination of animals, 4th bleed collection

Parameters of Animals used

Parameter	Rat	Mice
Strain	Sprague Dawley	Swiss Albino
Noof animals used	1	1
Amount of antigen for Primary Booster	Primary :150-200 µg Secondary: 100-150 µg/Injection	Primary:50-100 µg Secondary: 75 µg
Volume of antigen with adjuvant	0.5ml – 0.7ml	0.1- 0.2ml
Amount of Bleed	1-2 ml (test bleed) 2-4 ml (final bleed)	0.2-0.5 ml (test bleed) 0.5-1 ml (final bleed)

Serum was isolated from blood collected after the primary injection by incubating the blood for 1 hour, in 37°C water bath, and kept in 4°C O/N. Next day it was centrifuged (12,000 rpm for 10 minutes) and serum was collected. The antibody was checked and confirmed by performing western blot starting from 1:1000 dilution. Further characterization of the antibody was carried out.

2.20. Validation of raised EGFP antibody through Western blotting

2.20.1. Dot Blot

Protein

Protein concentration to be loaded are determined by using Bradford's reagent and quantified by UV-Visible spectroscopy. Protein samples of concentration ranging from 0.2, 0.4, 0.6 and 0.8 mg/ml along with BSA and lysozyme as controls were taken.

Blotting

An Immobilion-P Transfer PVDF Membrane (Millipore, Prod. No. IPV H00010, pore size 0.45 µm) is pre-wetted for 15 seconds in 100 % Methanol to allow membrane activation and then soaked in distilled water for 2 minutes followed directly by 5 minutes equilibration in TBS-T (20 mM Tris, 150 mM NaCl, 0.05 % Tween 20, pH 7.5). Prior to this the membrane was marked with grids on which protein will be loaded using llead pencil. Do not let PVDF membrane to dry out. As PVDF membrane is

hydrophobic therefore applying pre-wetting protocol with methanol is necessary. Protein was loaded on respective grids as previously marked. Loaded protein was allowed to dry. Non-specific sites are blocked with blocking solution (TBS-T with 2 % low fat milk powder or other blocking reagent) for 1hour at room temperature with continuous gentle agitation using rocker. Blocking solution is discarded and membrane is incubated with primary antibody (1:1000) in 10 ml of TBS-T containing 2% low-fat milk powder for 1hour, followed by 5minX1, 10minx2, 15minx2 wash with 1xTBS-T. Secondary antibody incubation was for 1hour. The membrane was washed with TBS-T 3 x 10. Dot blot reaction can be developed using ECL based reagents with appropriate secondary antibody, HRP (horseradish peroxidase) conjugated.

2.20.2. Cell culture:

HEK293T (human embryonic kidney) cells were maintained in a CO2 incubator at 37°C and 5% CO2concentration in high glucose Dulbecco's modified Eagle medium (DMEM) with L– glutamine (Gibco) supplemented with 10% fetal bovine Calf serum and 100 U/ml penicillin-streptomycin. Cells were maintained for a few passages and frozen as aliquots; for experiments, cells were sub-cultured no more than a month to ensure a low number of population doubling.

2.20.3. Preparation of sample from cell culture:

The cell culture dish placed on ice after 72hrs post transfection. The media was aspirated from culture and then the cells were washed with ice cold 1x Phosphate buffer saline aspirate (PBS) to remove residual media and removed. Now 100ul of ice-cold lysis buffer A [composition: 20mM Tris-Cl pH 7.8, 137mM NaCl and 1% NP-40 supplemented with 1X protease inhibitor cocktail (Roche)] was poured gently on culture dish and the cells were dislodging of the adherent cells off the dish by using a cold sterile plastic cell scraper, then the resultant cell suspension was gently transferred into a pre-cooled micro centrifuge tube kept on ice. Alternatively cells can also be done by trypsinization followed by wash with PBS prior to resuspension in lysis buffer. The tube was subjected to agitation often for by placing on vortex at very low speed. To clarify the cell suspension and extract proteins, the microcentrifuge containing cell suspension was centrifuged at 12000 x g for 10minutes in a 4°C precooled centrifuge. The tube was gently removed from centrifuge and placed on ice. The resultant supernatant was collected and placed in a fresh tube kept on ice, and

pellet was discarded. The protein concentration in sample was estimated using Bradford reagent using spectrophotometer. A 50µg protein sample was kept aside for loading and the remaining aliquots of protein was stored at -20℃ for further processing. The protein sample was mixed with equal volume of 2 x laemmli sample loading buffer and boiled at 100℃ for 5min. The protein sample was centrifuged at 16000xg for 5minutes.

2.20.4. Protein separation by Gel Electrophoresis:

As gel percentage selection depends on the protein of interest, ORF2 being 150kDa a 10% acrylamide gel was made with 5% stacking. Equal amounts of protein were loaded into the wells of the SDS-PAGE gel, along with molecular weight prestained marker. The electrophoresis was run at 100volts for 15min and then the voltage was increased to 120-140 for 1-2hrs to finish using Mini protein Tetra cell (Bio-Rad).

2.20.5. Electro transfer of proteins on to PVDF membrane:

After protein separation the gel was placed in 1x Transfer buffer. The PVDF membrane was measured and cut according to the required size of gel. Then the membrane was charged using methanol. Now the transfer sandwich was assembled avoiding air bubbles between membrane and gel. The entire cassette was placed in tank with1x transfer buffer with an ice block and run on ice for 100min at 100volts using Bio-Rad mini trans blot electrophoretic transfer cell

2.20.6. Blocking and Antibody Incubation and development of Blot:

The membrane was then blocked with blocking buffer (5% low fat milk in 1x TBS-T) and incubated at room temperature for 1hour. After blocking the membrane was incubated with polyclonal rabbit human α-L1 ORF1p (RRM) (1:33000) [Sur et al., 2017], α-GAPDH (1:6000) (Santa Cruz Biotechnology), α-FLAG (1:3000) (Sigma) overnight at 4℃ against target protein. Next day the membrane was washed for 1hour using 1x TBS-T (5x2, 10x2, 15x2min), incubated with conjugated secondary α-rabbit HRP and secondary α-mouse HRP (Jacksons Immuno Research Laboratories, USA). Western blots were developed using ECL western blotting detection reagent (Pierce) as per manufacturer's instructions. The bands were detected by exposing the blot on X-ray film (Hyper film from GE Healthcare).

2.21. Production of polyclonal antibody against human L1ORF2p using synthetic peptide

2.21.1 Production of Human Line1 Polyclonal Antibody in Mice and Rat Using Synthesized Peptide

1. Design of Peptide

The Endonuclease domain (EN) of human Line1 had been cloned and expressed earlier for production of antigen required for antibody production. The expressed protein was not sufficient for purification in soluble fraction, so a stretch of 14 amino acid peptide was chosen from EN domain. Then peptide sequence was analyzed for its hydrophobicity for single residue to predict solubility nature of peptide. Then the immunogenicity of peptide predicted using prediction servers namely. NHLBI-AbDesigner: Software to design peptide-directed antibodies", Proped, and NetMHCpan Server. After much analysis the peptide predicted and selected was "RSTRQKVNKDTQE"

2. Synthesis of Peptide

The selected stretch of amino acids RSTRQKVNKDTQE was sent to Biotech desk Pvt.Ltd for synthesis of peptide. The received peptide was in lyophilized form of 4mg. The purity of peptide analyzed by the HPLC report provided with it.

3. Conjugation of Peptide with mcKLH

As peptide alone is not efficient for inducing strong immune response due to its small molecular weight, it is required to conjugate the peptide to a higher molecular weight carrier protein. For this we had chosen KLH as a carrier protein due to its high molecular weight (390kDa) and also as it is derived from the limpet, a gastropod, it is phylogenetically distant from mammalian proteins, thus reducing false positives in immunologically-based research techniques in mammalian model organisms. For conjugation Imject™ EDC mcKLH Spin Kit (Thermo Fisher Scientific, 77671) was used.

Procedure for Hapten Conjugation using EDC

A. Conjugation Procedure

- Reconstituted one vial of carrier protein (mcKLH) by adding 200µl of ultrapure water to make a 10 mg/ml solution
- Dissolved up to 2 mg of peptide in 450 µl of Imject EDC Conjugation Buffer
- Added the 450 µl of peptide solution to the 200 µl carrier protein solution
- For mcKLH, dissolve one vial of EDC (10 mg) in 1 ml of ultrapure water and immediately added 50 µl of this solution to the carrier-hapten solution.
- Incubated reaction for 2 hours at room temperature

B. Conjugate Purification by Desalting

- If the conjugate is to be used for injection within one week, PBS may be used for purification. If the conjugate will be frozen, use the Imject Purification Buffer Salts for purification, which will preserve the product during freeze-thaw cycles
- Dissolved the contents of one bottle of Purification Buffer Salts by adding 10 ml degassed, ultrapure water to the bottle
- Twisted off the bottom closure of the desalting column and loosen the cap. Placed the column in a collection tube
- Now the column was centrifuged column at 1,000 × *g* for 2 minutes to remove storage solution. When using fixed-angle rotors, place a mark on the side of the column where the compacted resin is slanted upward
- Removed the cap and slowly added 1 ml of purification buffer to the column. Again centrifuged at 1,000 × *g* for 2 minutes to remove buffer. Repeated this step three additional times, discarding the buffer from the collection tube
- Placed the column in a new collection tube and slowly apply the sample to the center of the compact resin bed
- Centrifuged at 1,000 × *g* for 2 minutes to collect the sample
- (Optional) After the conjugate-containing fraction is collected, the non-conjugated hapten can be recovered by continuing to add buffer to the column and collecting additional fractions
- The 2µg/µl of total 700µl hapten-carrier conjugate can now be used for immunization. For further usage the conjugated peptide was stored at -20ºC.

Immunization and validation protocols were performed as given before for generation of EGFP antibody.

2.22. Cloning, Expression and Purification of pET30bEGFP-EN

2.22.1 Cloning of EGFP in pET30b

The primers 4F/4R were used to amplify and clone EGFP from pEGFP-N1 in bacterial expression vector pET30b (+) for fusion protein.

2.22.2 Cloning of EN in pET30b-EGFP

The amino acid stretch from 143 to 203 in Human Line-1ORF2 sequence was amplified using below primers. The amplified 180bp nucleotide sequence was then cloned at C-terminal of pET30b-EGFP in order to express fusion protein of hLine-1 ORF2 with EGFP tagged at N-terminal end.

Forward primer: 5'- AATAAGCTTATGGGAGACTTTAACACC-3' [ENp–pET30bf]
Reverse primer: 5'- AATCTCGAGTCAGGAATAGGTGTGGTGTGG-3' [ENp–pET30br]

2.23. Site Directed Mutagenesis (SDM)

Site directed mutagenesis is a highly versatile technique and *in vitro* procedure that can be used to introduce specific nucleotide substitutions or deletions in a tailored manner. It requires custom designed oligonucleotide primers to confer a desired mutation in a double-stranded DNA plasmid. Formerly, a method pioneered by Kunkel (Kunkel, 1985) that takes advantage of a strain deficient in dUTPase and uracil deglycosylase so that the recipient *E. coli* degrades the uracil-containing wild-type DNA. In brief, point-mutations can be introduced in to plasmids using primers containing a desired mutation in a PCR protocol that amplifies the entire plasmid template. The parent template is then removed using an endonuclease that depends on methylation (i.e. DpnI), and then the nuclease-resistant nicked plasmid (the PCR product) was transformed in to bacteria. The plasmids are isolated from the resulting colonies and are screened for the desired modification. Finally, the positive clones are sequenced to confirm the desired modification. Site-directed mutagenesis studies can be extremely useful for elucidating the function of a gene or protein or for creating variants of an enzyme with new and improved functions. It has proven to be valuable for analysing the structure–function relationship of proteins. The method provides a

means of introducing specific nucleotide changes into a gene. As a result, the effects of specific amino acid changes in a protein can be gauged, relative to the protein from the wild-type gene. The resulting structural changes may then be linked to observed changes in function, stability and/or activity. There are now many approaches available for generating site-directed mutants, but most prominent methods include the following PCR with modified primers.

1. PCR with modified primers- Primers with specific mutated nucleotide
2. Primer extension- extension uses nested primers to mutate a target region
3. Inverse PCR- This method uses two back-to-back primers to amplify the whole plasmid and the linear product is then ligated back to the circular form. The primer binding regions can be changed by altering the primer sequences to contain the desired mutation. Insertions can be made around the primer binding regions by adding flanking sequences to the primers, and deletions can be made by simply leaving a space between the two primers.

In this we have employed PCR technique with modified primers.

2.23.1. Methodology

1. Primers were designed with mutated nucleotide at 91(Glycine) position to Alanine, Aspartic acid, Arginine and Tyrosine respectively

2. The pET30b-EGFP plasmid DNA was subjected to PCR using above designed primers. Reaction mixture and PCR conditions used in SDM are mentioned below. After setting up the reaction, mixture was mixed by tapping gently and centrifuged briefly. Thermocycler (Applied Biosystems) was used to perform PCR

PCR Conditions:

- Step 1 (Initial Denaturation) : 94°C for 2 minutes
- Step 2 (Denaturation) : 94°C for 30 seconds
- Step 3 (Annealing) : 58°C for 15 sec

} Repeat for 15 Cycles

- Step 4 (Extension) : 72°C for 5 minutes
- Step 5 (Final Extension) : 72°C for 5 minutes
- Step 6 (Hold) : 4°C for infinite time

3. After the PCR was completed, entire amplified product was digested with 0.5µL DpnI restriction enzyme in order to digest the original wild type DNA. 2µL of the PCR product was transformed into *E.coli* DH5α competent cells as per the protocol described earlier. Plasmid DNA was isolated from the single colonies obtained after transformation.

Reagents	Amount
5X Phusion buffer	4µL
dNTP's (40mM)	1µL
Forward primer	1.0µL
Reverse primer	1.0µL
Phusion polymerase	2.5µL
Template DNA (200pg)	1µL
Nuclease free water	11.75µL

2.23.2. Confirmation of mutation by sequencing

Plasmid DNA isolated from the colonies obtained after transformation was sent for sequencing to GCC Biotech, India. Sanger's dideoxy chain termination method was employed for sequencing the plasmid DNA.

2.23.3. Characterization of EGFP and mutant proteins:

2.23.4. Protein quantification using UV absorption spectroscopy:

Tyrosine and Tryptophan amino acids absorb UV light at 280 nm with extinction coefficient ε=1490$M^{-1}cm^{-1}$ and 5500 $M^{-1}cm^{-1}$ respectively. Using Beer-Lambert's law protein concentration was estimated using the following equation.

$$A=\varepsilon Cl$$

Where A is absorbance, ε is extinction coefficient, C is concentration of the compound used and l is path of the light through the sample. Path length of quartz cuvette used is

1 cm. Both the proteins were diluted 50 times in 50mM Tris pH 8.0, 150mM NaCl and absorbance was recorded at 280nm against same buffer as blank.

2.23.5. UV absorption spectra of EGFP and mutant proteins:

Every protein has maximum absorbance of light at a particular wavelength. The mature chromophore of EGFP and mutant also has maximum absorbance at a particular wavelength *i.e*. 488nm. Hence both the proteins are scanned with varying wavelengths from 200nm to 700 nm using Cary 60 UV-Vis spectrophotometer, Agilent technologies. Proteins were taken at different concentrations ranging from 50µM to 2.5µM and their respective absorption spectra was recorded against a baseline set up with 50mM Tris pH 8.0, 150mM NaCl buffer. To record spectra quartz cuvette with 1 cm path length was used and entire experiment was carried out at room temperature (25±2^{o}C).

2.23.6. Fluorescence spectroscopy of EGFP and mutant proteins:

Molecules have various states referred to as energy levels. Fluorescence spectroscopy is primarily concerned with electronic and vibrational states. Generally, the species being examined has a ground electronic state (a low energy state) of interest, and an excited electronic state of higher energy. Within each of these electronic states are various vibrational states. In fluorescence spectroscopy, the species is first excited, by absorbing a photon, from its ground electronic state to one of the various vibrational states in the excited electronic state. Collisions with other molecules cause the excited molecule to lose vibrational energy until it reaches the lowest vibrational state of the excited electronic state. The molecule then drops down to one of the various vibrational levels of the ground electronic state again, emitting a photon in the process. As molecules may drop down into any of several vibrational levels in the ground state, the emitted photons will have different energies, and thus frequencies. Fluorescence spectra of both EGFP and mutant was recorded using Fluorolog fluorescence spectrophotometer with 10µM concentration of protein at different pH conditions ranging from 5 to 12 with increment of 1 pH unit using sodium phosphate buffer. The excitation of both the proteins was done at 488nm and emission was recorded from 400nm to 600nm with intensity of fluorescence recorded at 1nm increment.

2.23.7. CD spectra of EGFP and mutant proteins:

Circular Dichroism is a powerful biophysical technique which gives secondary structure information of proteins and nucleic acids. Using this method changes in secondary structure of proteins can be identified. Circular dichroism (CD) spectroscopy is a form of light absorption spectroscopy that measures the difference in absorbance of right- and left-circularly polarized light (rather than the commonly used absorbance of isotropic light) by a substance. It has been shown that CD spectra between 260nm and approximately 180nm can be analyzed for the different secondary structural types: alpha helix, parallel and antiparallel beta sheet, turn, and other. CD spectra of both the proteins were recorded by taking 100ng/mL of sample at room temperature and pH 8.0 against 50mM sodium phosphate, 150mM NaCl as blank using quartz cuvette.

2.24. Expression and Purification of fusion protein (pET30b-EGFP+EN)

The positive clone which was confirmed by sequencing transformed in BL-21 cells for protein expression. The culture of 400ml lysed, sonicated to isolate soluble and insoluble fractions. The affinity purification using Ni-NTA beads was proceeded with both soluble fraction. As in case of soluble fraction purification, other bands were also observed along with target protein, decided to purify with insoluble fraction.

2.25. Purification from Insoluble fraction (Pellet)

The pellet obtained after sonication was dissolved in 8M urea, 100Mm NaH2PO4, 10mM Tris-Cl (pH-8.0) and kept in ice for 30minutes.Then the lysed cells were centrifuged at 12,000rpm for 10minutes. The clarified supernatant was then taken and analyzed on SDS-PAGE. The soluble fraction was purified using affinity chromatography by Ni-NTA beads. The elution of protein was done with 300mM Imidazole.

2.25.1. Refolding of Eluted protein through Dialysis

The eluted protein was once again analyzed by SDS-PAGE and protein was quantified using Bradford's reagent. Now the quantified protein was subjected to step wise dialysis. Prior to this the dialysis membrane was activated by boiling in 2% w/v

NaHCO3 and 1mM EDTA for 10minutes and finally in1mM EDTA. The membrane then cooled down and the protein was loaded and kept in 6M Urea, 100Mm NaH2PO4, 10mM Tris-Cl (pH-8.0), 300mM Imidazole for 2hours at 4˚C on magnetic stirrer with continuous stirring. Then followed by gradual reduction of urea concentration done by 4M Urea, 100Mm NaH2PO4, 10mM Tris-Cl (pH-8.0), 300mM Imidazole for 2hours, 2M Urea, 100Mm NaH2PO4, 10mM Tris-Cl (pH-8.0), 300mM Imidazole for 2hours, 1M Urea, 100Mm NaH2PO4, 10mM Tris-Cl (pH-8.0), 300mM Imidazole for 2hours, 0.8M Urea, 100Mm NaH2PO4, 10mM Tris-Cl (pH-8.0), 300mM Imidazole for 2hours and finally in 0M Urea, 100Mm NaH2PO4, 10mM Tris-Cl (pH-8.0) for 4hours. The dialyzed protein was now concentrated using Amicon Ultra ® -0.5ml 3K (Millipore). The concentrated protein quantified again using Bradford's reagent using UV-Visible spectroscopy at 595nm.

2.26. Immuno Histo Chemistry (IHC) Protocol

Immunohistochemistry is a discrete and possible technique which can be used mainly to study protein expression in tissues. It's a combination of anatomical, immunological and biochemical techniques. Over years many aspects made IHC a feasible technique to identify and also quantify the expression of target protein especially in case of cancer. The steps involved in this protocol are given below.

1. Fixing and embedding the tissue

After surgery the fresh tissue was kept in 10%NBF (Neutral Buffered Formalin). The ideal fixation time will depend on the size of the tissue block and the type of tissue, but usually fixation time will be between 18-24 hours seems to be ideal for most applications.

2. Dehydration

The tissue needs to be dehydrated to remove the water from the tissue which is present- either bound to the tissue, or free in the tissue as it should be embedded in paraffin wax which is hydrophobic, therefore, most of the water in the tissue must be removed before it can be infiltrated with wax. This process is carried out by immersing tissue in a series of ethanol solutions of increasing concentrations i.e. by keeping in 80% Ethanol for 1hour 30mim, 90% Ethanol for 1hor 30mim and 100% Ethanol for 2hrs and again in 100% Ethanol for another 2 hours or were kept overnight. A series of

increasing concentrations is used to ensure that the water in the tissue is gradually replaced by the alcohol and to avoid excessive distortion of the tissue. Various components of the cell are also removed by this process. At the lower end of the ethanol concentrations, water soluble proteins are removed, whilst towards the 100% ethanol step, certain lipids may be dissolved.

3. Clearing

Although the tissue was almost dehydrated with 100% Ethanol but it is not possible to proceed straight to wax embedding cause ethanol and wax don't mix. This is where 'clearing' employed. The term 'clearing' refers to the property of the solvents used in this process, when tissue is immersed in it; it becomes transparent and clear as they have a relatively high refractive index. The solvent used for this intermediate stage was acetone and chloroform as they are miscible with both ethanol and paraffin wax. Following the dehydration, the tissue is immersed in Acetone1 for 1hour 30mim, Acetone2 for 1hour 30mim and in Chloroform1 for 1hour 30mim and finally in Chloroform2 1hour 30mim. All these steps should be done precisely in case of time duration as if kept for more time than required there is a chance of tissue getting more hard and unable to process further. Alternatively chloroform can be replaced by xylene but as it also removes fat residues in the tissue samples, this sometimes leads to shrinkage of tissue, here chloroform is preferred.

4. Infiltration and Block preparation

Infiltration is when the final chloroform is replaced with molten wax which infiltrates the tissue. This is achieved by immersion of tissue in equal proportion of chloroform and wax in 1:1 ratio and kept in for 3hrs, followed by pure wax for 3hous at 65°C. After this the molten wax was poured in mold after keeping the tissue in proper orientation.

5. Sectioning

The blocks were then chilled in a try of ice because the cold wax makes a clean cut compared to paraffin wax cut at room temperature. The paraffin blocks were then trimmed and sectioned using an automated microtome with 4-5µm thickness. The sections carefully collected were made to spread by keeping them over water maintained at 45°C. Then the sections were collected on coated glass slides. The glass slides were now kept at 65°C for 2hours before processing.

6. Antigen Retrieval

During formalin fixation, cross linking of proteins occur forming methylene bridges resulting in masking of antigenic sites. It is essential to unmask the antigen epitopes in order to allow the antibody binding Antigen retrieval methods break these methylene bridges and expose antigenic sites, allowing antibodies to bind. The two methods for antigen retrieval are heat induced epitope retrieval (HIER) and enzymatic retrieval**.** Hier was employed in this method. (Heat Induced Epitope Retrieval: HIER) or enzymatic digestion (Proteolytic Induced Epitope Retrieval: PIER). In this protocol, HIER had been used for antigenic retrieval.

7. Heat Induced Epitope Retrieval (HIER)

After adding the appropriate antigen retrieval buffer (10mM Tris, 1mM EDTA, p^H-9.0) to the container, the buffer was made warm using micro wave oven for 30 seconds. Prior to this the pressure cooker with lid open was placed on hotplate and turned it on full power. After the pressure cooker to come to a boil, de-paraffinized and rehydrated sections are now placed in buffer and the lid was closed tight. As soon as the cooker has reached full pressure, a 3min time lapse was observed until pressure was released during first release. Immediately turn off the hotplate and the pressure cooker was placed in an empty sink the pressure release valve was activated by keeping under the running coldwater. Once the pressure was completely out, the sections were allowed to cool for further processing.

8. Blocking

After antigen retrieval the sections were rinsed in distilled water to remove the traces of buffer. Then the sections were incubated with 1% BSA in 1x TBST for 1hour at room temperature.

9. Primary Antibody Incubation

The tissue sections were incubated with α-ORF1, α-ORF2 at 4℃ for 8 to 10hours. Then the samples were washed with 1xTBST for 5minx1, 10minx2, 15minx2. Then peroxide blocking was done using 3% hydrogen peroxide in order to reduce endogenous peroxidise activity therefore inhibiting background staining. The samples

were further incubated with secondary antibody for 1hour proceeded for washing as detailed above.

10. DAB substrate staining

3, 3′-Diaminobenzidine (DAB), a derivative of benzene, is an organic compound that is used in the staining of nucleic acids and protein. Although it is water soluble in its unoxidized form, it forms a water-insoluble brown precipitate when oxidized. In immunohistochemical procedures, the location of proteins can be detected using DAB as a substrate. A protein of interest is targeted by an antibody that is conjugated with a peroxidase enzyme, and in the presence of hydrogen peroxide, DAB is readily catalyzed to its oxidized form, forming a brown precipate. Counter staining with haematoxylin not exceeding 3 minutes, washed off with tap water, dehydrated using 50%, 75%, 90%, 100% Ethanol and finally with Xylene. The dehydrated slides were mounted with DPX mounting media. Images were captured using a light microscope (Leica Microsystems) equipped with a camera. Then sections were mounted and covered with glass slides and observed under micro scope. Signals were visualized by Intensity of DAB stained regions were measured with ImageRatio software [55] and plotted as percentage of expression.

Chapter 3

Results

Objective 1

3.1 To clone fragments of human LINE-1 ORF2p for the generation of antibody against hL1-ORF2p

3.1.1. Cloning of human LINE-1 ORF2 CysCysHisCys (CCHC) domain

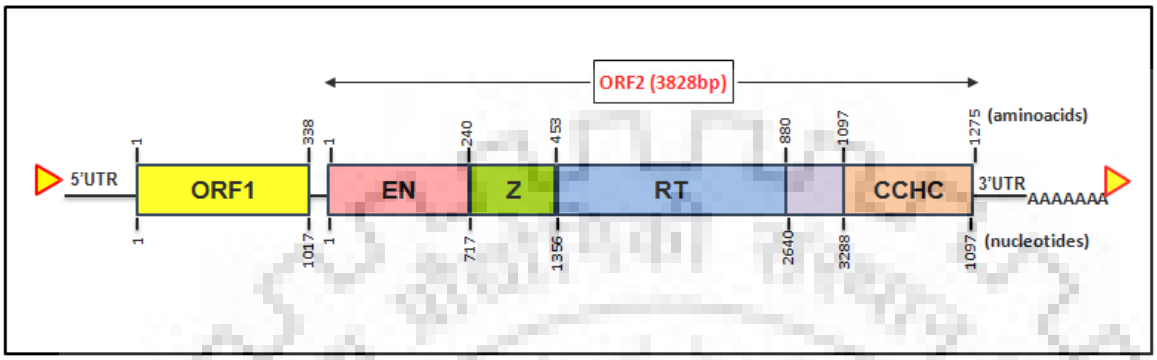


Figure R1: Schematic representation of amino acids and nucleotides map of human L1ORF1 and ORF2p. EN: endonuclease, RT: reverse transcriptase, CCHC: CysCysHisCys type zinc finger domain. Red triangle indicates Target site Duplication (TSD)

Human L1 encoded ORF2p is a 1275 amino acids protein (L1RP accession number AF: 148856.1) [Kimberland et al., 1999] with a predicted molecular weight of ~ 150 kDa [Appendix I] [Moran et al., 1996]. It has three partially characterized domains which are from N- terminal to C-terminal: Endonuclease (EN) (Amino Acids(AA) : 1-239 relative to L1RP accession number AF148856.1) , Reverse transcriptase (RT) (AA: 453-880) and CysCysHisCys (CCHC) type zinc finger domains(AA: 1096-1275) (Figure R1) [Clements et al., 1998; Moran JV et al., 1996; Feng et al., 1996; Malik et al., 1999; Piskareva et al.,2013] . The full length ORF2p (Mol wt ~150 kDa) expressed in very low amount and thus it is extremely tough to make enough protein to raise antibody against it. Hence, the C-terminal CCHC domain (Mol wt of 27kDa) of ORF2p was cloned and expressed. A disease causing human full length active L1 cloned in pBluescriptKS(-) {pBS-L1RP, a full length L1 inserted in intron 1 of the retinitis pigmentosa-2 (RP) gene was PCR amplified from the patient and cloned in pBSKS plasmid} [Schwahn U et al., 1998] was available in laboratory (a generous gift from Kazazian HH Jr. Johns Hopkins Univ, USA).

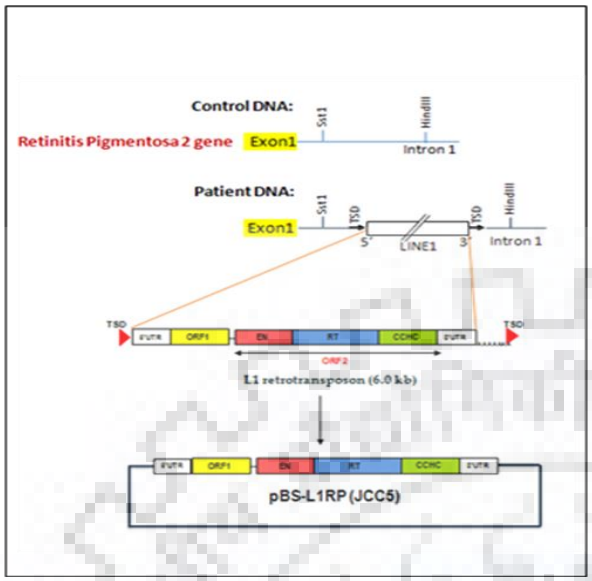


Figure R2: Schematic diagram of the full length L1 retrotransposon insertion in intron 1 of retinitis pigmentosa -2 (RP) gene in patient DNA. Black arrows flanking the L1 sequence represents 14 bp target site duplication (TSD). The complete L1 sequence was PCR amplified from a patient DNA and cloned in pBSKS(-) plasmid (JCC5 clone, Kimberland et al. 1999)was a kind gift from Kazazian Lab (Johns Hopkins University, USA).

The CCHC domain spanning nucleotides 3291-3825 was PCR amplified using pBS-L1RP as template with primers 1F and 1R (Table 1) and cloned in *EcoRI* and *XhoI* site of pET30a bacterial expression vector to get pET-CCHC (Figure R3). The clone was confirmed by Sanger sequencing (Figure) before setting the expression analysis.

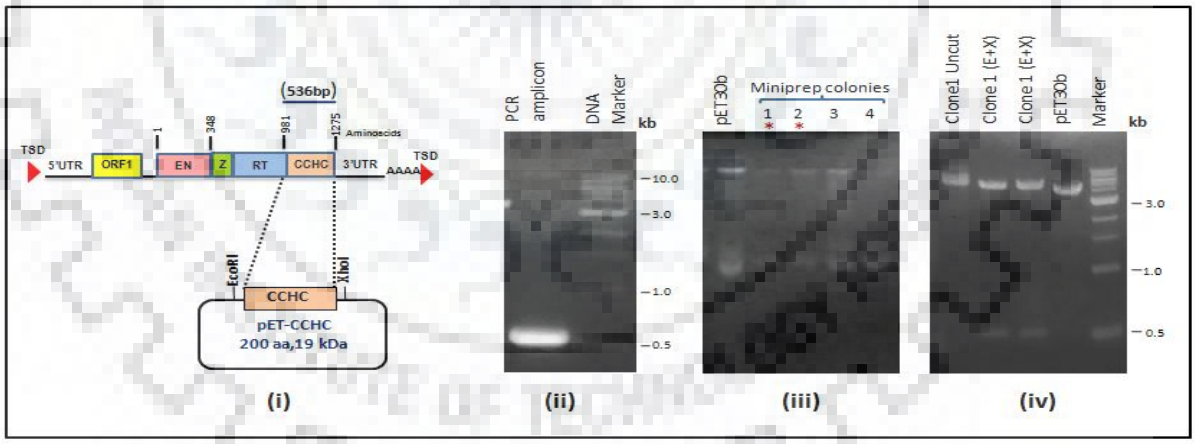


Figure R3: Cloning of CCHC fragment in bacterial expression vector (i) Scheme of L1 showing 0.537 kb fragment and its clone in pET30b expression vector. (ii) Primers were designed to amplify 0.5kb fragment. The PCR product (insert) respective to CCHC fragment of L1 amplified from pBS-L1RP resolved in 1.0% agarose gel. (iii) Agarose gel electrophoresis to screen the mini prep clones. (iv)Confirmation of Clone 1 and 2 by digesting with EcoR1 and Xho1 restriction enzymes which showed 0.5kb expected size of insert and vector respectively.

3.1.2. Expression Studies of recombinant human L1CCHC (hL1CCHC)

The expression of pET-CCHC was checked at 37°C for 3 hours with 0.4mM IPTG in *E.coli* expression cells. The total lysate from induced cells was analyzed and showed a distinct band at around 14kDa (Figure R4; panel i) when compared to uninduced lysate. The expected molecular mass of the pET-CCHC fragment is around 27kDa. In-order to confirm the identity pET-CCHC fragment with molecular mass of 14 kDa in induced cell lysate, the protein was first purified using Ni-NTA affinity followed by gel filtration chromatography and then subjected to matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) analysis. The mass spec results of the protein didn't match with human L1ORF2p CCHC domain and thus no further study was performed with pET-CCHC clone.

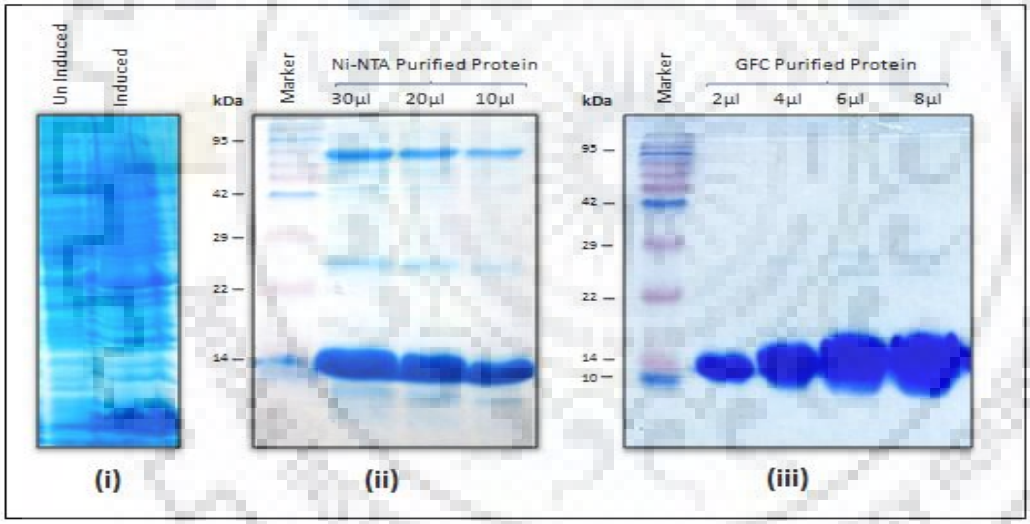


Figure R4: (i) Checking expression for human CCHC domain protein using positive clone with uninduced fraction. (ii) Purification of CCHC from soluble fraction by affinity purification. The Ni-agarose chromatography was performed as the clone contains six histidine amino acids (His-tag) at the N terminal. The elution showed significant amount of CCHC protein purified by Ni-agarose chromatography and gradual amount of protein was analyzed by SDS-PAGE (lane 2,3,4). (iii) Further the protein was also purified through gel filtration chromatography using HiLoad Superdex 75 pg preloaded column. The purified protein was loaded on 12% SDS-PAGE gel along with protein molecular weight marker.

3.2. Cloning of human LINE-1 (L1) Endonuclease (EN) domain

As expression analysis of CCHC domain didn't give desirable result, next human L1ORF2p EN domain was cloned. pBS-L1RP was use as template to PCR amplify the EN domain (1- 720bp) with primer set 2F/2R (Table 1). The amplified fragment was cloned in bacterial expression vector pET28C (Figure R5) to obtain pET-EN. The expression of EN was checked in BL21 strain of *E.coli* at 37°C for 3 hours with 0.4mM IPTG. Total lysate from induced cells were analyzed in SDS PAGE gel and compared with total uninduced lysate. A faint band at around 31 kDa with expected mol mass of pET-EN was observed in induced lysate (Figure R6).

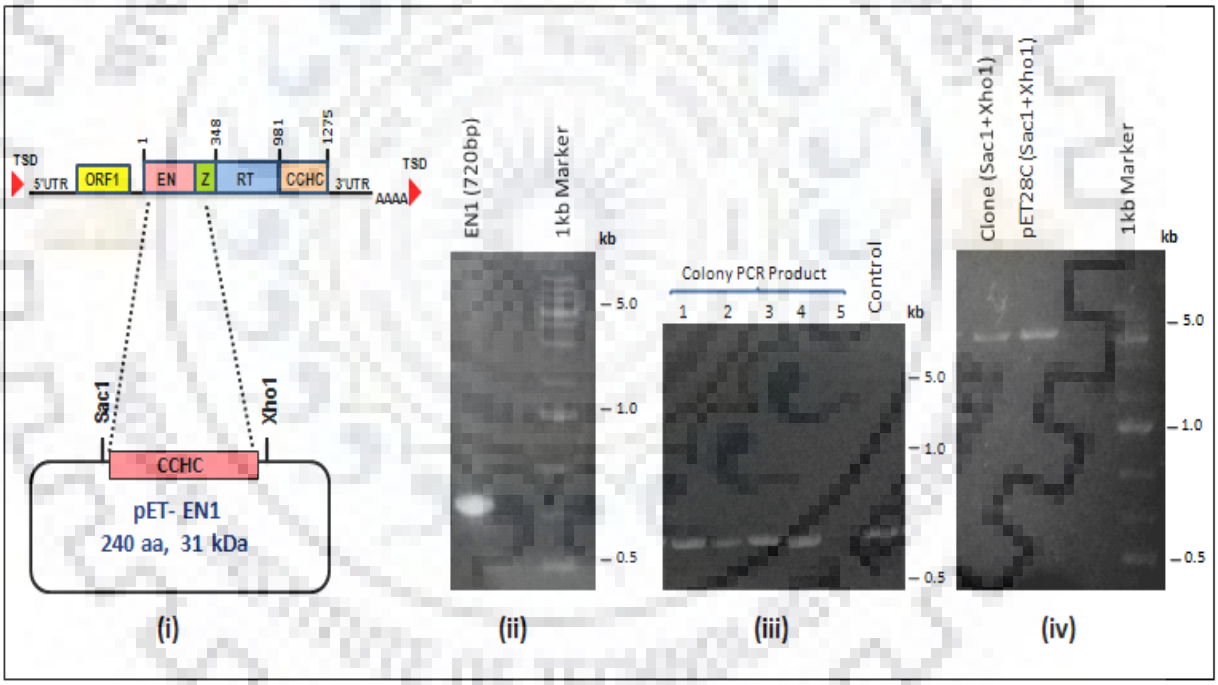


Figure R5: Cloning of hL1EN domain in bacterial expression vector (i) Schematic view of L1 showing 0.72 kb fragment and its clone in pET28C, a bacterial expression vector. (ii) Primers were designed to amplify 0.72 kb fragment. The PCR product (insert) resolved in 1.0% agarose gel. (iii) Agarose gel electrophoresis to screen the PCR amplicons obtained by colony PCR using primer set 2F/2R (Table 1). (iv) Confirmation of clone by digesting with *Sac1* and *Xho1* restriction enzymes.

In-order to confirm whether the protein with molecular mass of 31 kDa in induced cell lysate is recombinant human ORF2p, Western Blotting analysis was performed. The bacterial induced recombinant hL1EN contains six amino acids histidine tag at the N-terminal domain (Figure R6; panel ii) Using anti-His antibody, a band of around 31kDa was detected which confirmed the identity of recombinant human pET-EN in bacterial induced total cell lysate.

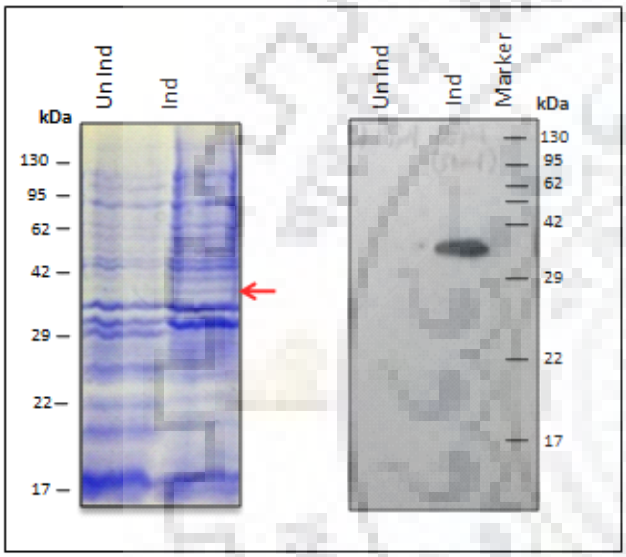


Figure R6: Expression and Western detection of human L1 EN domain protein in bacterial soluble lysate. (i) EN clone was transformed in BL-21 for protein expression. The clone was induced with 0.4 mMIPTG and the lysate was analyzed in SDS-PAGE along with control. (ii) Immunoblotting using anti-His antibody shows a specific band with MW 31 kDa (induced lane)

3.2.1. Expression studies of recombinant human L1EN

As the expression of pET-EN at 37C was extremely low, other temperatures were tried to optimize the expression of pET-EN clone. The expression was checked at 37ºC for 3hrs, 30°C for 6hrs, 25ºC for 10hrs and 16°C for 14hrs respectively. The total protein extracts were collected and suspended in lysis buffer and further subjected to sonication for cell disruption. The soluble and insoluble fractions were loaded and analyzed separately using SDS-PAGE (Figure R7; panel i). pET-EGFP construct (EGFP cloned in pET vector) was used as control protein for expression analysis. No significant expression was observed in other temperatures.

The expression was double check by purification total lysate (soluble fraction) obtained at 37°C using NI-NTA affinity chromatography. No special bands were obtained after elution suggesting that expression of pET-EN is extremely low at 37°C.

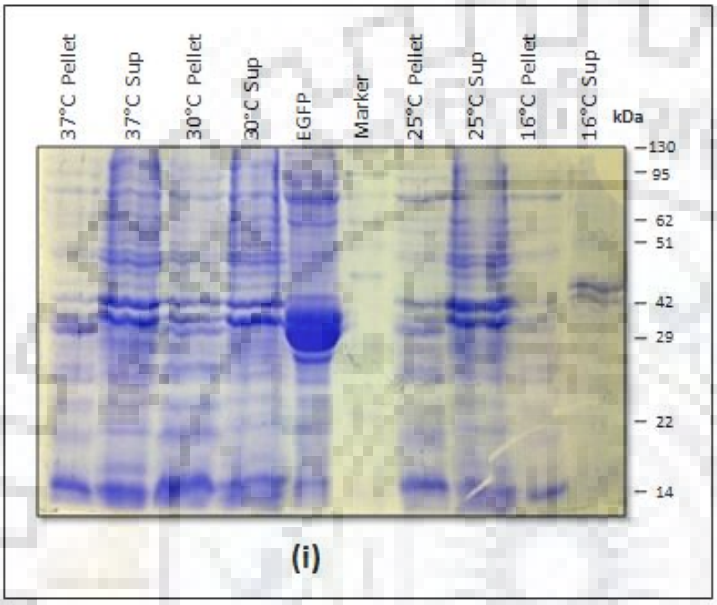


Figure R7: (i) Expression profile of human EN domain protein using positive clone at different temperatures: 37°C, 30°C, 25°C and 16°C. The soluble (sup) and insoluble (pellet) fractions were checked for all.

3.3. Cloning and Expression of human LINE-1 (L1) Endonuclease (EN2) domain

As full length EN may cause toxicity to host because of its endonuclease activity, thus a truncated EN fragment (nucleotides 265-720; amino acids 88-240 of L1ORF2) was tried to express in bacterial expression cells. The primers 3F/3R (Table 1) were designed to PCR amplify EN2 domain from pBS-L1RP. The truncated EN domain was amplified (0.45 kb) by PCR and cloned in bacterial expression vector pET28C (Figure R8).

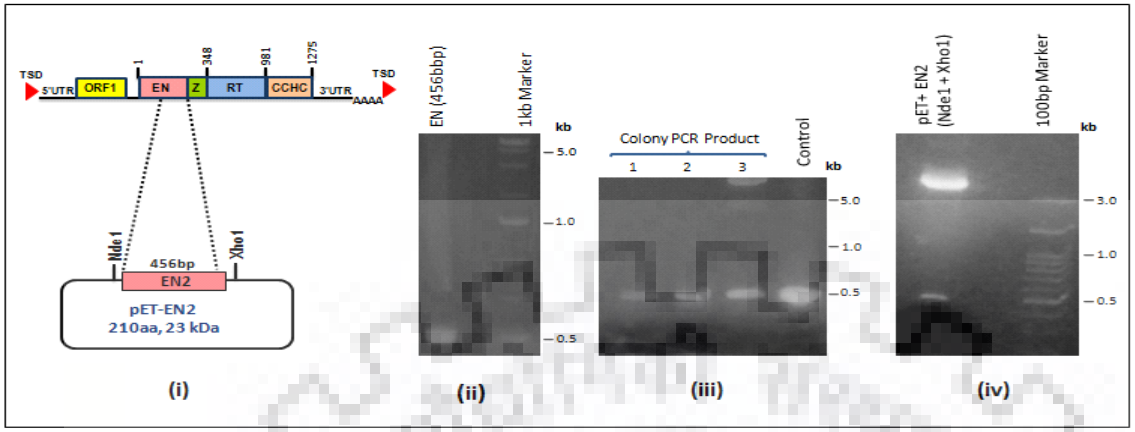


Figure R8: Cloning of hL1EN2 domain in bacterial expression vector (i) Schematic view of L1 showing 0.45 kb fragment and its clone in pET28C, a bacterial expression vector. (ii) Primers were designed to amplify 0.45 kb fragment. The PCR product (insert) resolved in 1.0% agarose gel. (iii) Agarose gel electrophoresis to screen the PCR amplicons obtained by colony PCR using the designed primers (3F/3R) along with control. (iv) Confirmation of Clone along with vector (pET28C) by digesting with *Nde1* and *Xho1* restriction enzymes which showed 0.45 kb expected size of insert and vector respectively.

The expression of EN was checked in BL21 strain of *E.coli* at 37°C for 3 hours with 0.4mM IPTG. Total lysate from induced cells was analyzed and no band (Expected MW of bacterial express EN2 ~21 kDa) was visible in induced lane (Figure R9).

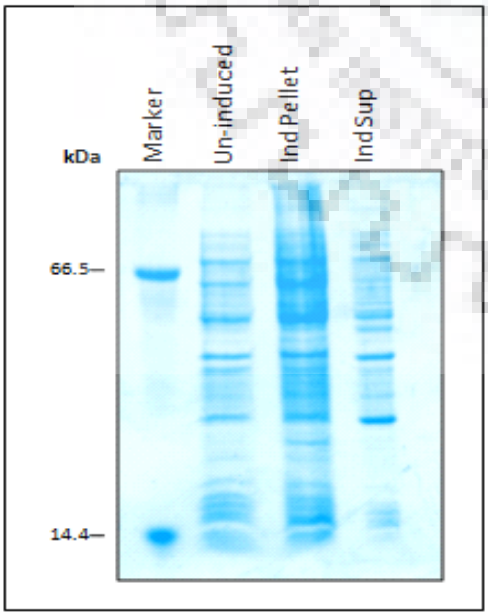


Figure R9: Expression of human L1 EN2 fragment in bacterial expression cells. Expression of EN2 domain protein in bacterial soluble lysate. EN2 clone was transformed in BL-21 for protein expression. The clone was induced with 0.4MmMIPTG. Total lysate along with uninduced were analyzed in SDS-PAGE

Objective 2

3.4. To generate in house antibody against human LINE-1 ORF2p protein using synthetic peptide

3.4. Production of Human LINE-1 ORF2p antibody using synthesized peptide

From the above study it is evident that the hL1ORF2p CCHC and EN domain didn't express well to get enough antigen for the generating hL1ORF2p antibody. Thus, synthetic peptide from hL1ORF2p sequence was designed for generating ORF2p antibody.

3.4.1. Design of Peptide

Next we used computational tool NHLBI-Ab designer to find out an optimum peptide which will generate antibodies against hL1ORF2p. The complete human L1 ORF2p sequence was screened using NHLBI-Ab [Pisitkunt et al. 2011] designer and the program showed list of peptides which can be good antigen to generate antibody against hL1ORF2p. Among those identified peptides, the sequence RSTRQKVNKDTQE was selected as same sequence with partial overlap was used before to synthesize antibody against hL1ORF2p [Schumann GG et al. 2004].

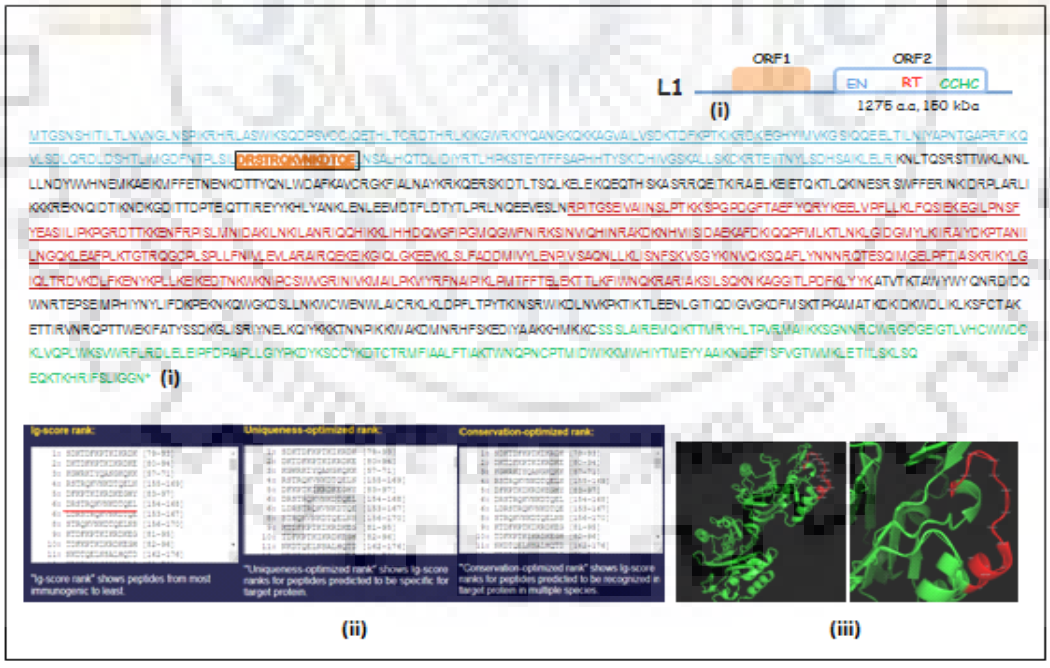


Figure R10: (i) Schematic representation of human L1ORF2p. The 1275 amino acid stretch corresponding to different domains represented in EN (blue), RT (red) and CCHC (green). (ii) List of peptide obtained using NHLBI-AB computational tool. The selected peptide is marked in red. (iii) Folding of selected peptide using PyMOL.

3.4.2. Conjugation of Peptide

As peptide alone is not efficient for inducing strong immune response due to its small molecular weight, it is required to conjugate the peptide to a higher molecular weight carrier protein. To perform this, we had chosen Keyhole limpet hemocyanin (KLH) as a carrier protein due to its high molecular weight (390kDa) and also as it is derived from the limpet, a gastropod, it is phylogenetically distant from mammalian proteins, thus reducing false positives in immunologically-based research techniques in mammalian model organisms. For conjugation of the peptide with the carrier protein, Imject™ EDC mcKLH Spin Kit (Thermo Fisher Scientific, 77671) was used (Figure R11). For conjugation of the peptide with the carrier protein, Imject™ EDC mcKLH Spin Kit (Thermo Fisher Scientific, 77671) was used (Figure R11).

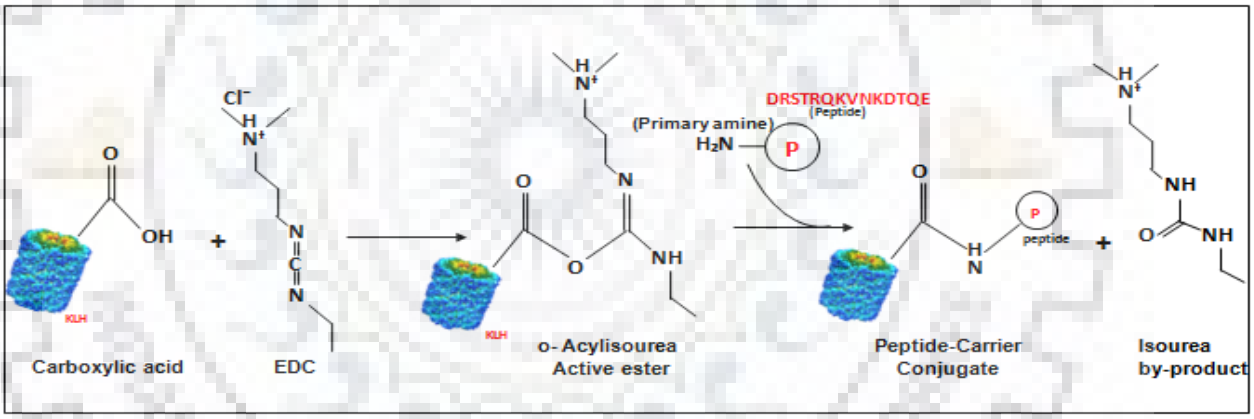


Figure R11. Schematic and step wise representation of Conjugation procedure. KLH (carrier protein) was linked with EDC (*N*-Ethyl-*N*'-(3-dimethylaminopropyl)carbodiimide hydrochloride), EDC reacts with available carboxyl groups on both the carrier protein and peptide hapten to form an active o-acylisourea intermediate which then reacts with a primary amine to form an amide bond and a soluble urea by-product. This reaction produces a Peptide – Carrier protein conjugate.

3.5. Generation of antibody against Enhanced Green Fluorescent Protein (EGFP) protein to check in-house antibody production protocol

In order to standardize the protocol of generation of an in house antibody, the procedure was first standardized by using purified enhanced green fluorescent protein (EGFP) as an antigen.

3.5.1. Cloning of Enhanced Green fluorescent protein

pEGFP-N1, a mammalian vector is available in our laboratory in which EGFP gene was cloned for animal cell culture, the EGFP gene was excised using restriction digestion using *BamH1* and *Not1*, then the insert was cloned in pET30b, a bacterial expression vector in the same site to maintain the same reading frame. The EGFP cloned in bacterial expression vector pET30b (pET-EGFP) was confirmed by digesting with BamHI and NotI.

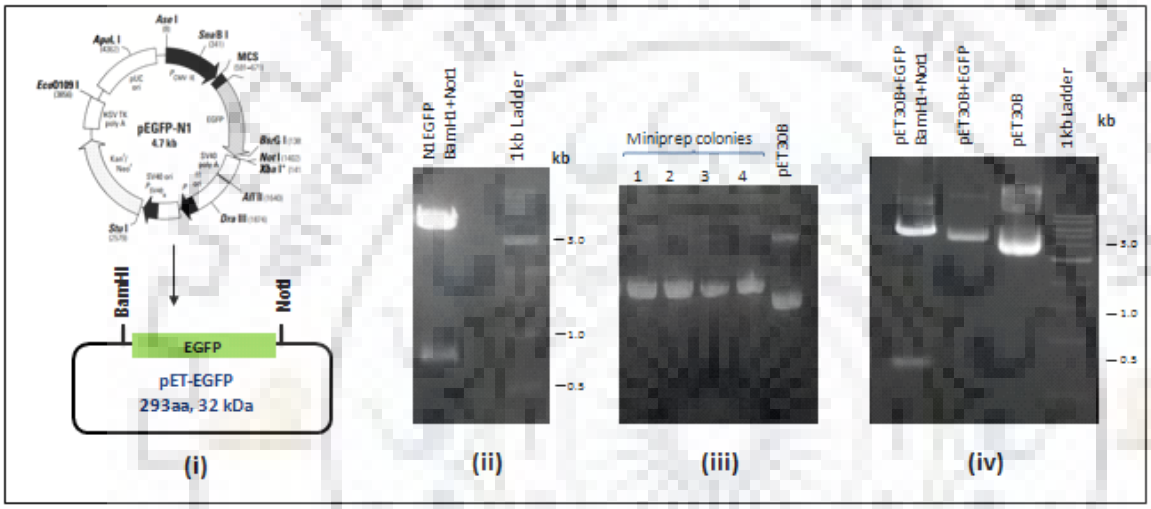


Figure R12: Cloning of EGFP gene in bacterial expression vector (i) Schematic representation of cloning EGFP, 0.72 kb fragment from pEGFP-N1 into pET30b expression vector. (ii) Insert released from digestion of pEGFP-N1 with Not1 and Xho1 (iii) Miniprep of colonies obtained after transformation through agarose gel electrophoresis. (iv) Confirmation of Clone by digesting with Not1 and Xho1 restriction enzymes which showed a fragment of 0.72kb, size of insert

3.5.2. Expression and Purification of Enhanced Green fluorescent protein

The expression of EGFP was checked in BL21 strain of *E.coli* at 37°C for 3 hours with 0.4mM IPTG. Total lysate from induced cells was analyzed and significant amount of EGFP protein was seen in induced lysate compared to uninduced sample. Both sup and pellet showed induced protein. Further, the protein production was optimized and found that at 25°C expression of protein was quite high when compared to other temperatures. The protein was purified using affinity chromatography by HisTrap column. The fractions were analyzed on SDS-PAGE. Further the affinity purified protein was subjected to

another step of purification using ÄKTAprime plus Size exclusion chromatography with Superdex 75pg column.

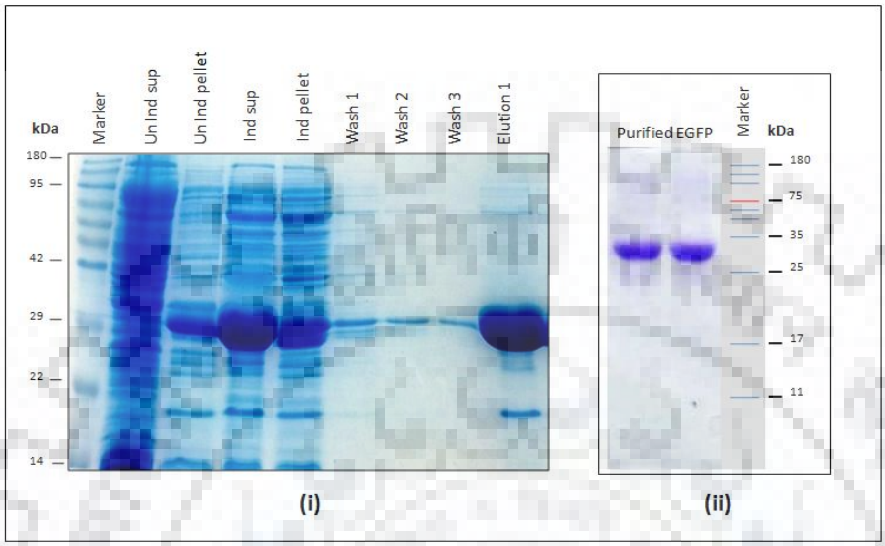


Figure R13: (i) Expression analysis of EGFP clone. The soluble fraction was purified using Ni-NTA affinity column. (ii) The fractions of EGFP collected from affinity chromatography were now further purified using Superdex 75pg pre loaded column through Gel filtration chromatography and the protein was analyzed on SDS-PAGE.

3.5.3. Generation of an In house polyclonal antibody against Enhanced Green fluorescent protein

The purified EGFP from *E.coli* expression cell was used as antigen and injected to rat to generate antibody against the protein. The SD rat (age 4 week) was maintained in the IITR animal facility as per institute guidelines. The pre-bleed was taken before injecting the antigen as control serum. The Western blot was performed using pre bleed and no band was detected using EGFP protein. Next, immunization was performed by injecting around 200 µg EGFP intraperitoneally, and a 90 day protocol was followed for generation of antibody (Figure R14). After 21 days, the first bleed was taken and the isolated serum was checked to find out antibody against EGFP by dot blot. The dot blot showed indication of antibody against EGFP where BSA and lysozyme protein was used as negative control. The final bleed after immunizing with third booster was validated by

immunoblotting. Again EGFP along with two control proteins (BSA and lysozyme) was separated in SDS-PAGE. The western blot with anti EGFP showed clear band in the EGFP Lane 3, no band was developed in the lane marked BSA and lysozyme. This validated the in house antibody generated protocol and the same protocol is followed for generating L1ORF2p antibody.

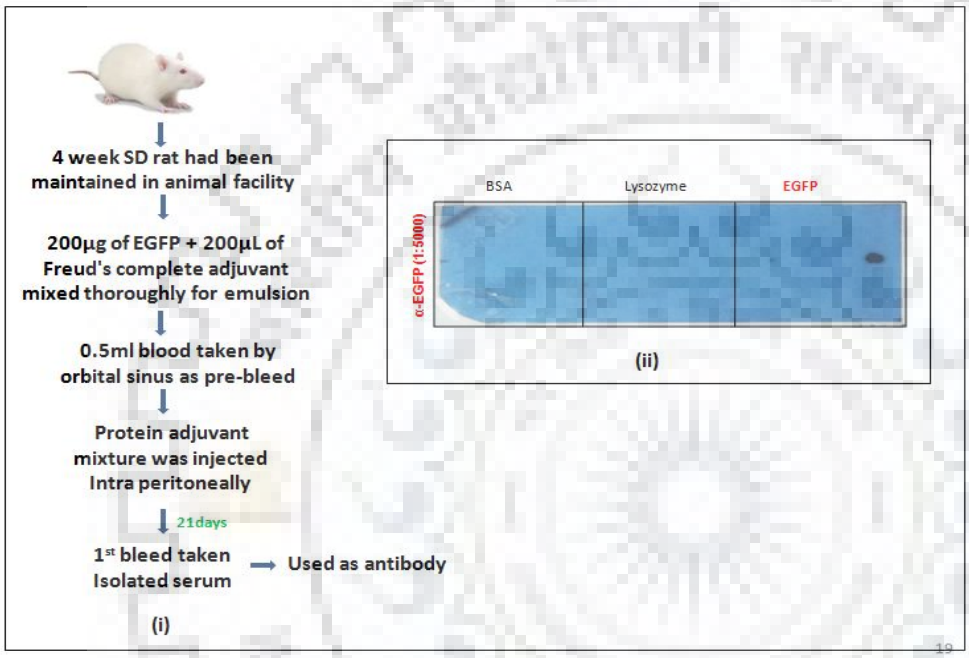


Figure R14: (i) Protocol for immunization of SD rat using EGFP protein for generation of an in house antibody. 200µg of EGFP along with Freud's complete adjuvant was injected peritoneally. After 21 days first bleed was taken and serum isolated then used as anti EGFP (ii) The antibody was validated using Dot blot, where activated PVDF membrane was spotted with BSA, lysozyme and EGFP in pre marked grids. The whole membrane was incubated with α-EGFP (1:1000). The immune response was observed only in case of EGFP but not in either BSA or lysozyme reveals the specificity and sensitivity of generated antibody.

Confirmation of inhouse generated EGFP antibody by western blotting. BSA, lysozyme and EGFP were separated in 12 % SDS-PAGE gel and western blotting was performed using α-EGFP and α-His antibodies.

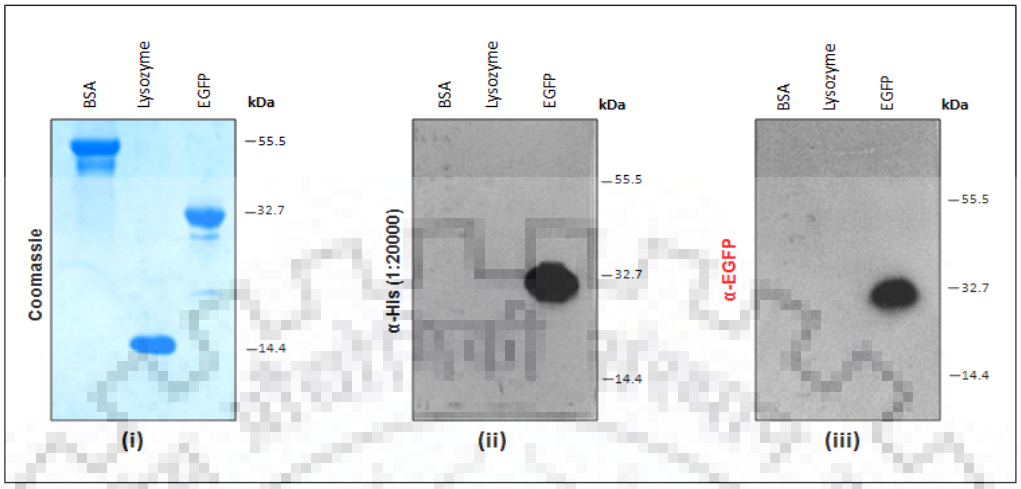


Figure R15: (i) Coomassie stained 12% SDS-PAGE showing BSA, lysozyme and EGFP. (ii) The Immuno blot of the same protein samples in diluted quantities stained with α-His gives specific band in case of EGFP as the construct have 6xHis tag at N-terminal and no response in case of BSA or lysozyme as they are synthesized. (iii) The replica blot having exact protein samples as (ii) but in this case it was stained with α-EGFP (1:10000) , equal amount of protein separated in 12% SDS-PAGE gel and stained with Coomassie used as loading control. The immune response developed only in case of EGFP but not in either BSA or lysozyme reveals the specificity and sensitivity of generated antibody.

3.5.4. Detection of L1 retrotransposition activity in HEK293T cells using in house generated EGFP antibody

Cell culture based L1 retrotransposition assay is a detection tool to investigate the jumping activity of an active L1. In this retrotransposition assay, the 3' UTR of a retrotransposon competent L1 is tagged with an indicator cassette designed to detect retrotransposition events. This retrotransposition indicator cassette consists of a selectable marker i.e. an EGFP reporter gene containing its own promoter and polyadenylation signal. The *EGFP* cassette is introduced at the 3'-UTR of L1 in opposite transcriptional orientations. The *EGFP* gene is interrupted by an intron in the same transcriptional orientation as the L1. The orientation is designed in such a way that only when a transcript initiated from the promoter driving L1 expression underwent retrotransposition, then only the EGFP does express. The expression of EGFP directly correlates the rate of retrotransposition *invitro*.This reporter cassette **p-cDNA-L1RP-EGFP**

was transfected in HEK293T and after 96 hrs post transfection; the cells were viewed under fluorescence microscope for EGFP fluorescence. Almost 2-3% EGFP positive cells were visible under fluorescence microscope. The same cells were lysed and the total lysate was checked for the expression of EGFP using in-house EGFP antibody. A single distinct band in lane marked pcDNA-L1RP-EGFP (Figure R17; panel iii) indicates that the in house anti-EGFP antibody is effective to detect the L1 retrotransposon activity in cell culture based retrotransposition assay.

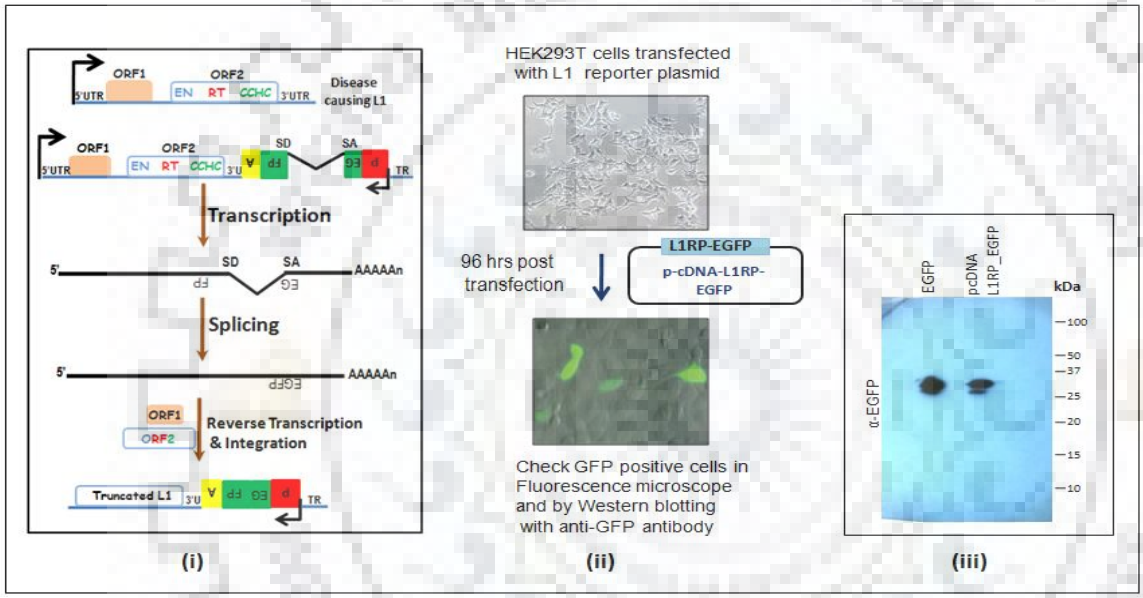


Figure R16: Detection of L1 retrotransposition events using inhouse EGFP antibody. (i) Scheme of cell culture based retro event using EGFP as a marker. (ii) 293T cells were transected with pCDNA and after 96 hrs post transfection EGFP positive cells were detected using fluorescence microscope. (ii) Cell lysate then separated in 12% SDS-PAGE gel and western blotting was performed using inhouse EGFP antibody.

3.6. Generation of an in house polyclonal antibody against human Line-1 ORF2p protein using synthetic peptide

Next to generate ORF2p specific antibody the KLH-ORF2p peptide conjugate as an antigen was injected to Swiss albino mice and Sprague Dawley rat. The animals were housed to IIT Roorkee animal facility as per institute guidelines. Around 50µg antigen per mice and 200µg per rat were injected as primary immunization followed by three booster

doses as per protocol (the detailed description in Material and Methods section). The first bleed taken after 21 days was checked for immune response by dot blot technique. Proteins were spotted in different amount in PVDF membrane and then incubated with 1st bleed serum as primary antibody. Simultaneously, EGFP was blotted and incubated with anti-EGFP which was treated as positive control. The signal in dot blot suggests that there is a immune response against ORF2p peptide. The dot blot results showed cross reactivity with antibody raised in rat, thus mice was used for further study. Next, the final bleed was collected and Western blot was performed to detect exogenous ORF2p in HEK293T lysate. HEK293T cells don't show endogenous ORF2p. An over-expressed mutant L1ORF2p construct pN13XD205 [Mandal et al, 2013] was transfected and the lysate was prepared after 4 days post transfection. The Western blot showed a band at around 150 kDa in lane D205A when it exposed for long time (around 3 minutes). A non specific band between 75-100 kDa was observed in the control lane. Although the conjugated hORF2 peptide showed antibody response and as the control lane also showed some non-specific bands another strategy was followed to make to hORF2p antibody (described below)

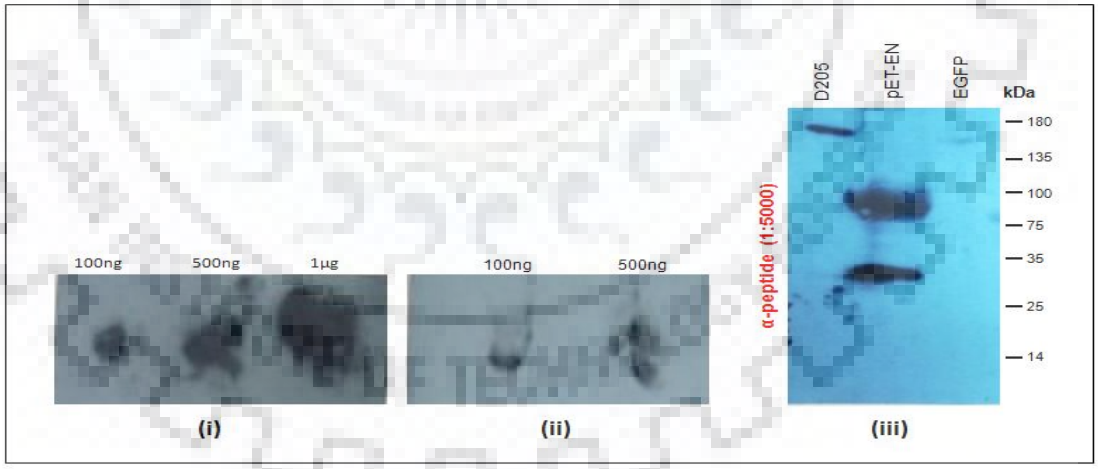


Figure R17: (i) Dot blot with 100ng, 500ng and 1µg of peptide showing immune response with peptide antibody at 1:5000 dilution. (ii) Dot blot of EGFP with α-EGFP (1:5000) (positive control). (iii) Validation of final bleed by immunoblotting with α- L1ORF2p. D205: total lysate from HEK293T cells after transfecting L1ORF2p EN mutant over-expressing construct (N1D205); pET-EN: total bacterial lysate from pET-EN induced clone; EGFP: total bacterial lysate from pET-EGFP clone.

3.7. Generation of an in house polyclonal antibody against human LINE--1 ORF2p protein using peptide tagged with EGFP as a carrier protein

Next to make ORF2p antibody a small stretch of EN domain was fused at the C-terminus of EGFP protein. The EGFP fused EN will be used as an antigen where EGFP acts as carrier protein. The cloning scheme of EGFP-EN in pET30b bacterial expression vector is shown in (Figure R19)

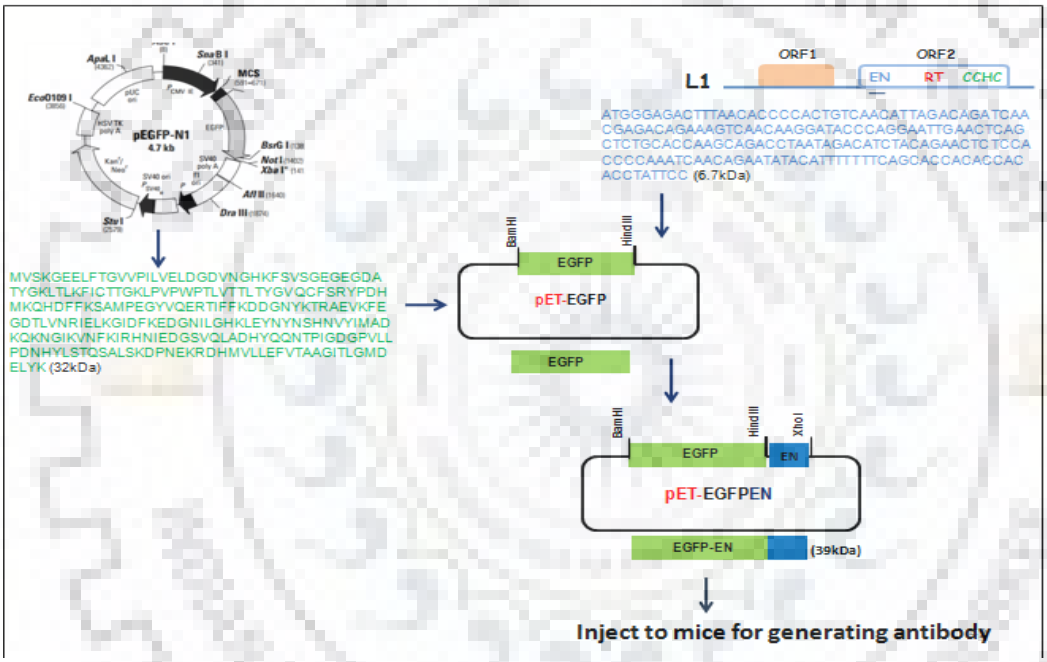


Figure R18: Strategy of making EGFP-EN fusion construct (i) Subcloning of EGFP gene from N1EGFP to pET30b followed by cloning of part of EN domain at the C-terminus of EGFP in same frame.

3.7.1. Cloning of EGFP in pET30b

To make EGFP-EN clone, the EGFP fragment was cloned first at the BamHI and HindIII sites of pET30b. The cloning is performed in such a way that EGFP coding sequence will not end with stop codon as portion of ORF2p EN sequence will be cloned next to it to make the fusion protein (EGFP-EN). For this purpose specific primers were designed to

amplify the EGFP coding part from p-EGFP N1 plasmid. The amplified fragment was cloned in pET30b bacterial expression vector. The clones were confirmed by digesting with BamHI and HindIII (Figure R20)

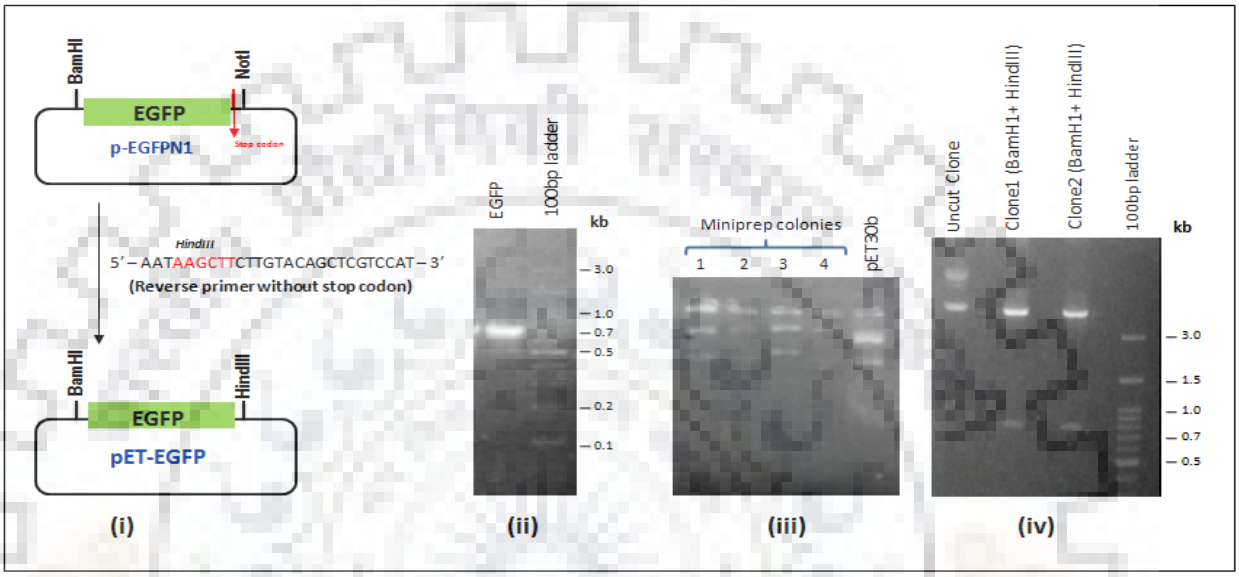


Figure R19: Cloning of EGFP fragment in bacterial expression vector (i) Schematic representation of subcloning of EGFP from pN1EGFP to pET-30b bacterial expression vector using specific primer set 4F/4R (Table 1). (ii) Primers were designed to amplify 0.75kb fragment. The PCR product (insert) of EGFP insert amplified from pEGFP-N1 was resolved in 1.0% agarose gel. (iii) Agarose gel electrophoresis to screen the mini prep clones (iv) Confirmation of Clone 1 and 2 by digesting with BamH1 and HindIII restriction enzymes showed 0.75 kb expected size of EGFP gene.

3.7.2. Expression of pET30b-EGFP

The two positive clones were transformed in BL-21 *E.coli* strain for protein expression. The protein was expressed by inducing at 0.4 O.D using 0.4mM IPTG along with uninduced. Surprisingly, the clone 2 didn't show any fluorescence suggest that either the protein is not expressing or a mutant version of EGFP protein which does not have any fluorescence is expressed. The whole cell lysate was prepared from both the clones and analyzed by SDS-PAGE gel electrophoresis. The result showed that a band of around 33 kDa corresponds to the EGFP molecular weight was present in both the lysate. This

suggests that there is some mutation/s introduced at crucial amino acid/s required for EGFP fluorescence in clone 2. To know about the mutation the clone 2 was send for Sanger sequencing. The result showed two nucleotides mutation at position 275 and 276 changed the amino acid Glycine to Alanine at 92 position. The fluorescence of wild type GFP depends on the formation of fluorophore by three amino acids Serine, Threonine and Glycine located at position 65, 66 and 67. The mutation in clone 2 which abolished the fluorescence is at position 92 which is situated far from the chromophore forming amino acids (Position 65, 66, 67). It is also important to mention that the role of this amino acid in the GFP fluorescence has not been explored yet.

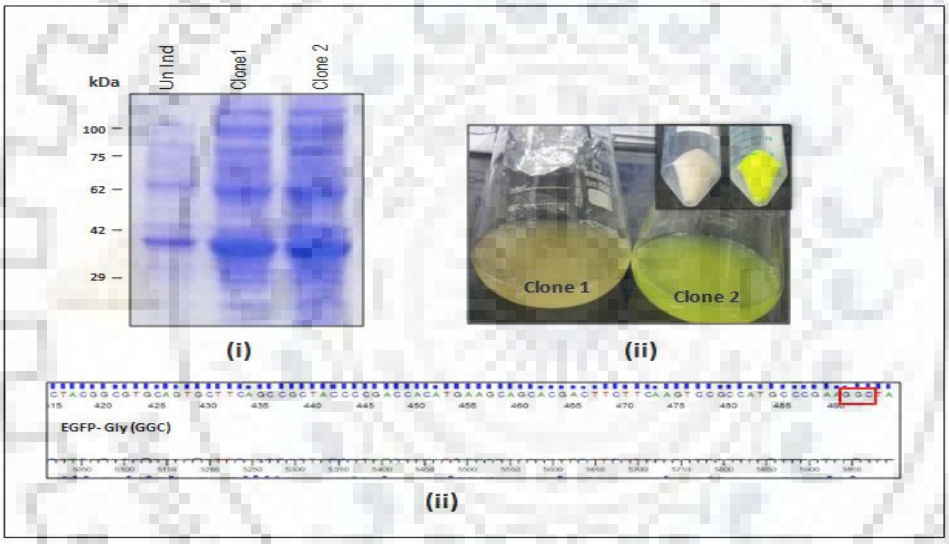


Figure R20: (i) SDS-PAGE showing protein expression in both clones 1 and 2 along with un-induced control. Both clones expressed equal amount of protein. (ii) Clone 1 induced culture didn't show any fluorescence. (iii) Sequencing results of clone 1 showed mutation at nucleotides at 275and 276 (which changed glycine to Alanine at position 92.

The *E.coli* cells expressing the mutant protein (pET-EGFP-G92A) didn't show any fluorescence when observed under fluorescence microscope. The E. coli cells carrying pET30b (vector only) and pET-EGFP was considered as negative and positive control respectively (Figure 21).

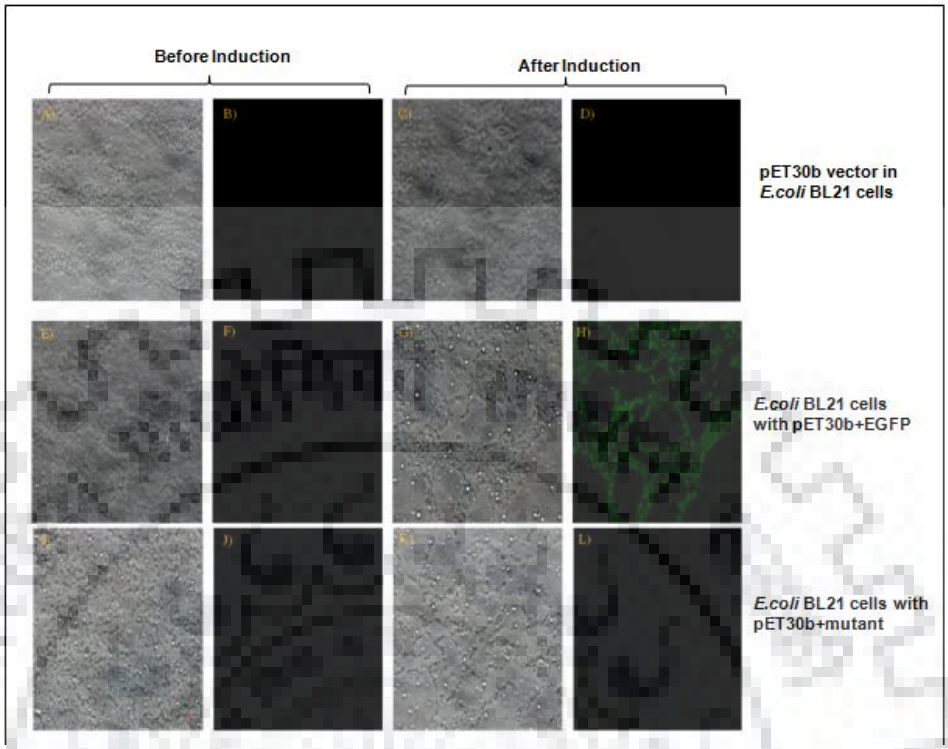


Figure R21: A,C,E,G,I,K- Phase contrast microscopy pictures. B,D,F,H,J,L- Fluorescence microscopy pictures. A-D: Negative control: E.coli BL21 carrying pET30b vector. E-H: E.coli BL21 carrying pET30b-EGFP. I-L: E.coli BL21 carrying pET30b-EGFP Mutant. A, B, E,F,I,J: Before induction. C,D,G,H,K.L: After Induction.

3.7.3. Mutation of Glycine at 92 to Asp, Arg and Tyrosine to study the effect on fluorescence of EGFP

Next to see the effect of other amino acids at position 92 glycine in wild type was changed to acidic (Aspartic acid), basic (Arginine) and aromatic (Tyrosine) amino acids to see the effect on EGFP fluorescence. The mutation was incorporated by site directed mutagenesis (SDM) using primer set where the wild type codon was replaced (Figure R22; panel i). The mutated clones were confirmed by Sanger sequencing (Figure R22; panel ii).

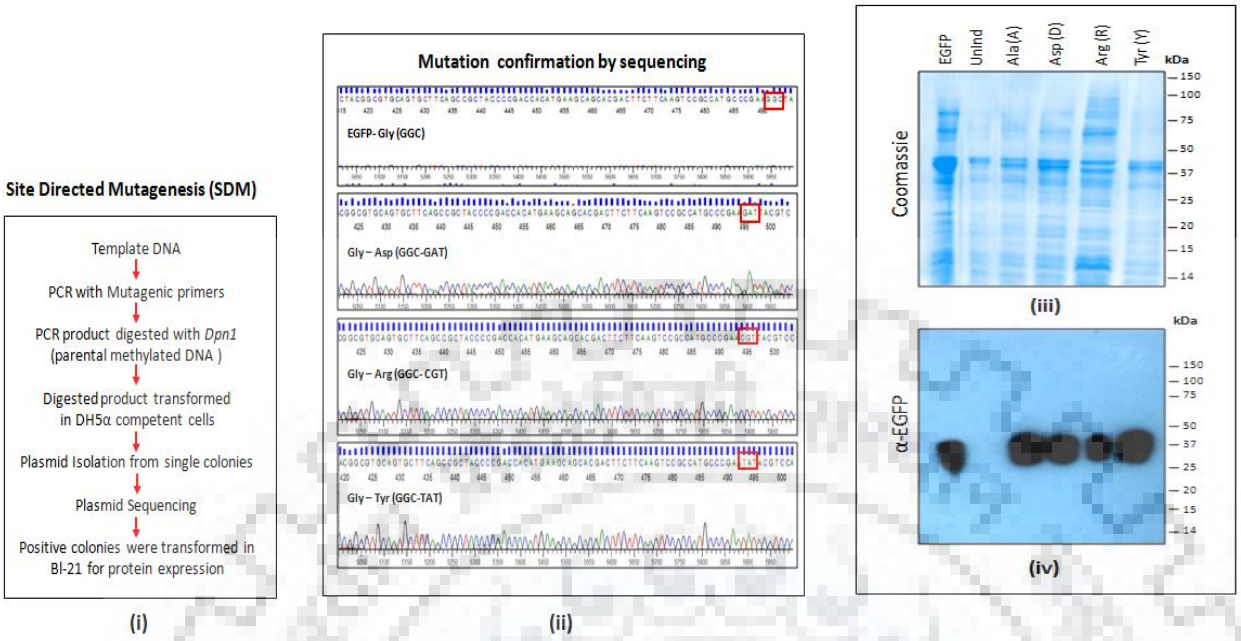


Figure R22: (i) Schematic representation showing procedure of site directed mutagenesis. (ii) Chromatogram showing mutation at 92 position where wild type glycine changed to Ala (GGC – GGC), Asp (GGC – GAT), Arg (GGC – CGT) and Tyr (GGC – TAT). (Iii). Loading control. UnINd: un-induced total lysate. (iv) Confirmation of EGFP (1:10000) mutated proteins by Western blotting.

The expression of the protein was also confirmed by Western blotting using anti-EGFP antibody (Figure R23). Total lysate from all mutated clones (pET-EGFP$^{92Gly\text{-}Asp}$, pET-EGFP$^{92Gly\text{-}Arg}$ and pET-EGFP$^{92Gly\text{-}Tyr}$) along with wild type EGFP was resolved in 10% SDS-PAGE gel and Western blot was performed with anti EGFP antibody (Figure R22; panel iv) . A single band at around 32 kDa in all 4 clones and also in the wild type control confirms their identity. The bacterial cells expressing those mutant EGFP proteins didn't show any fluorescence. All three mutated clones (pET-EGFP$^{92Gly\text{-}Asp}$, pET-EGFP$^{92Gly\text{-}Arg}$ and pET-EGFP$^{92Gly\text{-}Tyr}$) were checked for their expression in bacterial expression cells. The data showed that all three clones expressed in bacterial expression cells at 37˚C. The induced protein showed its presence both in supernatant as well as in pellet. The supernatant fraction was purified using Ni-agarose affinity chromatography as those proteins contain his tag at their N- terminal end. The data showed all three mutants are purified in homogeneity, although yields for pET-EGFP$^{92Gly\text{-}Tyr}$ clone were less in Figure R23.

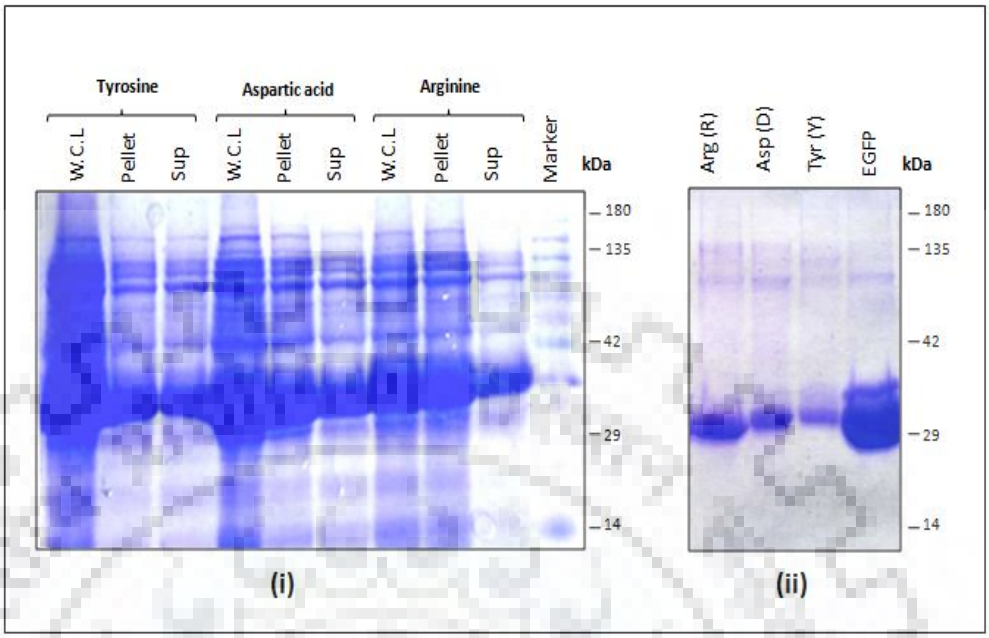


Figure R23: Expression and Purification of Mutant proteins. (i) Whole cell lysate (WCL), pellet and sup for all mutants were resolve in 12% SDS-PAGE gel after expressing at 16°C (ii) Affinity purified mutant proteins were analyzed in 12 % SDS-PAGE gel.

3.7.4. Biophysical Characterization of Mutated proteins

3.7.4.1. UV absorption spectroscopy

The EGFP has maximum absorbance at 490nm along with minor peak at 393nm. The same was observed for wild type EGFP in UV absorption spectroscopy. On the other hand all the mutants failed to show any absorption at 490 nm suggesting that fluorophore has not formed in those proteins. This also suggests that mutant proteins are not folded properly to form a twelve strand beta barrel structure which formed a hydrophobic core for the formation of fluorophore. The absorption peak at 280 nm for wild type and all mutants confirmed the presence of enough protein in all the samples employed for absorption spectroscopy. (Figure R24)

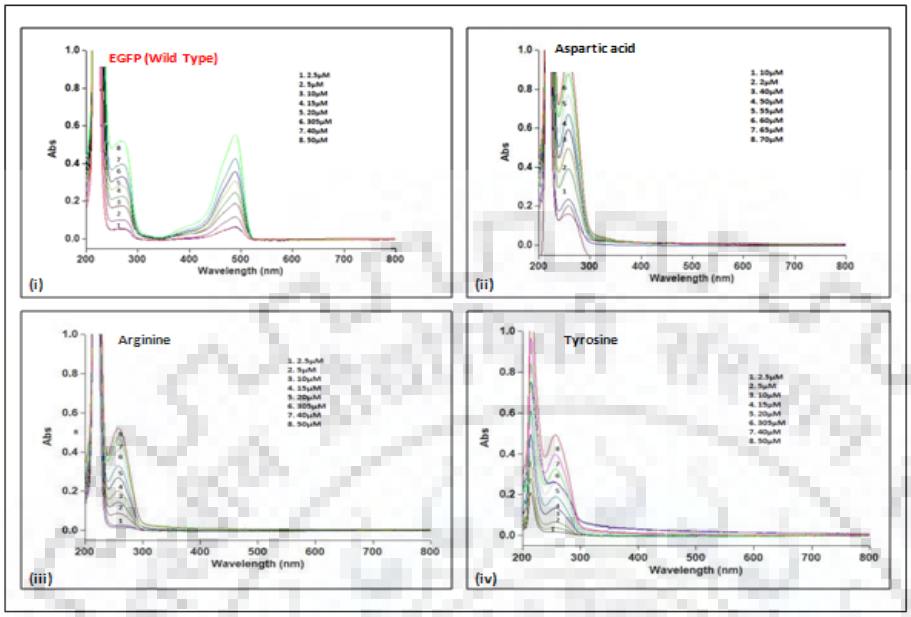


Figure R24: UV – Vis absorption spectra of EGFP and mutated proteins at 280nm with protein concentrations ranging from 2.5µM to 50µM. (i) UV absorption spectra of EGFP at 280nm, a specific absorption peak observed at 488nm conferred by mature chromophore responsible for fluorescence. (ii) UV absorbance spectra of Asp (D) (ii), Arg (R) (iii) and Tyr (R) (iv) where glycine was changed to Aspartic acid, Arginine and Tyrosine respectively. No absorption peak at 488nm reveals the absence of mature chromophore in those mutants.

3.7.4.2. Fluorescence spectroscopy

Next fluorescence spectroscopy of wild and mutant proteins was performed to study emission spectrum of EGFP and mutant proteins. At the same time I have checked if pH has any role in fluorescence of mutant proteins. Equal amount (10µM) of wild type and mutants, was dissolved in Tris-Cl buffer pH ranging from pH 5.0, 6.0, 7.0, 7.5 and pH 8.0 and incubated for 2 hrs at room temperature before recording the spectra. All the proteins were excited at 490nm and the emission spectra was recorded wavelength between 500nm to 700nm. Distinctive emission spectrum was observed at 560nm for wild type EGFP. No emission spectrum was observed for mutant proteins at any given pH (pH 5.0, 6.0, 7.0, 7.5 and pH 8.0). This suggests that mutated proteins probably did not absorb light at 490nm thus didn't show any emission spectra in different pH solution. This again suggests that mature fluorophore was not formed and thus no fluorescence observed in

mutated proteins. The spectra analysis also depicts that the mutant proteins did not fold into proper 3D structure even at other pH conditions and thus there was no fluorescence emission. The emission spectra of all mutated proteins along with wild type were recorded and shown in Figure R25.

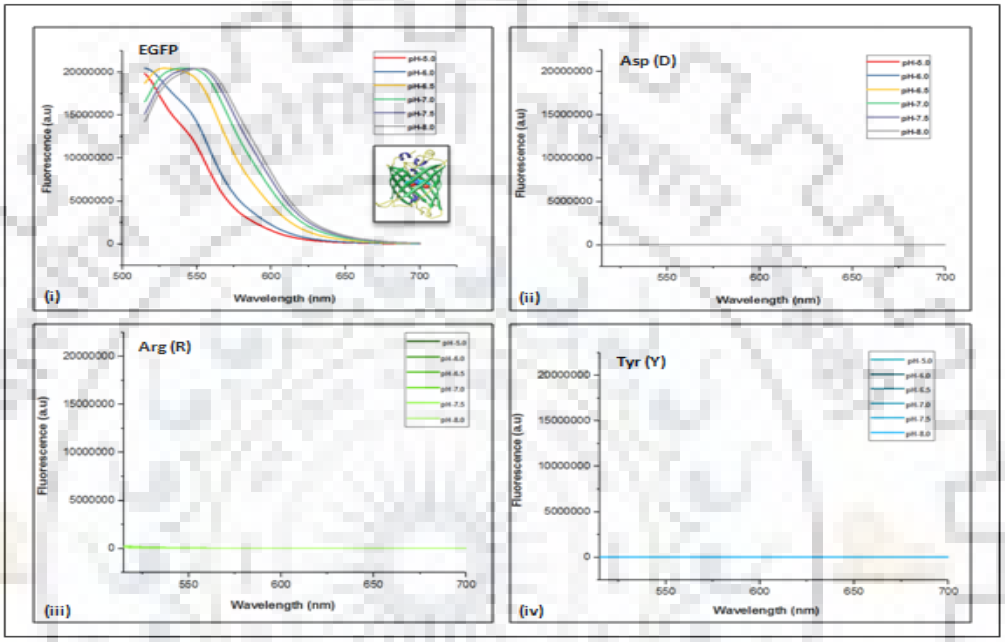


Figure R25: Fluorescence spectra of EGFP and mutated proteins excited at 490nm with protein concentration of 10µM at different pH ranging from 5.0 to 8.0. (i) Emission spectra of EGFP excited at 490nm, a specific emission peak observed at 560nm conferred by mature chromophore responsible for fluorescence. (ii) Emission spectra of of Asp (D) (ii), Arg (R) (iii) and Tyr (R) (iv) where glycine was changed to Aspartic acid, Arginine and Tyrosin respectively. No Emission peak at 560nm reveals the absence of mature chromophore in those mutants which is responsible for fluorescence

3.7.4.3. Fluorescence spectroscopy at 274 and 280nm.

Fluorescence based studies have been widely used to examine the interaction of a protein with different ligands (drugs, inhibitors etc). The aromatic amino acids (Trp, Phe and Tyr) residues in the proteins contribute to the intrinsic florescence. The fluorescent measurements generally provide basic information about the conformational changes and environment around the fluorophore. The fluorescence properties of EGFP were also analysed by fluorescence spectroscopy as explained in methods section. The protein was

checked for fluorescence emission at two different wavelengths 280 nm and 295 nm to check its properties. A foresaid red shift increased indicated that more exposure and interaction of Trp residues occurred with the fluorescence spectroscopy provides basic information about the conformational changes and environment around the fluorophore. Red shift in emission maximum indicates increased hydrophilic environment there by exposing the hydrophobic core of β- barrel which is responsible for structural dynamics and providing hydrophobic environment which is very much essential for chromophore formation and maturation there by leading to fluorescence as tabulated in given table. In case of mutant proteins the spectra had been shifted and increase of wavelength for sure reveals the unfolding of protein dynamics compared to wild type. Both the spectra of wild type with mutants were shown in Figure R26.

Sample	λ Excitation	λ Emission (max)	Wavelength (nm)	Shift in λ E(max)
EGFP-Gly	280			
EGFP- G92D	280	347	Increased	Red Shift
EGFP- G92R	295	347	Increased	Red Shift
EGFP- G92Y	280	347	Increased	Red Shift

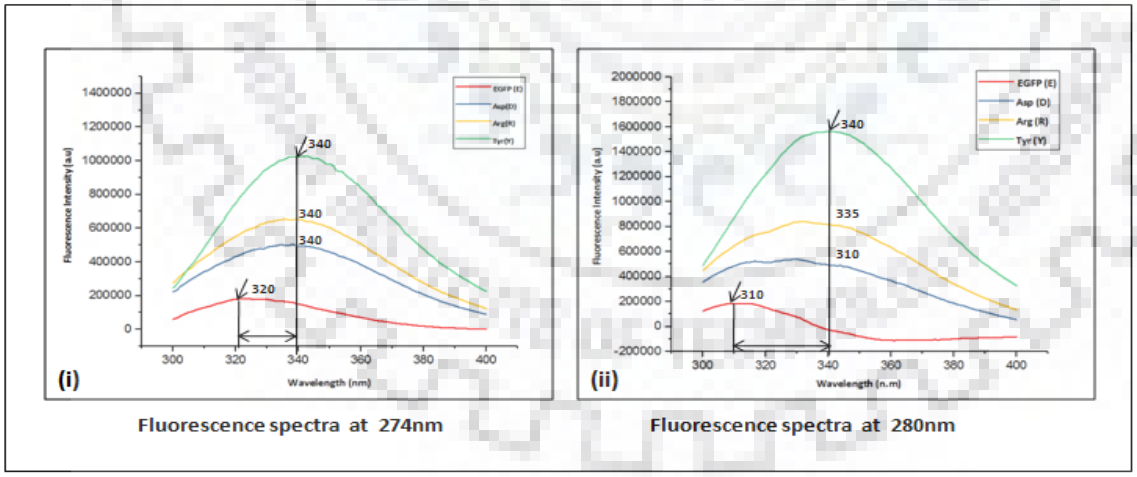
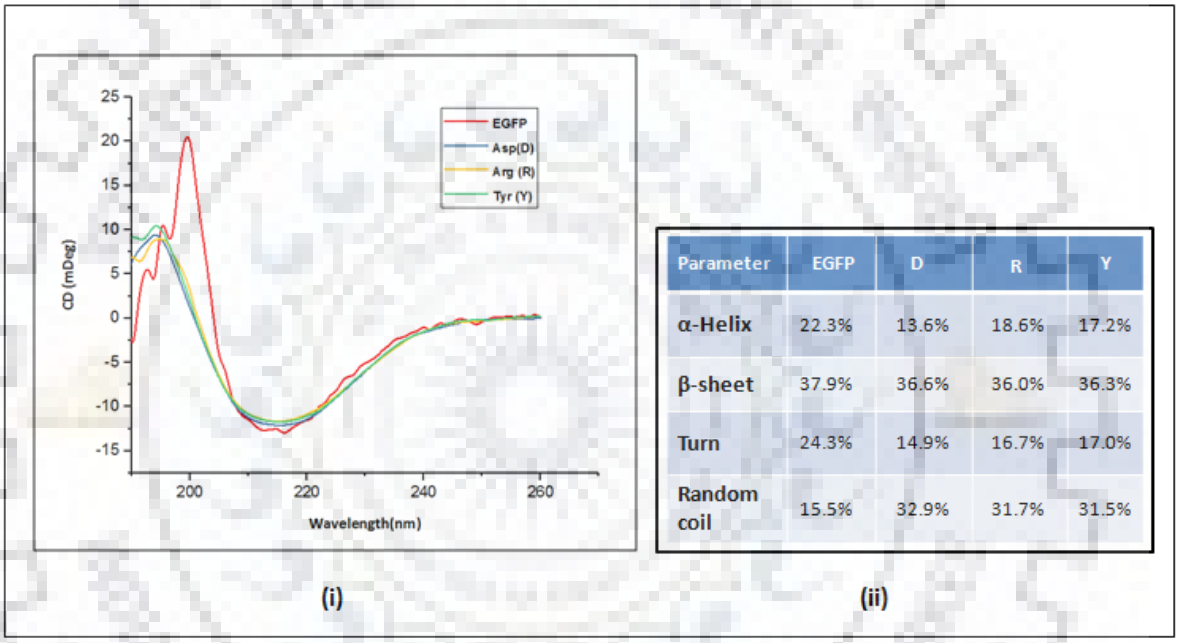


Figure R26: Fluorescence spectra of EGFP and mutated proteins excited at 280nm and 274nm with protein concentration of 10µM. (i) Emission spectra of EGFP and mutated proteins excited at 274 and 280 nm, a shift was found in case of mutants the spectra had a shift of 20nm due to increase in wavelength. This shift is called Red shift indicates the change in chromophore environment and increase of hydrophobicity.

3.7.4.4. CD (Circular Dichroism) Spectroscopy of EGFP and Mutant protein

Next CD spectroscopy was performed to analyze the effect of single aminoacid mutation at position 92 in secondary structure between EGFP and mutant proteins. The result (Figure R27) shows that there is significant difference in secondary structure, particularly at α-helix and β-sheets between wild type and mutant proteins. The predicted values of α-helix, β-sheets and random coil were tabulated along with the spectra.



Parameter	EGFP	D	R	Y
α-Helix	22.3%	13.6%	18.6%	17.2%
β-sheet	37.9%	36.6%	36.0%	36.3%
Turn	24.3%	14.9%	16.7%	17.0%
Random coil	15.5%	32.9%	31.7%	31.5%

Figure R27: CD (Circular Dichroism) spectra of EGFP and mutated proteins scanned from 190nm and 280nm with protein concentration of 20µM. (i) CD spectra of EGFP and mutated proteins. (ii) Secondary structure predicted values of α-helix, β-sheets and random coil corresponds to EGFP and mutated proteins.

3.8. To generate In house antibody against human LINE-1 ORF2p protein using EGFP tagged EN peptide (fusion protein)

3.8.1. Cloning of EN in pET30b+EGFP

A stretch of 180bp from ORF2 endonuclease (EN) domain (amino acid 143-202 of ORF2p) was PCR amplified using primer sets 4F/4R (Table 1) and template JCC5 (L1RP cloned in pBSKS+) and cloned in fusion with EGFP to make expression construct named pET-EGFP-EN. The clone was confirmed by restriction digestion (Figure R28).

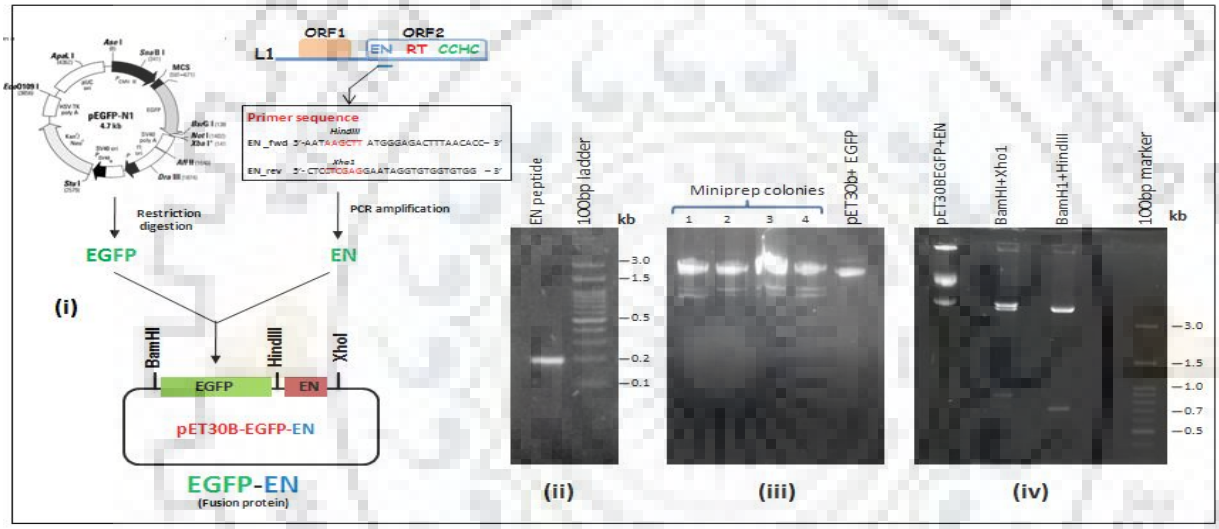


Figure R28: Cloning of EN peptide fragment in bacterial expression vector pET30b+EGFP
(i) Schematic representation of EGFP gene (720bp) from pEGFP-N1 and EN peptide from pBS-L1RP. A 180bp EN fragment was amplified using specific primers and cloned at the C-terminus of EGFP to get pET-EGFP-EN clone. (ii) EN peptide amplicon resolved in 1.2% agarose gel with 100bp DNA marker. (iii) Agarose gel electrophoresis to screen the mini prep clones. (iv)Confirmation of Clone by digesting with BamH1 and HindIII for 720bp fragment of EGFP and with HindIII and Xho1 to get EGFP+EN fragment of 930 bp.

3.8.2. Expression of EGFP+EN (fusion protein)

The positive clone pET-EGFP-EN was transformed in to E. coli expression cells (BL-21 strain) for protein production. The protein was expressed at 37C with 0.4 mM IPTG for 3hours. The induced lysate showed the expression of EGFP-EN with expected Mol wt of

around 39kDa when resolved in SDS-PAGE gel. The induced total lysate was separated to sup and pellet and resolved in SDS-PAGE. The result showed that induced EGFP-EN is present in equal amounts in both sup and pellet fraction.

3.8.3. Purification of EGFP-EN fusion protein from soluble fraction

The EGFP has 6XHis tagged sequence at the N-terminal, thus nickel agarose chromatography was employed to purify the EGFP-EN. The separation of total lysate into supernatant and pellet fraction showed almost 50 percent of protein was present in supernatant fraction, thus supernatant was used to purify EGFP-En using Ni-agarose chromatography. The supernatant was incubated with Ni-NTA beads for 1 hour at 4°C and then the bound protein was eluted with elution buffer containing 250mM Imidazole. The SDS PAGE analysis of eluted fraction showed non-specific bands as impurities at higher molecular mass along with specific protein (Figure R29; panel i)

3.8.4. Purification of EGFP+EN (fusion protein) from inclusion bodies

As the purified of EGFP+EN from sup fraction showed impurities, next pellet fraction was used to purify the EGFP+EN. Accordingly the pellet was first dissolved in buffer containing 8M urea and then affinity chromatography was performed using Ni-NTA beads. The protein was eluted using elution buffer contain 300 mM Imidazole in 8M urea buffer. The SDS-PAGE gel of purified protein showed that the protein is more than 90% pure with minimum contaminant bands (Figure R30; panel ii). As the purified EGFP-EN protein was in urea solution and thus was in denatured condition, the refolding of protein was performed by gradual removal of urea using dialysis. The purified protein was diluted and dialysis was performed in step to decrease the concentration of urea from 8M to 0M. Finally the dialyzed protein was concentrated and checked in SDS-PAGE gel (Figure R29; panel iii). The result showed a single purified band and thus purified protein can be used as an antigen for generating antibody against ORF2p.

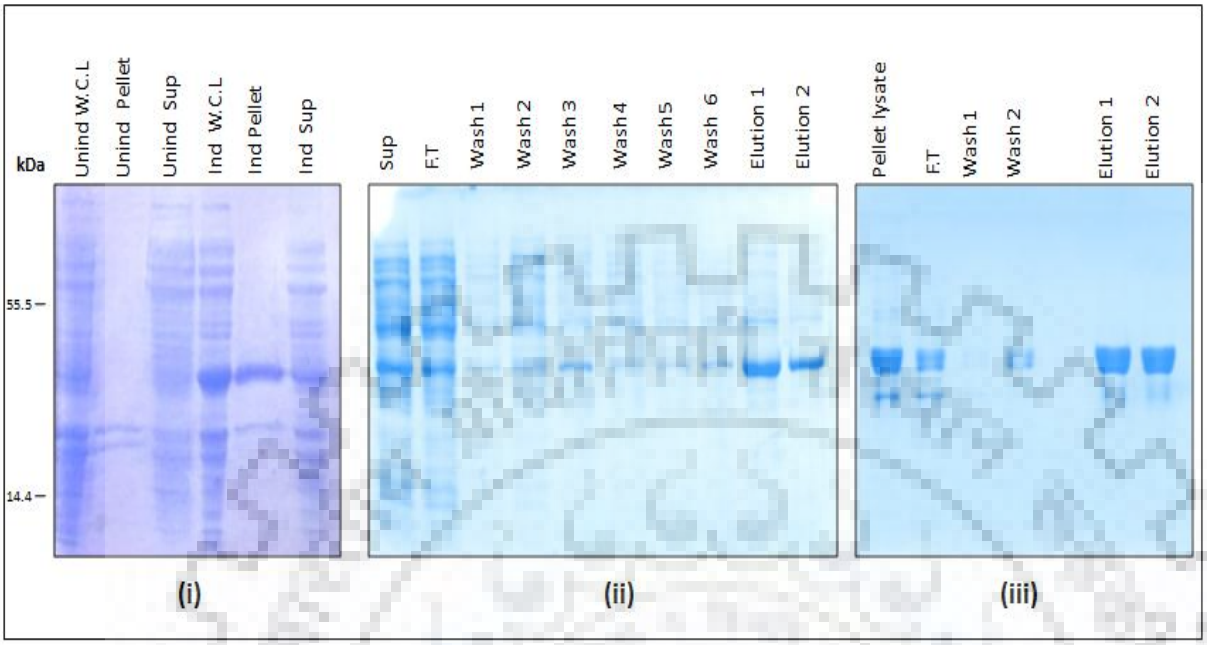


Figure R29: Expression and Purification from soluble and insoluble fractions of pET-EGFP+EN. (ii) Purification of fusion protein from soluble fraction. Lane 1- Soluble fraction taken for purification followed by flow through, then was washed with 50mM Imidazole (lanes 3,4 and 5). Wash 4,5 and 6 represents wash with 100mM Imidazole (lanes 6, 7 and 8). Elution 1 and 2 show higher molecular weight impurities along with protein of interest. (iii) Lane 1- protein extracted from insoluble fraction using 8M urea, lane 2-flow through after binding with Ni-NTA beads, lane 3 and 4 contains wash1 with 50mM and wash 2 100mM Imidazole. Lane 6 and 7 contains eluted samples after concentrating the protein.

3.8.5. Generation of EGFP-EN antibody and its validation

The fusion protein of EGFP+EN was now injected to mice for antibody production using the standard procedure at a concentration of 50µg/mice. After 21 days the first bleed was collected and the serum was separated from the blood isolated used as antibody. The serum isolated after 21 days (1st bleed) was check for antibody by employing dot blot. EGFP protein was spotted in the blot showed signal when serum was used a primary antibody. The spotted BSA didn't show any signal. This suggests that EGFP-EN is immunogenic in mice (Figure R30; panel i). Next, the same first bleed was checked by western blotting. 99RP-EGFP transfected 293T lysate was resolved in SDS-PAGE gel

and Western blot was performed to detect exogenous ORF2p. Bacterial induced total lysate expressing induced EGFP-EN was used as control. The blot didn't show any band of L1ORF2p at around 150 kDa in lane marked 99RP-EGFP.However the antibody can readily detect EGFP protein in the same sample (Figure R30; panel iii). It is possible that in the first bleed the titer of EN specific antibody is very less and thus unable to detect ORF2p. So mice were administered second and third booster antigen and wait for the 80 days to get the final bleed. Meanwhile an ORF2p antibody was available in the laboratory and was used to check its expression in OSCC samples

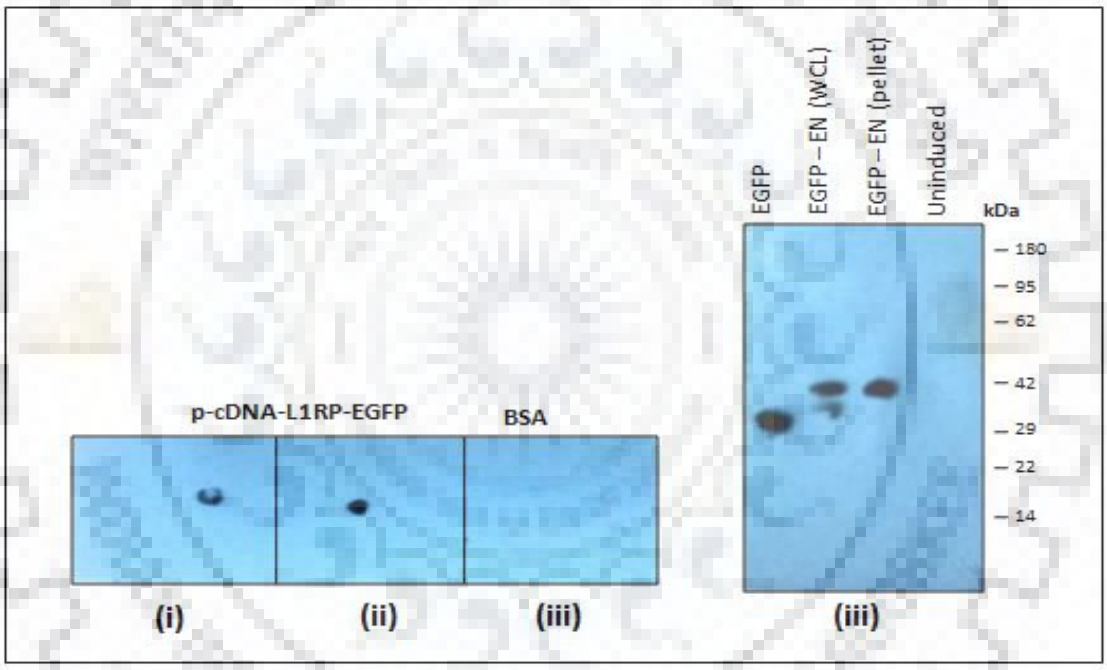


Figure R30: Validation of generated antibody against fusion protein (i) Dot blot showing transfected lysate of p-cDNA-L1RP-EGFP stained with α-ORF2 antibody gives immune response at 1:5000 dilution. (ii) Immune response of transfected p-cDNA-L1RP-EGFP cell lysate with generated antibody at dilution rate of 1:1000. (iii) Validation of final bleed using immunoblotting of EGFP and EGFP-EN with uninduced gave specific immune response and a distinct shift in case of EGFP-EN due to increased molecular mass of 7kDa due to EN peptide.

Objective 3

3.9. To investigate the human L1ORF2p expression profile in OSCC samples

3.9.1. Detection of L1ORF2p in OSCC samples

This study included a total number of 12operated sample from OSCC patients aged between 23 and 82 years, with oral squamous cell carcinoma. First, the histopathological study was performed by staining all samples with Hematoxylin-Eosin (Figure R31). The H and E staining showed neoplastic nature of the tissues with diffused nuclei; aberrant nuclear cytoplasmic ratio and distinct pearl like structure (Figure R31; panel C11).

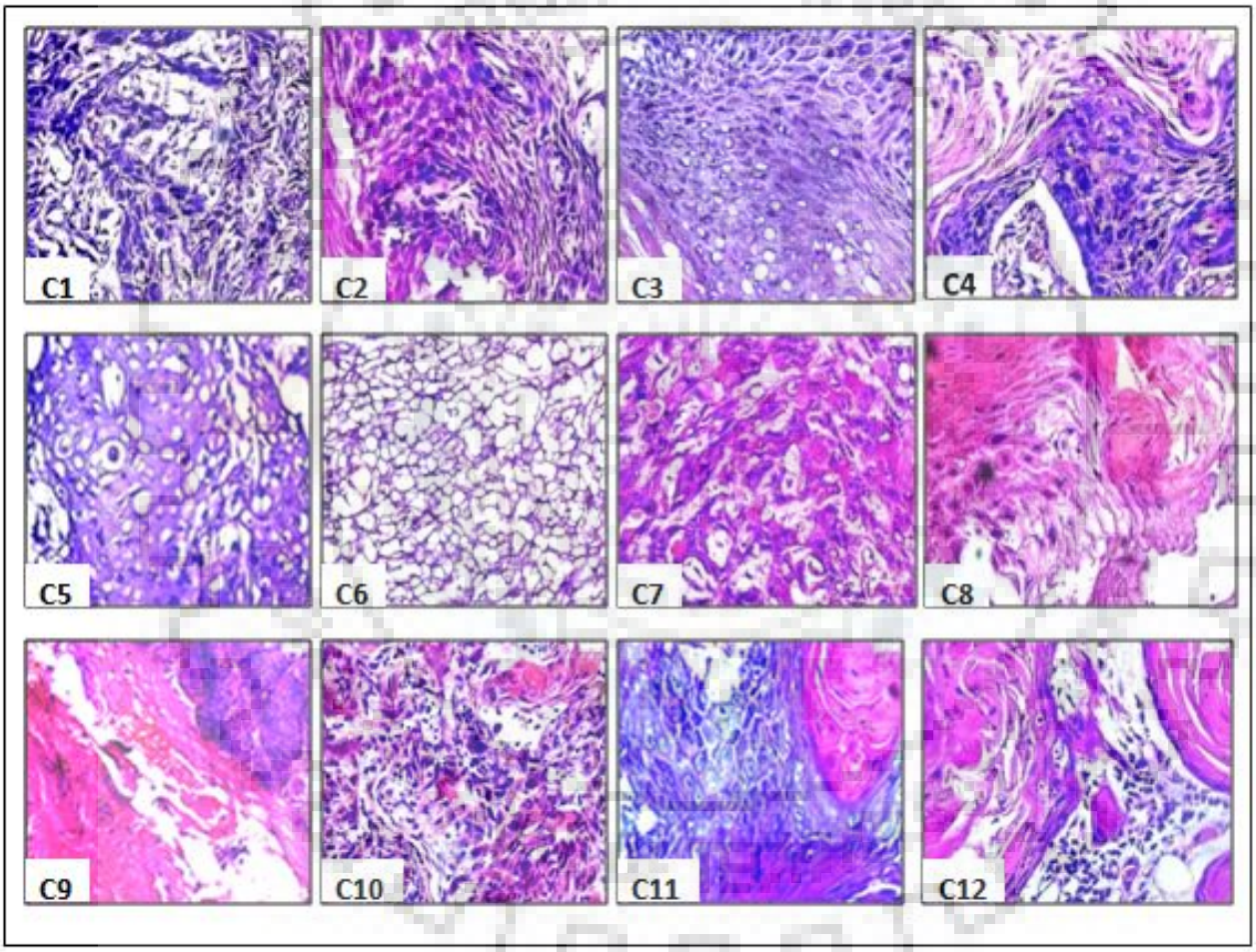


Figure R31: Histopathological Studies of collected Oral Squamous Cell Carcinoma tissue samples using Hematoxylin and Eosin staining. Hematoxylin is basic and positively charged dye that binds to negatively charged nucleic acids like DNA in nuclei, gives purple or blue color represents nuclear part of cell. Pink staining is due to Eosin as is acidic, negatively charged tend to attract cationic tissue sites like cytoplasm where proteins and other molecules are stained.

Next, all the OSCC samples were screened for the expression L1-ORF1p. Recently, our laboratory made an in-house L1 ORF1p antibody and the same antibody was used to detect ORF1p in those cancer samples (Sur et al. 2017). The IHC result showed that 8 out of 12 samples showed the presence of L1 ORF1p.

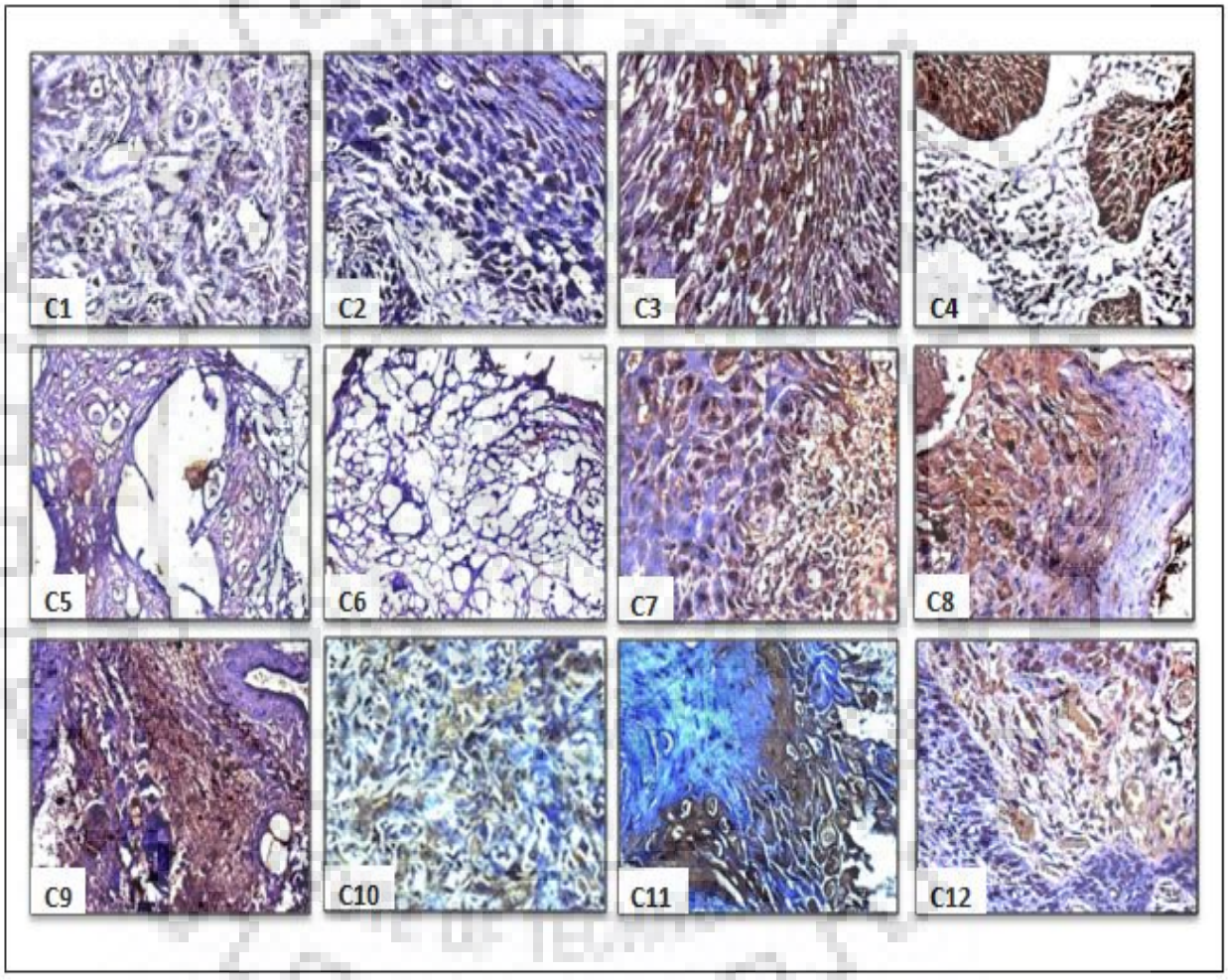


Figure R32: Immunohistochemistry of total twelve OSCC tissue samples with anti-ORF1p antibody to detect human ORF1p expressions. C3, C4, C7, C8, C9, C10, C11 and C12 samples were shown higher amount of expression when compared to other samples.

The samples which showed intense staining with anti-hL1-ORF1p were further checked for hL1-ORF2p expression. Recently, our lab has generated an in-house hL1-ORF2p antibody which showed good reactivity with human L1-ORF2p protein (Mukherjee et al, manuscript under preparation). The IHC study showed 5 out of 7 samples (C3, C4, C8, C9 and C12) was hL1-ORF2p positive. The staining of samples with CK-19 antibody was used as positive control and samples without incubation with primary antibody was treated as negative control. The detailed information regarding patients and collected samples are shown in table 2.

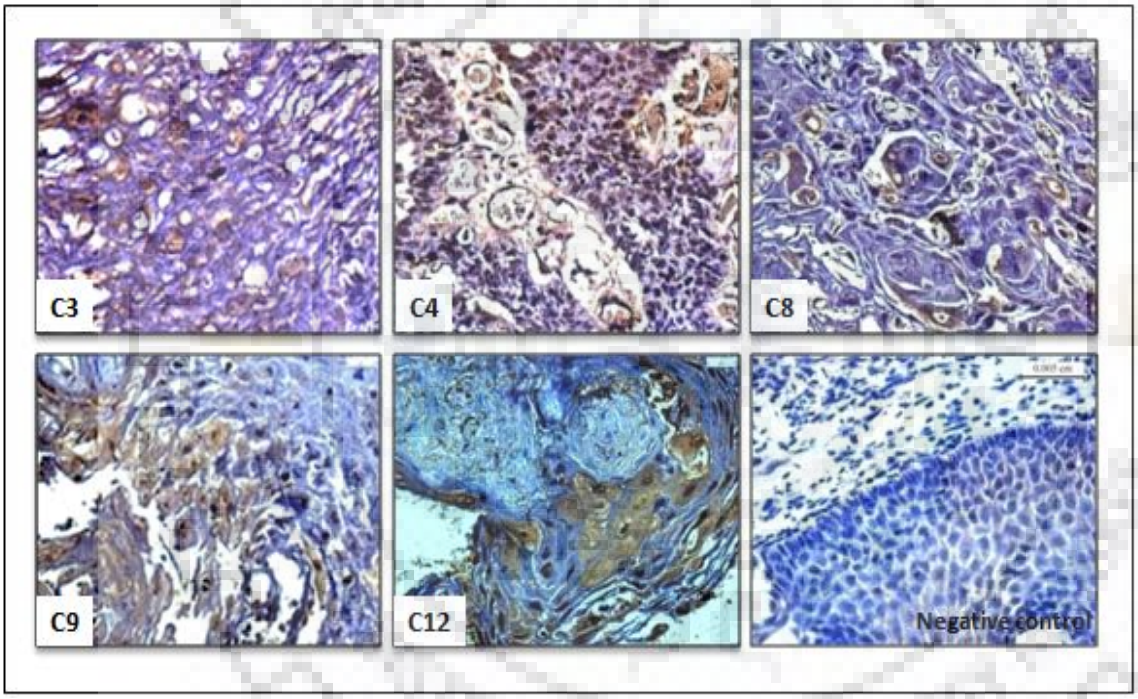


Figure R33: Immunohistochemistry of seven OSCC samples which are found positive for hORF1p, screened with in-house anti-hL1ORF2p antibody to detect human L1-ORF2p expression. Samples C3, C4, C8, C9, and C12 were shown higher amount of expression when compared to other samples. The results show the correlation between expression levels of ORF2p and ORF1p.

Chapter 4

Discussions

Expression of recombinant human L1ORF2p domains in bacteria

(i) C-terminal CCHC domain of human L1-ORF2p

Human L1 ORF2p (1275aa, MW around 150 kDa) [Ergun et al. 2004; Martin et al. 2001; Taylor et al. 2013 and Goodier et al. 2010] is the key protein required in the process of retrotransposition. It has three domains: an AP-like EN domain [Feng et al.1996] at the N terminal (1-239aa), Central reverse transcriptase (RT) domain [Mathias et al. 19991] and C terminal CCHC domain [Fanning and Singer 1987] all three domains are critical in the process of retrotransposition [Moran et al. 1996 and Feng et al. 1996]. Detection and purification of full length L1ORF2p is extremely tough both in-vivo and in-vitro. The N-terminal EN domain has been characterized in detail as this domain is expressed in bacterial expression system. However the details on central RT and C-terminal CCHC domains are not available because of difficulties in expression and purification of these domains. Although few studies are available on RT domain characterization [Mathias et al. 1991], regarding the characterization of CCHC domain, only one study is available [Piskareva et al., 2013]. To characterize the CCHC domain amino acids stretch 1096-1275 of hL1ORF2 was cloned in pET32 bacterial vector which produced thioredoxin-CCHC fusion protein of around 40 kDa with 6 His tag (pET-TRX-CTS). Purified TRX-CTS fusion protein showed non-specific RNA binding activities and Zn-knuckle structure does not affect the RNA binding. The CCHC domain present in all mammalian L1 elements, retroviral nucleocapsid protein and in other proteins that bind single stranded RNA. This observation suggests this domain has an important role in L1ORF2p -nucleic acid (NA) interaction. In addition to its role in NA binding, mutation analysis of retroviral CCHC domain suggests that this domain is important for reverse transcription, perhaps for unfolding structured RNA. Moran et al. [Moran et al.1996] showed that mutations of the key amino acids in the Zn knuckle structure ($C_{1143}xxxC_{1147}$ to $S_{1143}xxxS_{1147}$) completely abolished L1 retrotransposition in cell culture based retrotransposition assay. All these suggest that it is very important to study the CCHC domain in details. Here I have cloned the CCHC domain in pET30b expression vector. The recombinant protein has 14 and 33 amino acids extra from vector at the N-terminal and C- terminals respectively and the putative molecular weight is 19kDa. The cloned CCHC domain in bacterial expression vector

didn't get express using bacterial expression system. Previous studies showed that the human L1ORF2p expressed very less amount when expressed in bacteria. It is speculated that the ORF2p is toxic in E.coli because of its endonuclease activity. More studies are required for the cloning, expression and purification of CCHC domain.

(ii) N-terminal Endonuclease (EN) domain of hORF2p

The human L1ORF2p EN was the first retrotransposon domain to be isolated and studied. It was first identified in 1996 and also resembles APE (Apurinic/Apyrimidinic Endonuclease) which is primarily involved in DNA repair [Feng et al. 19996]. The EN domain of the human L1 ORF2p was further successfully crystallized and as hypothesized it is structurally similar to the APE [Weichenrieder et al. 2004]. The EN activity is known to required for efficient retrotransposition [Moran et al. 1996, Feng at al. 19996,Wei at al. 2001, Kulpa and Moran 2006]. However, in certain DNA repair-deficient cell lines, L1 retrotransposition can also take place in EN-independent manner [Morrish et al. 2002 and 2007]. In the same way Alu elements are also known to be able to retrotranspose in an EN-independent manner [Christian et al. 2016]. The EN domain is well conserved throughout variety of species; in contrary it can tolerate mutations in multiple conserved amino acids without significant loss in retrotransposition efficiency [Kines et al. 2016]. The ORF2p fragments that contain the active EN domain are found to be cytotoxic in mammalian cells through DNA double strand breaks (DSB) [Gasior et al. 2004; Kines et al. 2016; Hedges and Deininger 2007 and Wallace et al. 2008]. None of the EN protein expressed above was used for antibody production. Due to inability to express the human L1-ORF2p CCHC domain, next I have attempted to express human L1-ORF2p EN domain in bacterial expression system. The 730bp EN domain spanning 1-243 amino acids of ORF2p was cloned in pET28c bacterial expression vector. After induction with 0.4mM IPTG at 37C for 3 hours in *Escherichia coli* BL21 (DE3), a thin band at around 31kDa was observed when total lysate was analyzed in 12% SDS-PAGE gel. The 31 kDa band was also confirmed by Western blotting using anti-his antibody as the protein contains 6 amino acids histidine tag at the N-terminus. In general induction with 0.4-0.6 mM IPTG concentration at 37˚C for 3-4 hours is routinely used for expressing protein in bacterial system when IPTG inducible plasmid is used (for example pET-28c). In this condition,

the pET-EN showed very less expression, thus some other factors like different temperatures, IPTG concentration and different buffers were checked to see the effects on EN domain protein expression. In this study, I have used three different temperatures (16°C, 30°C and 37°C) and my data showed that the expression at different temperatures didn't influence the expression EN domain protein. The protein expression was also verified at different concentration of IPTG ranging from 0.02-0.8mM; none of them showed increased expression of EN domain protein. In some cases growth media used for expression of protein also affects the solubility of the protein [Singh et al., 2005]; therefore The EN domain was expressed in different types of media (for example SSC, and terrific broth). The expression analysis in all three media showed that the human EN domain protein expression was approximately same, but in case of terrific broth the expression was slightly higher than others when compared. The soluble fraction at 37°C was proceeded for affinity purification using Ni-NTA agarose beads. The elution fraction consists of a thin band corresponds to recombinant EN protein. But the amount of protein was very low and not enough for the production of antibody against then further we also tried to purify protein from inclusion bodies as expression was also observed in insoluble fraction. There are several reasons behind that expressed protein might form inclusion bodies in bacteria which are (a) high inducer condition (b) higher copy no of plasmid (c) expression under increased temperatures (d) strong promoter system and others [Singh et al., 2015; Upadhyay et al., 2016]. Being toxic to host organism, the expression of EN domain may result in very low levels and folding of the EN domain protein in bacterial system was not proper and thus, formed inclusion bodies. After purification from insoluble fraction using 8M urea using affinity chromatography, the eluted fraction contains larger amounts of EN proteins than soluble fraction. But during refolding by dialysis, the protein again formed inclusion bodies in dialysis tube. A truncated EN fragment was tried to express as complete EN domain might be toxic for bacteria. The induction didn't show significant expression in induced cell lysate.

Generation of Antibody against L1ORF2

The antibodies against human L1 proteins (ORF1p and ORF2p) are very important tolls to study the biology of L1 retrotransposon, especially to identify the expression

patterns of L1 encoded proteins in germ tissues, early embryonic stage, different anatomical regions of human brain and in different types of cancers. Very few laboratories have antibody against ORF2p and recently one antibody is commercially available [Ergün et al., 2004; Goodier et al., 2004; Sokolowski et al., 2014; De Luca et al. 2016 and Rangasamy D et al. 2012]. Although, two peptide antibodies reported by Goodier et al [Goodier et al. 2004] have showed significant detection of ORF2p in L1 transfected cells, the data showed that the antibody specially one which raised against the ORF2p C-terminal peptide is not very sensitive and also showed some extra bands apart from 150 kDa ORF2p. It is not clear whether those extra bands are truncated version of ORF2p or some other cellular proteins cross-react with ORF2p antibodies. Recently, De Luca et al. [De Luca et al. 2016] reported antibody against ORF2p which showed good reactivity in detecting ORF2p in mammalian cell lines and cancer tissues. As the expression of different domains of human L1ORF2p (EN and CCHC) in bacterial expression system was not successful, synthetic peptide from ORF2p sequence was tried to make human L1-ORF2p antibody. Online bio-informatics tool was used to find a suitable peptide from ORF2p sequence. The program selected many peptides, among them a 14 amino acid stretch from EN domain was picked up; the peptide was synthesized commercially conjugate with KLH carrier protein before injecting it to animals for antibody production. The inhouse antibody generation protocol was optimized by generating antibody against enhanced green fluorescent protein (EGFP) in rat. The EGFP was sub cloned and purified using bacterial expression vector. For this purpose EGFP was sub cloned in bacterial expression vector. The purified protein was then injected into SD rat for the generation of polyclonal antibody. The raised antibody was validated using dot and Western blotting techniques. After successful standardization of protocol with EGFP, human L1-ORF2p peptide-carrier protein-conjugate was injected to mice for the generation of antibody against hL1ORF2. The final bleed was validated using exogenous hL1-ORF2p where an over-expressed hL1-ORF2p clone was transfected to HEK293T cells [Mandal et al, 2013]. Although a specific band at around 150 kDa corresponds to human L1ORF2p molecular mass was detected, the antibody showed some non-specific bands. In order to avoid non specific results in further experiments, we decided to use EGFP as carrier protein for human L1-ORF2p antibody production. EGFP is well known for its robust

expression both in bacterial and mammalian hosts and has many examples where GFP is used a fusion protein. Recently GFP fused L1ORF0 enabled detection of endogenous ORF0 in induced pluripotent stem (iPS) cells of primates [Denli et al. 2015]. A 60 amino acids stretch from ORF2 EN domain was fused at the C-terminal of EGFP to make construct pET-EGFP-EN. The fusion protein EGFP-EN was expressed using bacterial expression system and purified by affinity chromatography. The purified protein was injected to mice for the generation of antibody. The first bleed showed good antibody response against EGFP. However the antibody didn't detect exogenous ORF2p. It's possible the antibody titer is less in first bleed and thus same experiment will be performed using fourth bleed which will be collected soon.

Glycine 92 is crucial for the Fluorescence of Enhanced Green Fluorescent Protein (EGFP)

During cloning of EGFP-EN construct a unique EGFP mutant (Gly92Ala) was found which didn't show any fluorescence. Wild type GFP folds in a beta barrel structure and formed a chromophore using amino acids 65-67 which are Ser-Tyr-Gly. In EGFP Ser 65 changed to Thr 65 which gives five-fold more fluorescence than wild type GFP. Apart from these three amino acids several others have been reported which influence chromophore formation and those are Gln69, Arg96, His148, Thr203, Ser 205, Glu222 and few others [Fu JL et al., 2015; Stepanenko et al., 2008; Campanini et al., 2013; Kummer et al., 2000; Shu X et al., 2007 and Arpino et al., 2012]. The data showed that Gly 92 which is present in the loop between beta strand 3 and the central alpha helix is very crucial for the formation of chromophore and was not reported before. GFP with single amino acid substitutions can be used as biosensors if the fluorescence reverts back in a particular environmental change like pH, and temperature. Accordingly, the fluorescence of the mutant protein (Gly92Ala) was measured at different temperature as well as at different pH. The chromophore of the wild type EGFP shows excitation and emission peak at 395 nm and at 508 nm respectively. The mutant protein didn't show any excitation and emission at any particular temperature and pH suggesting that mature chromophore is not formed in the mutant protein. In-order to know whether any other amino acid can replace Gly at that position, the Gly was replaced with amino acids Aspartic acid, Arginine and Tyrosine. All three mutants were failed to show any fluorescence suggesting that Gly92 is conserved at that particular position of EGFP.

Detection of hL1ORF2p expression in OSCC

Many findings reveal that L1 retrotransposition is quite common in human cancers [Lee et al. 2012; Solyom et al. 2012; Iskow et al. 2010 Rodic et al. 2014]. Likewise, it has been shown that about half of common cancers express human L1 ORF1p [Stransky et al. 2011]. Here, we build on these studies by characterizing L1 ORF2p expression in OSCC samples. In addition to a dearth of data for L1 expression in OSCC, we chose to focus on this subtype of head and neck cancer because of its high prevalence, particularly among Indian patients [International Agency for research on Cancer, WCR (world cancer research) and WHO (world health organization) report 2014, [WHO, Oral Health Publications].Here, I studied 12 OSCC samples obtained from Acharya Tulsi Cancer Hospital located at Bikaner, Rajasthan, India. Then samples were first screened for the presence of L1ORF1p and the results revealed that ~60% (7/12) are positive for L1 ORF1p expression. These data are in agreement with a previous characterization of head and neck cancer samples distinct from OSCC which identified ~61% positive for ORF1p [Gupta B et al., 2016]. The seven positive samples which showed ORF1p positive was screened for ORF2p using an inhouse ORF2p antibody generated from the L1RT domain (Mukherjee et al., manuscript under preparation). The data showed five out of seven are showed ORF2p positive. The localization studies of hL1-ORF1p showed its cytoplasmic localization; occasionally limited number cells do show nuclear localization [Goodier et al. 2007; Doucet et al. 2010; Horn et al. 2014]. We observed similar pattern of localization in those seven samples which showed ORF1p positive by IHC; three samples (C2, C3 and C4) showed nuclear localization of ORF1p. While ORF1p is absolutely required for retrotransposition *in cis* [Moran et al, 1996; Wei et al., 2001] it is unclear currently whether increased nuclear localization of ORF1p is associated with an increase in insertion frequency. Interestingly, studies of breast cancer using murine models and human samples have reported L1 proteins (ORF1p and ORF2p) in the nucleus in advanced stages of cancer [Harris et al., 2010; Chen et al. 2012; De Luca et al. 2016]. Importantly, it has been reported that the samples associated with nuclear localization of L1 proteins showed very poor clinical outcome [Chen et al. 2012]. Mutations in p53 and its aberrant expression are common in almost every type of cancer [Kastenhuber and Lowe 2017; Yue et al. 2017]. Recent reports have shown that p53 can restrain

retrotransposons [Wylie et al. 2016]. Rodic et al [Rodic et al. 2014] reported that up-regulation of ORF1p in cancer tissues is correlated with highly expressed mutant p53. Although future studies, including L1 insertion analysis (e.g. L1-seq), will address whether L1 contributes to OSCC genome evolution, alternative functions for ORF1p, including its RNA-binding activity independent of retrotransposition in OSCC should not be dismissed. In this study the OSCC samples positive for expression of both hORF1p and ORF2p were also stained with CK-19 antibody used as tumor biomarker mostly in epithelial cancers [Ogden GR et al. 1993; Frohwitter et al. 2016; Vaidya et al. 1996].

Bibliography

1. Adams, J.W., Kaufman, R.E., Kretschmer, P.J., Harrison, M. and Nienhuis, A.W., 1980. A family of long reiterated DNA sequences, one copy of which is next to the human beta globin gene. Nucleic acids research, 8(24), pp.6113-6128.

2. Agrawal, N., Frederick, M.J., Pickering, C.R., Bettegowda, C., Chang, K., Li, R.J., Fakhry, C., Xie, T.X., Zhang, J., Wang, J. and Zhang, N. 2011. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1.Science. 333 (6046), pp.1154-1157.

3. Alisch, R.S., Garcia-Perez, J.L., Muotri, A.R., Gage, F.H. and Moran, J.V., 2006. Unconventional translation of mammalian LINE-1 retrotransposons. *Genes & development*, *20*(2), pp.210-224.

4. Alves, G., Tatro, A. and Fanning, T., 1996. Differential methylation of human LINE-1 retrotransposons in malignant cells. *Gene*, *176*(1-2), pp.39-44.

5. Andraos, N., Tabor, S. and Richardson, C.C., 2004. The Highly Processive DNA polymerase of Bacteriophage T5 ROLE OF THE UNIQUE N AND C TERMINI. *Journal of Biological Chemistry*, *279*(48), pp.50609-50618.

6. Antelo, M., Balaguer, F., Shia, J., Shen, Y., Hur, K., Moreira, L., Cuatrecasas, M., Bujanda, L., Giraldez, M.D., Takahashi, M. and Cabanne, A., 2012. A high degree of LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer. *PloS one*, *7*(9), p.e45357.

7. Ariel, P. and Ryan, T.A. 2012. New insights into molecular players involved in neurotransmitter release. Physiology. 27(1), pp.15-24.

8. Arpino, J.A., Rizkallah, P.J. and Jones, D.D., 2012. Crystal structure of enhanced green fluorescent protein to 1.35 Å resolution reveals alternative conformations for Glu222. *PloS one*, *7*(10), p.e47132.

9. Athanikar, J.N., Badge, R.M. and Moran, J.V., 2004. A YY1-binding site is required for accurate human LINE-1 transcription initiation. Nucleic acids research, 32(13), pp.3846-3855.

10. Baba, Y., Watanabe, M., Murata, A., Shigaki, H., Miyake, K., Ishimoto, T., Iwatsuki, M., Iwagami, S., Yoshida, N., Oki, E. and Sakamaki, K., 2014. LINE-1 hypomethylation, DNA copy number alterations, and CDK6 amplification in esophageal squamous cell carcinoma. *Clinical Cancer Research*, *20*(5), pp.1114-1124.

11. Badge, R.M., Alisch, R.S. and Moran, J.V., 2003. ATLAS: a system to selectively identify human-specific L1 insertions. *The American Journal of Human Genetics*, *72*(4), pp.823-838.

12. Baillie, J.K., Barnett, M.W., Upton, K.R., Gerhardt, D.J., Richmond, T.A., De Sapio, F., Brennan, P.M., Rizzu, P., Smith, S., Fell, M. and Talbot, R.T., 2011. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*, *479*(7374), p.534.

13. Baillie, J.K., Barnett, M.W., Upton, K.R., Gerhardt, D.J., Richmond, T.A., De Sapio, F., Brennan, P.M., Rizzu, P., Smith, S., Fell, M. and Talbot, R.T., 2011. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*, *479*(7374), p.534.

14. Bao, X., Wu, H., Zhu, X., Guo, X., Hutchins, A.P., Luo, Z., Song, H., Chen, Y., Lai, K., Yin, M. and Xu, L., 2015. The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters. Cell research, 25(1), p.80.

15. Becker, K.G., Swergold, G., Ozato, K. and Thayer, R.E., 1993. Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. Human molecular genetics, 2(10), pp.1697-1702.

16. Bennett, E.A., Coleman, L.E., Tsui, C., Pittard, W.S. and Devine, S.E., 2004. Natural genetic variation caused by transposable elements in humans. *Genetics*, *168*(2), pp.933-951.

17. Bhattacharya, S., Bakre, A. and Bhattacharya, A., 2002. Mobile genetic elements in protozoan parasites. *Journal of genetics*, *81*(2), pp.73-86.

18. Biankin, A.V., Waddell, N., Kassahn, K.S., Gingras, M.C., Muthuswamy, L.B., Johns, A.L., Miller, D.K., Wilson, P.J., Patch, A.M., Wu, J. and Chang, D.K. 2012. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature. 491(7424), pp.399-405.

19. Boeke, J.D., Garfinkel, D.J., Styles, C.A. and Fink, G.R., 1985. Ty elements transpose through an RNA intermediate. Cell, 40(3), pp.491-500.

20. Boissinot, S., Chevret, P. and Furano, A.V., 2000. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Molecular Biology and Evolution*, *17*(6), pp.915-928.

21. Boissinot, S., Entezam, A., Young, L., Munson, P.J. and Furano, A.V., 2004. The insertional history of an active family of L1 retrotransposons in humans. *Genome research*, *14*(7), pp.1221-1231.

22. Branciforte, D. and Martin, S.L., 1994. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Molecular and cellular biology*, *14*(4), pp.2584-2592.

23. Brandizzi, D., Gandolfo, M., Velazco, M.L., Cabrini, R.L. and Lanfranchi, H.E., 2008. Clinical features and evolution of oral cancer: A study of 274 cases in Buenos Aires, Argentina. *Age (in years)*, *40*(3), p.9.

24. Brandizzi, D., Gandolfo, M., Velazco, M.L., Cabrini, R.L. and Lanfranchi, H.E., 2008. Clinical features and evolution of oral cancer: A study of 274 cases in Buenos Aires, Argentina. *Age (in years)*, *40*(3), p.9.

25. Britten RJKohne, D.E., 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. Science, 161, p.529540.

26. Campanini, B., Pioselli, B., Raboni, S., Felici, P., Giordano, I., D'Alfonso, L., Collini, M., Chirico, G. and Bettati, S., 2013. Role of histidine 148 in stability and dynamics of a highly fluorescent GFP variant. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1834*(4), pp.770-779.

27. Chen, L., Dahlstrom, J.E., Chandra, A., Board, P., Rangasamy, D. 2012. Prognostic value of LINE-1 retrotransposon expression and its subcellular localization in breast cancer. Breast Cancer Res Treat.136(1), pp.129-42.

28. Cho, Y.W., Hong, T., Hong, S., Guo, H., Yu, H., Kim, D., Guszczynski, T., Dressler, G.R., Copeland, T.D., Kalkum, M. and Ge, K. 2007. PTIP associates with MLL3-and MLL4-containing histone H3 lysine 4 methyltransferase complex. Journal of Biological Chemistry. 282(28), pp.20395-20406.

29. Christensen, S.M. and Eickbush, T.H., 2005. R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. *Molecular and cellular biology*, *25*(15), pp.6617-6628.

30. Christian, C.M., Sokolowski, M., Kines, K.J. and Belancio, V.P., 2017. Involvement of conserved amino acids in the C-terminal region of LINE-1 ORF2p in retrotransposition. *Genetics*, *205*(3), pp.1139-1149.

31. Clements, A.P. and Singer, M.F. 1998. The human LINE-1 reverse transcriptase: effect of deletions outside the common reverse transcriptase domain. Nucleic acids research. 26(15), pp.3528-3535.

32. Cordaux, R. and Batzer, M.A., 2009. The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics*, *10*(10), p.691.

33. Cost, G.J., Feng, Q., Jacquier, A. and Boeke, J.D. 2002. Human L1 element target-primed reverse transcription in vitro. The EMBO journal. 21(21), pp.5899-5910.

34. Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Marchetto, M.C., Muotri, A.R., Mu, Y., Carson, C.T., Macia, A., Moran, J.V. and Gage, F.H., 2011. Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proceedings of the National Academy of Sciences*, *108*(51), pp.20382-20387.

35. Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Yeo, G.W., Mu, Y., Lovci, M.T., Morell, M., O'Shea, K.S., Moran, J.V. and Gage, F.H., 2009. L1 retrotransposition in human neural progenitor cells. *Nature*, *460*(7259), p.1127.

36. Counter, C.M., Avilion, A.A., LeFeuvre, C.E., Stewart, N.G., Greider, C.W., Harley, C.B. and Bacchetti, S., 1992. Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *The EMBO journal*, *11*(5), pp.1921-1929.

37. Criscione, S.W., Zhang, Y., Thompson, W., Sedivy, J.M. and Neretti, N., 2014. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC genomics*, *15*(1), p.583.

38. De Luca, C., Guadagni, F., Sinibaldi-Vallebona, P., Sentinelli, S., Gallucci, M., Hoffmann, A., Schumann, G.G., Spadafora, C. and Sciamanna, I., 2016. Enhanced expression of LINE-1-encoded ORF2 protein in early stages of colon and prostate transformation. *Oncotarget*, *7*(4), p.4048.

39. De Luca, C., Guadagni, F., Sinibaldi-Vallebona, P., Sentinelli, S., Gallucci, M., Hoffmann, A., Schumann, G.G., Spadafora, C. and Sciamanna, I., 2016. Enhanced expression of LINE-1-encoded ORF2 protein in early stages of colon and prostate transformation. *Oncotarget*, *7*(4), p.4048.

40. Deininger, P.L., Moran, J.V., Batzer, M.A. and Kazazian Jr, H.H., 2003. Mobile elements and mammalian genome evolution. *Current opinion in genetics & development*, *13*(6), pp.651-658.

41. Denli, A.M., Narvaiza, I., Kerman, B.E., Pena, M., Benner, C., Marchetto, M.C., Diedrich, J.K., Aslanian, A., Ma, J., Moresco, J.J. and Moore, L. 2015. Primate-specific ORF0 contributes to retrotransposon-mediated diversity. Cell. 163(3), pp.583-593.

42. DePena, C.A., Van, P.T. and Lee, Y.Y., 1990. Lymphoma of the head and neck. *Radiologic Clinics of North America*, *28*(4), pp.723-743.

43. Dikshit, R., Gupta, P.C., Ramasundarahettige, C., Gajalakshmi, V., Aleksandrowicz, L., Badwe, R., Kumar, R., Roy, S., Suraweera, W., Bray, F. and Mallath, M., 2012. Cancer mortality in India: a nationally representative survey. *The Lancet*, *379*(9828), pp.1807-1816.

44. Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F. and Kazazian, H.H., 1991. Isolation of an active human transposable element. Science, 254(5039), pp.1805-1808.

45. Doucet, A.J., Hulme, A.E., Sahinovic, E., Kulpa, D.A., Moldovan, J.B., Kopera, H.C., Athanikar, J.N., Hasnaoui, M., Bucheton, A., Moran, J.V. and Gilbert, N., 2010. Characterization of LINE-1 ribonucleoprotein particles. PLoS genetics, 6(10), p.e1001150.

46. Doyle, G.A., Crist, R.C., Karatas, E.T., Hammond, M.J., Ewing, A.D., Ferraro, T.N., Hahn, C.G. and Berrettini, W.H., 2017. Analysis of LINE-1 elements in DNA from postmortem brains of individuals with schizophrenia. *Neuropsychopharmacology*, *42*(13), p.2602.

47. Ergün, S., Buschmann, C., Heukeshoven, J., Dammann, K., Schnieders, F., Lauke, H., Chalajour, F., Kilic, N., Strätling, W.H. and Schumann, G.G., 2004. Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. Journal of Biological Chemistry, 279(26), pp.27753-27763.

48. Esnault, C., Maestre, J. and Heidmann, T., 2000. Human LINE retrotransposons generate processed pseudogenes. *Nature genetics*, *24*(4), p.363.

49. Ewing, A.D., 2015. Transposable element detection from whole genome sequence data. *Mobile DNA*, *6*(1), p.24.

50. Ewing, A.D., Gacita, A., Wood, L.D., Ma, F., Xing, D., Kim, M.S., Manda, S.S., Abril, G., Pereira, G., Makohon-Moore, A. and Looijenga, L.H., 2015. Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome research*, *25*(10), pp.1536-1545.

51. Fanning, T. and Singer, M., 1987. The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. Nucleic acids research, 15(5), pp.2251-2260.

52. Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T. and Waki, K., 2009. The

regulated retrotransposon transcriptome of mammalian cells. *Nature genetics*, *41*(5), p.563.

53. Feng, Q., Moran, J.V., Kazazian Jr, H.H. and Boeke, J.D., 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell, 87(5), pp.905-916.

54. Freeman, J.D., Goodchild, N.L. and Mager, D.L., 1994. A modified indicator gene for selection of retrotransposition events in mammalian cells. *Biotechniques*, *17*(1), pp.46-48.

55. Fu, J.L., Kanno, T., Liang, S.C., Matzke, A.J. and Matzke, M., 2015. GFP loss-of-function mutations in Arabidopsis thaliana. *G3: Genes, Genomes, Genetics*, *5*(9), pp.1849-1855.

56. Garfinkel, D.J., 2005. Genome evolution mediated by Ty elements in Saccharomyces. Cytogenetic and genome research, 110(1-4), pp.63-69.

57. Gasior, S.L., Wakeman, T.P., Xu, B. and Deininger, P.L., 2006. The human LINE-1 retrotransposon creates DNA double-strand breaks. *Journal of molecular biology*, *357*(5), pp.1383-1393.

58. Gibbs RA, Weinstock GM, Metzker ML, et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 2004;428:493–521.

59. Gilbert, N., Lutz, S., Morrish, T.A. and Moran, J.V., 2005. Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Molecular and cellular biology*, *25*(17), pp.7780-7795.

60. Gilbert, N., Lutz-Prigge, S. and Moran, J.V., 2002. Genomic deletions created upon LINE-1 retrotransposition. *Cell*, *110*(3), pp.315-325.

61. Goodier, J.L., 2016. Restricting retrotransposons: a review. Mobile DNA, 7(1), p.16.

62. Goodier, J.L., Cheung, L.E. and Kazazian Jr, H.H. 2013. Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. Nucleic acids research. 41(15), pp.7401-7419

63. Goodier, J.L., Zhang, L., Vetter, M.R. and Kazazian, H.H. 2007. LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. Molecular and cellular biology. 27(18), pp.6469-6483.

64. Gupta, B., Johnson, N.W. and Kumar, N. 2016. Global epidemiology of head and neck cancers: a continuing challenge. Oncology. 91(1), pp.13-23.

65. Han, K., Lee, J., Meyer, T.J., Remedios, P., Goodwin, L. and Batzer, M.A., 2008. L1 recombination-associated deletions generate human genomic variation. *Proceedings of the national academy of sciences*, *105*(49), pp.19366-19371.

66. Hancks, D.C. and Kazazian Jr, H.H., 2012. Active human retrotransposons: variation and disease. *Current opinion in genetics & development*, *22*(3), pp.191-203.

67. Harris, C.R., Normart, R., Yang, Q., Stevenson, E., Haffty, B.G., Ganesan, S., Cordon-Cardo, C., Levine, A.J., Tang, L.H.2010. Association of nuclear localization of a long interspersed nuclear element-1 protein in breast tumors with poor prognostic outcomes. Genes Cancer. 1(2), pp.115-24.

68. Hattori, M., Kuhara, S., Takenaka, O. and Sakaki, Y., 1986. L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. Nature, 321(6070), p.625.

69. Haya-Fernández, M., Bagan, J.V., Murillo-Cortés, J., Poveda-Roda, R. and Calabuig, C., 2004. The prevalence of oral leukoplakia in 138 patients with oral squamous cell carcinoma. *Oral diseases*, *10*(6), pp.346-348.

70. Hayakawa, T., Satta, Y., Gagneux, P., Varki, A. and Takahata, N., 2001. Alu-mediated inactivation of the human CMP-N-acetylneuraminic acid hydroxylase gene. *Proceedings of the National Academy of Sciences*, *98*(20), pp.11399-11404.

71. Hedges, D.J. and Deininger, P.L., 2007. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *616*(1-2), pp.46-59.

72. Helman, E., Lawrence, M.S., Stewart, C., Sougnez, C., Getz, G. and Meyerson, M. 2014. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. Genome research. 24(7), pp.1053-1063.

73. Hirata, R.M., Jaques, D.A., Chambers, R.G., Tuttle, J.R. and Mahoney, W.D., 1975. Carcinoma of the oral cavity. An analysis of 478 cases. *Annals of surgery*, *182*(2), p.98.

74. Hohjoh, H. and Singer, M.F., 1996. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. The EMBO journal, 15(3), pp.630-639.

75. Hohjoh, H. and Singer, M.F., 1997. Ribonuclease and high salt sensitivity of the ribonucleoprotein complex formed by the human LINE-1 retrotransposon.

76. Holmes, S.E., Singer, M.F. and Swergold, G.D., 1992. Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. Journal of Biological Chemistry, 267(28), pp.19765-19768.

77. Huang, C.R.L., Burns, K.H. and Boeke, J.D. 2012. Active transposition in genomes. Annual review of genetics. 46, pp.651-675.

78. Indian Council of Medical Research. National Cancer Registry Programme - Biennial Report (1988- 89) of the National Cancer Registry Programme. New Delhi: Indian Council of Medical Research; 1992.

79. International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature, 409(6822), p.860.

80. Iskow, R.C., McCabe, M.T., Mills, R.E., Torene, S., Pittard, W.S., Neuwald, A.F., Van Meir, E.G., Vertino, P.M. and Devine, S.E., 2010. Natural

mutagenesis of human genomes by endogenous retrotransposons. *Cell*, *141*(7), pp.1253-1261.

81. Ivics, Z., Hackett, P.B., Plasterk, R.H. and Izsvák, Z., 1997. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell, 91(4), pp.501-510.

82. Kastenhuber, E.R. and Lowe, S.W. 2017. Putting p53 in context. Cell. 170(6), pp.1062-1078.

83. Kazazian, H.H., 2004. Mobile elements: drivers of genome evolution. *science*, *303*(5664), pp.1626-1632.

84. Kemp, S., Gallagher, G., Kabani, S., Noonan, V. and O'Hara, C., 2008. Oral non-Hodgkin's lymphoma: review of the literature and World Health Organization classification with reference to 40 cases. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, *105*(2), pp.194-201.

85. Keszler, A., Piloni, M.J., Paparella, M.L., Soler, D.M., Ron, P.C. and Narbaitz, M., 2008. Extranodal oral non-Hodgkin's lymphomas. A retrospective study of 40 cases in Argentina. *Acta odontologica latinoamericana: AOL*, *21*(1), pp.43-48.

86. Khan, H., Smit, A. and Boissinot, S., 2006. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome research, 16(1), pp.78-87.

87. Khazina, E., Truffault, V., Büttner, R., Schmidt, S., Coles, M. and Weichenrieder, O., 2011. Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. Nature structural & molecular biology, 18(9), p.1006.

88. Khazina, E., Truffault, V., Büttner, R., Schmidt, S., Coles, M. and Weichenrieder, O. 2011. Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. Nature structural & molecular biology. 18(9), pp.1006-1014.

89. Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.D.L., Coviello, G.M., Wright, W.E., Weinrich, S.L. and Shay, J.W., 1994. Specific association of human telomerase activity with immortal cells and cancer. *Science*, *266*(5193), pp.2011-2015.

90. Kines, K.J., Sokolowski, M., deHaro, D.L., Christian, C.M., Baddoo, M., Smither, M.E. and Belancio, V.P., 2016. The endonuclease domain of the LINE-1 ORF2 protein can tolerate multiple mutations. *Mobile DNA*, *7*(1), p.8.

91. Kitkumthorn, N., Tuangsintanakul, T., Rattanatanyong, P., Tiwawech, D. and Mutirangura, A., 2012. LINE-1 methylation in the peripheral blood mononuclear cells of cancer patients. *Clinica Chimica Acta*, *413*(9-10), pp.869-874.

92. Kulpa, D.A. and Moran, J.V., 2005. Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Human molecular genetics*, *14*(21), pp.3237-3248.

93. Kulpa, D.A. and Moran, J.V., 2006. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. Nature structural & molecular biology, 13(7), p.655.

94. Kummer, A.D., Wiehler, J., Rehaber, H., Kompa, C., Steipe, B. and Michel-Beyerle, M.E., 2000. Effects of threonine 203 replacements on excited-state dynamics and fluorescence properties of the green fluorescent protein (GFP). *The Journal of Physical Chemistry B*, *104*(19), pp.4791-4798.

95. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature2001;409:860–921.

96. Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L.J., Lohr, J.G., Harris, C.C., Ding, L., Wilson, R.K. and Wheeler, D.A., 2012. Landscape of somatic retrotransposition in human cancers. *Science*, *337*(6097), pp.967-971.

97. Lee, J., Kim, Y.J., Mun, S., Kim, H.S. and Han, K., 2015. Identification of human-specific AluS elements through comparative genomics. *Gene*, *555*(2), pp.208-216.

98. Leibold, D.M., Swergold, G.D., Singer, M.F., Thayer, R.E., Dombroski, B.A. and Fanning, T.G., 1990. Translation of LINE-1 DNA elements in vitro and in human cells. *Proceedings of the National Academy of Sciences*, *87*(18), pp.6990-6994.

99. Leibold, E.A., Laudano, A. and Yu, Y., 1990. Structural requirements of iron-responsive elements for binding of the protein involved in both transferrin receptor and ferritin mRNA post-transcriptional regulation. *Nucleic Acids Research*, *18*(7), pp.1819-1824.

100. Li, W., Prazak, L., Chatterjee, N., Grüninger, S., Krug, L., Theodorou, D. and Dubnau, J., 2013. Activation of transposable elements during aging and neuronal decline in Drosophila. *Nature neuroscience*, *16*(5), p.529.

101. Lindtner, S., Felber, B.K. and Kjems, J., 2002. An element in the 3′ untranslated region of human LINE-1 retrotransposon mRNA binds NXF1 (TAP) and can function as a nuclear export element. *Rna*, *8*(3), pp.345-356.

102. Luan, D.D., Korman, M.H., Jakubczak, J.L. and Eickbush, T.H., 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell*, *72*(4), pp.595-605.

103. Majumdar, S. and Rio, D.C., 2015. P transposable elements in Drosophila and other eukaryotic organisms. *Microbiology spectrum*, *3*(2).

104. Majumdar, S., Singh, A. and Rio, D.C., 2013. The human THAP9 gene encodes an active P-element DNA transposase. *Science*, *339*(6118), pp.446-448.

105. Majumder PP. 2014. Mutational landscape of gingivo-buccal oral cancer: new cancer genes and molecular subgroups identified.Mol Cytogenet. 21;7.

106. Malik, H.S., Burke, W.D. and Eickbush, T.H. 1999. The age and evolution of non-LTR retrotransposable elements. Molecular biology and evolution. 16(6), pp.793-805.

107. Mallet, Y., Avalos, N., Le Ridant, A.M., Gangloff, P., Moriniere, S., Rame, J.P., Poissonnet, G., Makeieff, M., Cosmidis, A., Babin, E. and Barry, B., 2009. Head and neck cancer in young people: a series of 52 SCCs of the oral tongue in patients aged 35 years or less. *Acta oto-laryngologica*, *129*(12), pp.1503-1508.

108. Mandal, P.K., Bagchi, A., Bhattacharya, A. and Bhattacharya, S. 2004. An Entamoeba histolytica LINE/SINE pair inserts at common target sites cleaved by the

109. Mandal, P.K., Ewing, A.D., Hancks, D.C. and Kazazian Jr, H.H. 2013. Enrichment of processed pseudogene transcripts in L1-ribonucleoprotein particles. Human molecular genetics. 22(18), pp.3730-3748

110. Martin, S.L. and Bushman, F.D., 2001. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. Molecular and cellular biology, 21(2), pp.467-475.

111. Martin, S.L., 1991. Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Molecular and cellular biology*, *11*(9), pp.4804-4807.

112. Martin, S.L., 2010. Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. RNA biology, 7(6), pp.706-711.

113. Mashberg, A., Merletti, F., Boffetta, P., Gandolfo, S., Ozzello, F., Fracchia, F. and Terracini, B., 1989. Appearance, site of occurrence, and physical and clinical characteristics of oral carcinoma in Torino, Italy. *Cancer*, *63*(12), pp.2522-2527.

114. Mathias, S.L., Scott, A.F., Kazazian, H.H., Boeke, J.D. and Gabriel, A., 1991. Reverse transcriptase encoded by a human transposable element. Science, 254(5039), pp.1808-1810.

115. McClintock, B. 1950. The origin and behavior of mutable loci in maize. Proceedings of the National Academy of Sciences. 36(6), pp.344-355.

116. McClintock, B. 1956, January. Controlling elements and the gene. Cold Spring Harbor symposia on quantitative biology (Vol. 21, pp. 197-216). Cold Spring Harbor Laboratory Press.

117. Miki, Y., Katagiri, T., Kasumi, F., Yoshimoto, T. and Nakamura, Y., 1996. Mutation analysis in the BRCA2 gene in primary breast cancers. *Nature genetics*, *13*(2), p.245.

118. Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B. and Nakamura, Y., 1992. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer research*, *52*(3), pp.643-645.

119. Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B. and Nakamura, Y. 1992. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer research. 52(3), pp.643-645.

120. Milian, A., Bagan, J.V. and Vera, F., 1993. Squamous cell carcinoma of the oral cavity: a follow up study of 85 cases and analysis of prognostic variables. *Bulletin du Groupement international pour la recherche scientifique en stomatologie & odontologie*, *36*(1-2), pp.29-35.

121. Minakami, R., Kurose, K., Etoh, K., Furuhata, Y., Hattori, M. and Sakaki, Y., 1992. Identification of an internal cis-element essential for the human Li transcription and a nuclear factor (s) binding to the element. Nucleic acids research, 20(12), pp.3139-3145.

122. Moran, J.V., DeBerardinis, R.J. and Kazazian, H.H., 1999. Exon shuffling by L1 retrotransposition. *Science*, *283*(5407), pp.1530-1534.

123. Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D. and Kazazian Jr, H.H., 1996. High frequency retrotransposition in cultured mammalian cells. Cell, 87(5), pp.917-927.

124. Morrish, T.A., Garcia-Perez, J.L., Stamato, T.D., Taccioli, G.E., Sekiguchi, J. and Moran, J.V. 2007. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. Nature. 446(7132), pp.208-212.

125. Morse, B., Rotherg, P.G., South, V.J., Spandorfer, J.M. and Astrin, S.M., 1988. Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. *Nature*, *333*(6168), p.87.

126. Mouse Genome Sequencing Consortium, 2002. Initial sequencing and comparative analysis of the mouse genome. Nature, 420(6915), p.520.

127. Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V. and Gage, F.H., 2005. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*, *435*(7044), p.903.

128. Muotri, A.R., Marchetto, M.C., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K. and Gage, F.H., 2010. L1 retrotransposition in neurons is modulated by MeCP2. *Nature*, *468*(7322), p.443.

129. Murata, A., Baba, Y., Watanabe, M., Shigaki, H., Miyake, K., Ishimoto, T., Iwatsuki, M., Iwagami, S., Sakamoto, Y., Miyamoto, Y. and Yoshida, N., 2013. Methylation levels of LINE-1 in primary lesion and matched metastatic lesions of colorectal cancer. *British journal of cancer*, *109*(2), p.408.

130. Myers, J.S., Vincent, B.J., Udall, H., Watkins, W.S., Morrish, T.A., Kilroy, G.E., Swergold, G.D., Henke, J., Henke, L., Moran, J.V. and Jorde, L.B., 2002. A comprehensive analysis of recently integrated human Ta L1 elements. *The American Journal of Human Genetics*, *71*(2), pp.312-326.

131. Neville, B.W. and Day, T.A., 2002. Oral cancer and precancerous lesions. *CA: a cancer journal for clinicians*, *52*(4), pp.195-215.

132. Nigumann, P., Redik, K., Mätlik, K. and Speek, M., 2002. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics, 79(5), pp.628-634.

133. Ogino, S., Nosho, K., Kirkner, G.J., Kawasaki, T., Chan, A.T., Schernhammer, E.S., Giovannucci, E.L. and Fuchs, C.S., 2008. A cohort study of tumoral LINE-1 hypomethylation and prognosis in colon cancer. *JNCI: Journal of the National Cancer Institute*, *100*(23), pp.1734-1738.

134. Oliver, A.J., Helfrick, J.F. and Gard, D., 1996. Primary oral squamous cell carcinoma: a review of 92 cases. *Journal of oral and maxillofacial surgery*, *54*(8), pp.949-954.

135. Ostertag, E.M. and Kazazian Jr, H.H., 2001. Biology of mammalian L1 retrotransposons. Annual review of genetics, 35(1), pp.501-538.

136. Ostertag, E.M., DeBerardinis, R.J., Goodier, J.L., Zhang, Y., Yang, N., Gerton, G.L. and Kazazian Jr, H.H., 2002. A mouse model of human L1 retrotransposition. *Nature genetics*, *32*(4), p.655.

137. Ostertag, E.M., Goodier, J.L., Zhang, Y. and Kazazian Jr, H.H., 2003. SVA elements are nonautonomous retrotransposons that cause disease in humans. *The American Journal of Human Genetics*, *73*(6), pp.1444-1451.

138. Pandey, R. and Mukerji, M., 2011. From 'JUNK'to just unexplored noncoding knowledge: the case of transcribed Alus. *Briefings in functional genomics*, *10*(5), pp.294-311.

139. Park, S.Y., Seo, A.N., Jung, H.Y., Gwak, J.M., Jung, N., Cho, N.Y. and Kang, G.H., 2014. Alu and LINE-1 hypomethylation is associated with HER2 enriched subtype of breast cancer. *PloS one*, *9*(6), p.e100429.

140. Patchsung, M., Boonla, C., Amnattrakul, P., Dissayabutra, T., Mutirangura, A. and Tosukhowong, P., 2012. Long interspersed nuclear element-1 hypomethylation and oxidative stress: correlation and bladder cancer diagnostic potential. *PloS one*, *7*(5), p.e37009.

141. Patel, K.A., Sethi, R., Dhara, A.R. and Roy, I., 2017. Challenges with osmolytes as inhibitors of protein aggregation: Can nucleic acid aptamers provide an answer?. *International journal of biological macromolecules*, *100*, pp.75-88.

142. Pérez-Mancera, P.A., Rust, A.G., Van Der Weyden, L., Kristiansen, G., Li, A., Sarver, A.L., Silverstein, K.A., Grützmann, R., Aust, D., Rümmele, P. and Knösel, T. 2012. The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma. Nature. 486(7402), pp.266-270.

143. Perrat, P.N., DasGupta, S., Wang, J., Theurkauf, W., Weng, Z., Rosbash, M. and Waddell, S., 2013. Transposition-driven genomic heterogeneity in the Drosophila brain. *Science*, *340*(6128), pp.91-95.

144. Pink, R.C., Wicks, K., Caley, D.P., Punch, E.K., Jacobs, L. and Carter, D.R.F. 2011. Pseudogenes: pseudo-functional or key regulators in health and disease?. Rna. 17(5), pp.792-798.

145. Pisitkun, T., Hoffert, J.D., Saeed, F. and Knepper, M.A., 2011. NHLBI-AbDesigner: an online tool for design of peptide-directed antibodies. *American Journal of Physiology-Cell Physiology*, *302*(1), pp.C154-C164.

146. Piskareva, O. and Schmatchenko, V. 2006. DNA polymerization by the reverse transcriptase of the human L1 retrotransposon on its own template in vitro. FEBS letters. 580(2), pp.661-668.

147. Piskareva, O., Denmukhametova, S. and Schmatchenko, V. 2003. Functional reverse transcriptase encoded by the human LINE-1 from baculovirus-infected insect cells. Protein expression and purification. 28(1), pp.125-130.

148. Piskareva, O., Ernst, C., Higgins, N. and Schmatchenko, V. 2013. The carboxyterminal segment of the human LINE1 ORF2 protein is involved in RNA binding. FEBS Open Bio. 3(1), pp.433-437.

149. Prak ET, Dodson AW, Farkash EA, Kazazian HH Jr. Tracking an embryonic L1 retrotransposition event. Proc Natl Acad Sci USA 2003;100: 1832–1837.

150. Prakash, T., Khandelwal, M., Dasgupta, D., Dash, D. and Brahmachari, S.K., 2004. CoPS: comprehensive peptide signature database. *Bioinformatics*, *20*(16), pp.2886-2888.

151. Raiz, J., Damert, A., Chira, S., Held, U., Klawitter, S., Hamdorf, M., Löwer, J., Strätling, W.H., Löwer, R. and Schumann, G.G. 2011. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic acids research. 40(4), pp.1666-1683.

152. Rapidis, A.D., Gullane, P., Langdon, J.D., Lefebvre, J.L., Scully, C. and Shah, J.P., 2009. Major advances in the knowledge and understanding of the epidemiology, aetiopathogenesis, diagnosis, management and prognosis of oral cancer. *Oral oncology*, *45*(4-5), pp.299-300

153. Ribeiro, A.C.P., Silva, A.R.S., Simonato, L.E., Salzedas, L.M.P., Sundefeld, M.L.M.M. and Soubhia, A.M.P., 2009. Clinical and histopathological analysis of oral squamous cell carcinoma in young people: a descriptive study in Brazilians. *British Journal of Oral and Maxillofacial Surgery*, *47*(2), pp.95-98.

154. Richardson, S.R., Doucet, A.J., Kopera, H.C., Moldovan, J.B., Garcia-Pérez, J.L. and Moran, J.V., 2015. The influence of LINE-1 and SINE retrotransposons on mammalian genomes. *Microbiology spectrum*, *3*(2).

155. Richardson, S.R., Morell, S. and Faulkner, G.J., 2014. L1 retrotransposons and somatic mosaicism in the brain. *Annual review of genetics*, *48*, pp.1-27.

156. Rodić, N., Sharma, R., Sharma, R., Zampella, J., Dai, L., Taylor, M.S., Hruban, R.H., Iacobuzio-Donahue, C.A., Maitra, A., Torbenson, M.S. and Goggins, M., 2014. Long interspersed element-1 protein expression is a hallmark of many human cancers. *The American journal of pathology*, *184*(5), pp.1280-1286.

157. Roy I, Kurganov BI. 2017. Protein aggregation. Int J Biol Macromol.;100:1-2.

158. Saito, K., Kawakami, K., Matsumoto, I., Oda, M., Watanabe, G. and Minamoto, T., 2010. Long interspersed nuclear element 1 hypomethylation is a marker of poor prognosis in stage IA non–small cell lung cancer. *Clinical Cancer Research*, *16*(8), pp.2418-2426.

159. Salas, L.A., Villanueva, C.M., Tajuddin, S.M., Amaral, A.F., Fernandez, A.F., Moore, L.E., Carrato, A., Tardón, A., Serra, C., García-Closas, R. and Basagana, X., 2014. LINE-1 methylation in granulocyte DNA and trihalomethane exposure is associated with bladder cancer risk. *Epigenetics*, *9*(11), pp.1532-1539.

160. Saleh, A.A., Gune, U.S., Chaudhary, R.K., Turakhiya, A.P. and Roy, I., 2014. Roles of Hsp104 and trehalose in solubilisation of mutant huntingtin in heat shocked Saccharomyces cerevisiae cells. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1843*(4), pp.746-757.

161. Sankaranarayanan, R., Garrote, L.F., Anta, J.L., Pisani, P. and Salva, A.R., 2002. Visual inspection in oral cancer screening in Cuba: a case-control study. *Oral Oncology*, *38*(2), pp.131-136.

162. SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z. and Bennetzen, J.L., 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science, 274(5288), pp.765-768.

163. Scott, A.F., Schmeckpeper, B.J., Abdelrazik, M., Comey, C.T., O'Hara, B., Rossiter, J.P., Cooley, T., Heath, P., Smith, K.D. and Margolet, L., 1987. Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics, 1(2), pp.113-125.

164. Scully, C. and Bagan, J., 2009. Oral squamous cell carcinoma overview. *Oral oncology*, *45*(4/5), pp.301-308.

165. Scully, C. and Bagan, J.V., 2009. Oral squamous cell carcinoma: overview of current understanding of aetiopathogenesis and clinical implications. *Oral diseases*, *15*(6), pp.388-399.

166. Scully, C., Malamos, D., Levers, B.G., Porter, S.R. and Prime, S.S., 1986. Sources and patterns of referrals of oral cancer: role of general practitioners. *Br Med J (Clin Res Ed)*, *293*(6547), pp.599-601.

167. Sequencing Consortium, 1998. Genome sequence of the nematode C. elegans: A platform for investigating biology. Science, 282, pp.2012-2018.

168. Shain, A. H. & Pollack, J. R. 2013.The spectrum of SWI/SNF mutations, ubiquitous in human cancers. PLoS One.8, e55119.

169. Shankar, R., Kataria, B. and Mukerji, M., 2009. Finding Alu in primate genomes with AF-1. *Bioinformation*, *3*(7), p.287.

170. Shapiro, J.A., 1979. Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proceedings of the National Academy of Sciences, 76(4), pp.1933-1937.

171. Sheen, F.M., Sherry, S.T., Risch, G.M., Robichaux, M., Nasidze, I., Stoneking, M., Batzer, M.A. and Swergold, G.D., 2000. Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition. *Genome research*, *10*(10), pp.1496-1508.

172. Shigaki, H., Baba, Y., Watanabe, M., Murata, A., Iwagami, S., Miyake, K., Ishimoto, T., Iwatsuki, M. and Baba, H., 2013. LINE-1 hypomethylation in gastric cancer, detected by bisulfite pyrosequencing, is associated with poor prognosis. *Gastric Cancer*, *16*(4), pp.480-487

173. Shin, W., Mun, S., Kim, J., Lee, W., Park, D.G., Choi, S., Lee, T.Y., Cha, S. and Han, K., 2019. Novel Discovery of LINE-1 in a Korean Individual by a Target Enrichment Method. *Molecules and cells*, *42*(1), p.87.

174. Shu, X., Leiderman, P., Gepshtein, R., Smith, N.R., Kallio, K., Huppert, D. and Remington, S.J., 2007. An alternative excited-state proton transfer pathway in green fluorescent protein variant S205V. *Protein Science*, *16*(12), pp.2703-2710.

175. Shugars, D.C. and Patton, L.L., 1997. Detecting, diagnosing, and preventing oral cancer. *The nurse practitioner*, *22*(6), pp.105-109.

176. Shukla, R., Upton, K.R., Muñoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., Ghisletti, S. and Sinha, S., 2013. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell, 153(1), pp.101-111.

177. Silverman Jr, S., 1988. Early diagnosis of oral cancer. *Cancer*, *62*(S1), pp.1796-1799.

178. Silverman S. Oral cancer. Semin Dermatol 1994;13:132–7.

179. Singh, A., Upadhyay, V., Upadhyay, A.K., Singh, S.M. and Panda, A.K. 2015. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. Microbial cell factories. 14(1), pp.41.

180. Smit, A.F., Tóth, G., Riggs, A.D. and Jurka, J., 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. Journal of molecular biology, 246(3), pp.401-417.

181. Sokolowski, M., Chynces, M., DeHaro, D., Christian, C.M. and Belancio, V.P., 2017. Truncated ORF1 proteins can suppress LINE-1 retrotransposition in trans. *Nucleic acids research*, *45*(9), pp.5294-5308.

182. Sokolowski, M., DeFreece, C.B., Servant, G., Kines, K.J. and Belancio, V.P., 2014. Development of a monoclonal antibody specific to the endonuclease domain of the human LINE-1 ORF2 protein. *Mobile DNA*, *5*(1), p.29.

183. Solyom, S., Ewing, A.D., Rahrmann, E.P., Doucet, T., Nelson, H.H., Burns, M.B., Harris, R.S., Sigmon, D.F., Casella, A., Erlanger, B. and Wheelan, S.,

2012. Extensive somatic L1 retrotransposition in colorectal tumors. *Genome research*, *22*(12), pp.2328-2338.

184. Speek, M., 2001. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Molecular and cellular biology, 21(6), pp.1973-1985.

185. Startek, M., Szafranski, P., Gambin, T., Campbell, I.M., Hixson, P., Shaw, C.A., Stankiewicz, P. and Gambin, A., 2015. NAR Breakthrough Article Genome-wide analyses of LINE–LINE-mediated. *Nucleic Acids Research*, p.1.

186. Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A. and Coulson, A., 2003. The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS biology, 1(2), p.e45.

187. Stepanenko, O.V., Verkhusha, V.V., Shavlovsky, M.M., Kuznetsova, I.M., Uversky, V.N. and Turoverov, K.K., 2008. Understanding the role of Arg96 in structure and stability of green fluorescent protein. *Proteins: Structure, Function, and Bioinformatics*, *73*(3), pp.539-551.

188. Stransky, N. et al. 2011.The mutational landscape of head and neck squamous cell carcinoma. Science.333, pp.1157–1160.

189. Sur, D., Kustwar, R.K., Budania, S., Mahadevan, A., Hancks, D.C., Yadav, V., Shankar, S.K. and Mandal, P.K., 2017. Detection of the LINE-1 retrotransposon RNA-binding protein ORF1p in different anatomical regions of the human brain. *Mobile DNA*, *8*(1), p.17.

190. Swergold, G.D., 1990. Identification, characterization, and cell specificity of a human LINE-1 promoter. Molecular and cellular biology, 10(12), pp.6718-6729.

191. Symer, D.E., Connelly, C., Szak, S.T., Caputo, E.M., Cost, G.J., Parmigiani, G. and Boeke, J.D., 2002. Human l1 retrotransposition is associated with genetic instability in vivo. *Cell*, *110*(3), pp.327-338.

192. Takasu, M., Hayashi, R., Maruya, E., Ota, M., Imura, K., Kougo, K., Kobayashi, C., Saji, H., Ishikawa, Y., Asai, T. and Tokunaga, K., 2007. Deletion of entire

HLA-A gene accompanied by an insertion of a retrotransposon. *Tissue antigens*, *70*(2), pp.144-150.

193. Tang, Z., Steranka, J.P., Ma, S., Grivainis, M., Rodić, N., Huang, C.R., et al. 2017. Human transposon insertion profiling:analysis, visualization and identification of somaticLINE-1 insertions in ovarian cancer. Proc.NatlAcad.Sci. USA. 114, pp.733–740.

194. Taylor, M.S., LaCava, J., Mita, P., Molloy, K.R., Huang, C.R.L., Li, D., Adney, E.M., Jiang, H., Burns, K.H., Chait, B.T. and Rout, M.P. 2013. Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. Cell, 155(5), pp.1034-1048.

195. Tchénio, T., Casella, J.F. and Heidmann, T., 2000. Members of the SRY family regulate the human LINE retrotransposons. Nucleic acids research, 28(2), pp.411-415.

196. Teugels, E., De Brakeleer, S., Goelen, G., Lissens, W., Sermijn, E. and De Grève, J., 2005. De novo Alu element insertions targeted to a sequence common to the BRCA1 and BRCA2 genes. *Human mutation*, *26*(3), pp.284-284.

197. Thayer, R.E., Singer, M.F. and Fanning, T.G., 1993. Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1-encoded protein. Gene, 133(2), pp.273-277.

198. Trelogan, S.A. and Martin, S.L., 1995. Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proceedings of the National Academy of Sciences*, *92*(5), pp.1520-1524.

199. Upadhyay P, Gardi N, Desai S, Chandrani P, Joshi A, Dharavath B, Arora P, Bal M, Nair S, Dutt A. 2017. Genomic HPV characterization of tobacco/nut chewing -negative early stage tongue tumors identify MMP10 as a candidate to predict metastases. Oral Oncology 73:56-64.

200. Usdin, K. and Furano, A.V., 1989. The structure of the guanine-rich polypurine: polypyrimidine sequence at the right end of the rat L1 (LINE) element. *Journal of Biological Chemistry*, *264*(26), pp.15681-15687.

201. Vincent, B.J., Myers, J.S., Ho, H.J., Kilroy, G.E., Walker, J.A., Watkins, W.S., Jorde, L.B. and Batzer, M.A., 2003. Following the LINEs: an analysis of primate genomic variation at human-specific LINE-1 insertion sites. *Molecular biology and evolution*, *20*(8), pp.1338-1348.

202. Wallace, M.R., Andersen, L.B., Saulino, A.M., Gregory, P.E., Glover, T.W. and Collins, F.S., 1991. A de novo Alu insertion results in neurofibromatosis type 1. *Nature*, *353*(6347), p.864.

203. Wallace, N.A., Belancio, V.P. and Deininger, P.L. 2008. L1 mobile element expression causes multiple types of toxicity. Gene. 419(1), pp.75-81.

204. Wang, H., Xing, J., Grover, D., Hedges, D.J., Han, K., Walker, J.A. and Batzer, M.A., 2005. SVA elements: a hominid-specific retroposon family. *Journal of molecular biology*, *354*(4), pp.994-1007.

205. Waterston RH, Lindblad-Toh K, Birney E, et al. Initial sequencing and comparative analysis of the mouse genome. Nature 2002;420:520–562.

206. Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D. and Moran, J.V., 2001. Human L1 retrotransposition: cis preference versus trans complementation. Molecular and cellular biology, 21(4), pp.1429-1439.

207. Wei, W., Morrish, T.A., Alisch, R.S. and Moran, J.V., 2000. A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. *Analytical biochemistry*, *284*(2), pp.435-438.

208. World Health Organization. Strengthening the prevention of oral cancer: the WHO perspective http://www.who.int/oral_health/publications/

209. Wylie, A., Jones, A.E., D'Brot, A., Lu, W.J., Kurtz, P., Moran, J.V., Rakheja, D., Chen, K.S., Hammer, R.E., Comerford, S.A. and Amatruda, J.F. 2016. p53

genes function to restrain mobile elements. Genes & development. 30(1), pp.64-77.

210. Yang, N. and Kazazian, H.H. 2006. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nature structural & molecular biology. 13(9), pp.763-771.

211. Yang, N., Zhang, L., Zhang, Y. and Kazazian Jr, H.H., 2003. An important role for RUNX3 in human L1 transcription and retrotransposition. Nucleic acids research, 31(16), pp.4929-4940.

212. Yoder, J.A., Walsh, C.P. and Bestor, T.H., 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends in genetics, 13(8), pp.335-340.

213. Zamudio-Bulcock, P.A., Everett, J., Harteneck, C. and Valenzuela, C.F. 2011. Activation of steroid-sensitive TRPM3 channels potentiates glutamatergic transmission at cerebellar Purkinje neurons from developing rats. Journal of neurochemistry. 119(3), pp.474-485.

214. Zimmerly, S., Guo, H., Eskest, R., Yang, J., Perlman, P.S. and Lambowitz, A.M., 1995. A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell*, *83*(4), pp.529-538.

Appendix

6.1. Appendix I

6.1.1. Human L1 ORF2p Protein sequence

MTGSNSHITI LTLNVNGLNS PIKRHRLASW IKSQDPSVCC IQETHLTCRD THRLKIKGWR KIYQANGKQK 70

KAGVAILVSD KTDFKPTKIK RDKEGHYIMV KGSIQQEELT ILNIYAPNTG APRFIKQVLS DLQRDLDSHT 140

LIMGDFNTPL SILDRSTRQK VNKDTQELNS ALHQTDLIDI YRTLHPKSTE YTFFSAPHHT YSKIDHIVGS 210

KALLSKCKRT EIITNYLSDH SAIKLELRIK NLTQSRSTTW KLNNLLLNDY WVHNEMKAEI KMFFETNENK 280

DTTYQNLWDA FKAVCRGKFI ALNAYKRKQE RSKIDTLTSQ LKELEKQEQT HSKASRRQEI TKIRAELKEI 350

ETQKTLQKIN ESRSWFFERI NKIDRPLARL IKKKREKNQI DTIKNDKGDI TTDPTEIQTT IREYYKHLYA 420

NKLENLEEMD TFLDTYTLPR LNQEEVESLN RPITGSEIVA IINSLPTKKS PGPDGFTAEF YQRYKEELVP 490

FLLKLFQSIE KEGILPNSFY EASIILIPKP GRDTTKKENF RPISLMNIDA KILNKILANR IQQHIKKLIH 560

HDQVGFIPGM QGWFNIRKSI NVIQHINRAK DKNHVIISID AEKAFDKIQQ PFMLKTLNKL GIDGMYLKII 630

RAIYDKPTAN IILNGQKLEA FPLKTGTRQG CPLSPLLFNI VLEVLARAIR QEKEIKGIQL GKEEVKLSLF 700

ADDMIVYLEN PIVSAQNLLK LISNFSKVSG YKINVQKSQA FLYNNNRQTE SQIMGELPFT IASKRIKYLG 770

IQLTRDVKDL FKENYKPLLK EIKEDTNKWK NIPCSWVGRI NIVKMAILPK VIYRFNAIPI KLPMTFFTEL 840

EKTTLKFIWN QKRARIAKSI LSQKNKAGGI TLPDFKLYYK ATVTKTAWYW YQNRDIDQWN RTEPSEIMPH 910

IYNYLIFDKP EKNKQWGKDS LLNKWCWENW LAICRKLKLD PFLTPYTKIN SRWIKDLNVK PKTIKTLEEN 980

LGITIQDIGV GKDFMSKTPK AMATKDKIDK WDLIKLKSFC TAKETTIRVN RQPTTWEKIF ATYSSDKGLI 1050

SRIYNELKQI YKKKTNNPIK KWAKDMNRHF SKEDIYAAKK HMKKCSSSLA IREMQIKTTM RYHLTPVRMA 1120

IIKKSGNNRC WRGCGEIGTL VHCWWDCKLV QPLWKSVWRF LRDLELEIPF DPAIPLLGIY PKDYKSCCYK 1190

DTCTRMFIAA LFTIAKTWNQ PNCPTMIDWI KKMWHIYTME YYAAIKNDEF ISFVGTWMKL ETIILSKLSQ 1260

EQKTKHRIFS LIGGN* 1275 aa

Green: EN domain (1-239 aa)

Red: RT Domain (453-880 aa)

Blue: CCHC domain (1096-1275 aa)

6.1.2. ORF2 amino acids and nucleotide sequence

```
atgacaggatcaacttcacacataacaatattaactttaaatataaatggactaaattct
 M  T  G  S  T  S  H  I  T  I  L  T  L  N  I  N  G  L  N  S
gcaattaaaagacacagactggcaagttggataaagagtcaagacccatcagtgtgctgt
 A  I  K  R  H  R  L  A  S  W  I  K  S  Q  D  P  S  V  C  C
attcaggaaacccatctcacgtgcagagacacacataggctcaaaataaaaggatggagg
 I  Q  E  T  H  L  T  C  R  D  T  H  R  L  K  I  K  G  W  R
aagatctaccaagccaatggaaaacaaaaaaaggcaggggttgcaatcctagtctctgat
 K  I  Y  Q  A  N  G  K  Q  K  K  A  G  V  A  I  L  V  S  D
aaaacagactttaaaccaacaaagatcaaaagagacaaagaaggccattacataatggta
 K  T  D  F  K  P  T  K  I  K  R  D  K  E  G  H  Y  I  M  V
aagggatcaattcaacaagaggagctaactatcctaaatatttatgcacccaatacagga
 K  G  S  I  Q  Q  E  E  L  T  I  L  N  I  Y  A  P  N  T  G
gcacccagattcataaagcaagtcctcagtgacctacaaagagacttagactcccacaca
 A  P  R  F  I  K  Q  V  L  S  D  L  Q  R  D  L  D  S  H  T
ttaataatgggagactttaacaccccactgtcaacattagacagatcaacgagacagaaa
 L  I  M  G  D  F  N  T  P  L  S  T  L  D  R  S  T  R  Q  K
gtcaacaaggatacccaggaattgaactcagctctgcaccaagcagacctaatagacatc
 V  N  K  D  T  Q  E  L  N  S  A  L  H  Q  A  D  L  I  D  I
tacagaactctccaccccaaatcaacagaatatacatttttttcagcaccacaccacacc
 Y  R  T  L  H  P  K  S  T  E  Y  T  F  F  S  A  P  H  H  T
tattccaaaattgaccacatagttggaagtaaagctctcctcagcaaatgtaaaagaaca
 Y  S  K  I  D  H  I  V  G  S  K  A  L  L  S  K  C  K  R  T
gaaattataacaaactatctctcagaccacagtgcaatcaaactagaactcaggattaag
 E  I  I  T  N  Y  L  S  D  H  S  A  I  K  L  E  L  R  I  K
aatctcactcaaagccgctcaactacatggaaactgaacaacctgctcctgaatgactac
 N  L  T  Q  S  R  S  T  T  W  K  L  N  N  L  L  L  N  D  Y
tgggtacataacgaaatgaaggcagaaataaagatgttctttgaaaccaacgagaacaaa
 W  V  H  N  E  M  K  A  E  I  K  M  F  F  E  T  N  E  N  K
gacaccacataccagaatctctgggacgcattcaaagcagtgtgtagagggaaatttata
 D  T  T  Y  Q  N  L  W  D  A  F  K  A  V  C  R  G  K  F  I
gcactaaatgcctacaagagaaagcaggaaagatccaaaattgacaccctaacatcacaa
 A  L  N  A  Y  K  R  K  Q  E  R  S  K  I  D  T  L  T  S  Q
ttaaaagaactagaaaagcaagagcaaacacattcaaaagctagcagaaggcaagaaata
 L  K  E  L  E  K  Q  E  Q  T  H  S  K  A  S  R  R  Q  E  I
actaaaatcagagcagaactgaaggaaatagagacacaaaaaacccttcaaaaaatcaat
 T  K  I  R  A  E  L  K  E  I  E  T  Q  K  T  L  Q  K  I  N
gaatccaggagctggtttttgaaaggatcaacaaaattgatagaccgctagcaagacta
 E  S  R  S  W  F  F  E  R  I  N  K  I  D  R  P  L  A  R  L
ataaagaaaaaaagagagaagaatcaaatagacacaataaaaaatgataaaggggatatc
 I  K  K  K  R  E  K  N  Q  I  D  T  I  K  N  D  K  G  D  I
accaccgatcccacagaaatacaaactaccatcagagaatactacaaacacctctacgca
 T  T  D  P  T  E  I  Q  T  T  I  R  E  Y  Y  K  H  L  Y  A
aataaactagaaaatctagaagaaatggatacattcctcgacacatacactctcccaaga
 N  K  L  E  N  L  E  E  M  D  T  F  L  D  T  Y  T  L  P  R
ctaaaccaggaagaagttgaatctctgaatcgaccaataacaggctctgaaattgtggca
```

```
 L  N  Q  E  E  V  E  S  L  N  R  P  I  T  G  S  E  I  V  A
ataatcaatagtttaccaaccaaaaagagtccaggaccagatggattcacagccgaattc
 I  I  N  S  L  P  T  K  K  S  P  G  P  D  G  F  T  A  E  F
taccagaggtacaaggaggaactggtaccattccttctgaaactattccaatcaatagaa
 Y  Q  R  Y  K  E  E  L  V  P  F  L  L  K  L  F  Q  S  I  E
aaagagggaatcctccctaactcattttatgaggccagcatcattctgataccaaagccg
 K  E  G  I  L  P  N  S  F  Y  E  A  S  I  I  L  I  P  K  P
ggcagagacacaaccaaaaaagagaattttagaccaatatccttgatgaacattgatgca
 G  R  D  T  T  K  K  E  N  F  R  P  I  S  L  M  N  I  D  A
aaaatcctcaataaaatactggcaaaccgaatccagcagcacatcaaaaagcttatccac
 K  I  L  N  K  I  L  A  N  R  I  Q  Q  H  I  K  K  L  I  H
catgatcaagtgggcttcatccctgggatgcaaggctggttcaatatacgcaaatcaata
 H  D  Q  V  G  F  I  P  G  M  Q  G  W  F  N  I  R  K  S  I
aatgtaatccagcatataaacagagccaaagacaaaaaccacatgattatctcaatagat
 N  V  I  Q  H  I  N  R  A  K  D  K  N  H  M  I  I  S  I  D
gcagaaaaagcctttgacaaaattcaacaacccttcatgctaaaaactctcaataaatta
 A  E  K  A  F  D  K  I  Q  Q  P  F  M  L  K  T  L  N  K  L
ggtattgatgggacgtatttcaaaataataagagctatctatgacaaacccacagccaat
 G  I  D  G  T  Y  F  K  I  I  R  A  I  Y  D  K  P  T  A  N
atcatactgaatgggcaaaaactggaagcattccctttgaaaaccggcacaagacaggga
 I  I  L  N  G  Q  K  L  E  A  F  P  L  K  T  G  T  R  Q  G
tgccctctctcaccgctcctattcaacatagtgttggaagttctggccagggcaatcagg
 C  P  L  S  P  L  L  F  N  I  V  L  E  V  L  A  R  A  I  R
caggagaaggaaataaagggtattcaattaggaaaagaggaagtcaaattgtccctgttt
 Q  E  K  E  I  K  G  I  Q  L  G  K  E  E  V  K  L  S  L  F
gcagacgacatgattgtttatctagaaaaccccatcgtctcagcccaaaatctccttaag
 A  D  D  M  I  V  Y  L  E  N  P  I  V  S  A  Q  N  L  L  K
ctgataagcaacttcagcaaagtctcaggatacaaaatcaatgtacaaaaatcacaagca
 L  I  S  N  F  S  K  V  S  G  Y  K  I  N  V  Q  K  S  Q  A
ttcttatacaccaacaacagacaaacagagagccaaatcatgggtgaactcccattcaca
 F  L  Y  T  N  N  R  Q  T  E  S  Q  I  M  G  E  L  P  F  T
attgcttcaaagagaataaaatacctaggaatccaacttacaagggatgtgaaggacctc
 I  A  S  K  R  I  K  Y  L  G  I  Q  L  T  R  D  V  K  D  L
ttcaaggagaactacaaaccactgctcaaggaaataaaagaggagacaaacaaatggaag
 F  K  E  N  Y  K  P  L  L  K  E  I  K  E  E  T  N  K  W  K
aacattccatgctcatgggtaggaagaatcaatatcgtgaaaatggccatactgcccaag
 N  I  P  C  S  W  V  G  R  I  N  I  V  K  M  A  I  L  P  K
gtaatttacagattcaatgccatccccatcaagctaccaatgactttcttcacagaattg
 V  I  Y  R  F  N  A  I  P  I  K  L  P  M  T  F  F  T  E  L
gaaaaaactactttaaagttcatatggaaccaaaaaagagcccgcattgccaagtcaatc
 E  K  T  T  L  K  F  I  W  N  Q  K  R  A  R  I  A  K  S  I
ctaagccaaaagaacaaagctggaggcatcacactacctgacttcaaactatactacaag
 L  S  Q  K  N  K  A  G  G  I  T  L  P  D  F  K  L  Y  Y  K
gctacagtaaccaaaacagcatggtactggtaccaaaacagagatatagatcaatggaac
 A  T  V  T  K  T  A  W  Y  W  Y  Q  N  R  D  I  D  Q  W  N
```

```
agaacagagccctcagaaataatgccgcatatctacaactatctgatctttgacaaacct
 R  T  E  P  S  E  I  M  P  H  I  Y  N  Y  L  I  F  D  K  P
gagaaaaacaagcaatggggaaaggattccctatttaataaatggtgctgggaaaactgg
 E  K  N  K  Q  W  G  K  D  S  L  F  N  K  W  C  W  E  N  W
ctagccatatgtagaaagctgaaactggatcccttccttacaccttatacaaaaatcaat
 L  A  I  C  R  K  L  K  L  D  P  F  L  T  P  Y  T  K  I  N
tcaagatggattaaagatttaaacgttaaacctaaaaccataaaaaccctagaagaaaac
 S  R  W  I  K  D  L  N  V  K  P  K  T  I  K  T  L  E  E  N
ctaggcattaccattcaggacataggcgtgggcaaggacttcatgtccaaaacaccaaaa
 L  G  I  T  I  Q  D  I  G  V  G  K  D  F  M  S  K  T  P  K
gcaatggcaacaaaagacaaaattgacaaatgggatctaattaaactaaagagcttctgc
 A  M  A  T  K  D  K  I  D  K  W  D  L  I  K  L  K  S  F  C
acagcaaaagaaactaccatcagagtgaacaggcaacctacaacatgggagaaaatttt t
 T  A  K  E  T  T  I  R  V  N  R  Q  P  T  T  W  E  K  I  F
gcaacctactcatctgacaaagggctaatatccagaatctacaatgaactcaaacaaatt
 A  T  Y  S  S  D  K  G  L  I  S  R  I  Y  N  E  L  K  Q  I
tacaagaaaaaaacaaacaaccccatcaaaaagtgggcgaaggacatgaacagacacttc
 Y  K  K  K  T  N  N  P  I  K  K  W  A  K  D  M  N  R  H  F
tcaaaagaagacatttatgcagccaaaaaacacatgaagaaatgctcatcatcactggcc
 S  K  E  D  I  Y  A  A  K  K  H  M  K  K  C  S  S  S  L  A
atcagagaaatgcaaatcaaaaccactatgagatatcatctcacaccagttagaatggca
 I  R  E  M  Q  I  K  T  T  M  R  Y  H  L  T  P  V  R  M  A
atcattaaaaagtcaggaaacaacaggtgctggagaggatgcggagaaataggaacactt
 I  I  K  K  S  G  N  N  R  C  W  R  G  C  G  E  I  G  T  L
ttacactgttggtgggactgtaaactagttcaaccattgtggaagtcagtgtggcgattc
 L  H  C  W  W  D  C  K  L  V  Q  P  L  W  K  S  V  W  R  F
ctcagggatctagaactagaaataccatttgacccagccatcccattactgggtatatac
 L  R  D  L  E  L  E  I  P  F  D  P  A  I  P  L  L  G  I  Y
ccaaatgagtataaatcatgctgctataaagacacatgcacacgtatgtttattgcggca
 P  N  E  Y  K  S  C  C  Y  K  D  T  C  T  R  M  F  I  A  A
ctattcacaatagcaaagacttggaaccaacccaaatgtccaacaatgatagactggatt
 L  F  T  I  A  K  T  W  N  Q  P  K  C  P  T  M  I  D  W  I
aagaaaatgtggcacatatacaccatggaatactatgcagccataaaaaatgatgagttc
 K  K  M  W  H  I  Y  T  M  E  Y  Y  A  A  I  K  N  D  E  F
atatcctttgtagggacatggatgaaattggaaaccatcattctcagtaaactatcgcaa
 I  S  F  V  G  T  W  M  K  L  E  T  I  I  L  S  K  L  S  Q
Gaacaaaaaaccaaacaccgcatattctcactcataggtgggaattga*
 E  Q  K  T  K  H  R  I  F  S  L  I  G  G  N
```

EN1Sac1Fwd: 5'- ATT*GAGCTC*A**a**atgacaggtcaacttcacac-3'

EN1Xho1Rev: 5'- AAT*CTCGAG*tcatctctgaatctgagggtgtgt - 3'

EN2Nde1Fwd: 5'- ATT*CATATG*tcaaaagagacaaagaaggcca-3'

EN2Xho1Rev: 5'- AAT*CTCGAG*tcatctctgaatctgagggtgtgt - 3'

ENpeptideHindIIIFwd: 5'- AAT*AAGCTT*atgggagactttaacacc-3'

ENpeptideXho1Rev: 5'- AAT*CTCGAG*tcaggaataggtgtggtgtgg-3'

CCHCEcoR1Fwd: 5'- AAA*GAATTC*Ttcatcactggccatcagagaa -3'

CCHCXho1Rev: 5'- AAA*CTCGAG* attcccacctatgagtgagaga -3'

ATT, AAA and ATT sequences are the overhangs at the beginning for proper placement of restriction enzyme at the end of PCR amplified insert for proper digestion

*GAGCTC i*s Sac1 site

CTCGAG is Xho1 site

CATATG is Nde1 site

AAGCTT is HindIII site

GAATTC is EcoR1 site

a is an extra nucleotide added for making right frame tca stop codon

6.1.3. pET 30b-CCHC (PCR amplified fragment using primer pair CCHCEcoR1Fwd and CCHCXho1Rev)

atgcaccatcatcatcattcttctggtctggtgccacgcggttctggtatgaaagaaaccgctgctg
ctaaattcgaacgccagcacatggacagcccagatctgggtaccgacgacgacgacaaggccatggcgat
atcggatcc**gaattc**ttcatcactggccatcagagaaatgcaaatcaaaaccactatgagatatcatctc
acaccagttagaatggcaatcattaaaaagtcaggaaacaacaggtgctggagaggatgcggagaaatag
gaacacttttacactgttggtgggactgtaaactagttcaaccattgtggaagtcagtgtggcgattcct
cagggatctagaactagaaataccatttgacccagccatcccattactgggtatatacccaaatgagtat
aaatcatgctgctataaagacacatgcacacgtatgtttattgcggcactattcacaatagcaaagactt
ggaaccaacccaaatgtccaacaatgatagactggattaagaaaatgtggcacatatacaccatggaata
ctatgcagccataaaaaatgatgagttcatatcctttgtagggacatggatgaaattggaaaccatcatt
ctcagtaaactatcgcaagaacaaaaaaccaaacaccgcatattctcactcaTAGGTGGGAATcaccacc
accaccaccactga**ctcgag**…

Red: pET 30b vector sequence

Black: CCHC domain

6.1.4. ET-hL1CCHC protein sequence

M H H H H H H S S G L V P R G S G M K E T A A A K F E R Q H M D S P D L G T D D D D K A M A I S D P N S S S L A I R E M Q I K T T M R Y H L T P V R M A I I K K S G N N R C W R G C G E I G T L L H C W W D C K L V Q P L W K S V W R F L R D L E L E I P F D P A I P L L G I Y P N E Y K S C C Y K D T C T R M F I A A L F T I A K T W N Q P K C P T M I D W I K K M W H I Y T M E Y Y A A I K N D E F I S F V G T W M K L E T I I L S K L S Q E Q K T K H R I F S L I G GN H H H H H H *

Red letters are extra amino acids coming from vector backbone. Black letters indicate CCHC protein sequence. There are extra 52 amino acids at the N-terminal region.

Total no. of amino acids translated including vector amino acids: 237

Molecular weight and PI of the protein is 27.2kDa and 8.8 respectively.

6.1.5. pET 28c-EN1 (PCR amplified fragment using primer pairs EN1Sac1Fwd and EN1Xho1Rev)

atgggcagcagccatcatcatcatcacagcagcggcctggtgccgcgcggcagccatatggctagc
atgactggtggacagcaaatgggtcggatccgaattcgagctcaaatgacaggatcaacttcacacata
acaatattaactttaaatataaatggactaaattctgcaattaaaagacacagactggcaagttggata
aagagtcaagacccatcagtgtgctgtattcaggaaacccatctcacgtgcagagacacacataggctc
aaaataaaaggatggaggaagatctaccaagccaatggaaaacaaaaaaaggcaggggttgcaatccta
gtctctgataaaacagactttaaaccaacaaagatcaaaagagacaaagaaggccattacataatggta
aagggatcaattcaacaagaggagctaactatcctaaatatttatgcacccaatacaggagcacccaga
ttcataaagcaagtcctcagtgacctacaaagagacttagactcccacacattaataatgggagacttt
aacaccccactgtcaacattagacagatcaacgagacagaaagtcaacaaggatacccaggaattgaac
tcagctctgcaccaagcagacctaatagacatctacagaactctccacccccaaatcaacagaatataca
ttttttttcagcaccacaccacacctattccaaaattgaccacatagttggaagtaaagctctcctcagc
aaatgtaaaagaacagaaattataacaaactatctctcagaccacagtgcaatcaaactagaactcagg
attaagtgactcgag

Red: pET 28c vector sequence

Black: EN domain

6.1.6. pET-hL1EN1 protein sequence

M G S S H H H H H H S S G L V P R G S H M A S M T G G Q Q M G R I R I R A Q M T G S T S H I T I L T L N I N G L N S A I K R H R L A S W I K S Q D P S V C C I Q E T H L T C R D T H R L K I K G W R K I Y Q A N G K Q K K A G V A I L V S D K T D F K P T K I K R D K E G H Y I M V KG S I Q Q E E L T I L N I Y A P N T G A P R F I K

Q V L S D L Q R D L D S H T L I M G D F N T P L S T L D R S T R Q K V N K D T Q
E L N S A L H Q A D L I D I Y R T L H P K S T E Y T F F S A P H H T Y S K I D H I V
G S K A L L S K C K R T E I I T N Y L S D H S A I K L E L R I K

Red letters are extra amino acids coming from vector backbone. Black letters indicate EN protein sequence. There are extra 38 amino acids at the N-terminal region.

Total no. of amino acids translated including vector amino acids: 278

Molecular weight and PI of the protein is 31.4kDa and 9.8 respectively.

6.1.7. pET 28c-EN2 (PCR amplified fragment using primer pairs EN2Nde1Fwd and EN2Xho1Rev)

atgggcagcagccatcatcatcatcacagcagcggcctggtgccgcgcggcagcca
tatggctagcatgactggtggacagcaaatgggtcggatccgaattcgagctcaaatca
aaagagacaaagaaggccattacataatggtaaagggatcaattcaacaagaggagcta
actatcctaaatatttatgcacccaatacaggagcacccagattcataaagcaagtcct
cagtgacctacaaagagacttagactcccacacattaataatgggagactttaacaccc
cactgtcaacattagacagatcaacgagacagaaagtcaacaaggatacccaggaattg
aactcagctctgcaccaagcagacctaatagacatctacagaactctccaccccaaatc
aacagaatatacatttttttcagcaccacaccacacctattccaaaattgaccacatag
ttggaagtaaagctctcctcagcaaatgtaaaagaacagaaattataacaaactatctc
tcagaccacagtgcaatcaaactagaactcaggattaagtgactcgag

Red: pET 30b vector sequence

Black: CCHC domain

6.1.8. pET-hL1EN2 protein sequence

M G S S H H H H H H S S G L V P R G S H M A S M T G G Q Q M G R I R I R A Q I
K R D K E G H Y I M V K G S I Q Q E E L T I L N I Y A P N T G A P R F I K Q V L S
D L Q R D L D S H T L I M G D F N T P L S T L D R S T R Q K V N K D T Q E L N S
A L H Q A D L I D I Y R T L H P K S T E Y T F F S A P H H T Y S K I D H I V G S K A
L L S K C K R T E I I T N Y L S D H S A I K L E L R I K*

Red letters are extra amino acids coming from vector backbone. Black letters indicate EN protein sequence. There are extra 38 amino acids at the N-terminal region.

Total no. of amino acids translated including vector amino acids: 189

Molecular weight and PI of the protein is 21.2kDa and 9.8 respectively.

6.1.9. pET 30b-EGFP (PCR amplified fragment using primer pairs EGFPBamH1Fwd and EGFPHindIIIRev)

atgcaccatcatcatcattcttctggtctggtgccacgcggttctggtatgaaaga
aaccgctgctgctaaattcgaacgccagcacatggacagcccagatctgggtaccgacg
acgacgacaaggccatggcgatatcggatccaccggtcgccaccatggtgagcaagggc
gaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacgg
ccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccc
tgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccacc
ctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgactt
cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacg
acggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgc
atcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctgga
gtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatca
aggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccac
taccagcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccactacct
gagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgc
tggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtaa

Red: pET 30b vector sequence

Black: EGFP gene

6.1.10. pET- EGFP protein sequence

M H H H H H H S S G L V P R G S G M K E T A A A K F E R Q H M D S P D L G T D
D D D K A M A I S D P P V A T M V S K G E E L F T G V V P I L V E L D G D V N G
H K F S V S G E G E G D A T Y G K L T L K F I C T T G K L P V P W P T L V T T L T
Y G V Q C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E R T I F F K D D G N
Y K T R A E V K F E G D T L V N R I E L K G I D F K E D G N I L G H K L E Y N Y N
S H N V Y I M A D K Q K N G I K V N F K I R H N I E D G S V Q L A D H Y Q Q N T P
I G D G P V L L P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G
I T L G M D E L Y K*

Red letters are extra amino acids coming from vector backbone. Black letters indicate EN protein sequence. There are extra 54 amino acids at the N-terminal region.

Total no. of amino acids translated including vector amino acids: 293

Molecular weight and PI of the protein is 32.2kDa and 5.7 respectively.

6.1.11. pET 30b-EGFP+EN (PCR amplified fragment using primer pairs ENHindIIIFwd and ENXho1Rev)

atgcaccatcatcatcattcttctggtctggtgccacgcggttctggtatgaaagaaacc
gctgctgctaaattcgaacccagcacatggacagcccagatctgggtaccgacgacgacaca
aggccatggcgatatcggatccACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCA
CCGGGGTGGTGCCCATCCTGGTCGACTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTC

CGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAG
CCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGT
CCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTT
CGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA
GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCA
GCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAA
CCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGT
CCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGaagct
tATGGGAGACTTTAACACCCCACTGTCAACATTAGACAGATCAACGAGACAGAAAGTCAACAA
GGATACCCAGGAATTGAACTCAGCTCTGCACCAAGCAGACCTAATAGACATCTACAGAACTCT
CCACCCCAAATCAACAGAATATACActcgagcaccaccaccaccaccactgagatccggctgc
taa

Red: pET 30b vector sequence

Black: EGFP gene

Blue: EN peptide

6.1.12. pET- EGFP+EN protein sequence

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTD
DDDKAMAISDPPVATMVSKGEELFTGVVPILVELDGDVNG
HKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTT LT
YGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGN
YKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYN
SHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTP
IGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEF VTAA
GITLGMDELYKKLMGDFNTPLSTLDRSTRQKVNKDTQELN
SALHQADLIDIYRTLHPKSTEYTFFS AP HH TYS*

Red letters are extra amino acids coming from vector backbone. Green and blue letters indicate EGFP and EN protein sequence. There are extra 54 amino acids at the N-terminal region.

Total no. of amino acids translated including vector amino acids: 355

Molecular weight and PI of the protein is 39.9kDa and 5.9 respectively.

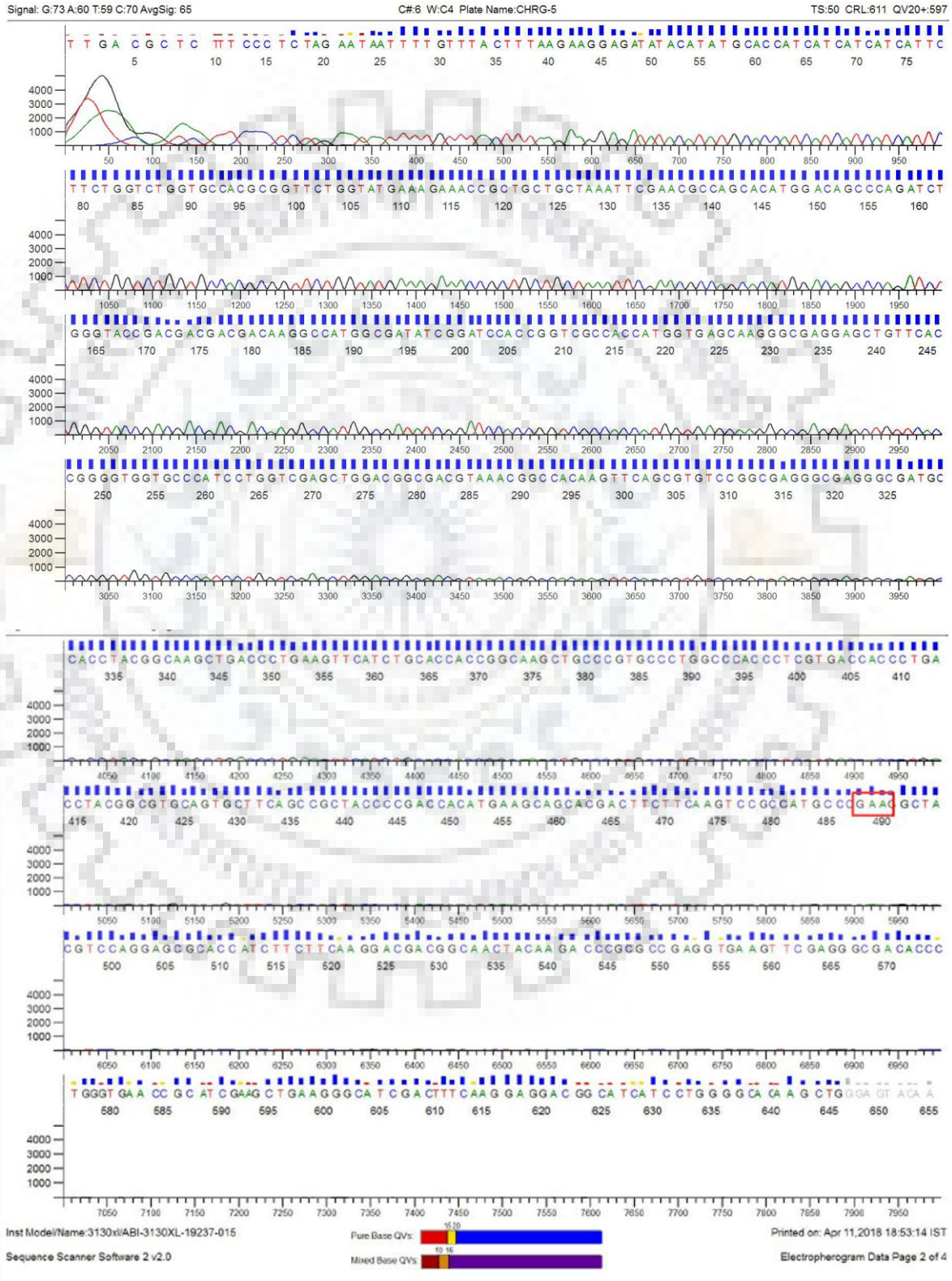
6.2. Appendix II

6.2.1. Table 1- Primers used in this study

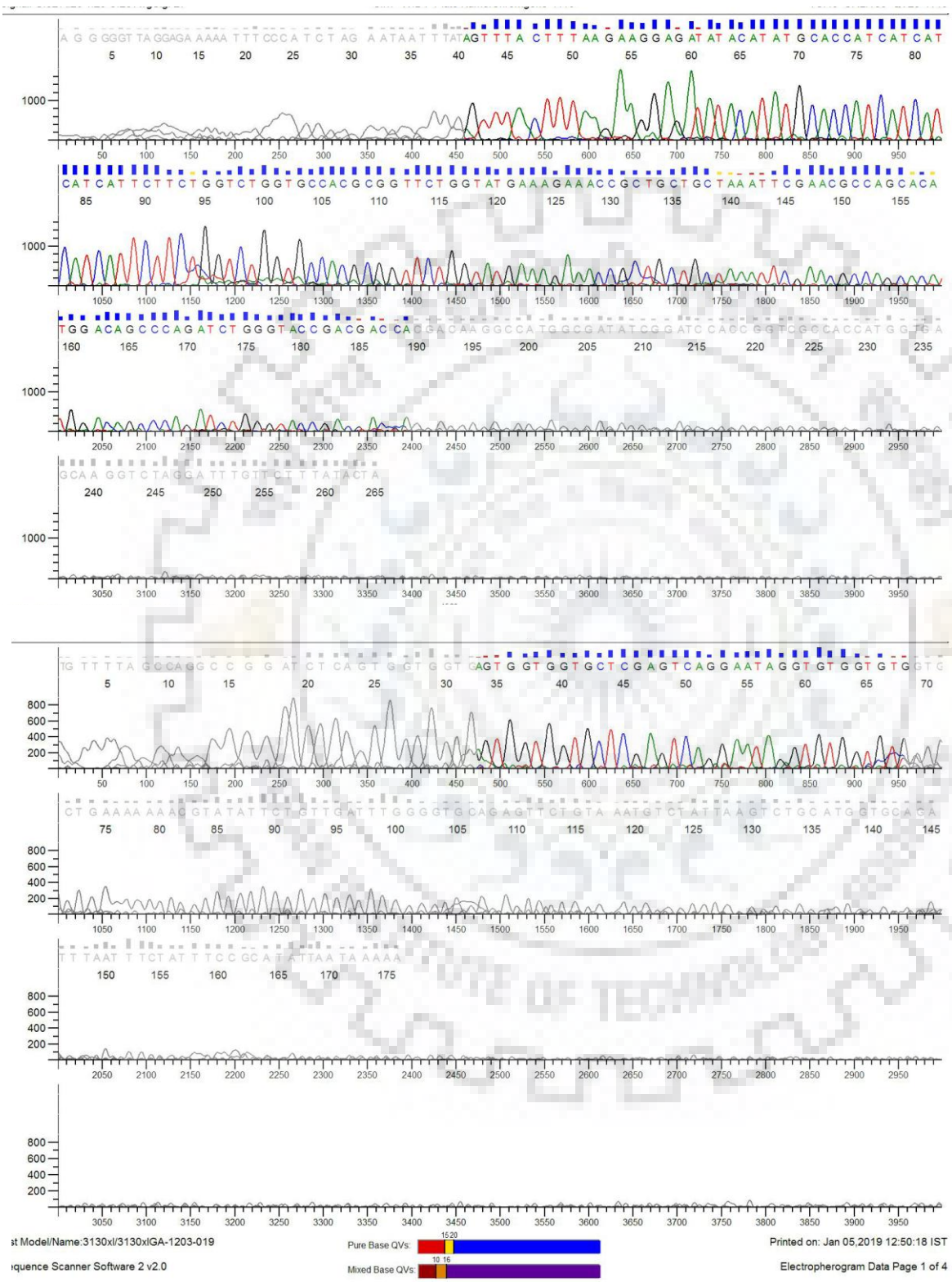
S.no	Primer name	Primer sequence
1	1F	5'- AAAGAATTCTTCATCACTGGCCATCAGAGAA -3'
2	1R	5'- AAACTCGAGATTCCCACCTATGAGTGAGAGA -3'
3	2F	5'-ATTGAGCTCAAATGACAGGTCAACTTCACAC-3'
4	2R	5'-CTCGGATCCTCACTTAATCCTGAGTTCTAG-3'
5	3F	5'-ATTCATATGTCAAAAGAGACAAAGAAGGCCA-3'
6	3R	5'-CTCGGATCCTCACTTAATCCTGAGTTCTAG-3'
7	4F	5'- AATGGATCCACCGGTCGCCACC -3'
8	4R	5'- AATAAGCTTCTTGTACAGCTCGTCCAT-3'
9	5F	5'- AATAAGCTTATGGGAGACTTTAACACC-3
10	5R	5'- AATCTCGAGTCAGGAATAGGTGTGGTGTGG-3'
11	6F	5'-CCATGCCCGAAGCATACGTCCAGGAGCGC-3'
12	6R	5'-CTCCTGGACGTATGCTTCGGGCATGGCGGAC-3'
13	7F	5'- CCATGCCCGAAGATTACGTCCAGGAGCGC -3'
14	7R	5'-CTCCTGGACGTATGCTTCGGGCATGGCGGAC-3'
15	8F	5'- CCATGCCCGAACGTTACGTCCAGGAGCGC -3'
16	8R	5'-CTCCTGGACGTATGCTTCGGGCATGGCGGAC-3'
17	9F	5'- CCATGCCCGAATATTACGTCCAGGAGCGC -3'
18	9R	5'-CTCCTGGACGTATGCTTCGGGCATGGCGGAC-3'

6.3. Appendix III

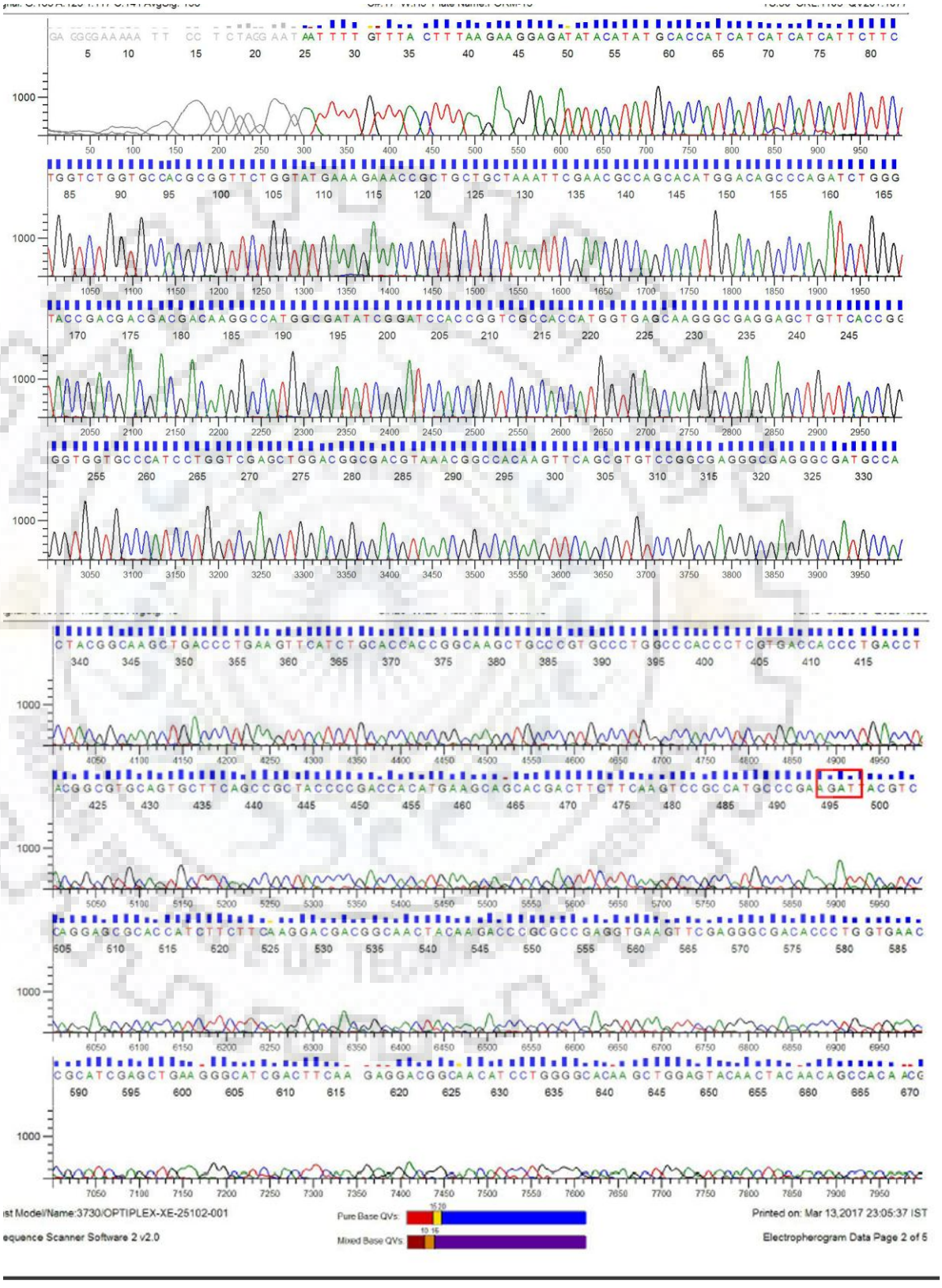
6.3.1. Sanger sequencing chromatogram showing cloned sequencing result of pET30b-EGFP



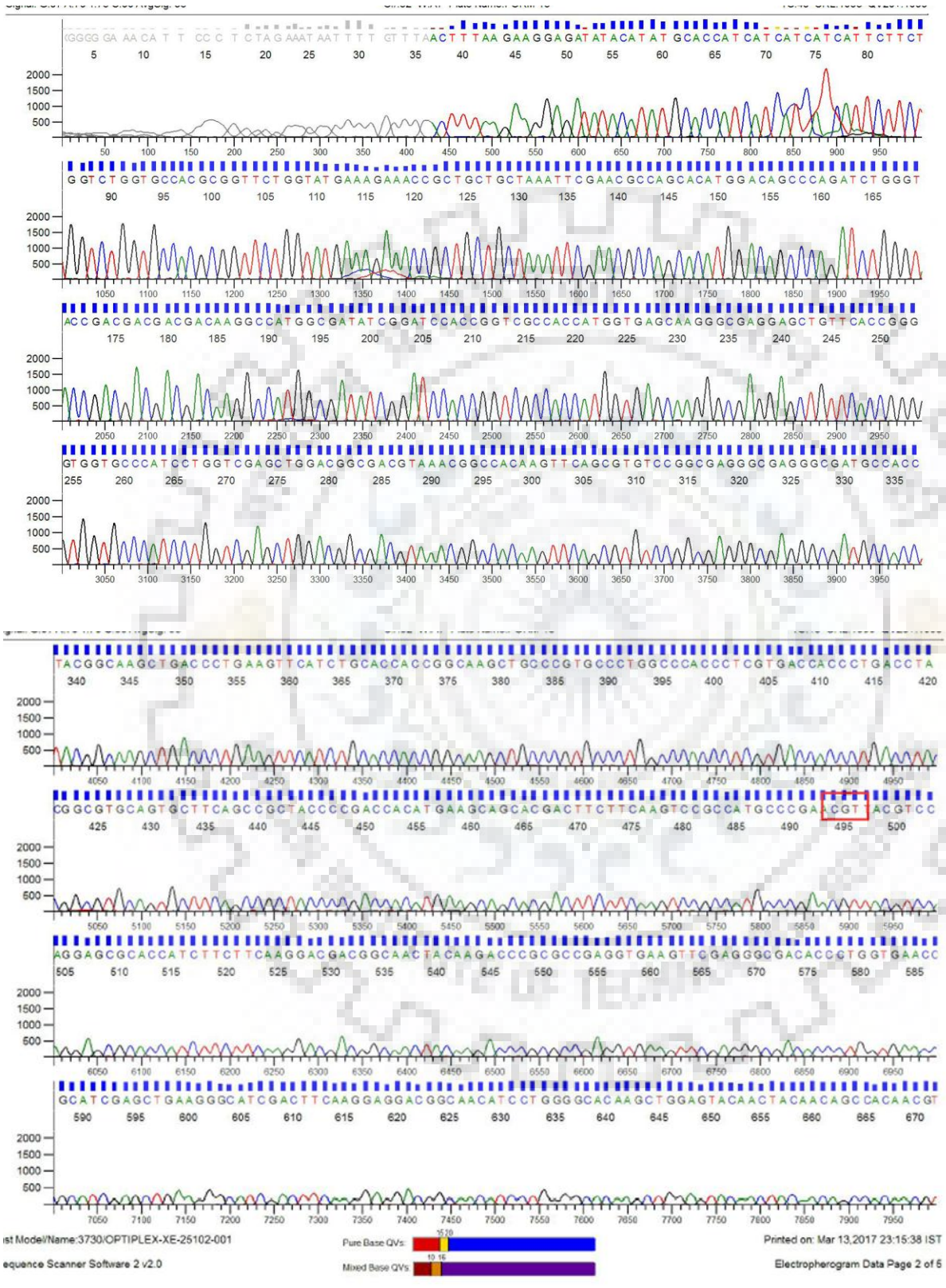
6.3.2. Sanger sequencing chromatogram showing cloned sequencing result of pET30b-EGFP+EN



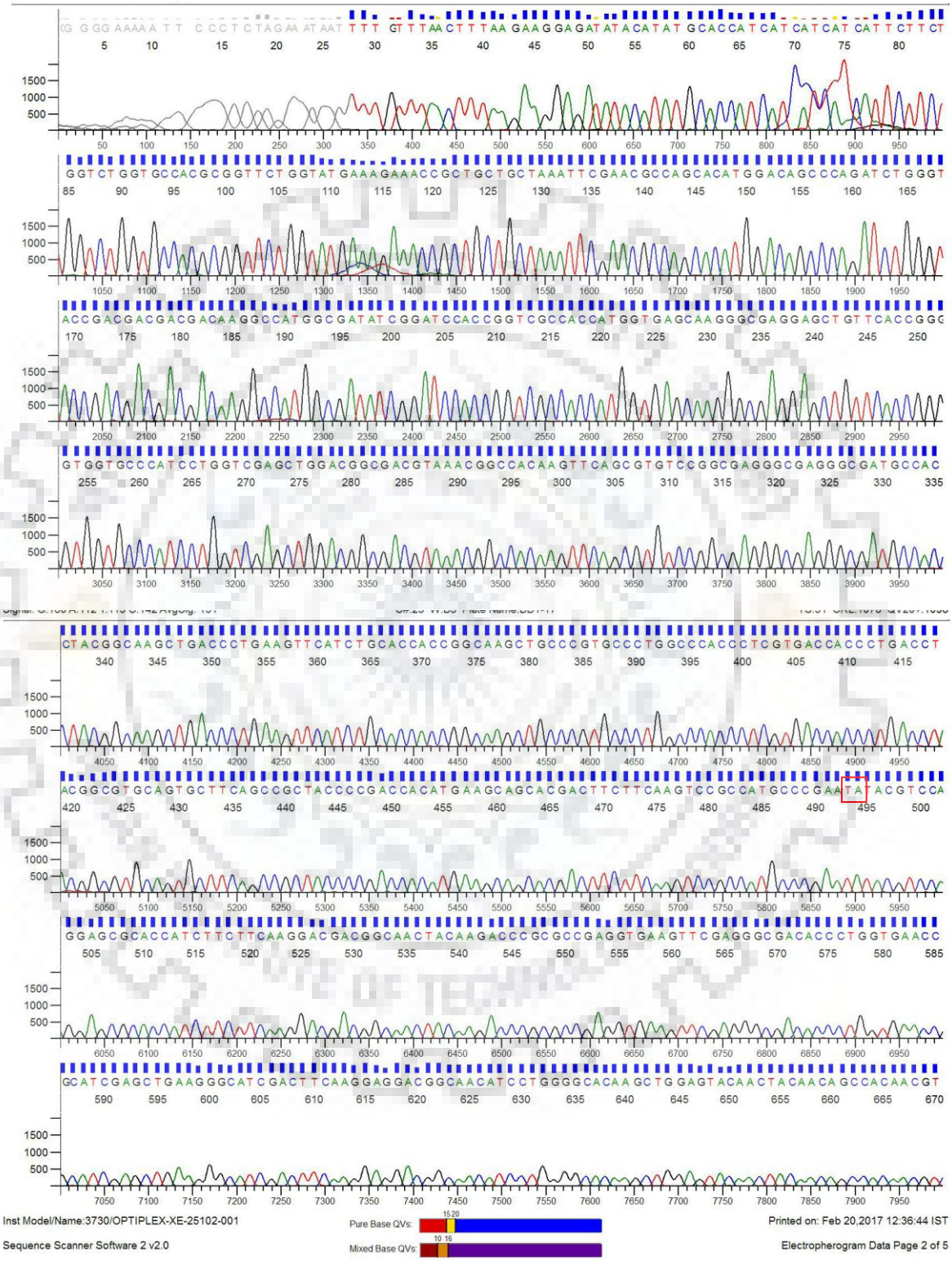
6.3.3. Sanger sequencing chromatogram showing Glycine to Aspartic acid in EGFP (GGC- GAT) at 92 position



6.3.4. Sanger sequencing chromatogram showing Glycine to Arginine in EGFP (GGC- CGT) at 92 position



6.3.5. Sanger sequencing chromatogram showing Glycine to Tyrosine in EGFP (GGC- TAT) at 92 position



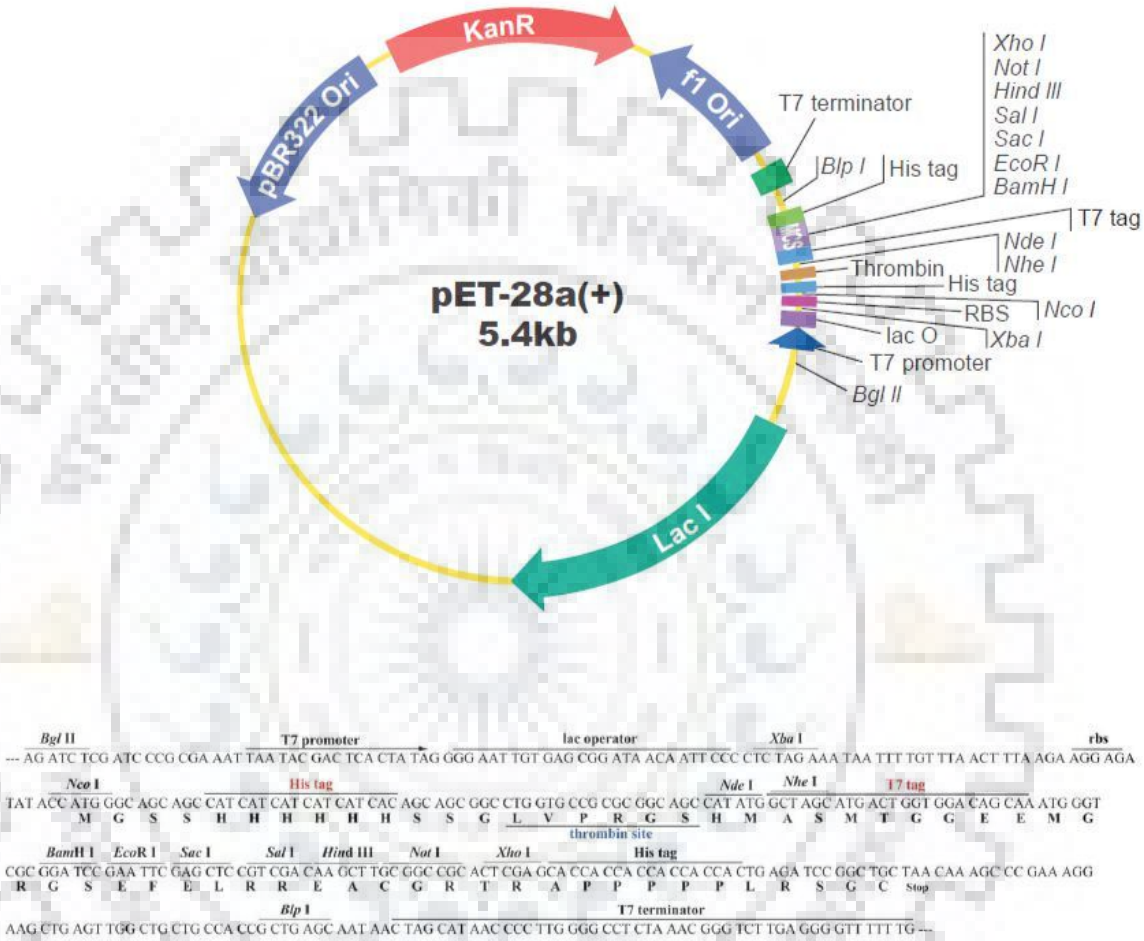
6.4. Appendix IV

6.4.1. Supplementary Table 2: The details of patients used in this study

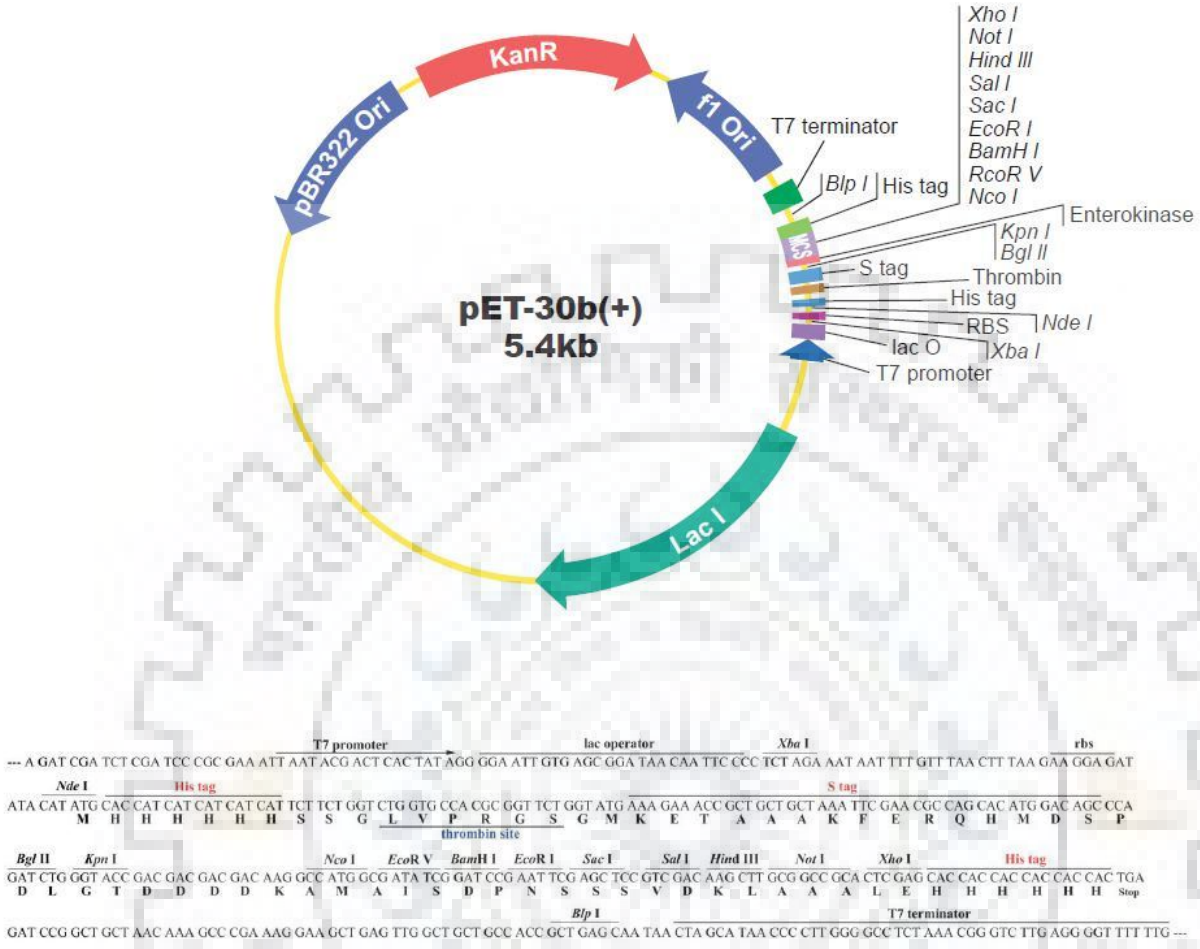
Serial no.	Age	Sex	Tissue type	ORF1p	ORF2	Grade	Stage
1	45	Male	tongue	negative	ND	1	I
2	70	Male	buccal mucosa	negative	ND	1	II
3	29	Male	buccal mucosa	positive	positive	2	II
4	35	Male	buccal mucosa	positive	positive	1	II
5	53	Male	buccal mucosa	negative	ND	1	I
6	58	Male	buccal mucosa	negative	ND	1	III
7	46	Male	buccal mucosa	negative	ND	1	II
8	54	Male	tongue	positive	positive	1	I
9	67	Male	buccal mucosa	positive	positive	2	III
10	42	Male	tongue	positive	ND	1	I
11	45	Female	buccal mucosa	positive	ND	2	I
12	45	Male	buccal mucosa	positive	positive	2	I

6.5. Appendix III

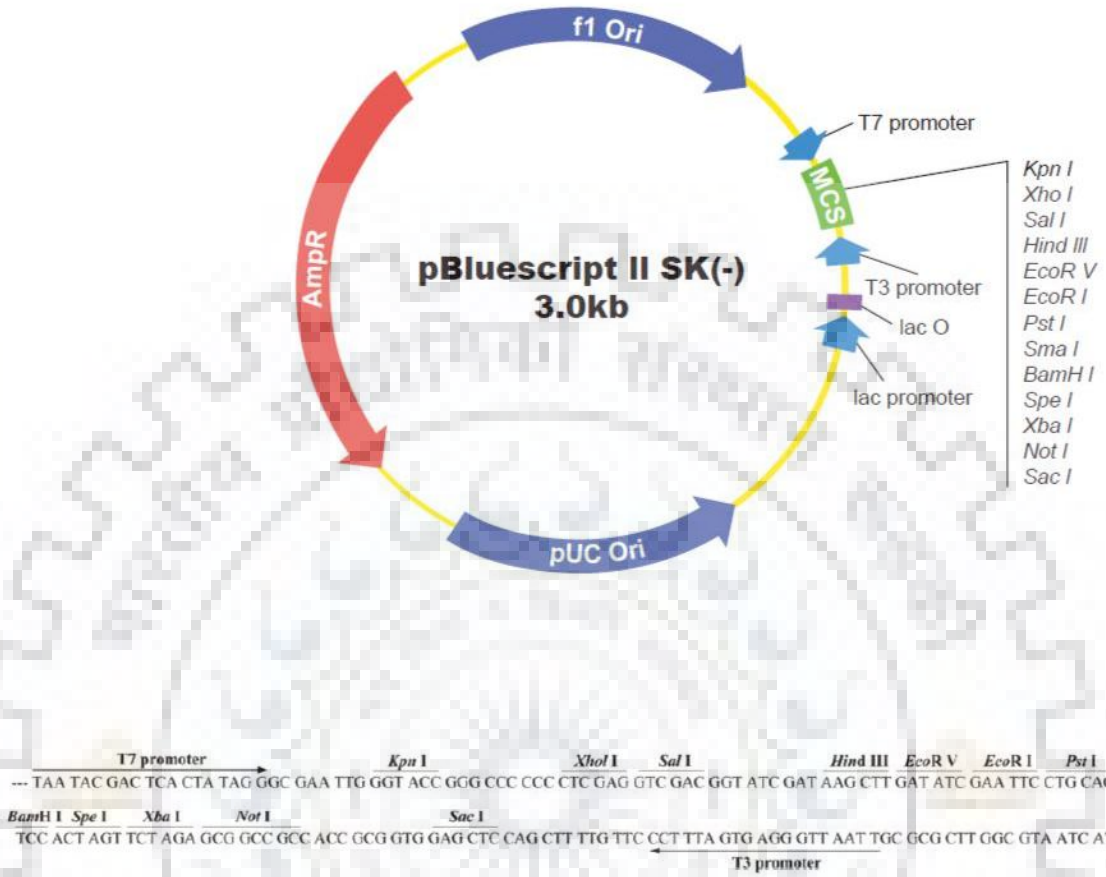
6.5.1. Plasmids used in this study



Plasmid map of pET-28a cloning and expression vector



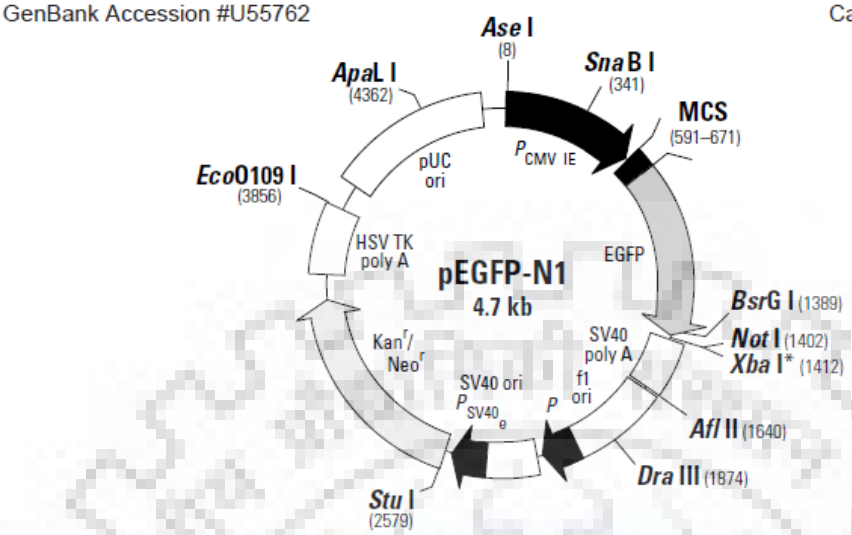
Plasmid map of pET-30b cloning and expression vector

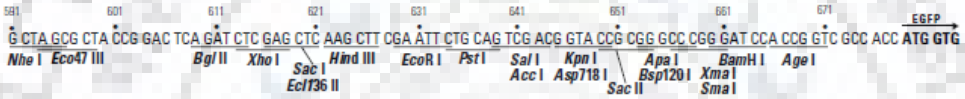


Plasmid map of pBS cloning vector

pEGFP-N1 Vector Information PT3027-5

GenBank Accession #U55762 Catalog #6085-1





Restriction Map and Multiple Cloning Site (MCS) of pEGFP-N1 Vector. (Unique restriction sites are in bold.) The *Not* I site follows the EGFP stop codon. The *Xba* I site (*) is methylated in the DNA provided by CLONTECH. If you wish to digest the vector with this enzyme, you will need to transform the vector into a *dam*+ and make fresh DNA.

Plasmid map of pEGFP-N1 cloning vector

6.6. Appendix VI

6.6.1. Poster Abstract I

Indo-US conference on 'Transcription, Chromatin Structure DNA Repair and Genomic Instability'

Indian Institute of Science, Bangalore, India. March 2018

Title: Generation of human LINE-1 retrotransposon encoded ORF2p antibody and its detection in cancer.

Sofia Pilli, Koel Mukherjee, Debpali Sur, Savita Budania and Prabhat K Mandal

Department of Biotechnology, Indian Institute of Technology, Roorkee, Uttrakhand, India

ABSTRACT

Oral Squamous Cell Carcinoma (OSCC) is very deadly and common cancer in India due to excessive use of tobacco. In order to find out factors contributing OSCC onset and progression; we analyzed the activity of a retrotransposon called Long INterpersed Element (LINE1 or L1) in OSCC samples from Indian patient. Retrotransposons are sequence which moves within a genome thus can disrupt essential genes. L1 is the only active retrotransposon with almost 500,000 copies present in human genome. An active L1 is 6 kb in length transcribe from an internal promoter and encodes two proteins designated as ORF1p and ORF2p. The jumping of L1 RNA requires the formation of L1-ribonucleo protein particles (L1-RNPs) where both the proteins bind with L1RNA in *cis*. In general, L1 is not transcribed and thus silent in normal somatic cells due to methylation of L1 promoter but is highly active in few cancers. Recent work from our laboratory showed L1 promoter hypomethylation and ORF1p expression in OSCC cancer samples. However, no studies have been performed to detect L1 encoded ORF2p as there is no commercial antibody available for this protein which is notoriously difficult to express. Here, we have synthesize few peptides from ORF2p and injected to rats to generate antibody against human ORF2p. The antibody will be used to detect whether ORF2p is expressed in ORF1p positive OSCC samples. Overall this study will generate a valuable reagent like ORF2p antibody to study the biology of human L1 retrotransposon.

6.6.2. Poster Abstract II

National Conference on Protein Structure and Dynamics in Health and Agriculture

Jamia Millia University, New Delhi, India. November 2017

Title: Generation of Human LINE1 retrotransposon encoded ORF2p antibody and its detection in human tissues

Sofia Pilli, Koel Mukherjee, Debpali Sur, Savita Budhania and Prabhat K Mandal

ABSTRACT

Retrotransposons are sequences which move from one place of the genome to another place using RNA as an intermediate and the process is called retrotransposition. The Long Interspersed Element 1 (LINE1 or L1), a type non-LTR retrotransposon is responsible for almost half of the human DNA in mass. It is still actively jumping in different parts of normal human brain and in certain types of cancer. An active L1 is 6 kb in length contains two open reading frames designated as L1-ORF1p and L1-ORF2p. The L1- ORF1p encodes a 40 kDa protein with single stranded nucleic acid binding activities where as ORF2p encodes a 150 kDa protein with demonstrated reverse transcriptase (RT) and endonuclease (EN) activities. Both proteins are critical for the process of retrotransposition. Recently we made the antibody of ORF1pand detected significant ORF1pexpression in different parts of normal human brain and oral cancer samples. Our data suggest that L1 retrotransposon might be very active in those tissues tested, although we don't know whether L1-ORF2p is also expressing in those tissues. Very little is known regarding humanL1-ORF2p as this protein is notoriously difficult to express. The L1-ORF2p has three distinct domains and N-terminal EN, central RT and C- terminal CCHC type of DNA binding domains. No commercial antibody of L1-ORF2p is available in the market. By employing bio-informatic approaches we have selected few fragments from human L1ORF2p, cloned in bacterial expression vector and checked its expression. Among all fragments, only two ORF2p fragments showed significant expression which we are purifying in homogeneity in order to raise antibody against that fragments. The ORF2p antibody if generate will be a valuable reagent to study the biology of human L1 retrotransposon.