INVESTIGATING HUMAN L1 RETROTRANSPOSON ACTIVITY IN ORAL SQUAMOUS CELL CARCINOMA PATIENTS

Ph.D. THESIS

by

SAVITA BUDANIA

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE – 247667 (INDIA)
DECEMBER, 2017

INVESTIGATING HUMAN L1 RETROTRANSPOSON ACTIVITY IN ORAL SQUAMOUS CELL CARCINOMA PATIENTS

A THESIS

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

of

DOCTOR OF PHILOSOPHY

in

BIOTECHNOLOGY

by

SAVITA BUDANIA

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE – 247 667 (INDIA)
December, 2017

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **"INVESTIGATING HUMAN L1 RETROTRANSPOSON ACTIVITY IN ORAL SQUAMOUS CELL CARCINOMA PATIENTS"** 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, 2013 to December, 2017 under the supervision of Dr. Prabhat K Mandal, Assistant Professor and Dr. Ramasare Prasad, 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 Institute.

Signature of the candidate

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

Signature of Supervisor **Signature of Supervisor**

The Ph.D. viva-voce Examination of Savita Budania, Research scholar has been held on 29 may 2018.

Chairman, SRC **Signature of external examiner**

This is to certify that the student has made all the corrections in the thesis.

Signature of supervisors **Head of the department**

Dated:

Abstract

Mobile genetic elements also known as jumping genes or transposable elements were discovered by Barbara McClintock in 1956 during her experiments with maize crop. These are the piece of DNA sequences which can move around; can jump from one place to another place of the genome in the same cell. During that time this was a very unconventional concept opposing the classical genetics theory that genes are static and have a particular fixed locus in a specific chromosome like beads in a string. Over subsequent years the same kind of phenomenon discovered in other organism including mouse and human, changes the concept that some DNA sequences can move within a genome and thus can cause mutation if it disrupts an essential gene.

Genome sequencing of human, mouse and other higher eukaryotic organisms revealed that around 1-2% of the total genome encodes proteins whereas different kind of repetitive sequences occupied more than half of the genome. These repeats are mainly belong to the transposable element which can be classified as 1) DNA transposons which move as such either by cut and paste or copy and paste mechanism and 2) Retrotransposons which use RNA as an intermediate to jump from one place of the genome to another place. Retrotransposons are further divided into LTR and Non-LTR retrotransposons on the basis of long terminal repeats (LTR) present or not. In human genome the DNA transposons (~3% of total genome) and the LTR transposons (~9%) are no longer active. The non-LTR transposons are of great interest as they are actively transposing in the human genome. There are two types of non-LTR transposons, autonomous (transposes using their own machinery) and non-autonomous (required machinery of autonomous elements for their transposition). Long Interspersed Element 1 or LINE1 are the only active autonomous non-LTR retrotransposon and occupies around 20% of the human genome. Transposable elements have been studied in many organisms since its discovery by Barbara McClintock but in human it was in 1988 when Kazazian *et al.* (Nature 1988, V332, 164-166) noticed a haemophilia A patient resulting from de-novo insertion of LINE-1 sequence without any pedigree for the disease. Sequencing of the factor VIII gene from patient showed LINE1 inserted in the exon 14 and which was the actual cause for the disease. Subsequently, Kazazian laboratory cloned the L1 sequence which disrupted factor VIII gene from patient DNA and then showed the element is highly

active in cell culture based retrotransposon assay. These observations confirmed for the first time that the transposable element is active in the recent day human genome.

An active L1 is 6.0 kb in length, containing a 900 base pairs (bp) 5´-untranslated region (5'-UTR) with an internal promoter activity, two open reading frames designated as ORF1p and ORF2p separated by a small 63 bp inter ORF spacer sequence and followed by a ~200 bp 3'-UTR. Although the functions of ORFs are poorly understood, both proteins are critical in the process of retrotransposition. ORF1 encoded protein showed *in vitro* single stranded nucleic acid binding and nucleic acids chaperone activities whereas ORF2 encodes a protein with reverse transcriptase (RT) and endonuclease (EN) activities.

It was a general believe that LINE-1 retrotransposons are only active in germ cells (sperm and ovum) and at early stage of development. It is also believed that LINE-1 as a parasite is active in germ cells for its propagation to the next generation. But recent high throughput sequencing analysis revealed that L1 is also active in certain parts of normal brain and in few cancers. The activity of L1 is high in those cancers which are epithelial origin. 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. No study has been performed to see the activity of L1 retrotransposons in oral cancer samples. **In this study, L1 retrotransposon activity has been investigated in oral cancer samples obtained from Indian patient.**

The thesis has been divided into four 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 activity in different types of cancer along with literature about oral cancer is included.

Chapter 2 comprises the materials and methods used in the research work, Those includes recipes for reagents, solutions, protocols for cloning, expression and purification of proteins, protocols for antibody generation, immunohistochemistry, western blotting and methylation studies and others.

Chapter 3 contains details of the results obtained in the study. The main objective of the study was "To find out human L1 retrotransposon activity in oral

cancer samples obtained from Indian patient". The main objective was answered by performing following sub-objectives.

i) To make antibody against human LINE1 proteins (ORF1p and ORF2p) and characterization of reverse transcriptase (RT) activity encoded by RT domain of ORF2p

ORF2 protein of human LINE1 contains three domains:- Endonuclease (EN) domain, Reverse transcriptase (RT) domain and CCHC domain. In the present study different size L1ORF2 fragments containing RT domain was cloned in a bacterial expression vector and its expression was checked in *E.coli* expression cells. The results showed that RT domain protein was expressed enough in bacterial expression system and due to mis-folding the protein formed inclusion bodies. Next, the RT domain protein was solubilized using urea and purified by Ni-agarose chromatography. After purification the refolding of the protein showed formation of inclusion bodies. Checking soluble fraction showed less 1% induced RT was still in soluble fraction. The purified protein from soluble fraction showed significant RT activity on L1 RNA template.

To get the antibody against ORF2p, the partially purified RT domain protein was separated in denatured SDS-PAGE gel and the band corresponds to RT domain protein was injected to rabbit. Immunoblot analysis using partially purified RT domain protein didn't show any band suggesting that injected protein was not immunogenic to rabbit.

Since human L1 RT domain protein formed inclusions bodies and didn't make any antibody in rabbit, next I tried to clone, express and purify human L1 ORF1p to generate antibody against it. The human ORF1 fragment was sub-cloned in bacterial expression vector and expression studies showed significant expression of ORF1p (~40kDa) in bacterial system. Although, the protein was expressed in significant amount, it was not purified in homogeneity both in Ni-agarose, as well as in gel filtration chromatography. Simultaneously, antibody against RRM domain (30 kDa) of ORF1p was being tried in the laboratory and RRM domain antigen showed good antibody response in rabbit. So detection of ORF1p in OSCC samples (described below) was performed using ORF1p RRM domain specific antibody [α-hORF1p (RRM)].

ii) To investigate the L1 promoter methylation status and LINE1 retrotransposon activity in OSCC samples.

OSCC samples were collected from Acharya Tulsi Regional Centre for Cancer and Treatment Bikaner, Rajasthan. All the experiments using cancer samples 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 hematoxylene and eosin staining. Next, the samples were proceeded to make paraffin block. Slides made from these blocks were proceeded for immunohistochemistry with ORF1p RRM domain specific antibody [α-hORF1p (RRM)]. Around 60% samples showed ORF1p positive suggesting human L1 retrotransposon pathway is highly active in OSCC samples in the cancer tissues compared to normal.

Epigenetic silencing of the L1 5'-UTR by DNA methylation is a common means to inactivate L1 expression and ultimately retrotransposition. Epigenetic alterations are frequent in cancers; indeed, several studies have reported L1 promoter Hypomethylation in a variety of cancers. To date, the methylated state of the L1 5'-UTR in OSCC remained unexamined; therefore bisulfite conversion analysis of genomic DNA across nine paired normal-cancer tissues followed by PCR, subcloning of amplicons, and Sanger sequencing to ascertain the methylation level of the L1 promoter were performed. Specifically, a 363 bp region of the L1 promoter (nucleotide sequence 209-572, L1HS from Repbase) was amplified which contains 20 CpG sites and the resultant amplicons were sequenced. Investigating L1 promoter methylation status, showed significant hypomethylation of L1 promoter in cancer tissues compared to its normal counterpart. Overall, the data shows 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 prospectives of the work has been discussed.

Acknowledgements

I would like to express my profound gratitude to all the wonderful people I had the opportunity to work with during my Ph.D. work. First and foremost I would like to take immense pleasure in expressing my deep sense of gratitude to my supervisor **Dr. Prabhat K. Mandal** for providing me the opportunity to work in his lab. I also want to thank 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 has always kept my spirits high.

I would like to thank Professor R. Prasad for the constant support for doing my experiments. I also express my sincere gratitude towards Professor Partha Roy and Dr. A. K. Sharma for their kind and valuable suggestions. I want to express my gratitude to the current head of the department Prof Partha Roy, who has always been helpful and available for having an open door and encouraging attitude.

I take this opportunity to sincerely acknowledge the Department of Biotechnology (DBT), Government of India for providing me the financial support.

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. Shilpi Saxena (Pathologist, Military hospital, Roorkee) for helping me with the suggestions whenever needed and providing her lab facilities for the experimental work.

I would like to express my thanks and hearty wishes to my labmates Sofia, Debpali, Koel and Keyur for their help and support during the experiments. My warmest thanks to my colleagues and friends at IIT Roorkee and outside it, specially Dr. Manju, Swati Choudhary, Jyoti, Pooja, Anchal, Himanshu, Kaushik, Apurva, Anjlika, Raj kishore, Rashmi, Poonam, Manish, Krishankant and Dr Ashwani for making my tenure at IIT Roorkee memorable.

I am specially thankful for my husband Dr. Vinay Tomar for his constant encouragement and support. He has supported me in the most stressful times and

always cheered me to make me strong. I also extend my thanks to each member of my in-laws family, father in law, Shri Om Singh, Mother in law, Smt. Jaswati Devi, Brother in laws Mr. Vipin Tomar and Dr. Amit Tomar, sister in laws Smt. Sushma and Dr. Payal who supported me in every possible way to see the completion of this work.

Finally I dedicate my thesis to my lovely parents **Sh. Devraj Budania** and **Smt. Durga Rani** who always believed me and supported me in every way they could. I thank my younger siblings Pradeep and Kavita for their unconditional love and affection.

Savita Budania

Publication

1. Sur, D., Kustwar, R.K.*, **Budania, S*.,**Mahadevan, A., Hancks, D.C., Yadav, V., Shankar, S.K., Mandal, P.K. 2017. Detection of the LINE-1 retrotransposon RNA-binding protein ORF1p in different anatomical regions of the human brain.Mob DNA. 22;8:17 **(* Equal contribution)**.
2. **Budania S**, Sur D, Nangal J, Saxena S, Biswas M, Prasad R, Yadav V, Hancks D, and Mandal PK. Extensive LINE-1 retrotransposon promoter hypomethylation and ORF1p expression in oral squamous cell carcinoma. The Febs Journal (Under Revision from December 2017).
3. **Budania S** and Mandal PK. Characterization of Human LINE1 encoded Reverse Transcriptase. (Manuscript under preparation).

Conferences Attended:

1. Characterization of Human LINE1 encoded Reverse Transcriptase. **Savita Budania**, Ramasre Prasad and Prabhat K Mandal. March 2014, **5th Annual International Conference on Advances in Biotechnology (BIOTECH 2015), Indian Institute of Technology Kanpur, Kanpur India March 2015**

2. L1 retrotransposon activity in head and neck cancer patients. **Savita Budania**, Debpali Sur, Jitendra Nangal, Prabhat K Mandal. March 2017, **"Biotech Day 2017", Department of Biotechnology Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India**

3. Extensive LINE-1 promoter hypomethylation and ORF1p expression in Oral squamous cell carcinoma **Savita Budania**, Debpali Sur, Shilpi Saxena, Jitendra Nangal, Manash Biswas, Ramasare Prasad, Vijay Yadav, Dustin C. Hancks, and Prabhat K. Mandal. October 2017. **"The Mobile Genome: Genetic and Physiological Impact of Transposable Element" EMBO EMBL Symposium, Heidelberg, Germany.**

LIST OF ABBREVIATION

μg: microgram

Ac: Activator

APEs: apurinic/apyrimidinic type endonucleases

BSA: Bovine Serum Albumin

CC: Coiled Coil

CTD:CarboxyTerminal Domain

CTD:CarboxyTerminal Domain

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: millimeter

mM: millmolar

MW: Molecular Weight

NaCl: sodium chloride

NRTIs: Nucleoside Reverse-Transcriptase Inhibitors

oC: 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

SDS-PAGE: Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis

SVA: (SINE-R/VNTR/Alu)

TPRT:Target-Primed Reverse Transcription

TBST: TrisBuffered 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

Abstract	**i**
Acknowledgements	**v**
List of Publications	**vii**
List of Abbreviations	**ix**
List of Contents	**xiii**
List of Figures and Tables	**xvii**
1. INTRODUCTION	
1.1 Transposable elements and its discovery	**1**
1.2 Types of transposable elements in mammalian genome	**2**
1.3 Structure of Human L1 transposable element	**6**
1.3.1 Human L1 promoter	**7**
1.3.2 Human L1 encoded proteins	**7**
1.3.2.1 L1-ORF1p	**7**
1.3.2.2 L1-ORF2p	**8**
1.3.2.3 L1-ORF0	**9**
1.4 Mechanism of LINE 1 retrotransposition	**10**
1.5 Restriction against human mobile elements:	**11**
1.6 Effects of L1 on Human Genome	**14**
1.7 Brief overview on cancer:	**15**
1.8 Head and Neck Squamous Cell Carcinoma (HNSCC):	**16**
1.9 The symptoms of HNSCC	**16**
1.10 Risk Factors for OSCC:	**17**
1.10.1 Smoking:	**17**
1.10.2 Alcohol use	**17**
1.10.3 HPV infection	**17**
1.11 Mutational Landscape of genes altered in OSCC	**18**
1.12 TP53 mutations and human cancer	**19**
1.13 LINE 1 and cancer	**19**
1.14 Objectives of the study	**20**
2. MATERIALS AND METHODS	
2.1 Sources of materials	**21**
2.2 Organisms and growth conditions	**21**
2.3 Culture media	**21**
2.3.1 Luria Broth (LB)	**21**
2.3.2 SOB medium	**22**
2.3.3 LB Agar	**22**
2.4 Preparation of solutions	**22**
2.4.1 Acid salt Buffer	**22**
2.4.2 INOUE solution	**22**

2.4.3 Reaction buffer (5X) for RT assay	**23**
2.4.4 Lysis buffer A:	**23**
2.5 Plasmid DNA isolation	**23**
2.5.1 Mini prep DNA isolation (Alkaline lysis method)	**23**
2.5.2 Midi prep DNA isolation (Alkaline lysis method)	**24**
2.6 Transformation of *E.coli* cells	**24**
2.6.1 Preparation of competent cells by The Inoue Method	**24**
2.6.2 Transformation of competent cells *E.coli* cells	**25**
2.7 Restriction enzyme digestion:	**25**
2.8 Insert isolation	**25**
2.9 Antarctic Phosphatase (AnP) treatment	**25**
2.10 Polymerase chain reaction:	**26**
2.11 Ligation of vector and insert:	**26**
2.12 Agarose gel electrophoresis:	**27**
2.13 SDS-PAGE Analysis of proteins:	**27**
2.14 Desalting and equilibration of protein samples:	**27**
2.15 Protein estimation:	**27**
2.16 Cloning, expression and purification of ORF1 protein:	**27**
2.17 Cloning, expression and purification of RT (SV1) protein:	**28**
2.18 Cloning, expression and purification SV2 and SV3 protein	**29**
2.19 Preparation of Alu DNA template for making Alu RNA:	**29**
2.20 Reverse Transcriptase (RT) Activity Assay:	**30**
2.21 Genomic DNA Extraction from cancer and normal tissues:	**31**
2.22 LINE-1 promoter methylation analysis:	**31**
2.23 Cell culture:	**32**
2.24 Protein extraction and immunoblots:	**32**
2.25 Immunohistochemistry (IHC):	**33**
2.26 Production of polyclonal antibody:	**33**
2.27 Statistical analysis:	**34**
3. RESULTS	
3.1 Objective 1: To over-express human LINE-1 ORF2p protein in a bacterial expression vector and investigate its reverse transcriptase activity. To make antibody against human LINE1 (L1)	**37**
3.1 .1 Cloning of human L-1RT domain:	**38**
3.1.2 Expression Studies of recombinant human L-1 RT (hL1RT):	**39**
3.1.3 Solubility studies of recombinant human L1RT:	**40**
3.1.4 RT Activity analysis of recombinant human L1 RT domain protein:	**47**
3.1.5 Generating antibody against human L1-ORF2p:	**49**
3.1.6 Sub-cloning of Human LINE-1 ORF1p in bacterial expression vector for its expression and purification in order to generate antibody:	**51**
3.2 Objective 2: To investigate the methylation status of Human LINE-1	**53**

retrotransposon promoter and LINE-1 ORF1p expression in OSCC samples	
3.2.1 Loss of DNA methylation at CpGs within the L1 5'-UTR in oral cancer samples:	**54**
3.2.2 Detectable expression of L1 ORF1p in oral cancer samples:	**60**
3.2.3 Aberrant TP53 expression in oral carcinoma	**62**
4. DISCUSSION	
4.1 Recombinant human ORF2p RT domain protein formed inclusion bodies	**63**
in bacterial expression system	
4.2 Antibodies against L1 encoded proteins (ORF1p and ORF2p) are	**66**
important tools to study the biology of L1 retrotransposon:	
4.3 ORF1p expression is common in OSCC:	**67**
4.4 L1 protein expression is associated with hypomethylation	**68**
of the L1 promoter in OSCC	
4.5 Tumor suppressor p53 protein might have some	**68**
role in L1 activation in OSCC	
5. BIBLIOGRAPHY	**69-84**
6. APPENDIX	
6.1. APPENDIX I	**85**
6.1.1. Human L1 ORF2 protein sequence	**85**
6.1.2. ORF2 amino acids and nucleotide sequence	**85**
6.1.3. p-BS-hL1RT	**88**
6.1.4:SV1: pET 28a-RT (Swapping of EcoRI- NotI fragment	**89**
from pBS to pET28a)	
6.1.5. pET-hL1RT protein sequence	**90**
6.1.6. SV2: pET 30b-RT_{SV2} nucleotide sequence	**90**
6.1.7. pET 30b-RT_{SV2} protein sequence	**91**
6.1.8. SV3: pET 30b-RT_{SV3} nucleotide sequence	**92**
6.1.9. pET 30b-RT_{SV3} protein sequence	**93**
6.2. APPENDIX II	**94**
6.2.1. The details of patients used in this study are in the following table:	**94**
6.3. APPENDIX III	**95**
6.3.1. >L1 Promoter (5'-UTR) sequence (L1.3, Accession # 19088.1)	**95**
6.3.2. Alignment of consensus L1HS with four clones obtained from C1	**96**
6.3.3. Alignment of consensus L1HS with four clones obtained from N1	**97**
6.3.4. Sanger sequences of bisulfite treated L1 5'UTR clone sequences	**98**
from 8 paired normal cancer tissues used for methylat ion analysis	
of L15'UTR in OSCC.	
6.3.5. Calculations for graphs for figure 20	**150**
6.4. APPENDIX IV	**152**

6.4.1. Plasmid maps used in this study:	**152**
6.5. APPENDIX V	**155**
6.5.1. Poster Abstract I	**155**
6.5.2. Poster Abstract II	**156**
6.5.3. Poster Abstract III	**157**

List of Figures

Figures	Page No.
1.1 Components of human genome	2
1.2 Different types of human transposable elements	3
1.3 Structure of different types of human transposable elements	4
1.4 Non-LTR retrotransposons in the human genome	5
1.5 Mechanism of L1 retrotransposition	10
1.6 Effects of retrotransposition	13
1.7 Oral squamous cell carcinoma	16
R1 Schematic representation of amino acids and nucleotides map of human L1 ORF1p and ORF2p.	38
R2: Schematic diagram of the full length L1 retrotransposon insertion in intron 1 of retinitis pigmentosa -2 (RP) gene in patient DNA	39
R3. Cloning of SV1 clone	40
R4. Checking expression of human RT domain protein using SV1 clone	41
R5. Expression of RT protein at different temperatures	42
R6. Human L1 encoded ORF2p solubillity prediction by ccSol Omics webserver	42
R7. Cloning of different length RT domain	43
R8. Cloning of SV2 fragment in bacterial expression vector	43
R9. Cloning of SV3 fragment in bacterial expression vector	44
R 10. Solubilizing RT domain protein (SV1 clone) present in inclusion bodies using different concentration of urea	45
R 11. Solubilizing RT domain protein (SV1 clone) present in inclusion bodies using different concentration of urea as described for SV1 clone	46
R12. Western detection of human L1 RT domain protein in bacterial soluble lysate	47
R13. Preperation of template for RT assay	48
R14. RT activity assay	
R15. Making antibody against ORF2p	50
R16. Cloning of human L1 ORF1 fragment in bacterial expression vector	52
R17. Isolation of paired normal tumor genomic DNA from cancer patients for L1 promoter methylation study	54
R18. DNA methylation study	55
R19. Bisulfite analysis of paired normal cancer tissues	56
R20. Analysis of methylation status in paired normal cancer tissues	58
R21. Human L1 ORF1 protein expression in operated oral cancer samples	59
R22. ORF1p immunohistochemistry analysis across oral cancer samples using human α-L1 ORF1p(RRM) antibody	60
R23. Quantification of ORF1p expression in OSCC samples	61
R24. L1 ORF1p expression in OSCC correlates with aberrant expression	62

of TP53	

List of Tables

2.1 List of primers used in the study	35
3.1 Blast results of methylated clone sequences	57

1.1. Transposable elements and its discovery:

Transposable elements also called as transposons or mobile DNA are piece of DNA sequence that have inherent ability to move from one genomic location to another location and thus have enormous role to alter the structure and function of the genome [Kazazian and Moran 2017; Huang et al., 2012; Moran and Gilbert 2002]. These sequence are repeated in nature (interspersed repeat), and occupied a significant fraction in almost every prokaryotic and eukaryotic genomes. The seminal discovery that a piece of DNA sequence can move from one place of the genome to another was first reported by Barbara McClintock when she was working with maize and named the particular piece of DNA segment as "controlling element". She identified a specific chromosomal breakage event that always occurred at the same locus on the maize chromosome number 9, which she named as the Dissociator (Ds) locus [McClintock, 1950]. After working several years she discovered that Ds sequence can change position within the chromosome and the movement of Ds sequence requires the presence of another sequence which she named as Activator (Ac) [McClintock, 1950 and McClintock, 1956].Today we know that Ac is an autonomous transposon with 4.5 kb in length and encodes a protein named transposase and Ds is a transposition incompetent derivatives of Ac (non-autonomous transposons). The Ds element transposes or move from one location to another when supplied with Ac-encoded transposase. As per McClintock words, the concept of transposon did not fit easily within the framework of genetics at that time. The decades of gene mapping data had shown that genes were arranged in linear array in the chromosomes and each gene had fixed position relative to each other, which made it hard for researchers to accept that genes could move within the genome. But slowly from the year 1960 onwards the similar phenomena discovered in bacteriophages [Taylor, 1963], bacteria [Shapiro, 1969] and eventually in Drosophilla [Engels and Preston, 1981] and thus, the scientific community gradually accepted that the transposons what McClintock discovered were not particular in maize but in fact widespread across species. Now because of the advent of sequencing we now know that transposons constitute more than 65% of our genome and ~ 85% of the maize genome [Lander et al., 2001; Messing et al., 2004].

1.2. Types of transposable elements in mammalian genome:

Genome sequencing of human, mouse and other higher eukaryotic organisms revealed that only 1-2% of the total genome encodes proteins whereas different kind of repetitive sequences occupy more than half of the genome [de Koning et al., 2011; Rodic and Burns 2013] (Figure 1.1). These repetitive sequences are mainly belong to the group of mobile DNA or jumping gene. There are two major groups of jumping genes or transposable elements in

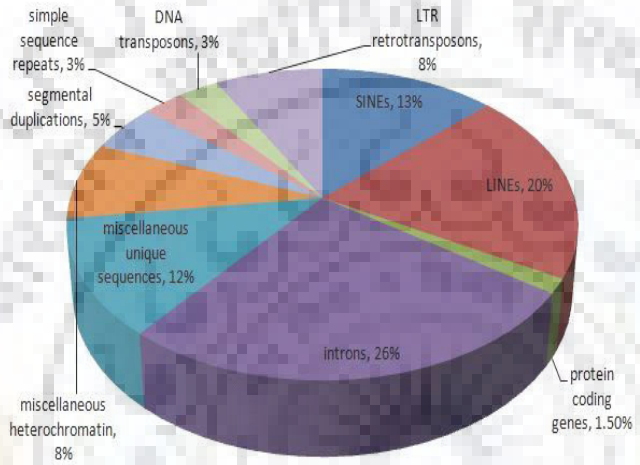


Figure P1: The pie chart showing only around 1.5% of human genome is responsible for protein coding where as different types of repeat sequences occupied more than 50% of the genome. Among all repeats Long Interspersed Elements (LINEs) occupied the highest proportion which is around 20% of the human genome and then Short Interspersed Element (SINE) which occupied around 13 %. DNA transposons and LTR retrotransposons occupied around 35 and 8% of human genome respectively.

mammalian genome (i) DNA Transposons and (ii) Retrotransposons [Ostertag and Kazazian 2001] (Figure 1.2). Structurally, DNA transposons are discrete pieces of DNA segment, surrounded by inverted terminal repeats (Figure 1.2). These inverted terminal repeats are actually binding site for the enzyme transposase which is encoded by the transposon itself and helps the transposon sequence to move from one genomic position to anoother position in the same genome. Transposons generally move via “cut and paste” mechanism in which the transposon is excised from the donor site and inserted into the new target site to make a insertion [Craig N 2002]. Mammalian DNA transposons which occupy around 3% and 1% of human and mouse DNA respectively have structural similarities with bacterial transposons [Lander et al 2001; Mandal and Kazazian 2008].

The majority of the transposons identified in human and mouse genome are belong to hAT and Tc-1 family,remnants or fossils of ancient elements and is unlikely that any transpositionally active element exists in recent day human and mouse genome [Lander et al 2001; Ostertag and Kazazian 2001] .

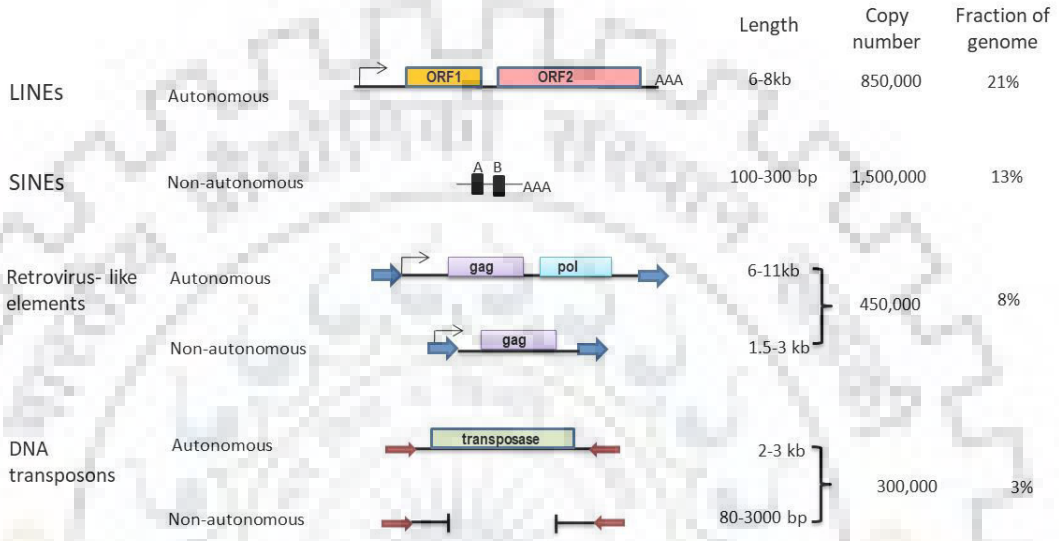


Figure 1.2: Different types of human transposable elements, their length, copy number and fractions occupied as per human genome sequencing project (Lander et al., 2001). The copy number of SINE in human genome is highest around 1,500,000 per haploid genome followed by LINE which is around 850,000. DNA transposon and LTR retrotransposon (Retrovirus-like elements) which are currently inactive in human occupied very small fraction compared to non-LTR retrotransposons.

In contrast to transposons, retrotransposons move by copy and paste mechanism and encodes a protein which has reverse transcriptase (RT) activity [Richardson et al, 2015]. Retrotransposable elements can be classified in two types: autonomous retrotransposonandnon-autonomous retrotransposons (Figure 1.2; Figure 1.3). Elements are considered autonomous if retrotransposition using its own protein machinery. There are two classes of autonomous retotransposons: Long Terminal Repeat (LTR) retrotransposons and non-LTR retrotransposons. Mammalian LTR retrotransposons are structurally similar to LTR retroviruses, only difference is the absence of lack envelop gene in LTR retrotransposon.The LTR retrotransposons occupy ~ 8% and ~10% of human and mouse genome respectively [Lander et al 2001; Mandal and Kazazian 2008].

The major class of LTR retrotransposon present in both the organisms genome belong to endogenous retroviruses (ERV) and Mammalian apparent long terminal repeat retrotransposon (MaLR). In mouse these elements are still active where as in human not a single active element has been found yet. The non-LTR retrotransposon class contains Long Interspersed Elements (LINEs) which include inactive elements, such as L2 and active element such as L1 in human and mice [**Ostertag and Kazazian 2001]**. Aproximately 20% of the human and mouse genome is composed of non-LTR retrotransposon (detailed description of human L1 retrotransposonis given below under subheading structure of human L1 transposable element) [Lander et al., 2001].

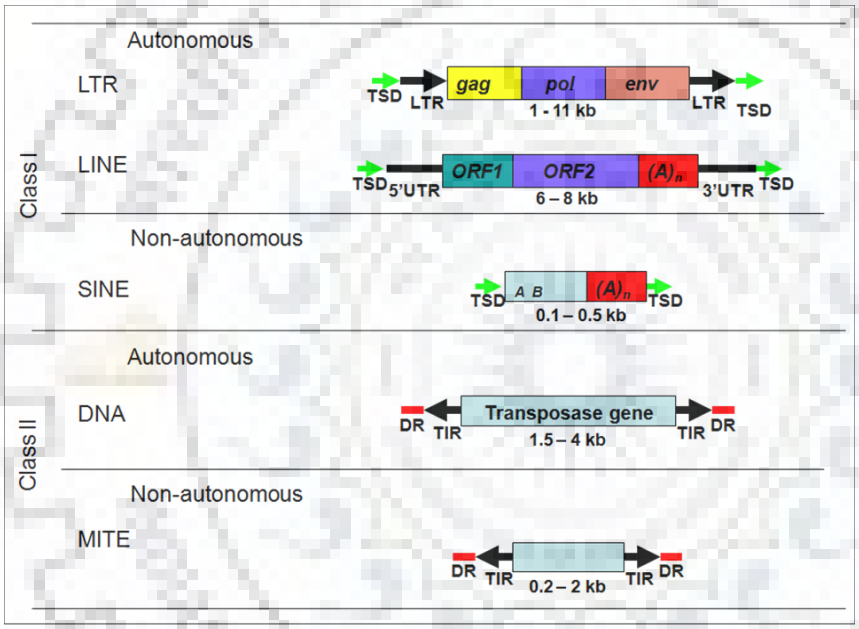


Figure 1.3: Structure of different types of transposable elements. Both class I and II are made with autonomous and non-autonomous types. Autonomous class 1 elements again are divided into two types based on the presence of long terminal repeat sequence called LTR. The LTR-retrotransposons are like retroviruses 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 depends on autonomous elements machinery for its jumping. Class II elements are DNA transposons again divided into two types autonomous and non-autonomous. Autonomous 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-autonomous transposable elements are those which use the machinery of autonomous elements for their movement in the genome. Alu also known as Short INterspersed Elements (SINEs) are the most abundant element in this category comprising 13-15% of human genome with more than 1.5 million copies [Lander et al., 2001; Cordaux and Batzer, 2009; Belancio et al., 2009; Pandey and Mukerji 2011]. Human Alu is 320 bp in length harbour an internal RNA polymerase III promoter, polIII terminator (TTTT) and polyA sequence of varying length at the 3'UTR and does not encode any proteins [Fuhrman et al.,1981]. Indeed, most SINEs

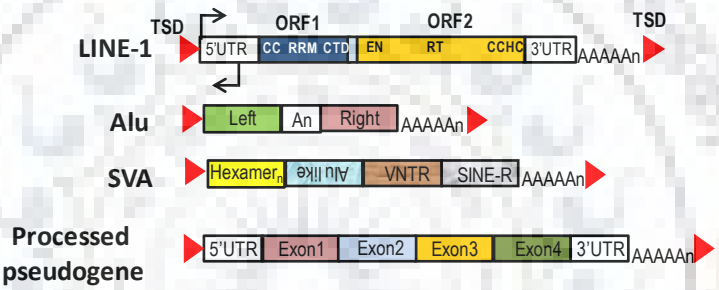


Figure 1.4: Non-LTR retrotransposons in the human genome. Details of the structure and abbreviation for LINE-1 UTR: untranslated regions (white box); sense and anti sense internal promoter black arrow. ORF1: blue box includes a coiled coil domain (CC), RNA recognition motif (RRM) and a C-terminal domain (CTD). Inter ORF spacer (light blue box between ORF1 and ORF2); ORF2 is yellow box includes endonuclease domain, reverse transcriptase domain (RT) and a cystine rich domains (CCHC); Poly A tract (stretches of A nucletides downstream of 3'UTR). Human Alu contain left and right monomer separated by a A-rich linker sequence and ends with poly A sequences. For human SVA hexameric CCCTCT repeat yellow box; inverted Alu like repeat light blue box; GC rich VNTR brown; SINE-R sequence homology with HERV-K10 (white box). Processed pseudogenes: spliced cellular mRNA with UTRs (white box) and coding ORF where four exons are fused (no intron present between two exons)

survive by sharing the 3' end with a resident LINE element. The promoter regions of all known SINEs are derived from tRNA sequences, with the exception of a single monophyletic family of SINEs derived from the signal recognition particle component 7SL [Weiner, 1980; Ulluand Tschudi, 1984]. This family, which also does not share its 3' end with a LINE, includes the only active SINE in the human genome, the Alu element. The human genome contains three distinct monophyletic families of SINEs: the active Alu, and

the inactive MIR and Ther2/MIR3 [Lander et al., 2001]. Although, Alu elements contain pol III promoter the transcriptional robustness of Alu is largely dictated by the presence of upstream enhancer elements [Chu et al., 1995, Conti et al., 2015]. The most prominent SINE retrotransposons in mouse genome are B1, B2 and B4 elements [Goodier et al., 2001;Dewannieux and Heidmann, 2005]. Other non-autonomous retrotransposonsin human genome includes SINE-R (short interspersed element of HERV origin)-VNTR (variable- number tandem repeat)-Alu (SVA) and processed pseudogenes (PP) [Ostertag et al., 2003; Richardson et al., 2014; Hancks and Kazazian 2016]. The SVA is the youngest active hominid specific non-autonomous composite retrotransposon of around 2 kb in length and almost 2700 copies of it are present in human genome [Wang et al., 2005; Schumann et al., 2010; Hanck et al., 2011; Raiz et al., 2012].Processed pseudogenes (PP) which are derived from cellular mRNA by L1 retrotransposon machinery, contain all the hallmark of L1-mediated retrotransposition (target site duplication, 5'-truncation, end in a polyA sequence of variable length). Recent estimate suggests that there are more than 8000 PPs derived from 3000 protein coding genes [Pei et al., 2012; Mandal et al., 2013]. The ribosomal proteins RNA have maximum number of PP than any protein sequences [Gonçalves et al., 2000].

1.3. Structure of Human L1 transposable element:

There are almost 500,000 LINE 1 sequences present in the human genome and most of them are 5'-truncated and non-functional (Figure 1.4) [Moran and Kazazian 2017; Faulkner and Garcia-Perez, 2017]. Although most of them are inactive, around 12000 are full length and out of them around 100 are actively retrotransposing in any given human genome [Brouha et al., 2003]. 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; Baoet al., 2015].

1.3.1. Human L1 promoter:

A full length LINE 1 sequence contains approximately 900bp long 5'UTR which contains bi-directional promoter activities [Ostertag et al 2001; Speek M 2001; Yang and Kazazian 2006]. The consensus LINE1 sequence contains a polyA signal AATAAA, which is required for RNA polymerase II termination suggesting LINE1 is transcribed by Pol II. The LINE 1 promoter is enriched with CpG islands; a consensus L1 sequence contains around 40 CpG dinucleotide at the promoter region. These CpG dinucleotides are heavily methylated in normal non-dividing cells leading to silencing of L1 transcription [Bao et al 2015; Moutri et al, 2010] [Yoder et al., 1997]. However, in many cancers and early stage of development the CpG dinucleotides in L1 promoters show hypomethylation resulting in L1 transcription and followed by translation of L1 encoded proteins, L1-ORF1p and L1-ORF2p [Shukla et al., 2013].There are several transcription factors that 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 bp 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.

1.3.2. Human L1 encoded proteins:

1.3.2.1. L1-ORF1p: Human L1 ORF1 encodes a 40 kDa protein (338 amino acid in length) termed ORF1p comprised of 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]. *In-vitro* studies have demonstrated that human ORF1p is a non-specific single-stranded (ss) nucleic acid binding protein with nucleic acid chaperone activity. Both human and mouse ORF1p bind to ssRNA and ssDNA with high affinity [Martin and Bushman, 2001; Khazina et al.. 2011]. ORF1p is mainly a cytoplasmic protein and shows *cis* preference, i.e., it binds with its own transcript [Wei et al., 2001; Kulpa and Moran 2005]. It is hypothesized that ORF1p helps in formingL1-RNP by binding with L1RNA and ORF2p and some cellular factors required for retrotrotransposition (Khazina et al, 2011, Martin et al, 2003; 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 the retrotransposition of engineered L1 in the cell culture based retrotransposoition assay [Moran et al., 2006; Goodier et al., 2007; 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; Martin 2010].

1.3.2.2. L1-ORF2p: Human L1-ORF2 is a 1267 amino acids protein of molecular weight around 150kDa with three reported conserved domains: the domains are (i) an N-terminal endonuclease domain (EN) (1-239 amino acids) [Feng et al.,1996](ii) followed by a central reverse transcriptase domain (RT) (453-880 aa) [Mathias et al.,1991, Malik et al., 1999; Singer and Clements, 1998] and (iii) a C-terminal zinc knuckle like domain (1096-1275 aa) [Moran et al., 1996;Piskareva et al., 2013].The L1 EN domain sequence and structure shows similarity to apurinic/apyrimidinic type endonucleases (APEs). Computational analysis and *in vitro* studies revealed that L1 EN nicks at consensus site which is 5'-TTTT/AA-3' on 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 activity 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 Mg^{2+} ion is preferred over Mn^{2+} [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 processed psuedogene (PP) in the human genome [Mandal et al., 2013;Pink et al., 2011]. These enriched mRNA are part of L1-RNP as they 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 purificationallowed 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 endonucleaseactivity. Importantly, the EN mutant of ORF2p has been found to be expressed in significantly high amounts, have allowed to detect the protein in Coomassie stained denatured SDS-PAGE gel after affinity purification [Mandal et al, 2013].

A third conserved domain of ORF2p is a 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 C domain in L1 retrotransposition.

1.3.2.3. L1-ORF0: Recently on the antisense strand of 5'-UTR apart from antisense promoter [Speek, 2001],a novel ORF named ORF0 has been identified [Denli et al., 2015]. The ORF0 is 70 amino acids in length and has splice donor sites which has the

ability to form fusion protein with downstream exons. Recent experiments showed that overexpression of ORF0 *in trans* increased L1 retrotransposoition in cell culture based retrotransposition assay [Denliet al., 2015].

1.4. Mechanism of LINE 1 retrotransposition:

It is thought that L1 retrotransposition occurs by a process termed as target primed reverse transcriptase (TPRT), originally described for R2Bm, a site specific non-LTR retrotransposon from the silkworm *Bombyxmori* [Luan et al., 1993; Luan and Eickbush,

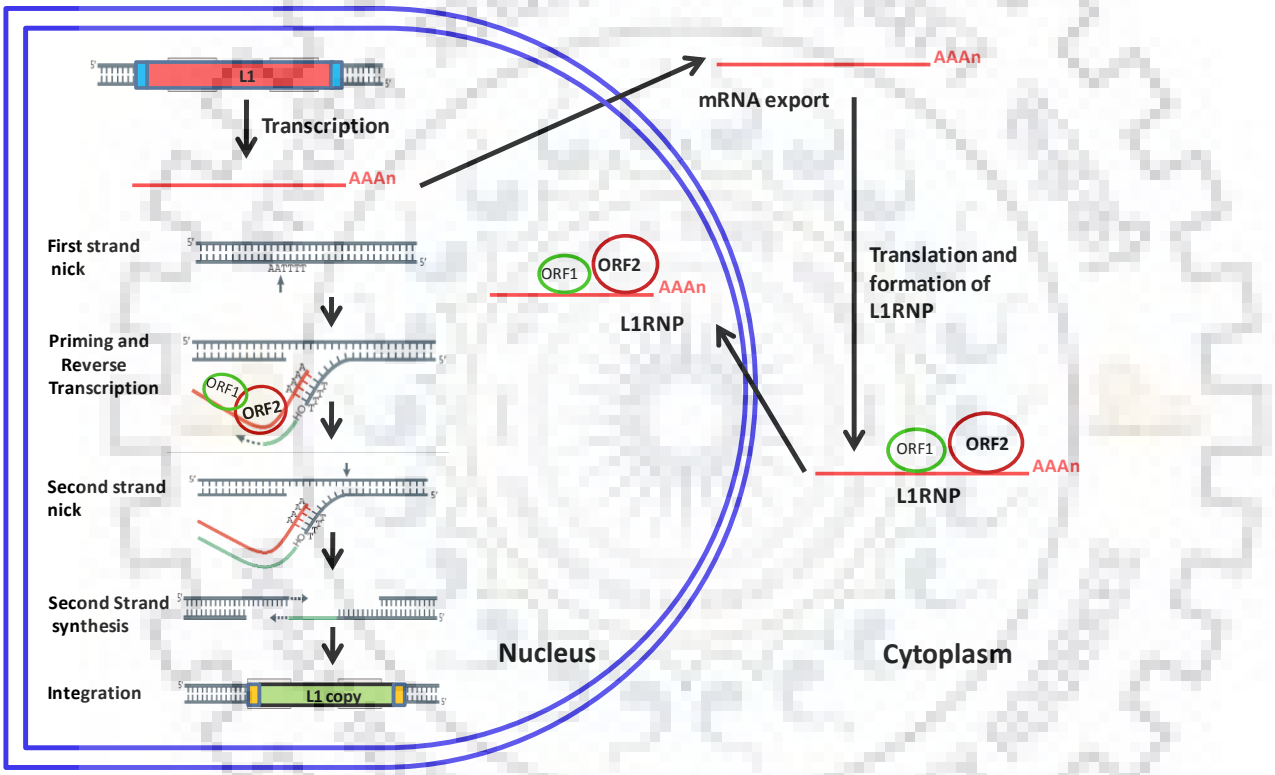


Figure 1.5: Mechanism of L1 retrotransposoition- An active L1 present in one chromosmal locus transcribed to mRNA, exported to cytoplasm where both ORFs are translated. Both ORF1 and ORF2 along with some cellular factors bind with L1 RNA and form L1-RNPs, the retrotransposition intermediate which enter into nucleus by an Unknown mechanism. The L1RNPs then search the target site where *de novo* isertion occurs by target-site primed reverse transcription mechanism (TPRT) . The ORF2p endonuclese activity makes a nick at a consensus sequence 5'-TTTT/A-3' (sign "/ " indicats nicking site). The ORF2p RT activity then used exposed 3'-OH group to synthesize first strand cDNA using L1 RNA as template. The second strand nick, second strand synthesis and repair makes a new copy to a new chromosomal locus. The L1 copy mostly truncated ends with Poly A tail and flanked by target site duplication.

1995; Cost et al., 2002]. It is hypothesized that following transcription from a chromosomal locus, a full length bicistronic L1-RNA exported to cytoplasm where both ORFs are translated (Figure 1.5). The newly synthesized ORF1p and ORF2p in the cytoplasm now interacts 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 enters the nucleus where a new L1 copy synthesize at the site of insertion via a coupled reverse transcription and integration mechanism termed as TPRT [Luan and Eickbush, 1995; Cost et al., 2002].. In this process ORF2p nicks the bottom strand of a target site ata degenerate consensus sequence (5'-TTTT/AA-3') that generates a free 3'-OH,which acts 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].

1.5. Restriction against human mobile elements:

Identification of more than 100 diseases causing retrotransposon insertions revealed that activation of retrotransposons are detrimental and thus organisms have developed array of defence mechanism to restrict the retrotransposition [Hancks and Kazazian 2016].The human L1 promoter contains around 40 CpG dinucleotides where methylation is established at very early stage of development and maintained throughout the life of an organism [Bestor and Bourc'his 2004a; Bestor and Bourc'his2004b] . It is thought that primarily the methylation of L1 promoter controls the L1 expression in somatic cells. Knock out of DNA methyl transferase proteins which are responsible for methylation of those CpGisland in L1 promoter increases L1 retrotransposition in mice [Bourc'his and Bestor, 2004b].

Post-transcriptional gene silencing mediated by small RNA (miRNA, SiRNA and piRNA) is another strategy to limit the movement of LINE1 in the genome. A large number of

endogenous retrotransposon-related small RNAs of sizes consistent with siRNAs, miRNAs and piRNAs have been detected in cells [Smalheiser et al., 2005, Tam et al., 2008; Malone and Hannon 2009] and all these RNAs showed some effects on restriction of transposons and retrotransposons in diverse organisms including human. Recently, Hamdorf et al. [2015] showed a specific miRNA, mir-128 inhibits L1 retrotransposition by binding with L1RNA. Further study showed that mir-128 also represses L1 retrotransposition by down regulating the nuclear import factor TNPO1 [Idica et al., 2017;;Hamdorf et al., 2015].

The APOBEC3 cytidine deaminases was the first host factor identified from human cells showed potent inhibitory effects on L1 retrotransposition [Ikeda et al., 2011]. It is important to mention that APOBEC3G which was identified earlier showed restriction on HIV infectivity [Sheehy et al., 2002]. Studies demonstrate that the human APOBEC3 gene family encodes 7 proteins which can catalyze deamination of cytidine to uridine residues in a single strand DNA substrate [Bogerd et al., 2006]. Among seven APOBEC3 genes, APOBEC3A and APOBEC3B showed robust inhibition of L1 and Alu retrotransposition in cultured cells retrotransposition assay [Chen et al., 2006;Bogerd et al., 2006; Schumann, 2007; Muckenfuss et al 2006;Wissing et al., 2011].

The other factors which showed inhibition on L1 retrotransposition are Trex1 [Li et al, 2017], SAMHD1 [Zhao et al., 2013],MOV10 [Arjan-Odedra et al., 2012; Goodier et al, 2012, Li et al., 2013] ZAP [Goodier et al., 2015; Moldovan et al., 2015] and others. Mutation in Trex1 and SAMHD1 causes Aicardi-Goutiers syndrome (AGS), a rare childhood disorder that can lead to neurodevelopmental deficiencies [Crow et al., 2006; Rice et al., 2009]. Trex1 is a 3'-5' DNA exonuclease whereas SAMHD1 is a triphosphohydrolase that can reduce intracellular dNTP pools. Although, both proteins showed inhibition of L1 retrotransposition in cultured cells, detailed studies are required to determine the exact mechanism by which Trex1 and SAMHD1 inhibit L1 retrotransposition.

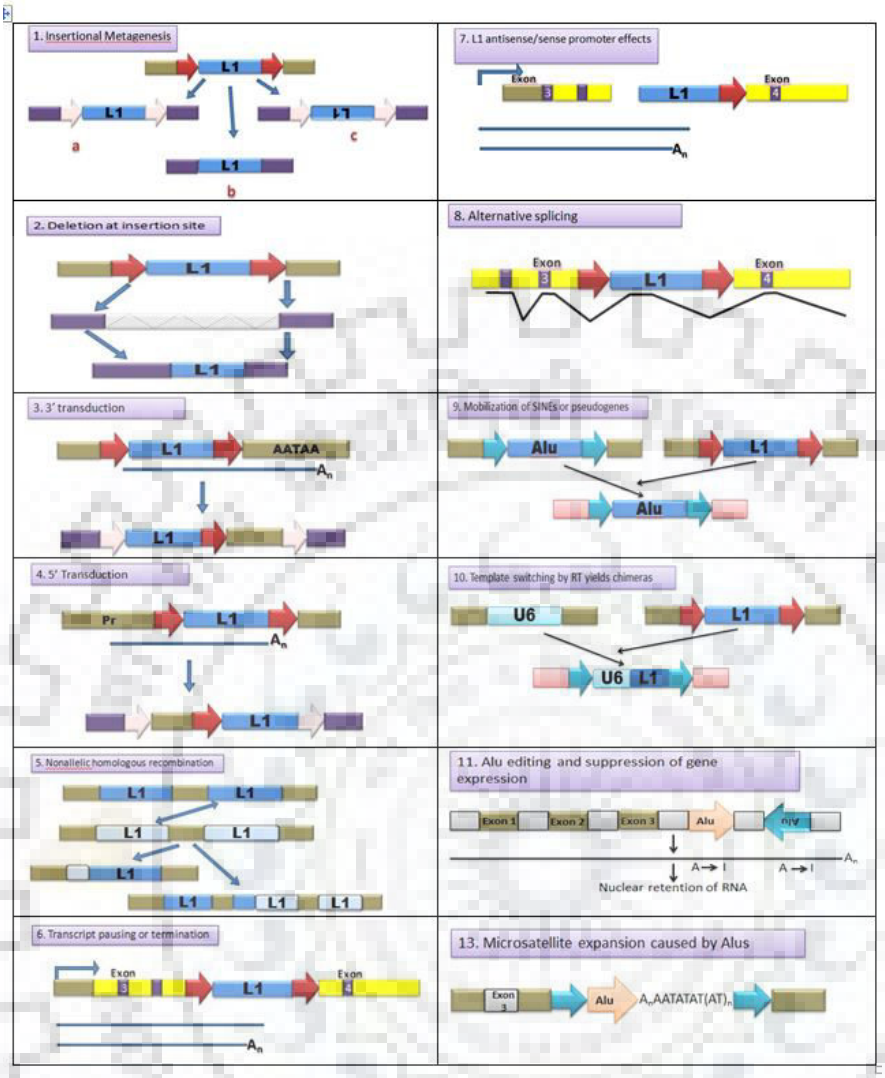


Figure 1.6: Effects of Retrotransposition: 1 and 2: L1 Insertion in new location disrupts the gene. Insertion may be full length, 5'- truncation, deletion or inversion. 3 and 4: transduction of flanking sequence (either 3' or 5') upon retrotransposition. 5: Unequal cross over can lead to deletion or duplications 6: transcriptional pausing due to poly A signal inside L1 sequence 7: The antisense promoter of L1 can produce transcript of genes upstream of L1 5'UTR. 8: Alternate splice site can lead to exon shuffling 9: L1 can change the chromatin structure, thereby altering gene expression 10: L1 machinery can retrotranspose Alu , Sva and cellular mRNA leading to further increase of genome size. 11:Template switching of L1 RT from L1 RNA to other RNA such as U6 RNA can produce chimeric insertion. 12. Alu element expansion have been occasionally associated with disease.

Moloneyleukemia virus 10 (MOV10) is an RNA helicase initially identified as an inhibitor of retroviruses including HIV-1 [Zheng et al., 2012] showed decrease in retrotransposition of human retrotransposons (L1, Alu and SVA) in cultured cells [Arjan-Odedra et al., 2012; Goodier et al., 2012, Li et al., 2013]. Studies revealed that MOV10 protein co-localizes with L1-ORF1p, binds with L1RNA and mutation of helicase domain of MOV10 reduces its ability to restrict retrotransposition [Goodier et al., 2012; Li et al., 2013]. More studies are required to know the exact mechanism by which MOV10 inhibits L1 retrotransposition.

Another family of proteins known as KRAB zinc-finger (KZNF) protein family recruits KAP1 and other associated proteins to LINE1 and SVA transposons resulting in the inhibition of their expression in embryonic stem cells [Jacobs et al., 2014]. Compared to older LINE1 elements, L1Hs shows that it does not contain the binding site present in the 5' UTR region allows it to escape from ZKZNF mediated repression. There are some cellular RNA binding proteins such as hnRNPL,RNaseL [Zhang et al., 2014] showed inhibition of engineered LINE-1 in cultured cells. Some of these proteins colocalizes with L1ORF1p and L1RNA in stress granules, suggesting that these proteins do not allow L1RNA to form L1-RNPs, the proposed retrotransposon intermediate, thus might play import role in L1 retrotransposon inhibition [Goodier et al., 2013;Goodier 2016]. As full length L1 RNA contains multiple splice sites and poly A signal sequence, it is thought that proteins involved in RNA splicing and end maturation may inhibit LINE-1 retrotransposition [Kazazian and Moran 2017].

1.6. Effects of L1 on Human Genome:

To date around 124 retrotransposon associated diseases have been documented and the list is sure to grow [Hancks and Kazazian 2016]. All these diseases are due to insertion mutation when L1, Alu and SVA retrotransposed to a cognate gene (Figure 1.6). Apart from simple insertion mutation, activity of L1 alters human genome in several ways. Often LINE-1 insertion is associated with 5'-transduction, 3'-transduction, large deletion, duplication, inversions, exon suffling, chimeric gene formation and others [Ostertag and Kazazian 2001; Goodier and Kazazian 2008]. L1 elements occasionally help in repairing double-stranded breaks in DNA. They insert at the site of double-strand breaks via an

endonuclease- independent pathway [Morrish et al., 2007].LINE-1 retrotransposons can also effect the expression of adjacent genes. LINE1 has an antisense promoter activity between 400bp to 600bp of 5' UTR region [Speek., 2001] which affects expression of genes reside close to the 5'-UTR (Figure 1.6).

1.7. Brief overview on cancer:

Cancer is defined as a group of diseases where cells start to divided abnormally and spread into neighbouring tissues. It is considered as the emperor of all maladies, an enemy that attacks silently to a person. As per statistics one in seven death is due to the cause of cancer and the number of cancer causing death is more than the death due to AIDS, tuberculosis and malaria combined [International Agency for Research on Cancer. World cancer report 2014]. The way the number of cancer patients are increasing world-wide day by day is alarming. As per 2016 WHO 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. It is important to be mentioned that when countries are grouped according to income, cancer showed the second leading cause of death following cardiovascular diseases in high income countries and third leading cause of death in low and middle income countries (following cardiovascular diseases and infectious and parasitic disease) [World Health Organization. Strengthening the prevention of oral cancer: the WHO perspective http://www.who.int/oral_health/publications/].

. Today's society has already realized the threat posed by numerous types of cancer. There are more than 100 type of cancers identified which are named as per the organs or tissues where the cancers develop. Among man the five most common site of cancer diagnosed includes lung, prostate, colorectal, stomach and liver. Among women those five sites are breast, colorectum, lung, cervix and stomach. All these 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 higher incident rate (43.3 per 100 000)[International Agency for Research on Cancer. World cancer report 2014].

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

Head and neck cancers also called as 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, 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% [Agrawal et al., 2011]. 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]

1.9. The symptoms of HNSCC:

The major symptoms of HNSCC are i) swelling or sore that does not heal (the most common symptom ii) red or white patch in the mouth iii) lump, bump, or mass in the head

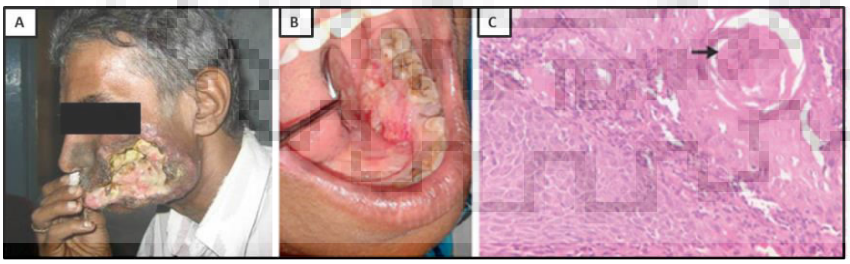


Figure 1.7: Oral squamous cell carcinoma. **A)** patient with severe form of OSCC **B)** picture showing OSCC inside mouth. **C)** Haematoxylene eosin stained section of OSCC, Keratin pearl (black arrow) and irregular nucleus are the characteristics of OSCC.

or neck area, with or without pain iv) persistent sore throat v) foul mouth odour not explained by hygiene vi) change in voice vii) frequent nose bleed, unusual nasal discharge, nasal obstruction and difficult breathing viii) difficulty in chewing, swallowing, or moving the jaw or tongue and many more [WHO tobacco 2013].

1.10. RISK Factors for OSCC:

1.10.1 Smoking: Tobacco and alcohol drinking are the two major risk factors causing OSCC. Tobacco is associated in causing oral cancer has been known since many years. [Niblock et al.,1902]. It is the strongest risk factor among both sexes causing oral cancer [Jayalekshmi et al., 2009]. Tobacco use increases the risk factor 27 times more than non users. In india, tobacco is consumed in variety of ways. Chewing betel-quid locally known as 'paan' is very popular which is prepared by using betel leaf (Piper betle), slaked lime (mostly calcium hydroxide), areca nut (Areca catechu) and tobacco. [Muwonge et al., 2008] Use of bidi (small handrolled cigarette) is very frequent in rural areas of India comparatively [Reibel et al., 2003; Dikshit et al., 2012].

1.10.2 Alcohol use: Most of the cultures worldwide consume different kinds of alcoholic beverages. In India most common alcoholic beverages made from barley and sugarcane. It has been seen that drinking alcohol affects most of the organs in the aero digestive tract like oral cavity, pharynx, stomach, colorectum etc. Alcohol drinking in combination with smoking tobacco increases the risk factor for OSCC. Recent studies have shown that ethanol in the alcoholic beverages is the most significant carcinogen among other components like acetaldehyde, ethyl carbamateetc [Lachenmeier et al., 2012].

1.10.3 HPV infection: The human papilloma virus (HPV) infection is very common and the causative factor of cervical cancer in women. More recently the HPV has also emerged as an additional risk factors for the development of HNSCC Two subtype of HPV; HPV-16 and HPV 18 have been detected in head and neck cancer patients. However, HPV associated HNSCC cancer showed improved survival rate than those which is caused by use of tobacco and alcohol; suggesting that HNSCC due to HPV infection have distinct biological features [Upadhyay et al., 2017; Murthy et al., 2017].

Studies also showed that around 2% head and neck cancer patients are positive with herpes simplex virus. [Herrero et al., 2003].

1.11. Mutational Landscape of genes altered in OSCC:

An understanding of cancer biology is critical to develop rationally designed therapy and offer preventive options. For decades, relevant knowledge have been sought by describing how cancer cells differ from normal cells, something being elucidated far more rapidly in the post-genomic era. Specification of the human genome enabled the identification of how various types of tumour cells differ from normal cells in multiple parameters, including relevant somatic mutations and altered gene expression, which is often determined through epigenetic changes such as alterations in DNA methylation patterns. Mutation or epigenetic change may mediate, amongst other effects, altered metabolism or modified intracellular signalling in response to growth-altering stimuli. In parallel, the role of cancer stem cells and the tumour microenvironment have been recognized. Accordingly, description of how inflammation, growth of new blood vessels and modification of the immune response mediate tumour growth now enables options for cancer prevention and treatment to be identified.

Similar to all other type of cancers, OSCC is thought to be initiated and progressed by a series of genetic alterations [Agrawal et al., 2011; Stransky et 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; Shainet al., 2013; Cho et al., 2007]. Also some genes that are altered in HNSCC like *CASP8*, *TP53*, *FAT1*, *NOTCH1* are also

altered on OSCC. Patients consuming tobacco either by smoking or directly chewing were found to have high C>G transversions mutations [Agrawal et al., 2011; Stransky et al., 2011; India Project Team of the ICGC, 2013].

1.12. TP53 mutations and human cancer

Mutations in *TP53* and its aberrant expression are common in almost every type of cancer [Kastenhuberand Lowe 2017; Yue et al., 2017]. Recent reports have shown that *TP53* can restrain retrotransposons and that this activity is evolutionarily conserved [Wylie et al., 2016]. Rodic et al. [2014] reported that up-regulation of ORF1p in cancer tissues is correlated with highly expressed mutant *TP53* [Ardeljan et al., 2017].

1.13. LINE 1 and cancer

In the year 1988 it was Kazazian's laboratory which first of all reported disease caused by an insertion of LINE1 element into a functional gene [Kazazian et al., 1988]. They shown insertion of LINE1 into factor VIII gene causing haemophilia in patients having no pedigree for the disease. Since then, in more than 100 diseases it has been confirmed that LINE1 disruption of genes is another cause of these diseases. 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. This insertion leads to disruption of APC gene which is a tumour suppressor gene [Groden et al., 1991].With the advancement of sequencing techniques, recently one more insertion in the same exon of APC has been identified but at different position [Scott et al. 2016].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 were 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 [Helmanet al., 2014], Prostate [Lee et al., 2012] and ovarian cancers [Tang et al., 2017; Lee et al. 2012].

1.14. Objectives of the Study:

It was a general believe that LINE-1 retrotransposons are only active in germ cells (sperm and ovum) and in early stages of development. It is also believed that LINE-1 as a parasite is active in germ cells for its propagation to the next generation. But recent high throughput sequencing analysis revealed that L1 is also active in certain parts of normal brain and in few cancers. The activity of L1 is high in those cancers which are epithelial in origin. 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 cancer is very deadly and highly prevalent in India due to excessive use of tobacco. No study has been performed to see the activity of L1 retrotransposons in oral cancer samples. **In this study, L1 retrotransposon activity in oral cancer samples obtained from Indian patient has been investigated.**

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

i) To over-express human LINE-1ORF2p protein in a bacterial expression vector and investigate its reverse transcriptase activity.

ii) To investigate the methylation status of Human LINE-1 retrotransposon promoter and LINE-1 ORF1p expression in OSCC samples.

2.1 Sources of materials:

All *E. coli* strains used for recombinant DNA work was available in the laboratory .The plasmids used for cloning and expression studies were also available in the laboratory. Restriction enzymes and other molecular biological reagents were purchased from New England Biolabs (NEB, UK), Roche Biochemicals (Germany), Amersham Pharmacia (USA), Promega (USA), Sigma (USA) and Himedia (India).All solutions were prepared in double distilled water unless stated otherwise. Autoclaving was done at a pressure of 15 lbs per square inch for 20 min

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, Uttarakhand, India).

2.2 Organisms and growth conditions:

Cells from frozen glycerol stock 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 3-4hrs to obtain log phase cultures.

2.3 Culture media:

2.3.1 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.

2.3.2 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 sterilised by autoclaving.

2.3.3 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 and kanamycin was added to a final concentration of 50µg/ml after cooling the LB agar to around 55°C and plates poured.

2.4Preparation of solutions:

2.4.1 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.

2.4.2 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.

2.4.3 Reaction buffer (5X) for Reverse transcriptase assay:

Composition (in molarity)

Tris-cl (M.W. 154.24)	250mM
$MgCl_2$ (M.W. 197.9)	20mM
KCl (M.W.74.55)	400mM
DTT (M.W.154.25)	50mM

Volume make up by DEPC treated water.

2.4.4 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
B-ME	0.05%

2.5 Plasmid DNA isolation:

2.5.1 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 centriguge tubes. The DNA was precipitated from the supernatant by adding 0.6 vol 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 RNAse A. The tube was incubated at 37°C for 30 minutes for RNAse digestion and 2 µl was loaded on a 0.8% agarose gel to quantify the amount of DNA.

2.5.2 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.

2.6. Transformation of *E.coli* cells:

2.6.1. 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 OD 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 + 375µl 100 % DMSO. Immediately 200 µl aliquots were snap freezed by immersing in liquid nitrogen and stored at -80°C until use.

2.6.2. 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.

2.7 Restriction enzyme digestion:

Restriction enzyme digestions were carried out in usually in 30-40 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, 1 mM EDTA and 0.03% Bromophenol Blue) into agarose gel and run for appropriate time.

2.8 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.

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 are listed in **Table 1**

PCR amplification was carried out either in 20 or 50 ul reaction volume with the desired number of cycles. The reaction mixture was prepared as per following table

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

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.

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 optimium separation. The DNA was visualised 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. Desalting and equilibration of protein samples:

Protein samples were desalted by dialysis against the appropriate buffer using a dialysis membrane of 3kDa molecular weight cutoff. Samples were dialysed against 400 volumes of desired buffers at 4°C with minimum two changes.

2. 15. Protein estimation

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. 16. Cloning, expression and purification of ORF1 protein:

ORF1 fragment (length 1014 bp) was cloned in EcoR1-Not1 restriction sites of vector pET30b. Following primers were designed to amplify the specific fragment from PBS-L1RP clone available in the laboratory.

Forward primer: 5'-AATGAATTCGATGGGGAAAAAACAG-3' [ORF1-HcF(EcoR1)]

Reverse primer: 5'-AATGCGGCCGCTTACATTTTGGCATG-3' [ORF1-HcR(Not1)]

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). The ORF1 protein was found to elute at an imidazole concentration of 250mM. Fractions containing ORF1protein were identified by 10% SDS-PAGE and concentrated by using concentrator. The purified protein was stored in aliquots of 50µl at -20°C.

2.17. Cloning, expression and purification of human L1 ORF2p RT domain protein (SV1 clone) protein:

A single colony of E. coli cells containing pET-hL1$RT_{SV1}$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. Thereafter, cells were harvested by centrifugation and the resuspended in 2 ml buffer A [50mM NaH2PO4 (pH-8.0), 300mM NaCl, 20mM Imidazole (pH-8.0), 10mM β-mercaptoethanol, 0.5% Triton X-100 and 2mM phenylmethylsulfonyl fluoride (PMSF)]. Next, cells were lysed by 3 cycles of freezing and thawing, followed by sonication on ice. 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).

The RT protein was found to elute at an imidazole concentration of 250mM. Fractions containing RT were identified by SDS-PAGE and concentrated by dialyzing against buffer B [50mM NaH2PO4 (pH-8.0), 300mM NaCl, 20mM Imidazole (pH-8.0), 10mM β-mercaptoethanol, 0.5% Triton X-100 and 2mM phenylmethylsulfonyl fluoride (PMSF)]. The purified protein was stored in aliquots of 25µl in 50% glycerol at -20°C.

2.18. Cloning, expression and purification SV2 and SV3 protein:

Following primers were used to amplify SV2 and SV3 fragments of RT domain using pBS-L1RP as template

Primers to clone SV2 fragment:
Forward primer: 5ˈ-ACTATTGAGCTCGAAGGAAATAGAGACACAA-3ˈ [RT1042Fwd Sac1)]
Reverse primer: 5ˈ-ACTATTGCGGCCGCTCAGTTTTCTTCTAGGGT-3ˈ [ORF1-HcR(Not1)]

Primers to clone SV3 fragment:
Forward primer :5ˈ-ATTGAGCTCGCAGGAAGAAGTTGAATCTCTG-3ˈ [RT1327F(Sac1)]
Reverse primer: 5ˈ-ACTATTGCGGCCGCTCAGTTTTCTTCTAGGGT-3ˈ [ORF1HcR(Not1)]

Both SV2 and SV3 PCR amplified fragments were cloned in vector pET30b using Sac1- Not1 restriction enzymes. The expression and purification protocol of pET-hL1RT$_{SV1}$ was used to purify RT domain protein of pET-hL1RT$_{SV2}$ and pET-hL1RT$_{SV3}$clones.

2.19. Preparation of Alu DNA template for making Alu RNA:

pBS-Alu DNA (Alu cloned in pBS plasmid) was used to amplify 257 bp Alu sequence using following two primers. Bold sequence denotes SP6 promoter primer sequence.

Alu59sp6fwd: **ATTTAGGTGACACTATAG**AGATCACGAGGTCAGGAG

Alu296A$_{10}$rev: TTTTTTTTTTGAGACGGAGTCTCGC

A PCR reaction was performed in a 25 µl reaction volume containing 1 µl pBS-Alu template (5- 10 ng DNA) , 12.5 µl 2X Go Taq green master mix (Promega), 1ul Alu59sp6fwd (20 µm), 1ul Alu296A_{10}rev (20 µM) and 9.5 µl water.The following PCR condition was used to amplify the product one cycle at 94°C for 30 sec followed by 30 cycles at 94°C for 20 sec, 58°C for 20 sec and 72°C for 20 seconds and finally one cycle at 72°C for 2 minutes. The product was resolved in 2.0 % Agarose gel. The band corresponds to 252 bp was gel excised and DNA was extracted using Qiagen gel extraction kit. Around 100 ng PCR purified DNA template was used to synthesize Alu RNA. The transcription reaction was performed as follows in a 20 ul volume (template DNA - 2 ul (~100 ng), 10X rNTPs mix- 2 µl , 5X SP6 RNA polymerase buffer - 4µl , SP6 polymerse- 1 µl and nuclease free water -11 µl). The reaction was incubated at 37°C for one hour. After that, 1 µl RNase free DNase (10 U/µl) was added to remove template DNA and the tube was incubated for 15 minutes at 37°C. The reaction was stopped by adding 2 µl 0.2 M EDTA (pH 8.0). Next, 2 µl reaction was mixed with 13 µl 1X DNA gel loading buffer and incubated at 65°C for 15 minutes to denature RNA. Thereafter,the tube was chilled on ice for 5 minutes. The in-vitro synthesized RNA was run on a 1.5% neutral agarose gel to check the quality and quantity of the RNA (The yield of RNA should be around 200 ng/µl). The synthesized RNA was stored at -70°C until use.

2.19. RT activity assay:

The RT reaction was set up in in a 20 µl reaction volume containing 2µl Alu RNA, , 4 µl 5X RT buffer [250 mMTris-HCl (pH 8.3), 400 mMKCl, 20 mM $MgCl_2$], 1 µl dNTPs (10 mM), 1 µl LEAP primer (50 µM), 0.5 µl RNasein (20U/µl), 2 µl DTT (0.1M), 1µl purified L1 RT domain protein (either SV2 or SV3 clone) and 9 µl nuclease free water. The reaction was incubated at 37°C for 1 hour. Next a PCR reaction was performed in 25 µl total volume containing 1 µl RT reaction, 0.5 µl linker specific reverse primer (Link Rev primer) (10 µm), 0.5 µl forward primer (Alu74F) (10 µM), 12.5 µl 2X Go Taq green master mix (Promega) and 10.5 µl water. The PCR condition was as follows: one cycle at 94°C for 30 sec followed by 35 cycles at 94°C for 20 sec, 56°C for 20 sec and 72°C for 20 seconds and finally one cycle at 72°C for 2 minutes. The product was resolved in a 1.5% agarose gel

LEAP primer: GCGAGCACAGAATTAATACGACTGGTTTTTTTTTTT

Alu 74 F: GATCGAGACCATCCTGGCTAACACG
Link Rev: GCGAGCACAGAATTAATACGACTGG

2.21. Genomic DNA Extraction from cancer and normal tissues:

Around 100 mg tissue was washed with PBS followed by homogenization of the sample in liquid nitrogen using a mortar and pestle .From homogenized tissue, genomic DNA was extracted using the Blood &Tissue DNeasy mini kit (Qiagen) as per manufacturer instruction; DNA was eluted in 100 µl TE buffer. The integrity of the DNA was checked by running on a 0.6% agarose gel.

2.22. LINE-1 promoter methylation analysis:

Bisulfite conversion of genomic DNA (1µg) isolated from tissue specimens was performed using the Epitect kit (Qiagen) following manufacturer's instructions. A 363bp sequence within the L1 5'UTR region (nucleotide number 209-572; L1HS, RepBase) [22], which contains twenty CpG dinucleotides, was amplified using For: 5'-AAGGGGTTAGGGAGTTTTTTT-3' and Rev: 5'- TATCTATACCCTACCCCCAAAA-3'.Briefly, the 50µl PCR reaction was set up using 2X GoTaq green (Promega) and 200ng template bisulfite treated DNA. Untreated genomic DNA was used as a negative control. No PCR amplification with the untreated DNA template suggested that the genomic DNA was 100% converted by bisulfite treatment. PCR amplification was as follows: one cycle at 94°C for 30 sec followed by 30 cycles at 94°C for 20 sec, 54°C for 30 seconds and 72°C for 60 seconds and finally one cycle at 72°C for 5 minutes. The PCR products were resolved in a 1.2% agarose gel. Bands were excised, gel extracted, and subcloned in the pGEM-T vector (promega) followed by transformation and blue-white screening. Plasmid DNA was extracted from positive colonies (white) using mini-prep DNA kit (Qiagen). Clones were first checked by electrophoresis in 1.2% agarose gel and five clones from each sample sent for Sanger sequencing using T7 promoter and SP6 universal primers. The sequenced clones were first characterized using Repeat masker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker), which allowed for subfamily annotation. Most of the sequences belonged to either the L1-HS or L1P1 subfamily.

2.23. Cell culture:

HEK293T (human embryonic kidney) MCF7 (breast cancer) and DU145 (prostate cancer) cells were maintained in a CO2 incubator at 37°C and 5% CO_2 concentration 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 (Gibco. Thermo Fisher Scientific). 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.

2.24. Protein extraction and immunoblots:

Whole cell lysate was prepared from MCF-7 and Du145 cells using lysis buffer A [composition: 20mM Tris-Cl pH 7.8 ,137mM NaCl and 1% NP-40 supplemented with 1X protease inhibitor cocktail (Roche)]. The lysate was centrifuged at 2500xg for 5 minutes at 4°C and the supernatant was transferred to a new 1.5 mL tube which was stored at -70°C until further use. For preparation of cancer and normal tissue lysate, around 150 -200 mg frozen tissue was placed in liquid nitrogen and crushed using a mortar and pestle. Next, the sample was transferred to a 1.5 ml tube containing 250 µl of cold RIPA buffer [150 mMNaCl, 1% NP-40, 0.5% Na-deoxycholate,0.1% SDS, 50 mMTris-Cl pH-8.0 with protease inhibitor cocktail (Roche)]. The crushed tissue was then passed through an 18-gauge needle 5-8 times followed by incubation on ice for 45 minutes with intermittent mixing. Finally, the lysate was centrifuged at 12000 x g for 10 minutes at 4°C; supernatant was transferred to a new tube and stored at -70°C until further use. The Bradford reagent (Bio-Rad) was used to estimate the protein concentration. Protein lysate was separated by SDS-PAGE (Mini protein Tetra cell (Bio-Rad)) and wet transferred by applying 100V for 75 minutes (Bio-Rad mini trans blot electrophoretic transfer cell) to nitrocellulose membrane (Millipore). Protein was detected using the following primary antibody. Polyclonal rabbit human α-L1 ORF1p (RRM) (1:33000) [Sur et al., 2017], α-GAPDH (1:6000) (Santa Cruz Biotechnology), . Secondary α-rabbit HRP(code no. 111-035-003) and secondary α-mouse HRP (code No. 115-035-003) were purchased from 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).

2.25. Immunohistochemistry (IHC):

Paraffin-embedded normal and cancer tissue sections on glass slides were deparaffinized, rehydrated in descending grade of ethanol solutions before proceeding for antigen retrieval. The antigen retrieval step was adapted from "abcam protocol" available at (http://www.abcam.com/protocols/immunocytochemistry-immunofluorescence-protocol). Briefly, antigen retrieval was performed in a common household vegetable steamer (pressure cooker) using Tris-EDTA antigen retrieval buffer (10 mMTris base, 1 mM EDTA solution, 0.05% Tween 20, pH-9.0), then the slides were washed 2 X 5 minutes each in TBST (1X TBS containing 0.025% Triton-X100) and then blocked in blocking solution (1% BSA in 1XTBST) for 1 hour at room temperature. Thereafter, slides were incubated with polyclonal rabbit α-ORF1p (RRM) antibody (1:500 diluted in blocking reagent) at 4oC overnight in humid chamber. The next day, slides were washed with 1XTBST and treated with 0.3% hydrogen peroxide to quench any peroxidise present within the tissue. Slides were then incubated with secondary antibody ((1:500 dilution goat α-rabbit HRP (Jacksons Immuno Research)) for an hour at room temperature. The slides were washed again 3 X 10 minutes at room temperature with gentle agitation. Signals were visualised by adding 3-3´-Diaaminobenzidinetetrahydrochloride (DAB substrate) solution to the slides and counterstained with haematoxylin, (Himedia) dehydrated with ascending order of ethanol and mounted with DPX mounting media. Images were captured using a light microscope (Leica Microsystems) equipped with a camera. Intensity of DAB stained regions were measured with ImageRatio software [55] and plotted as percentage of expression.

2.26. Production of polyclonal antibody:

B) Immunization protocol:

The Rabbit was maintained in Jawaharlal Nehru University (JNU, New Delhi) animal facility as per institute animal use guidelines. The recombinant antigen [(Band excised human ORF2RT domain protein (SV1 clone)] was mixed with equal volume of Complete Freund's Adjuvent (CFA) and passed through a glass syringes to make an emulsion. The emulsion was injected subcutaneously as per following schedule. For subsequent immunization (Booster), Incomplete Freund's Adjuvant was used for subsequent immunization (Booster doses). Blood was collected and kept at 37°C for

one hour and then kept to 4°C overnight. Next day, the coagulated blood was centrifuged at 9000xg/4ºC/20 minutes and the supernatant containing the serum was collected and stored in aliquots at -20°C.

Parameter:

Parameter	**Rabbit**
Strain	NZW
Number of animal used	1
Amount of antigen Primary Booster	Primary : 150-200 µg Secondary: 100-150 µg
Volume of antigen with adjuvant	1.5 ml
Amount of Bleed	3-4 ml (test bleed) 10-12 ml (final bleed)

Immunization schedule:

Day	**Activity**
2 days before immunization	Pre bleed (checked by western for non- specific signal)
0	Primary immunization
15	Test bleed (1)
24	1st booster
38	Test bleed (2)
46	2nd booster
60	Final bleed

2.27. Statistical analysis:

The methylation value was calculated as mC/(hmC+mC) for all examined CpGs for a particular patient where hmC=hypomethylated Cytosine and mC=methylated Cytosine. The methylation index for the LINE-1 elements in paired tumor and normal tissues was calculated as mean value of mC/(hmC+mC) for all examined CpG dinucleotides. The One-Sample Kolmogorov-Smirnov test was used to evaluate fitness to normal distribution of continuous parameters. Paired t-test was used to determine if there was a statistically significant change in the methylation status of LINE-1 in OSCC tumor versus paired normal. All analyses were performed using the sigma plot 13 package (manufacturer or website. A p-value less than 0.05 ($p < 0.05$) was considered statistically significant.

List of Primers used in the study:

S. #	Name	Sequence (5'-3')
1	ORF1-HcF(EcoR1)	AATGAATTCGATGGGGAAAAAACAG
2	ORF1-HcR(Not1)]	AATGCGGCCGCTTACATTTTGGCATG
3	SV2RTSacFwd	ACTATTGAGCTCGAAGGAAATAGAGACACAA
4	SV2RTNotRev	ACTATTGCGGCCGCTCAGTTTTCTTCTAGGGT
5	SV3RTSacFwd	ATTGAGCTCGCAGGAAGAAGTTGAATCTCTG
6	Alu59sp6fwd	ATTTAGGTGACACTATAGAGATCACGAGGTCAGGAG
7	Alu296A10rev	TTTTTTTTTTGAGACGGAGTCTCGC
8	LEAP primer	GCGAGCACAGAATTAATACGACTGGTTTTTTTTTTT
9	Alu 74 F	GATCGAGACCATCCTGGCTAACACG
10	Link Rev	GCGAGCACAGAATTAATACGACTGG
11	Bisulfite Fwd	AAGGGGTTAGGGAGTTTTTTT
12	Bisulfite Rev	TATCTATACCCTACCCCCAAAA

Objective 1

3.1 To over-express human LINE-1 ORF2p protein in a bacterial expression vector and investigate its reverse transcriptase activity.

3.1.1. Cloning of human LINE-1 (L1) Reverse Transcriptase (RT) domain

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) : 1-239relative 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] .

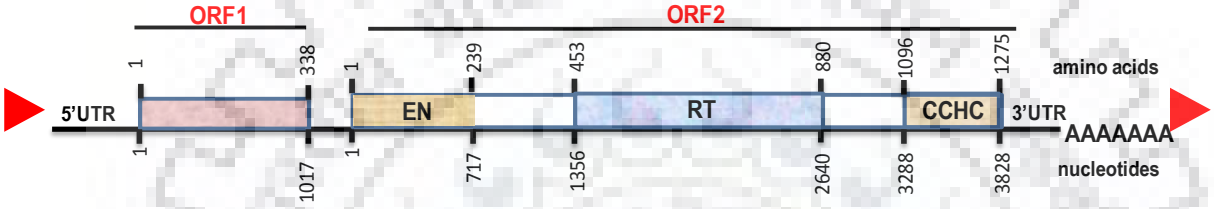


Figure R1: Schematic representation of amino acids and nucleotides map of human L1 ORF1p and ORF2p. EN: endonuclese, RT: reverse transcriptase, CCHC: CysCysHisCys type zinc finger domain. Red triangle indicates Target Site Duplication (TSD).

Although much has been learned for EN domain, our understanding on RT and CCHC domain is incomplete. Here, I have cloned Human L1RT domain to investigate the human L1 encoded RT activity in details. A disease causing human L1 cloned in pBluescript KS(-) (pBS-L1RP) {JCC5; 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 Uni., USA). The *EcoRI- BamHI* restriction map of ORF2 sequence is shown in Figure R2. In-order to clone the RT domain the *EcoRI-BamHI* fragment (1.4 kb) (Figure R2) was first cloned in *EcoRI- BamHI* site in pBS and then to pET 28a for expressing in bacterial system (Figure R3). The resultant clone (pET-hL1RT or SV1) encompasses ORF2p amino acids 474-987 and contain 34 amino acids extra at the N-terminal and 20 amino acids extra at the C-terminal which are part of pET28a expression vector (Appendix 1). The predicted molecular weight and PI of the SV1 RT domain protein is 61.2 kDa and 9.1 respectively (Appendix 1).

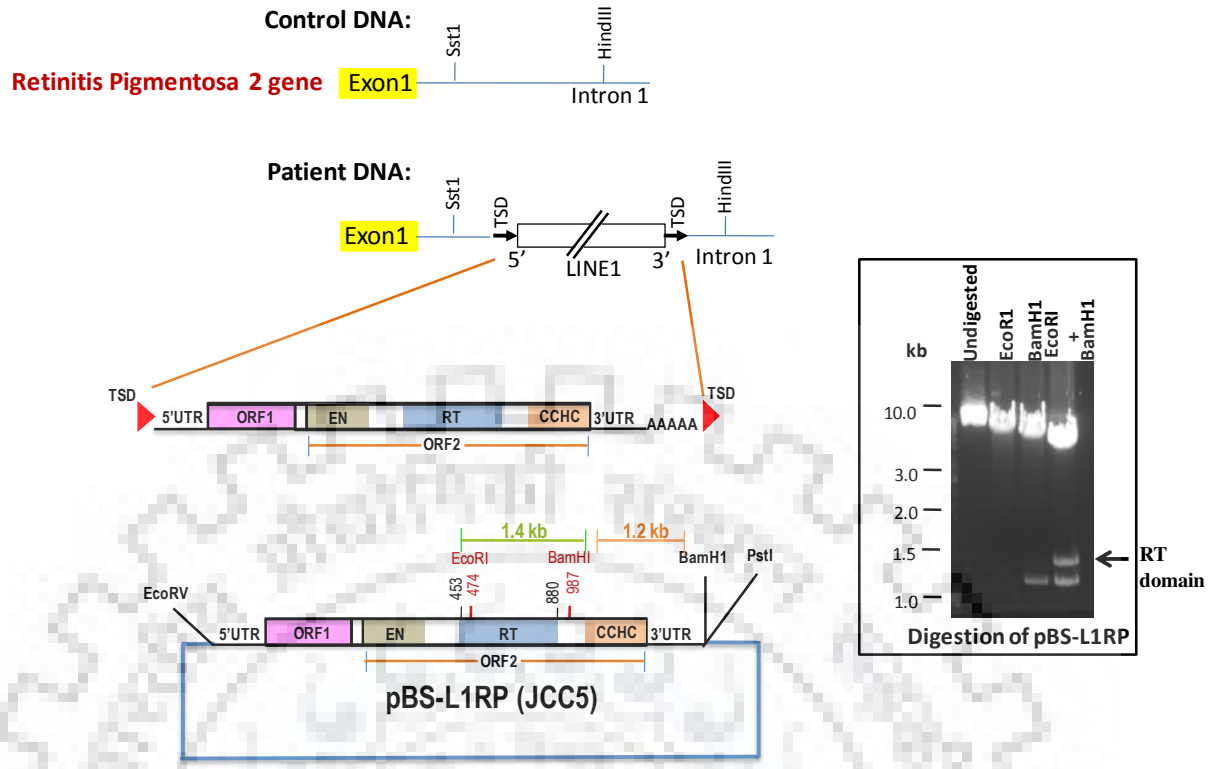


Figure R2: Schematic diagram of the full length L1 retrotransposon insertion in intron 1 of retinitis pigmentosa -2 (RP) gene in patient DNA. Black arrows flanking the L1 sequence represents 14 bp target site duplication (TSD). The complete L1 sequence was PCR amplified from a patient DNA and cloned in pBS KS(-) plasmid (JCC5 clone, Kimberland et al. 1999)was a kind gift from Kazazian Lab (Johns Hopkins University, USA). The JCC5 clone was digested with *EcoRI- BamHI* in-order to clone the 1.4 kb RT domain in bacterial expression vector.

3.1.2. Expression Studies of recombinant human L1RT (hL1RT) :

The expression of SV1 was checked in BL21 strain of *E.coli* at 37°C for 3 hours with 0.4mM IPTG. Total lysate from induced cells was analysed and showed significant expression of human L1 ORF2p RT (hL1RT) domain at around 61kDa (Figure R4A). As a control we took human L1 ORF1p cloned in pET 30b expression vector. In-order to confirm whether the protein with molecular mass of 61 kDa in induced cell lysate is recombinant human ORF2p, Western Blotting analysis was performed. The bacterial induced recombinant hL1RT contains six amino acids histidine tag at the N-terminal domain (Figure R4B) (Appendix 1).Using anti-His antibody, a band of around 61kDa was detected which confirmed the identity of recombinant hORF2p in bacterial induced total cell lysate. The human L1ORF1p cloned in bacterial expression vector (pET30b) was taken as positive control for immunoblotting experiment.

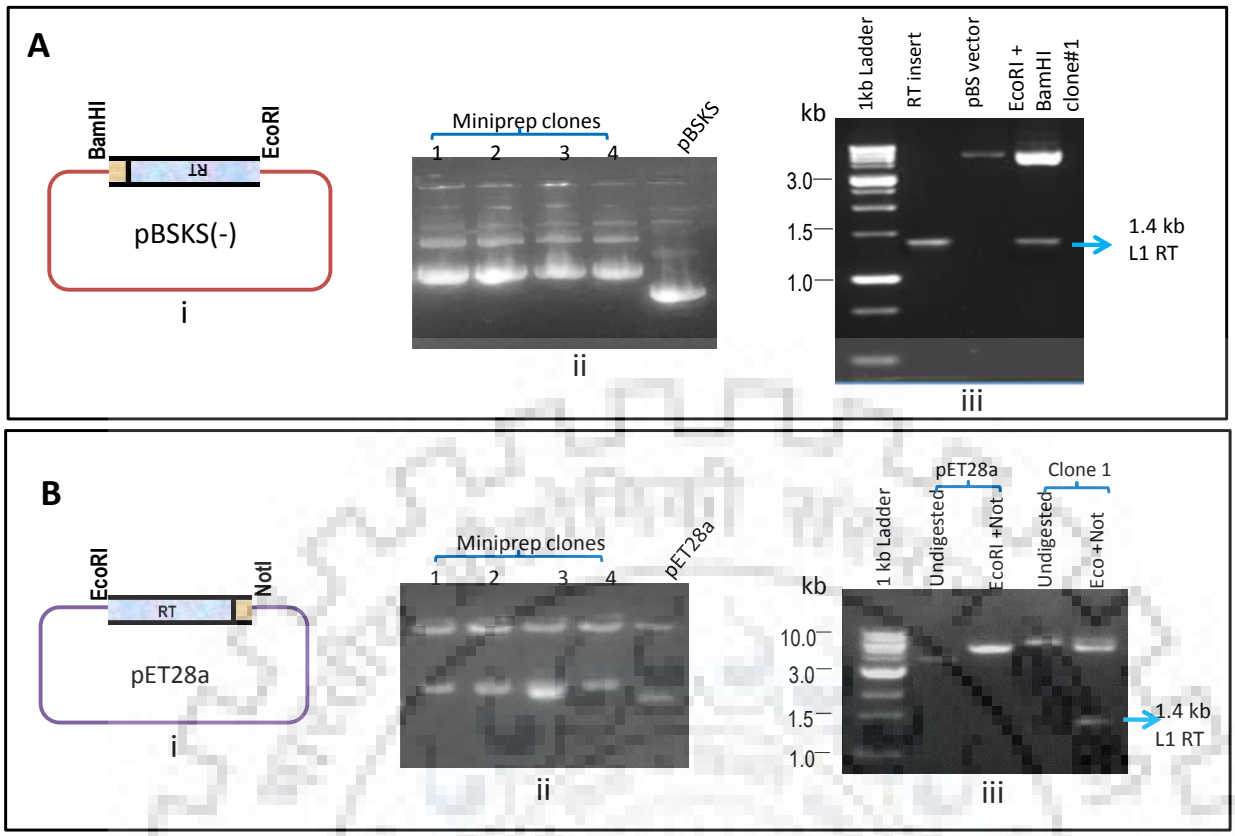


Figure R3: **Cloning of SV1 A.** (i) Schematic drawing of 1.4 kb *EcoRI-BamHI* human RT domain fragment in pBSKS(-). (ii) Mini prep clones showed retardation compared to empty vector. (iii)Restriction digestion of clone 1 confirm the insert of right size.
B.(i) Scheme of RT domain fragment cloned in *EcoRI-NotI* site of pET28a bacterial expression vector. (ii) Checking mini prep clones by agarose gel electrophoresis to find out right clone having insert. (iii) Restriction digestion of clone 1 with *EcoRI and NotI* restriction enzyme followed by agarose gel electrophoresis to confirm the right clone.

3.1.3 Solubility studies of recombinant human L1RT:

In-order to check whether the expressed RT from SV1 clone is in soluble fraction or form inclusion bodies, the supernatant and pellet of total cytoplasmic fraction was analysed in denatured SDS-PAGE gel. The results showed that the RT domain protein present only in the pellet fraction suggesting that at 37°C the protein mis-folded and formed inclusion bodies. The induced protein which is not folded properly and thus insoluble at 37°C might show proper folding if expressed at lower temperature. Thus, we analysed the expression of SV1 clone at 16°C and 30°C. Analysis of supernatant and pellet fraction by denatured SDS-PAGE revealed that although both

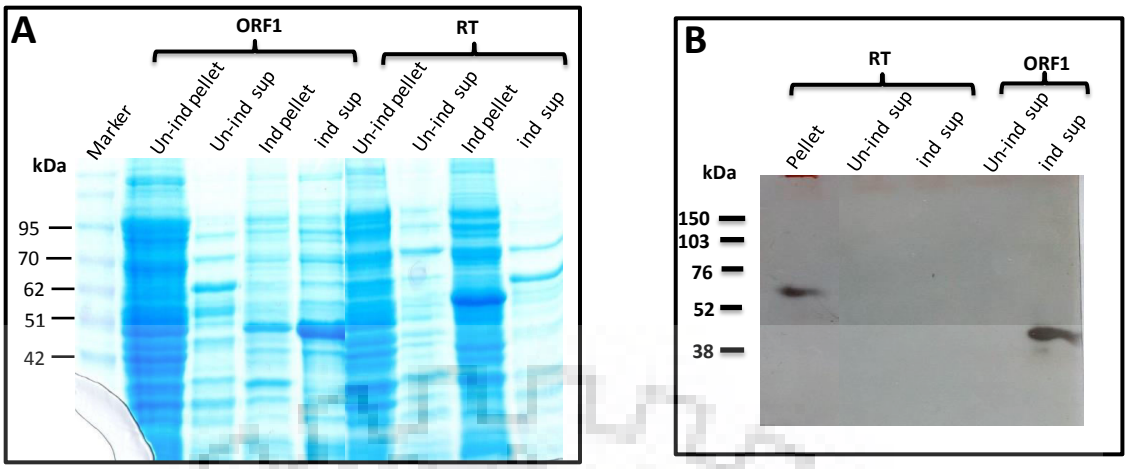


Figure R4: Checking expression for human RT domain protein using SV1 clone . **(A)** Both supernatant and pellet of induced cell lysate was analysed for the presence of human RT domain protein. The human L1 ORF1p cloned in pET30b was used as positive control. **(B)** Western blot confirmation of recombinant RT domain protein using anti-his antibody. A single band at around 61 kDa corresponds to the molecular weight of recombinant RT was detected only in pellet fraction. No protein was detected in supernatant. Human ORF1p was taken as control for Western technique.

the temperature favoured significant expression of RT from SV1 clone, the recombinant protein again was mis-folded and thus formed inclusion bodies (Figure R5).As the RT domain protein coming from SV1 clone was not folded properly, a bioinformatics analysis was performed to find out extra stretches of amino acids in ORF2 sequence either at N-terminal or C-terminal which might make human RT soluble (Figure R6). ccSOL*omics* a web server was used to predict the solubility of recombinant human RT protein in *E.coli* [Agostini et al., 2014]. The analysis predicted two clones SV2: amino acids 348-981 and SV3: amino acids 443-981 may be soluble if expressed in bacterial expression system (Figure R6 and Figure R7).

Next PCR primers were designed to clone both SV2 and SV3 in pET30b vector (Figure R8; Figure R9). Both the clones (pET30b-hRT_{SV2} or pET30b-hRT_{SV3}) were confirmed by sequencing before doing their expression studies (Appendix I). The expression studies showed only SV3 had significant expression in Coomassie stained SDS-PAGE gel as compared to control and at 37°C [Figure R11] but again the recombinant RT domain protein formed inclusion bodies in that specific temperature. Next, another two temperature i.e. 16°C and 30°C was checked to examine SV3 RT domain protein expression and solubility. The results showed although both the temperature favoured RT expression from SV3 clone, the protein in *E.Coli* host was

not folded properly thus formed inclusion bodies. As the expression studies of human RT domain protein from two clones (SV1 and SV3) resulted in aggregation of the expressed protein into the inclusion bodies, it was decided to purify the RT domain protein by solubilising the Inclusion bodies using urea followed by purification and

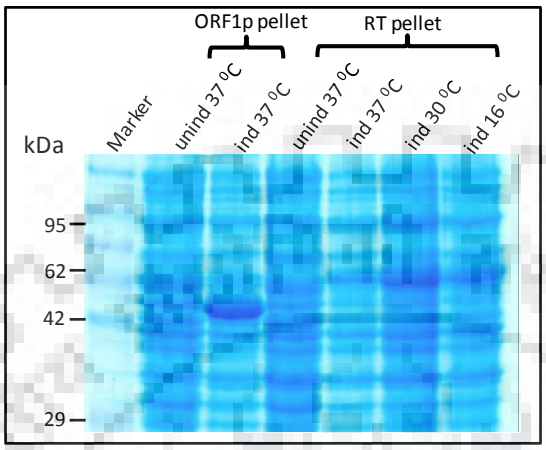


Figure R5: Expression of RT domain protein at different temperature. 10% denatured SDS-PAGE gel electrophoresis showing solubilization of the induced RT protein (SV1 clone) in three different temperature. Expression of RT protein (SV1 clone) in all three different temperature showed that protein formed inclusion bodies. Human L1 ORF1p cloned in pET30b vector was taken as positive control.

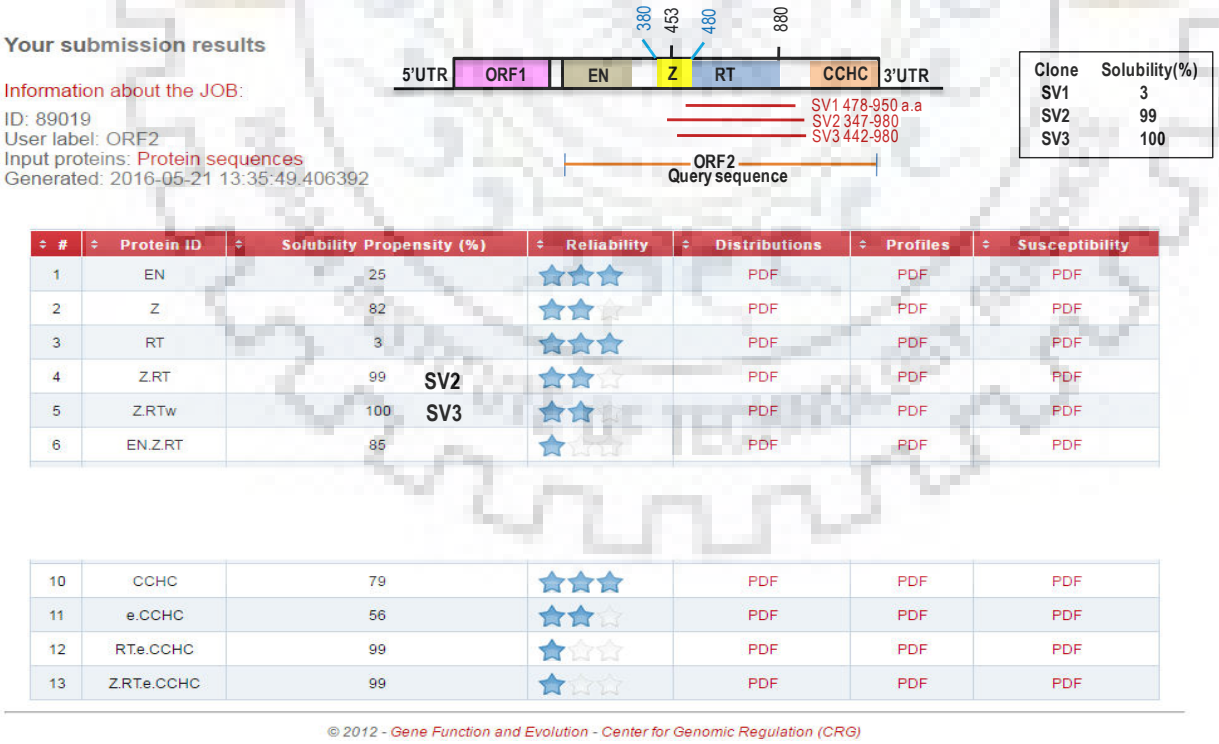


Figure R6: Human L1 encoded ORF2p solubility prediction by ccSol *Omics* webserver [Agostini et al., 2014]. Result showed RT domain with some extra amino acids at C-terminal and N-terminal might be more soluble in comparison to only RT domain

refolding to produce the bioactive molecules. The pellet was first solubilised in denaturing buffer containing different concentration of urea. The result showed that

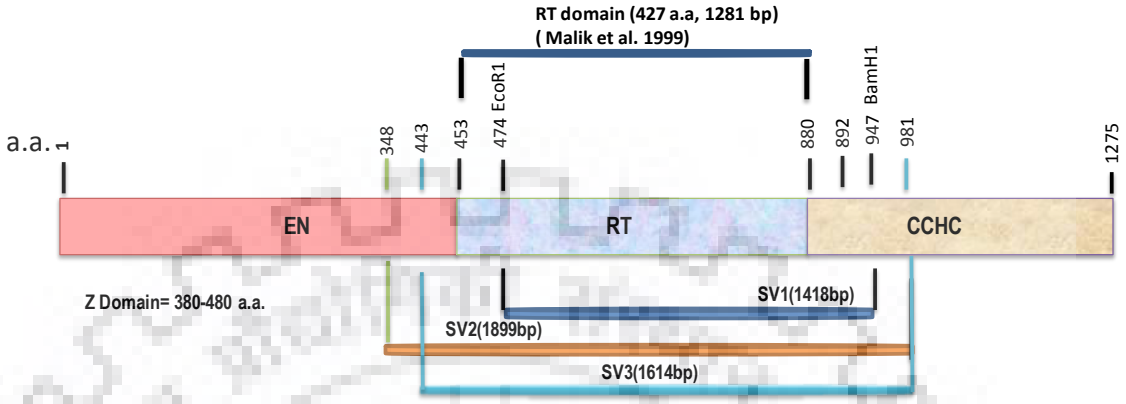


Figure R7: Scheme of three different RT clones. Three different length fragments (SV1, SV2 and SV3) encompassing RT domain selected to clone and express in bacterial expression vector.

RT domain protein (SV1 clone) was not soluble in the denaturing buffer with 2M urea concentration; however the protein came to the soluble fraction when urea concentration increases to 4M and 8M in denaturing buffer (Figure R10).

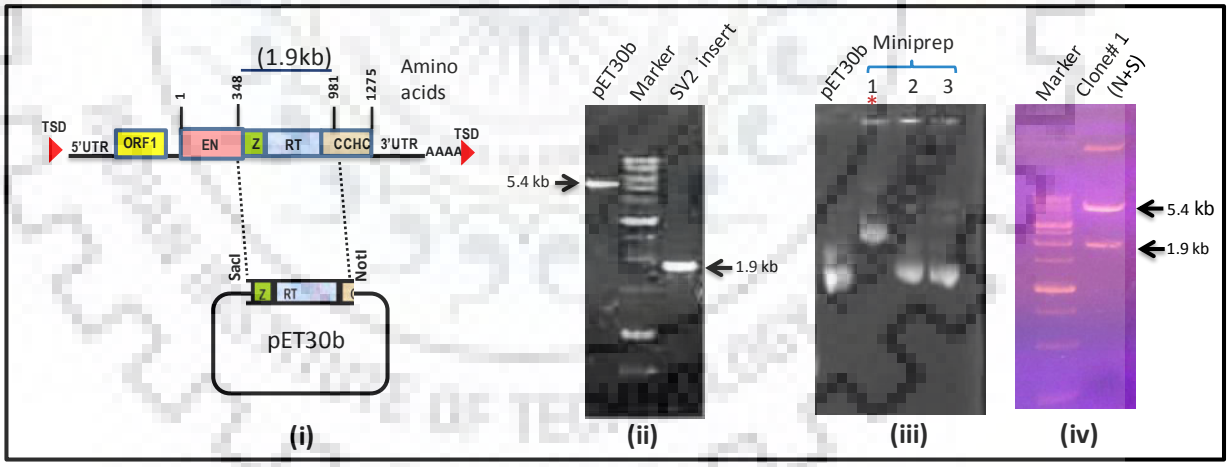


Figure R8: **Cloning of SV2 fragment in bacterial expression vector** (i) Scheme of L1 showing 1.9 kb fragment and its clone in pET30b expression vector. (ii) Primers were designed to amplify 1.9 kb fragment. The PCR product (insert) and NotI-SacI linearized pET30b (vector) resolved in 1.0% agarose gel. (iii) Agarose gel electrophoresis to screen the mini prep clones showing "Clone 1" might have required insert. iv. Confirmation of Clone 1 by digesting with NotI and SacI restriction enzyme which showed 1.9 kb and 5.4 kb bands the expected sizes of insert and vector respectively.

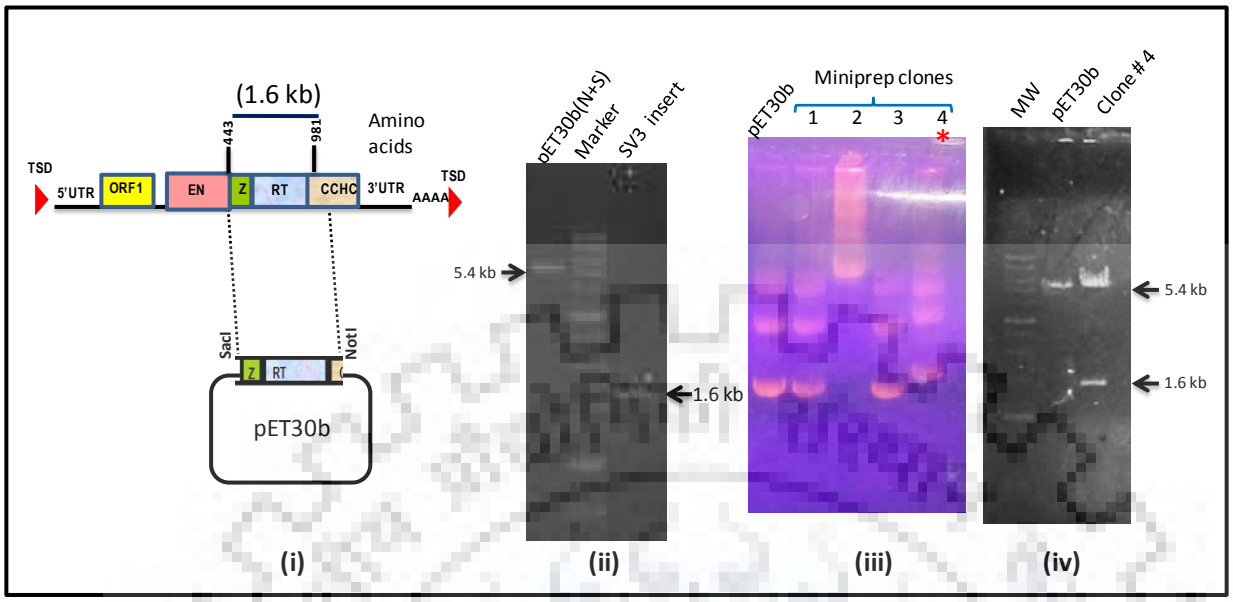


Figure R9: **Cloning of SV3 fragment in bacterial expression vector** (i) Schematic view of L1 showing 1.6kb fragment and its clone in pET30b expression vector. (ii) Primers were designed to amplify 1.6 kb fragment. The PCR product (insert) and NotI-SacI linearized pET30b (vector) resolved in 1.0% agarose gel. (iii) Agarose gel electrophoresis to screen the mini prep clones showing "Clone 4" might have required insert. (iv) Confirmation of Clone 4 by digesting with NotI and SacI restriction enzyme which showed 1.6 kb and 5.4 kb bands the expected sizes of insert and vector respectively.

The soluble fraction with 4M urea was then used to purify His tag fused RT domain protein using Ni- agarose chromatography. Analysis of the purified protein by SDS-PAGE gel electrophoresis followed by Coomassie staining revealed two bands in the elution fraction (Figure R10 Panel ii). One band corresponds to 61kDa, the RT domain protein. The second band around 48kDa was a major contaminant co-eluted with the RT domain protein. Analysis of the flow through (FT; not bound to the Ni-agarose beads during incubation) part showed that significant amount of the protein didn't bind to the Ni agarose. Also, a large fraction of the protein after binding to the Ni-agarose beads did not come to the elution (Figure R10 Panel ii). The recombinant human L1-ORF1p (cloned in pET30b) and vector induced (pET28a) lysates was used as control for the experiment (Figure R10 Panel ii)

Next, dialysis was performed to refold the recombinant RT domain protein by removal of urea from the eluted fraction in order to check the reverse transcriptase activity. The dialysis of the RT domain protein in the refolding buffer without any urea showed protein got precipitated in dialysis tube within an hour of incubation at cold

room suggesting that removing urea from the protein in one step crashing out protein from solution. To counter this problem, the dialysis was performed by sequential decrease of urea concentration in the dialysis buffer (refolding buffer). The result showed that the protein was in solution till the urea concentration was 1M in the

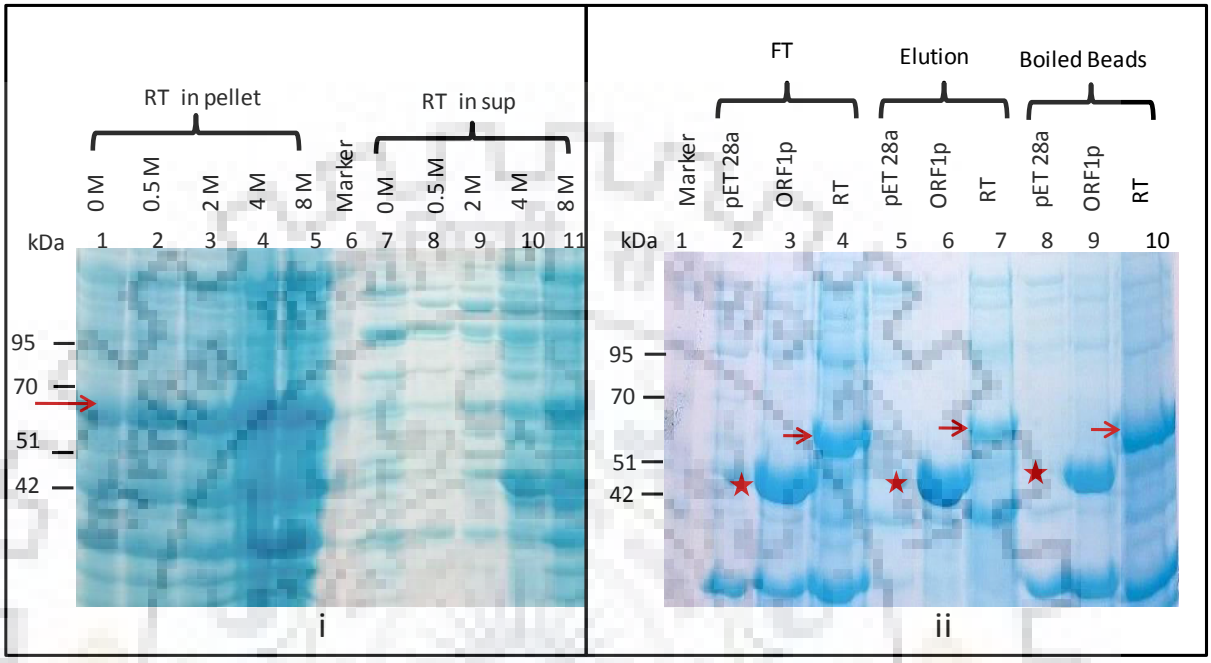


Figure R10: Solubilizing RT domain protein (SV1 clone) present in inclusion bodies using buffer containing different concentration of urea. **Panel i:** The induced pellet was dissolved in denaturing buffer and then centrifuged to separate supernatant and pellet. fractions. Panel 1 Lanes 1-5 showed a significant amount of RT protein is still in the pellet fraction even when 8M urea was used. Lane 6-11 showed the proteins which came to the soluble fraction after using different concentration of urea . The RT present in inclusion bodies was not soluble till urea concentration increased upto 2mM (Lanes 6, 7 and 8). The RT domain protein started coming in the soluble fraction at urea concentration 4 mM (lane 9) and showed maximum solubility at 8M concentration (lane 10). **Panel ii:** Purification of urea soluble RT from inclusion bodies by affinity purification. The Ni-agarose chromatography was performed as the RT proteins contains six histidine amino acids (His-tag) at the N terminal. The elution showed significant amount of RT protein purified by Ni-agarose chromatography (lane 7) . Another protein MW 40 kDa also eluted with RT. Analysis of flow through (FT) fraction revealed that a substantial amount of protein didn't bind with the beads (lane 4). Also analysis of boiled beads showed that significant amount of RT protein was still bound to the beads even after even using 250mM imidazole concentration in elution buffer (lane 10) . Induced pET 28a and human ORF1p cloned in pET30a was used as control samples.

dialysis buffer. The dialysed protein containing 1M urea in refolding buffer was then used to see if it show any RT activity on Alu RNA template (described below).The *in vitro* RT assay showed that the human RT in elution buffer with 1M urea was not active (data not shown as the gel was completely blank. The same strategy of purification was used to purify RT domain protein expressed from SV3 clone (Fig. R11)

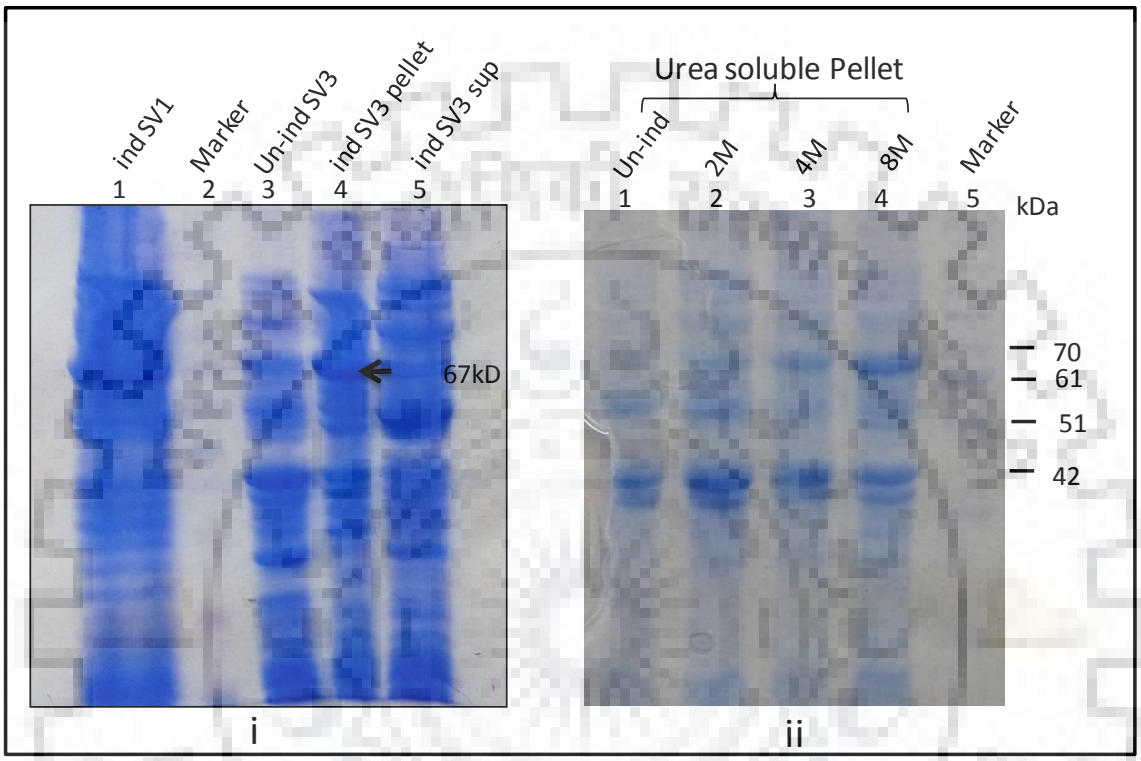


Figure R11: Solubilizing RT domain protein (SV3clone) present in inclusion bodies using buffer containing different concentration of urea as described for SV1 clone (Figure R10). **Panel i:** SDS gel showed that RT domain protein coming from SV3 clone formed inclusion bodies and thus present only in pellet fraction (lane 4). **Panel ii**: The induced pellet was dissolved in buffer with different concentration of urea and showed buffer containing 8M urea made RT protein soluble from inclusion bodies (lane 4).

Although we have checked that almost total induced RT from all three clones formed inclusion bodies after expressing in bacteria, it might possible that still a little amount of protein is presentin the soluble fraction. The SV1, SV2 and SV3 clones were induced with 0.4mM IPTG at 37°C and after lysis, pellet and supernatant was separated. The supernatant from all three clones were used to purify RT protein using Ni-agarose chromatography. The elution fractions were separated on denatured SDS-PAGE gel and western blotting was performed using anti-His antibody. The result

showed less than 1% of expressed RT was still in soluble fraction (Figure R12).The data showed elution fraction from clone SV3 contained more RT protein compared to other two clones. The SV2 clone showed some degradation; hence purified RT from this clone was not used for reverse transcriptase assay. The purified RT protein from SV1 and SV3 was then used to check reverse transcriptase activity.

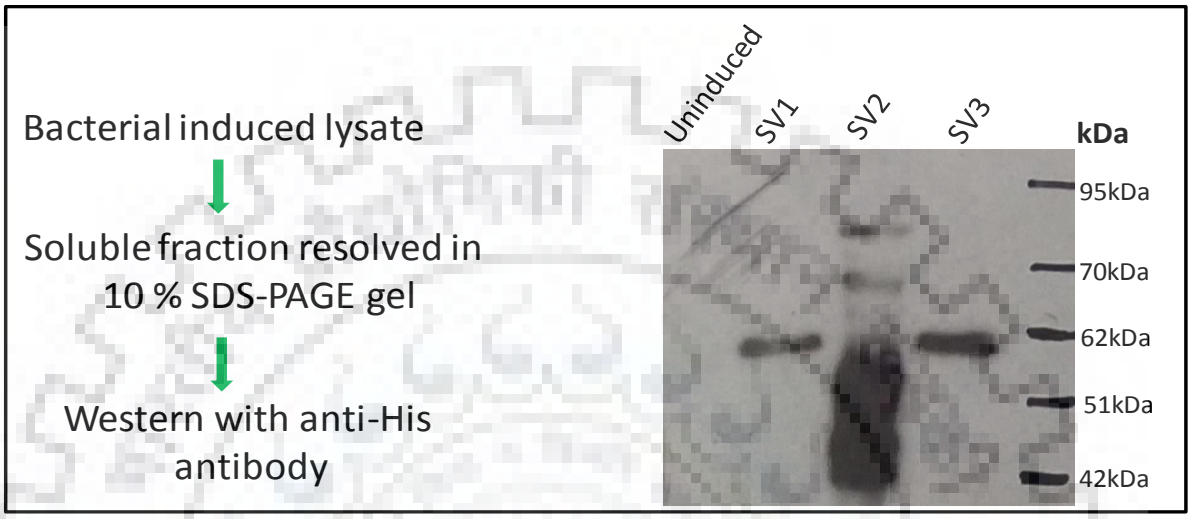


Figure R12: Western detection of human L1 RT domain protein in bacterial soluble lysate. RT clones (SV1, SV2 and SV3) were induced with IPTG , total lysate was seperated by centrifugation into soluble fraction and pallet. The soluble fraction was passed through Ni-agarose chromatography. The elations were checked for the presence of RT by immunoblotting using anti-His antibody.

3.1.4. RT Activity analysis of recombinant human L1 RT domain protein:

To find out if recombinant ORF2p RT domain proteins (purified from SV1, SV2 and SV3 clones) show RT activity, we carried out reverse transcriptase assay referred to as LINE Element Amplification Protocol (LEAP) originally designed to show human ORF2p mediated RT activity [Kulpa et al., 2006].Less than 1% RT protein present in soluble fraction of induced lysate from all three RT clones were purified using Ni-agarose beads. The bound protein was eluted using 100 ul elution buffer [Elution 1 (E1)] ; the step was repeated one more time to get all the RT protein bound with the beads and the fraction was labelled as elution 2 (E2).Briefly the purified RT domain protein was incubated with a primer that contains a unique linker sequence (length 20 nt) at the 5' end followed by 12-nt poly (T) sequence and with an *in-vitro* synthesized Alu RNA (Figure R13). If the RT domain protein is active, elongation will occur which

can be detected by carrying out PCR with an Alu specific forward primer and a linker specific reverse primer (Figure R14). Analysis of PCR reaction resolved on a agarose gel demonstrated a LEAP product at around 250 base pairs corresponds to the expected size. The MMLV reverse transcriptase was used as a positive control which showed the same LEAP product on agarose gel (Figure R14). The negative control without any RT domain protein added in the LEAP reaction didn't show any product after PCR. These assay confirmed that the soluble fraction of recombinant human RT RT domain protein purified from clones SV1 and SV3 using bacterial expression system is biological active.

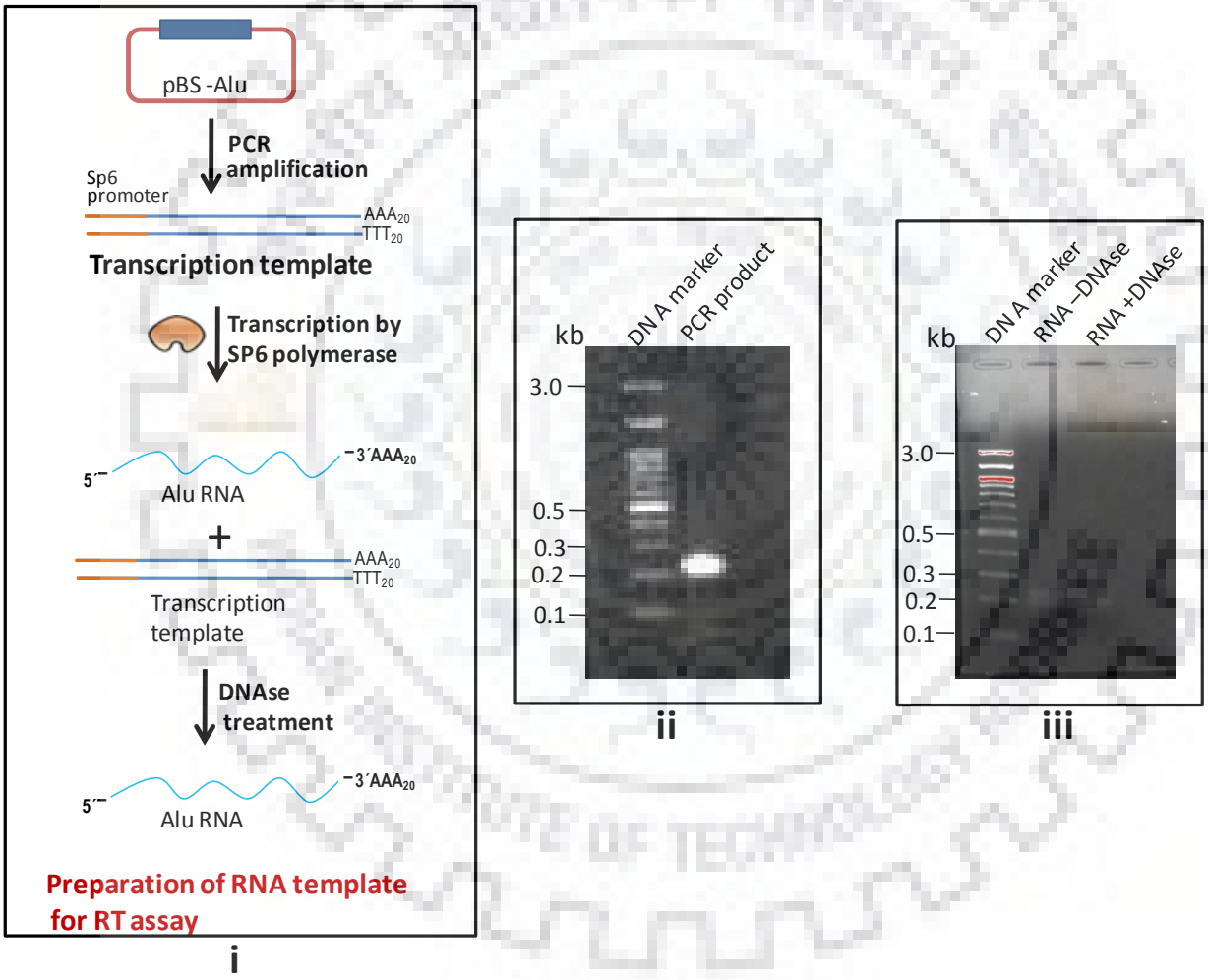


Figure R13 : Preparation of RNA template for RT assay: **Panel i:** Scheme of RNA template preparation for RT assay: The human Alu sequence cloned in pBS KS(-) plasmid (pBS-Alu) was amplified using primer set Alu 59SP6Fwd and Alu296A$_{10}$Rev . **Panel ii:** The resultant PCR product is 296 bp in length contains SP6 promoter at the 5'-site and 10 A nucleotides at the 3'-end resolved in 1.5% agarose gel.. **Panel iii:** *In vitro* RNA synthesis was carried out using SP6 polymerase, the template DNA was removed by DNAse treatment and the synthesized RNA was checked by resolving in 1.2% agarose gel.

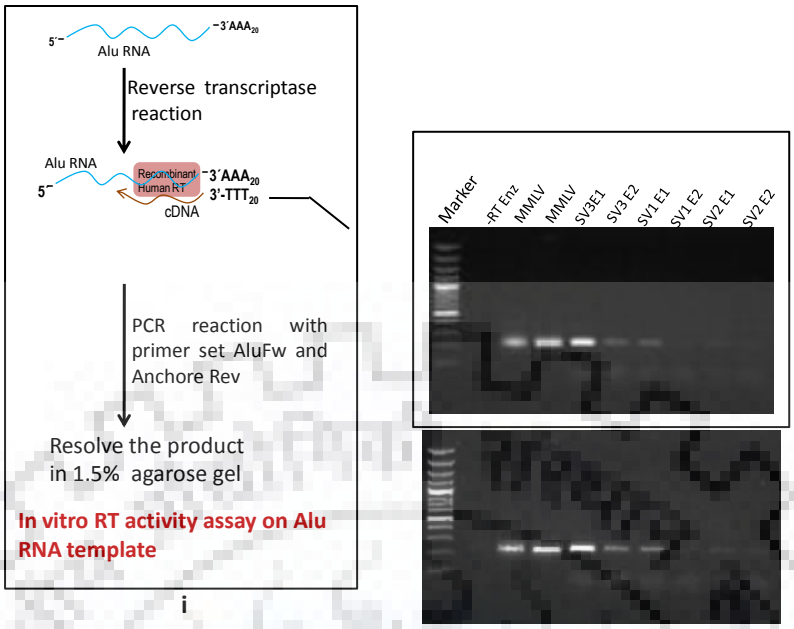


Figure R14: **RT activity assay. Panel i:** Scheme of RT activity assay. *In vitro* synthesized Alu RNA was incubated with purified human LINE-1 RT protein and an anchored primer which ends with 10T nucleotides at the 3'-end. If human LINE-1 RT protein is active, it will synthesize cDNA on Alu RNA template. Around one-tenths volume of cDNA was used as template to PCR amplify with primer pair (Alu Fwd and Linker Rev) . The PCR amplified product was then resolved in 1.5% agarose to check the product of RT. **Panel ii:** The cDNA reaction without any RT enzyme added was used as negative control. Replacing human L1 RT (SV1 or SV3) with MMLV RT in the reaction mix was used as positive control. No activity was detected from RT obtained from SV2 clones. SV3E1 and SV3E2 represents elution 1 (E1) and elution 2 (E2) respectively. Same acronym is used for clone SV1 and SV2.

1.5.

Generating antibody against human L1-ORF2p:

In-order to make antibody against ORF2p, the RT domain protein (SV1 clone) was purified from inclusion bodies by dissolving urea followed by passing through Ni-NTA agarose column. The purified protein was resolved in 10% denatured SDS-PAGE gel (Figure R15). The band corresponds to RT protein was then excised from the gel and mixed in homogeneity with Freund's complete adjuvant and injected to rabbit. After two subsequent booster administrations, the serum was checked for ORF2p antibody. The immunoblot didn't show any band suggesting that the injected RT might not immunogenic to rabbit or due to some experimental fault. Since, the aim was to generate ORF2p antibody in-order to detect L1 activity in cancer samples and as it was not successful, next antibody against ORF1p was planned carried out.

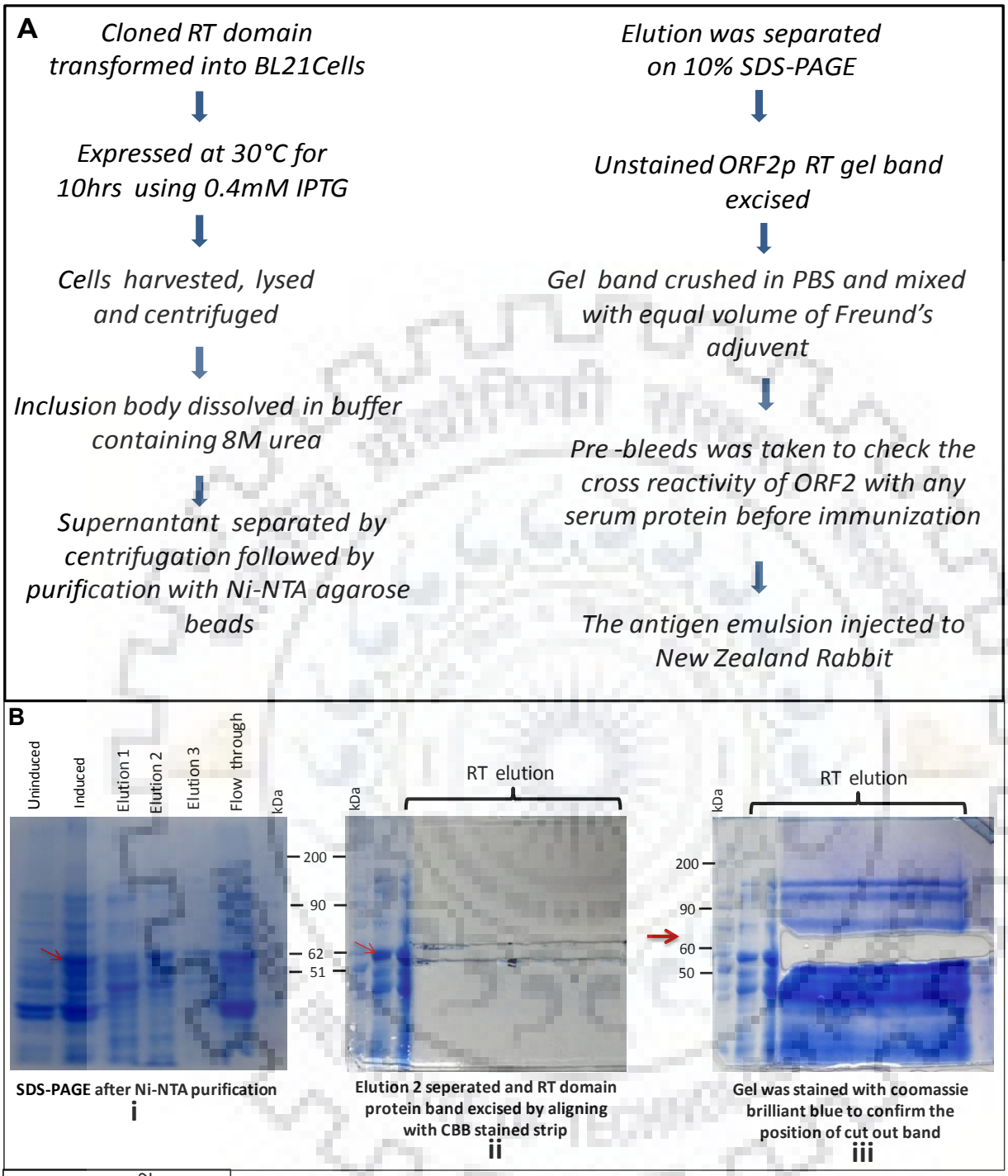


Figure R15: A) Scheme of making antibody against ORF2p using bacterial expressed RT domain protein purified from SV1 clone. **B)** Purification of RT protein for antibody production. **Panel i -** 10% denatured SDS-PAGE gel showed expression and partial purification of RT from inclusion bodies of SV1 clone. **Panel ii -** Purified protein was separated on 10% denatured SDS-PAGE gel and the unstained gel was align with a strip of stained gel containing MW marker and purified RT to excise the band of interest. The excised band was mixed with Complete Freund's Adjuvent and then injected to rabbit. **Panel iii -** The excised gel was stained to confirm that the right band was excised for immunization. **C) -** Western blot analysis of total lysate obtained from induced pET30a and pET-hRTSV1 clone using rabbit immune sera as primary antibody.

3.1.6. Sub-cloning of Human LINE-1 ORF1p in bacterial expression vector for its expression and purification in order to generate antibody:

Human ORF1p protein often used to check the activity of L1 retrotransposons in cells and tissues. In 2013 when the project was started, no commercial antibody against ORF1p was available in the market. Today, only one antibody against human L1ORF1p is available in the market [Rodic et al., 2014] (Anti-LINE-1 ORF1 antibody; EMD Millipore Cat. # MABC1152). In order to express and purify recombinant human ORF1p in bacterial expression system, the human L1 ORF1p was PCR amplified (nucleotides 907-1923 as per L1RP accession number AF:148856.1) [Kimberland et al.,1999] from pBS-L1RP and then sub-cloned to pET30b expression vector. Induction studies of pET-ORF1p clone showed significant expression of ORF1p in BL-21*E.coli* cells. Next, the total bacterial induced lysate was incubated with Ni-agarose beads and the bound ORF1p was eluted by incubating beads with elution buffer containing 250 mM imidazole. The purified protein showed multiple lower molecular weight bands possibly the degradation products of intact L1ORF1p which could not get rid off after repeating the experiments multiple times. In parallel, our laboratory was attempting to make ORF1p antibody using only the RRM domain of ORF1p and the experiment was successful [Sur et al, 2017]. The anti-RRM ORF1 antibody raised in rabbit showed distinct single band at around 40 kDa when Western blot analysis was performed using total lysate from MCF breast cancer cell line (Figure R 22A). It is important to mention that MCF cell lines expressed significant amount of endogenous ORF1p [Chen et al., 2012]. So his antibody (anti-RRMhORF1p) was used for investigating L1 retrotransposon activity in oral cancer samples.

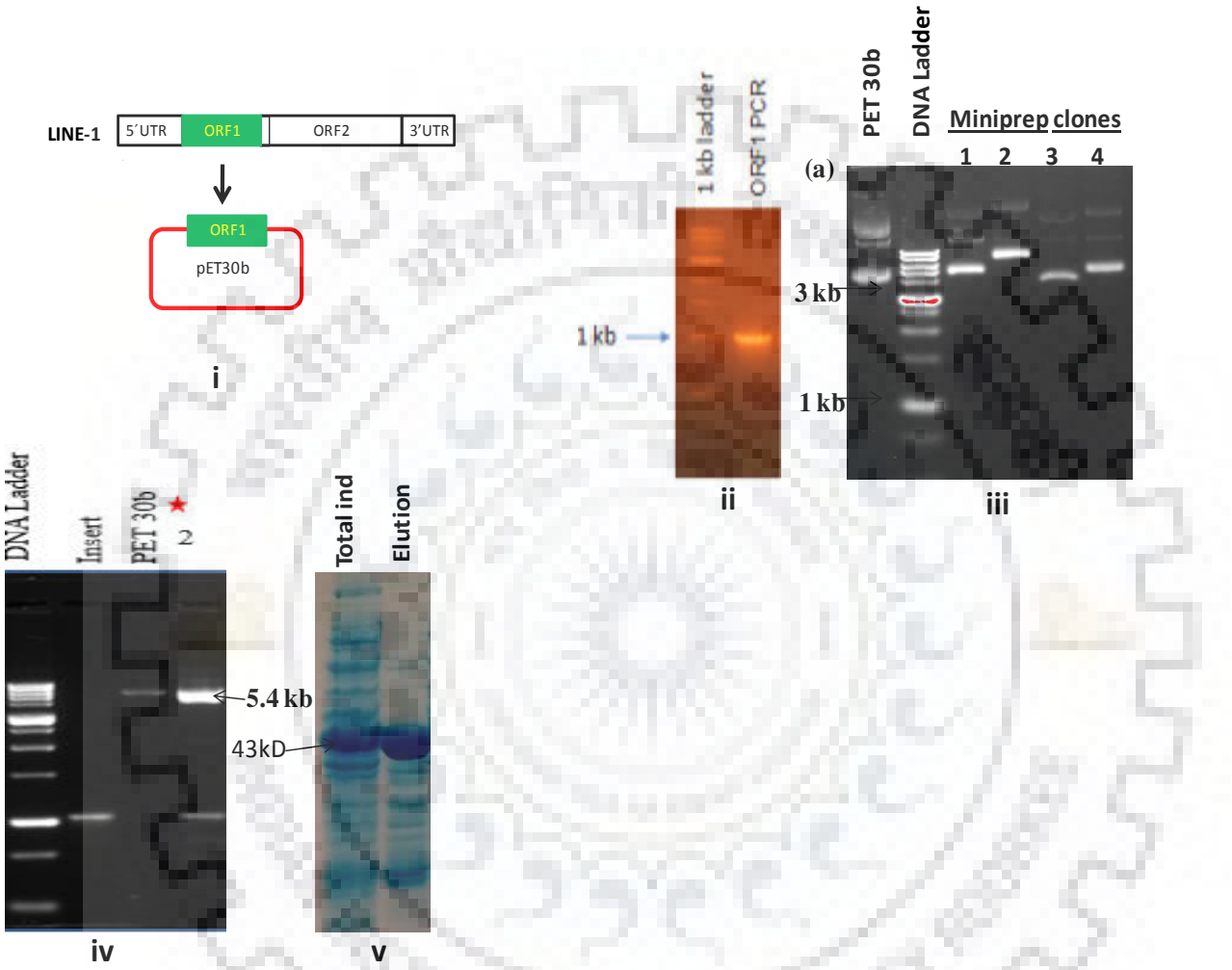


Figure R16: **Cloning of human L1 ORF1 fragment in bacterial expression vector** (i) Scheme of L1 showing ORF1 fragment and its clone in pET30b expression vector. (ii) Primers were designed to amplify 990 bp ORF1p fragment. The PCR product (insert) resolved in 1.0% agarose gel. (iii) Agarose gel electrophoresis to screen the mini prep clones showing "Clone 2" might have required insert. (iv) Confirmation of Clone 2 by digesting with NotI and SacI restriction enzyme which showed 1kb and 5.4 kb bands the expected sizes of insert and vector respectively. (v) SDS PAGE gel showed total induced (lane 1) and Ni-agarose purified ORF1p (lane 2) form *E.coli* lysate.

Objective 2

3.2. To investigate the methylation status of Human LINE-1 retrotransposon promoter and LINE-1 ORF1p expression in OSCC samples

3.2.1. Loss of DNA methylation at CpGs within the L1 5'-UTR in oral cancer samples

Epigenetic silencing of the L1 5'-UTR by DNA methylation is a common means to inactivate L1 expression and ultimately retrotransposition. Epigenetic alterations are frequent in cancers; indeed, several studies have reported L1 promoter Hypomethylation in a variety of cancers [van Hoesel et al., 2012; Daskalos et al., 2009; Saito et al., 2010]. To date, the methylated state of the L1 5'-UTR in OSCC remained unexamined; therefore, we performed bisulfite conversion analysis of genomic DNA across eight paired normal-cancer tissues followed by PCR, subcloning of amplicons, and sanger sequencing to ascertain the methylation level of the L1 promoter

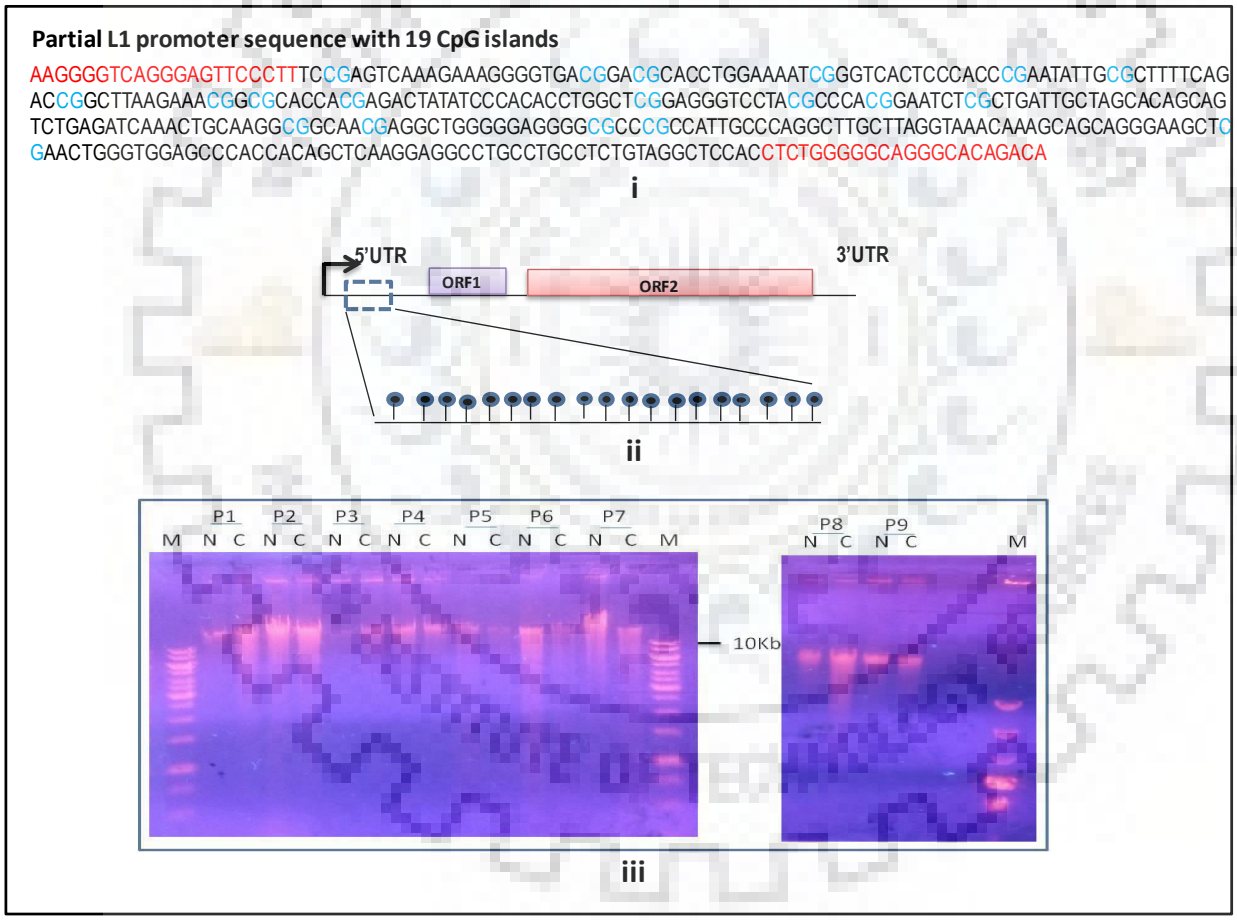


Figure R17: **Isolation of paired normal tumor genomic DNA from cancer patients for L1 promoter methylation study** (i) Partial L1 5'UTR sequence (nucleotide number 209-572) containing 19 CpG dinucleotides. (ii) Scheme of L1 showing the positions of 19 CpG dinucletides in L1 5'UTR. (iii) Total genomic DNA was purified from 9 matched normal cancer tissue and separated in 0.6% agarose gel. P1 Patient 1; M- DNA MW marker; N- normal; C- Cancer.

(Figure R17 and Figure R18). The details of patients used in this study are provided in Appendix 2. Specifically by taking bisulfite converted genomic DNA (normal or paired cancer) as template, I amplified a 363 bp region of the L1 promoter (nucleotide sequence 209-572, L1HS from Repbase [Bao et al. 2017] which contains 20 CpG sites and the resultant amplicons were sequenced (Figure R18 and Figure R19). For each sample pair, (e.g. OSCC/matched normal) five independent clones were sequenced (Figure R19B) (clone sequences were included in Appendix 3).Consistently, our sequence analysis uncovered that most of the L1 amplicons belonged to the major active L1 subfamily, L1-Hs (Table 3.1), suggesting that we were tracking potentially active L1s in the tumor tissues.

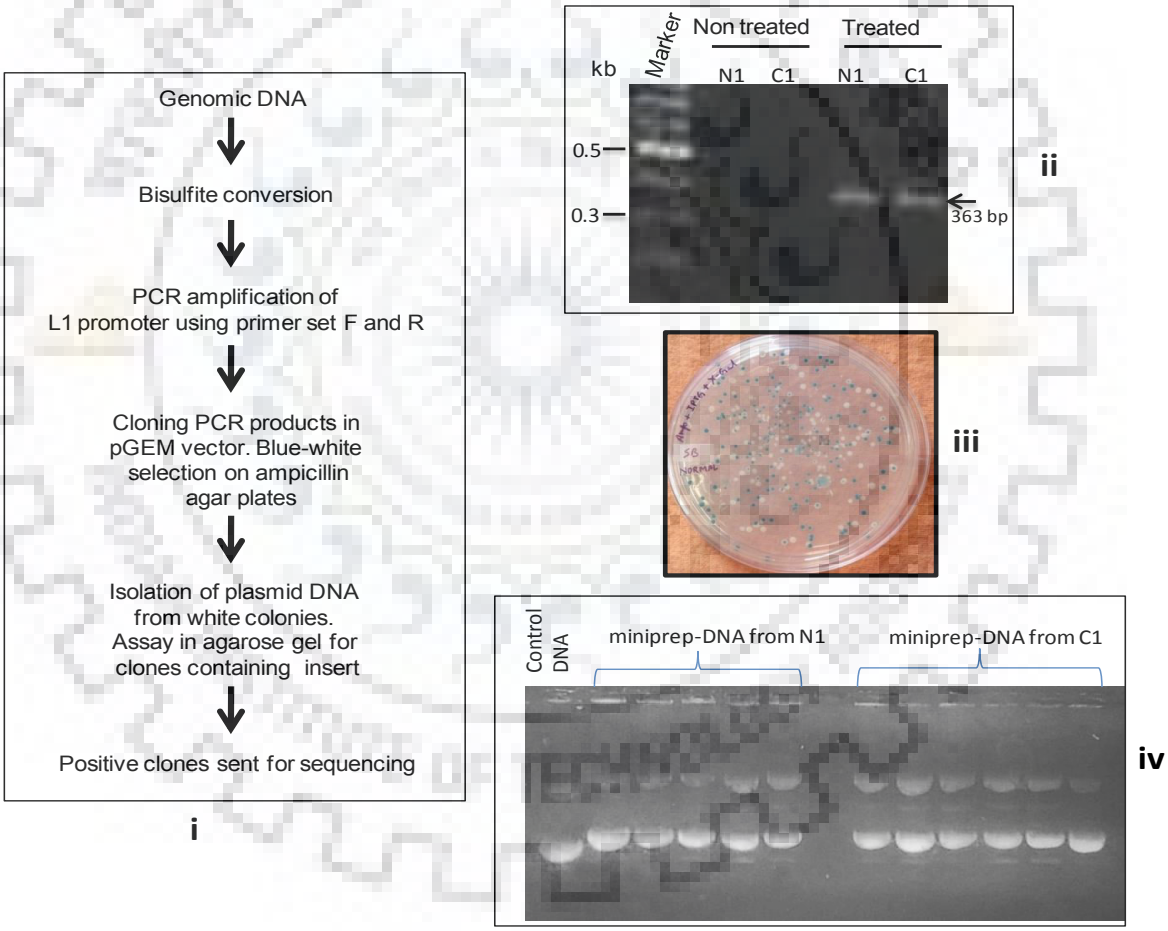


Figure R18: DNA methylation study (i)Flow chart of L1 5'-UTR DNA methylation study. (ii) Bisulfite-treated genomic DNA was used as template to amplify a 363bp fragment from L1 5'-UTR containing 19 CpG dinucleotides. Untreated DNA did not show any band with bilsufite-treated specific primers. N1- Normal, C1 Cancer. (iii) PCR products from both matched normal and cancer were subcloned into pGEM-T (Promega) and screened by blue-white selection on agar plates containing IPTG X-Gal. (iv) Plasmid DNA from white colonies was isolated and checked in agarose gel for the presence of insert.

Of the eight paired samples analyzed, significant hypomethylation was detected in five of the tumor tissues (Sample S1, S2, S3, S4, and S15) relative to matched normal tissue (Figure R19B; Figure R20A). When comparing normal to matched OSCC tissue, S15 exhibited the largest difference in hypomethylation (more than 2-fold) in the cancer tissue whereas S7 and S14 displayed no detectable difference in methylation level between the cancer and matched normal tissue (Figure R20A).

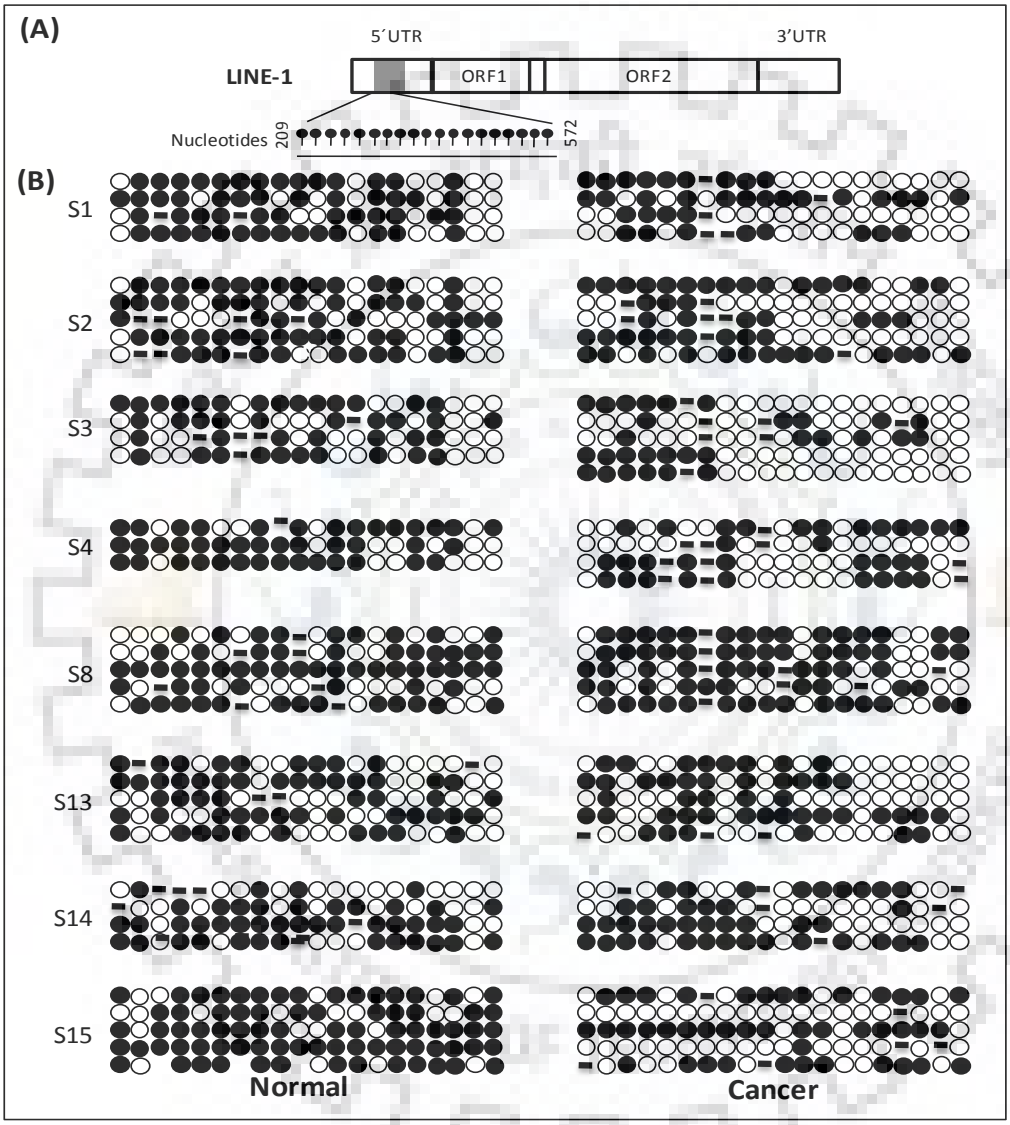


Figure R19: Bisulfite analysis of paired normal cancer tissues: A) Scheme of a full-length active human L1 containing a 5'-UTR encoding an internal promoter, ORF1p and ORF2p and a 3'-UTR. The positions of 20 CpG residues analysed in this study are shown as lollipops. B) Bisulfite analysis of eight paired normal cancer tissues to determine DNA methylation levels of the L1 promoter in OSCC tissues. An average of 5 clones were Sanger sequenced for each patient sample. The position of each CpG residue is relative to the sequence of L1-Hs (Repbase) [Bao et al. 2015]. Open and closed circles denote unmethylated and methylated cytosines, respectively. Horizontal dashes indicate mutated CpG site.

Table 3.1: Blast results of methylated (paired cancer normal) clone sequences. For both cancer and paired normal samples, the region of the L1 promoter (shaded yellow) amplified after bisulfite treatment was cloned. Three to five representative clones were sequenced from each cancer and paired normal samples and then blast using Rep Base [Bao et al., 2015] to find out in which L1 subfamily those clones are belonged to.

	Cancer		Paired Normal	
	Clone #	**Matched L1**	**Clone #**	**Matched L1**
Sample 1	C1a	L1HS	N1a	L1HS
	C1b	L1HS	N1b	L1HS
	C1c	L1P1	N1c	L1P1
	C1d	L1P1	N1d	L1HS
	C1e	No sequence	N1e	No sequence
Sample 2	C2a	L1PA2	N2a	L1HS
	C2b	L1P1	N2b	L1HS
	C2c	L1P1	N2c	L1P1
	C2d	L1HS	N2d	L1HS
	C2e	L1HS	N2e	L1P1
Sample 3	C3a	L1HS	N3a	L1HS
	C3b	L1P1	N3b	L1HS
	C3c	L1P1	N3c	L1P1
	C3d	L1HS	N3d	L1P1
	C3e	L1HS	N3e	No sequence
Sample 4	C4a	L1P1	N4a	L1HS
	C4b	L1P1	N4b	L1HS
	C4c	L1P1	N4c	L1P2
	C4d	L1P1	N4d	No sequence
	C4e	No sequence	N4e	L1P1
Sample 8	C8a	L1HS	N8a	L1HS
	C8b	L1HS	N8b	L1HS
	C8c	L1P1	N8c	L1HS
	C8d	L1HS	N8d	L1HS
	C8e	L1P1	N8e	L1HS
Sample 13	C13a	L1HS	N13a	L1HS
	C13b	L1HS	N13b	L1HS
	C13c	L1HS	N13c	L1P1
	C13d	L1HS	N13d	L1HS
	C13e	L1HS	N13e	L1P1
Sample 14	C14a	L1P1	N14a	No sequence
	C14b	L1P1	N14b	L1P1
	C14c	L1HS	N14c	L1P1
	C14d	L1P1	N14d	L1HS
	C14e	L1HS	N14e	L1HS
Sample 15	C15a	L1P1	N15a	L1HS
	C15b	L1HS	N15b	L1P1
	C15c	L1HS	N15c	L1HS
	C15d	L1HS	N15d	L1HS
	C15e	L1P1	N15e	L1HS

Further interrogation of specific CpGs within the amplified L1 5'-UTR sequence indicates that not all sites lose methylation in OSCC. Specifically, some positions are more hypomethylated (site: 4, 5, 8, 9, 10, 12, 14, 15 and 18), while several (site: 2, 3, 11, 13, 16, 17 and 19) show no differences in methylation state between the cancer and paired normal (Figure R20B). Quantification of methylation at L1-5'-UTR showed distinct differences when comparing normal methylation index = 64.78± 2.7) to OSCC (49.46 ± 3.9) (Figure R20C). The calculation is included in the appendix. These data demonstrate that L1 promoter CpG dinucleotide hypomethylation is common in OSCC tissue samples.

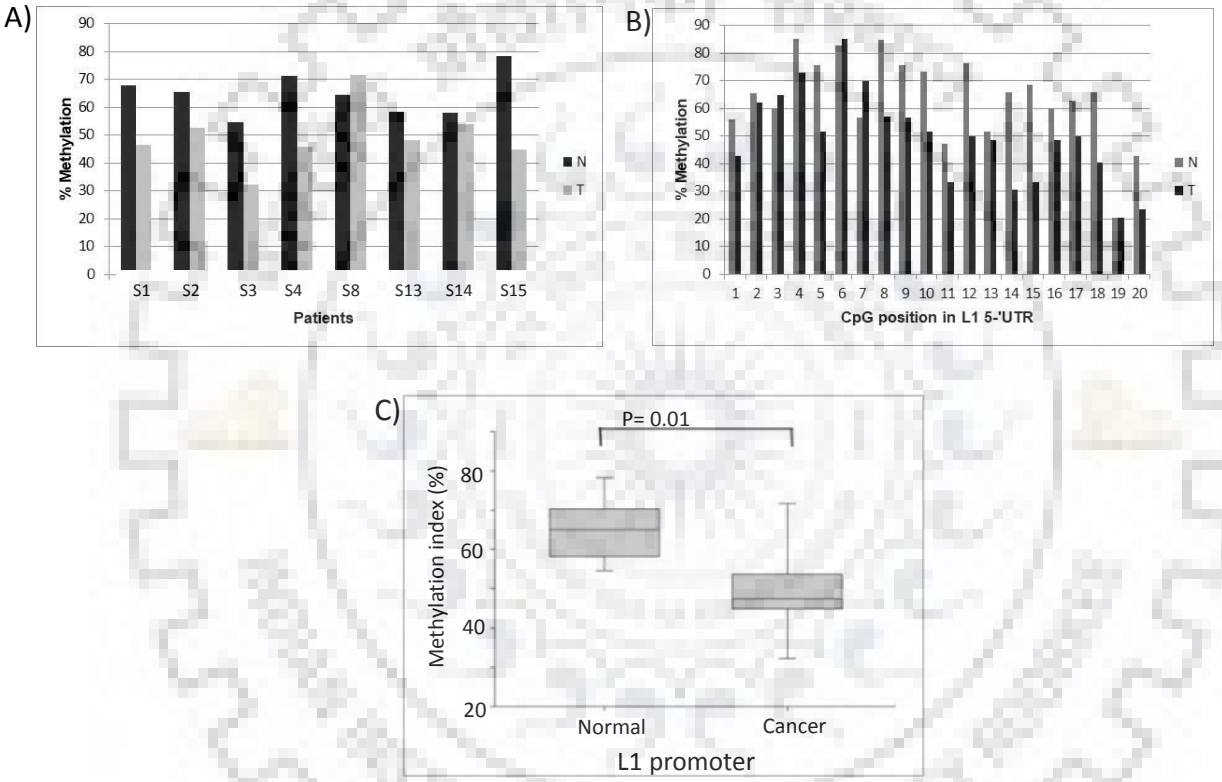


Figure R20: Analysis of methylation status in paired normal cancer tissues: A) % methylation of 8 OSCC patient tissues relative to matched samples. B) Methylation levels for each of the twenty CpG dinucleotides in the L1 promoter assayed across the matched tissues. C) Methylation index for the L1 promoter of eight OSCC tissues compared to matched normal tissue. The p-value was calculated by 2-tailed paired t-test.

To determine whether the loss of epigenetic silencing in the 5'-UTR is associated with L1 protein expression – a requirement for the production of new insertions, we assayed tissue samples using a polyclonal antibody specific to the RRM domain of ORF1p (Suret al. 2017). As a control, western blot analysis was carried out using

lysates from a cell line known to express L1 (MCF-7) and one not known to (Du145 prostate cancer) (Figure R21A)[Chen et al., 2012]. Indeed, we observe a robust band at 40 kDa – the predicted size of ORF1p – in MCF-7 cells but not Du145. Next, we performed immunocytochemistry on MCF-7 and Du145 cell lines using the same ORF1p antibody as an additional control. In agreement with our western blot data, we observe staining almost exclusively in MCF-7 cells (Figure R21B). Similar to previous reports, we observe ORF1p primarily in the cytoplasm [Goodier et al., 2007; Doucet et al., 2010; Horn et al., 2014].

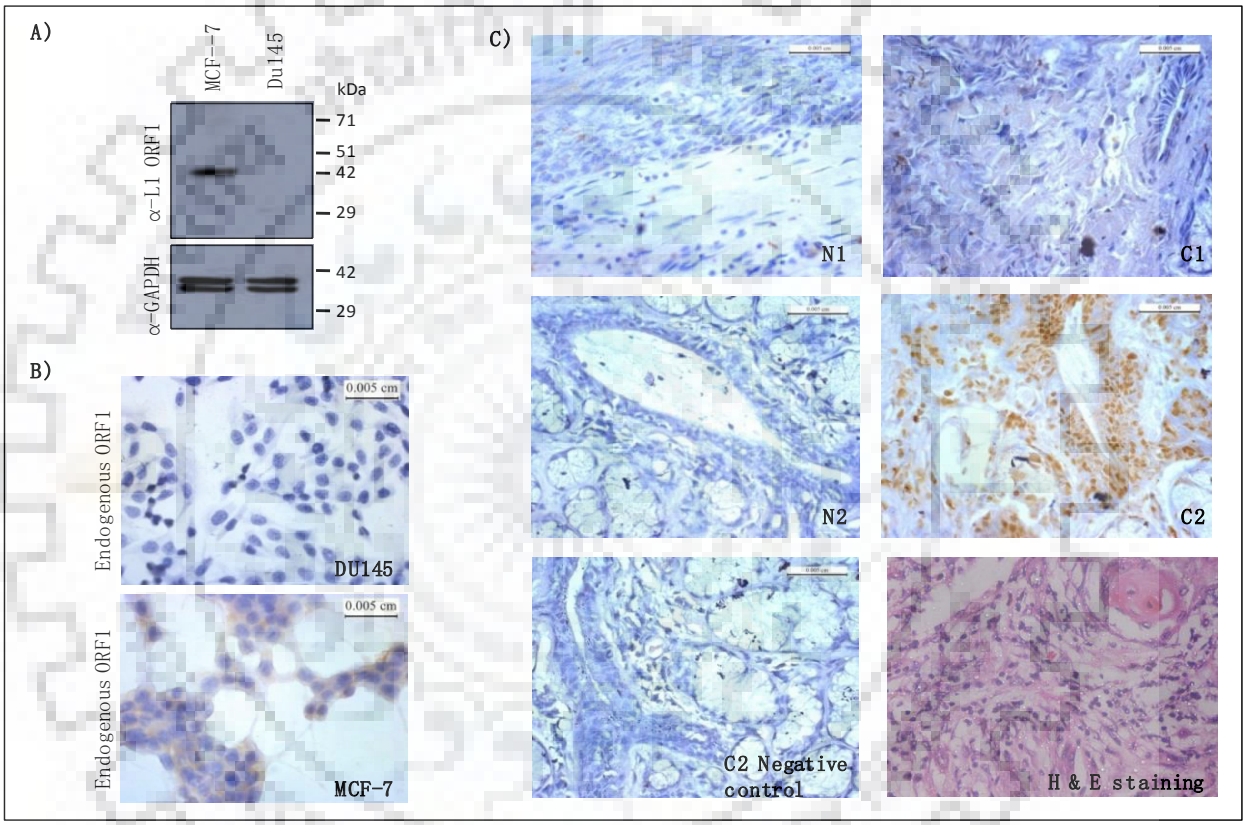


Figure R21: Human L1 ORF1 protein expression in operated oral cancer samples. A) Validation of human α-L1 ORF1p (RRM) antibody by immunocytochemistry in untransfected Du145 and MCF-7 cancer cell lines. **B)** Western blot analysis for endogenous ORF1 in Du145 and MCF-7 cells (panel 1). Immunoblot with α-GAPDH serves as a loading control (panel 2). **C)** IHC staining of oral cancer and matched normal tissues using α-L1 ORF1p (RRM) antibody. All images were collected at 40X magnification. N-Normal; C-Cancer; number indicates patient number (e.g. C1- Patient 1 oral cancer tissue; N1- patient 1 matching normal tissue). The cancer tissue section from patient 2 not treated with primary antibody served as negative control (bottom left panel); Hematoxylin and eosin staining of sample C2 served as positive control (bottom right panel).

The five tissue samples displaying hypomethylated L1 5'-UTRs were further characterized by immunohistochemistry to determine whether loss of epigenetic silencing led to increased L1 protein expression. Indeed, in three of these tissues (S2, S3 and S4) we observe significant staining using an antibody specific for the RRM domain of ORF1p (Figure R21C and Figure R22), supporting an association between hypomethylation of the L1 5'-UTR and L1 protein expression in OSCC.

3.2.2. Detectable expression of L1 ORF1p in oral cancer samples

Next, we wanted to test how frequent L1 ORF1p expression was across OSCC samples. Thus, we carried out IHC on a total of twelve post-operated oral cancer samples. The neoplastic nature of all cancer samples used in this study was confirmed by hematoxylin and eosin staining (H&E staining).A representative of this staining is shown in figure R21C (right bottom panel). Using the anti-ORF1 antibody we detected

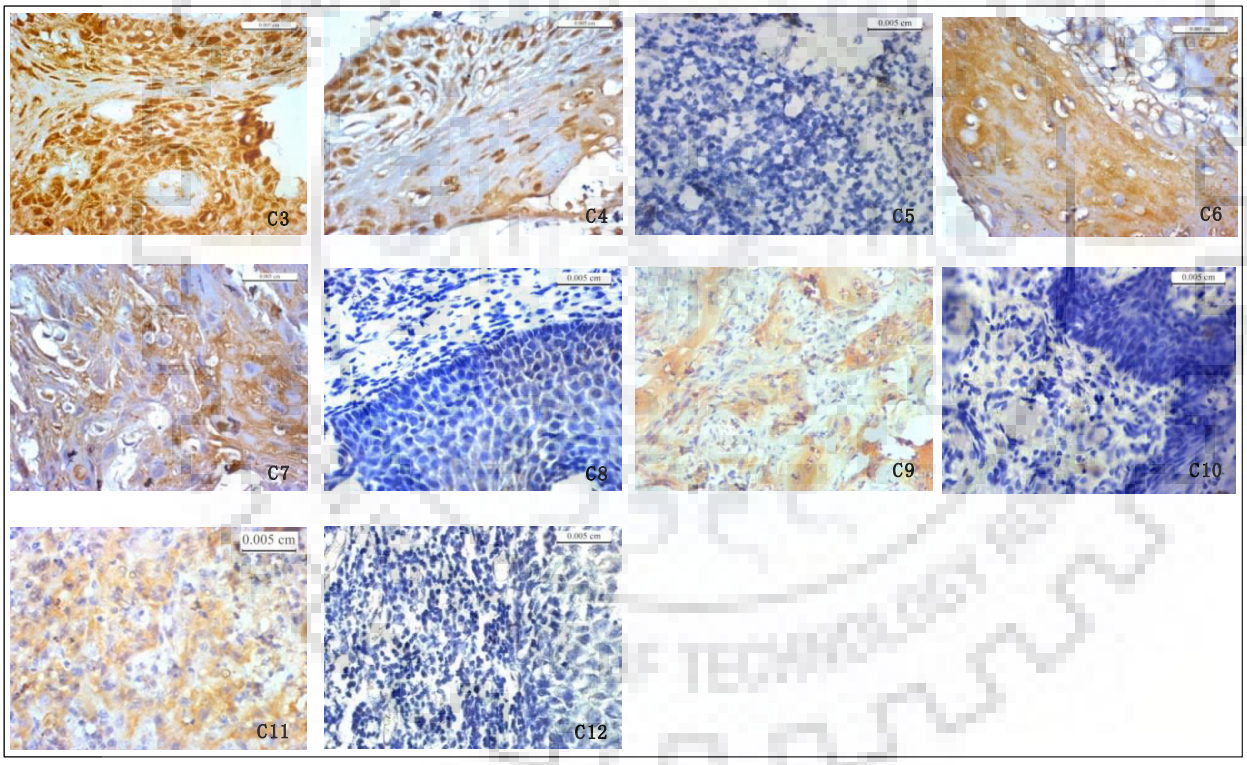


Figure R22: ORF1p immunohistochemistry analysis across oral cancer samples using human α-L1 ORF1p (RRM) antibody. IHC was performed for ten more OSCC samples to detect ORF1p expression. All pictures were taken in 40X magnification.

some level of expression in over half of our samples (7 out of 12; 58%) (Figure R21C, Figure R22). Supporting cancer specific expression in OSCC, we do not observe any staining using anti-ORF1p in normal oral tissues tested (Figure R21C). Incubation of a cancer tissue section from patient 2 (C2) without primary antibody functioned as a

negative control (Figure R21C, left bottom panel). Of the 12 samples screened here, we found sample C3 (Figure 3A) to display the most intense stain with moderate expression in four samples (C2, C4, C6, and C7) (Figure R21C and Figure R22)and low expression in two other samples (C9 and C11) (Figure R22 and Figure R23A).Furthermore, ORF1p positive tissues revealed that in sample C3 (high expression) ORF1p appeared predominantly in the nucleus (Figure R22). Likewise, sample C2 and C4 (moderate expression) also showed nuclear staining withanti-ORF1p (Figure R21C and Figure R22). In contrast, in low expressing samples our IHC indicates that ORF1p is mainly cytoplasmic. DAB signal intensities for all the positive samples were calculated using immunoratio software (Figure 23A).To complement our IHC analysis, we performed Western blot analysis using total protein lysate from patient matched normal and OSCC tissue. Consistent with the IHC for sample 3, we detected robust ORF1p band in cancer lysate (lane C3) (Figure R23B) while no band was in the lane loaded with lysate from normal tissue (lane N3) (Figure 23B). Protein lysate from MCF-7 cells served as a positive control (Figure R23B). These data suggest that L1 ORF1p expression is may be the norm in oral squamous cell carcinoma.

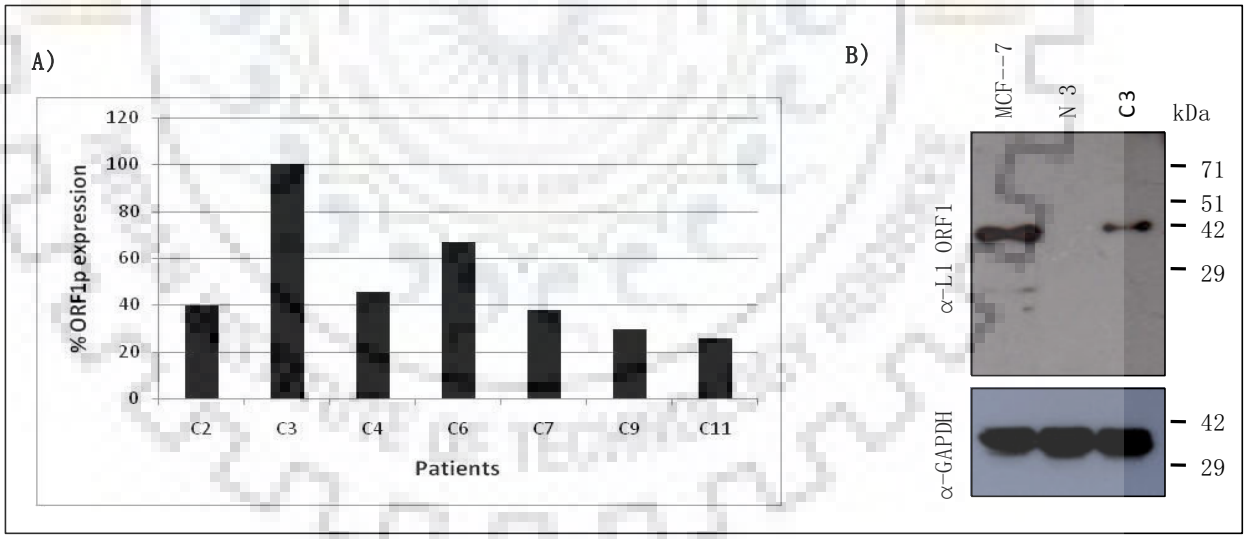
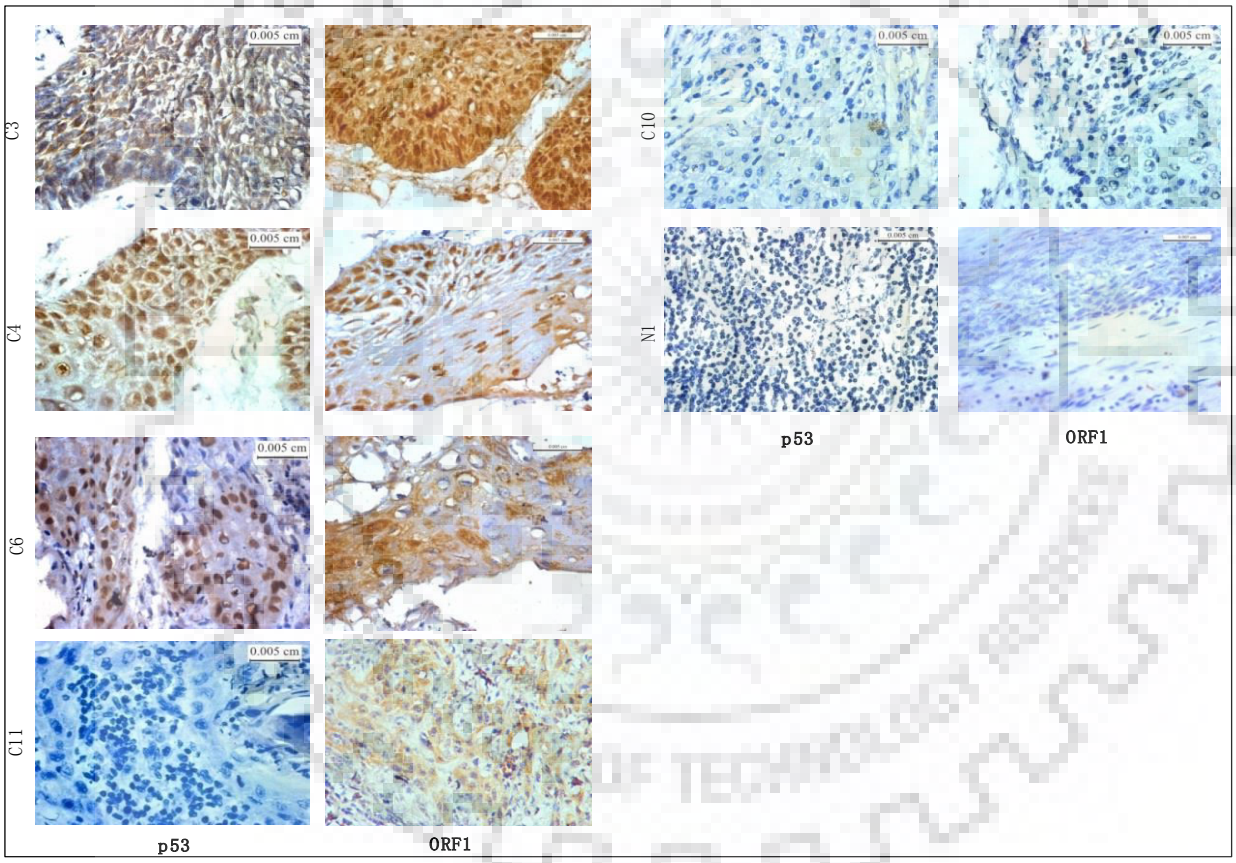


Figure R23: Quantification of ORF1p expression in OSCC samples: A) Quantification of DAB signal intensities, a measure of ORF1p expression in different cancer samples. Values were calculated and plotted using ImmunoRatio software [Tuominen et al. 2010]. **B)** Detection of L1ORF1p by western blot analysis in whole tissue lysate obtained from sample C3. The corresponding matched normal (N3, lane 2) served as a negative control and the total lysate from the MCF-7 cell line (lane 1) was used as a positive control.

3.2.3. Aberrant TP53 expression in oral carcinoma

Our knowledge of the mechanisms regulating L1 retrotransposition activity are incomplete. A recent study reported a positive correlation between TP53 mutation and L1 protein expression in several types of tumors [Rodic et al., 2014; Wylie et al., 2016). Indeed Rodic et al. (Wylie et al., 2016) demonstrated that the aberrant expression of TP53 is highly associated with L1 expression in lung, ovarian and pancreatic carcinoma. Analyzing TP53 expression in ORF1p positive oral carcinoma samples, we detected TP53 by IHC in three out of the four of the IHC positive samples (C3, C4, C6 and C11) but not an ORF1p negative sample (C10) (Figure R24). Among four ORF1p positive samples tested, three samples displayed (C3, C4 and C6) very high expression of TP53. Furthermore, no ORF1p or TP53 was detected in normal



buccal mucosa tissue from patient one (N1) (Figure R24).These data suggest that L1 ORF1 protein expression in OSCC is often associated with aberrant TP53 expression.

Figure R24: L1 ORF1p expression in OSCC correlates with aberrant expression of p53. Five oral cancer samples (C3, C4, C6, C10 and C11) were tested for p53 expression by IHC. Among those five samples, four (C3, C4, C6 and C11) were ORF1p positive and one (C10) was negative. Three (C3, C4 and C6) out of five samples showed significant p53 expression. Normal tissue from patient one (N1) served as a negative control for both ORF1p and p53 staining.

4.1 Recombinant human ORF2p RT domain protein formed inclusion bodies in bacterial expression system:

It is hypothesized that L1 element formed a new copy in the genome likely by a coupled reverse transcription and integration process termed as target primed reverse transcription(TPRT) which was originally demonstrated for R2 element found in arthropods and partially characterized for human L1 element [Luan et al., 1993; Luan and Eickbush 1995; Cost et al., 2002]. The reverse transcription of L1mRNA to cDNA is the most crucial steps in L1 amplification process [Moran et al, 1996]. Here, the human L1 reverse transcriptase activity has been characterized by cloning different fragments from ORF2 sequence which contain the central RT domain. The SV1 clone which is 21 amino acids shorter (474-987) at the N-terminal than predicted RT domain (453-880) [Malik et al. 1999] showed robust expression in bacterial expression system. Previous studies to characterize the human L1 RT domain activity used either full length or truncated ORF2 (both from 5' and 3' ends) didn't show such robust expression of recombinant ORF2p which has shown in the present study[Clements et al, 1998; Piskareva et al, 2003; Piskerva et al., 2006]. The expressed RT protein is readily detected in Commassie stained denatured SDS-PAGE gel from total induced lysate. Although, the amount of expressed protein was significantly high, the folding of the RT domain protein in bacterial system was not proper and thus, formed inclusion bodies. It is generally observed that when only a domain, instead of full length clone(complete protein) expressed in *E. coli*, the high-level expression of only domain often formed inclusion bodies (Singh et al, 2015). Probably, same things happened when RT domain expressed alone instead of full length ORF2p which failed to express in *E. coli*. There are several reasons because of that a protein might form inclusion bodies when expressed in bacteria which are i) high inducer condition ii) high copy no of plasmid iii) expression under increased temperature iv) strong promoter system and others (Singh et al., 2015; Upadhyay et al., 2016). Some of the factors like different temperatures, inducer condition (IPTG), buffer pH were checked to see the solubility of the RT domain protein. In this study, three different temperatures (16°C, 30°C and 37°C) were used and the data showed that in all three temperatures the RT domain protein showed robust expression

but due to misfolding the RT domain protein formed inclusion bodies. In general 0.4-0.6 mM IPTG concentration is routinely used for expressing protein in bacterial system when IPTG inducible plasmid is used (for example pET-30b). In-order to verify different concentration of IPTG, the concentration ranging from 0.02- 0.8mM was checked on human L1 RT domain protein solubility, although the protein was expressing in significant high amount in all these concentrations, the protein was not folded properly and thus formed inclusion bodies. Sometimes growth media used for expression affects the solubility of the protein [Singh et al., 2005]. The expression analysis in different types of media (normal LB, SSC, and terrific broth) showed that the human RT domain protein was present only in inclusion bodies. In the next step, few more RT domain fragments were cloned after performing some bioinformatics analysis using ORF2 protein sequence. It has seen that presence of some patches of amino acids and if amino acids composition of protein are more hydrophobic in nature, the protein likely formed inclusion bodies when expressed in bacteria where a particular chaperone is absent to facilitate the folding (Upadhya et al., 2016). There are multiple software available [Agostini et al., 2014; Chang et al., 2016; Hebditch et al, 2017] in the literature which predicts soluble properties of a particular stretch of peptide sequence. Analyzing SV1 clone (amino acids 474-987 of human ORF2) using ccSOL omics webserver [Agostini et al., 2014] showed that the stretches of amino acids present in SV1 clone have high tendency to misfold and thus can form inclusion bodies. Taking complete human ORF2 protein sequence as query the same software [Agostini et al., 2014] showed a clone termed SV3 (amino acids 443-981) might be soluble in bacterial expression system. The present study showed although SV3 equally expressed like SV1 in bacterial expression system, the protein coming from SV3 clone was not present in soluble fraction. Trying to purify RT domain protein from inclusion bodies showed that presence of chaotropic agent like urea in soluble buffer make the RT domain protein soluble. More than 80% protein present in pellet comes to soluble fraction when urea with concentration 8M was used. Trying to refold the protein by slowly removing the urea through dialysis showed that the protein formed misfolded structure thus, formed precipitate in dialysis bag when urea concentration decreased less than 1M in dialysis buffer. Employing dialysis in buffers with different pH (ranged from pH:

4-10) did not help; in all these pH protein formed white precipitate in the dialysis bag when the urea concentration decreases less than 1M. Next, purification of RT domain protein in denatured condition (RT protein in solubilizing buffer containing 8M urea) using Ni-agarose chromatography (as recombinant human RT contains six-Histidine residues at the N-terminal end) showed partially purified RT domain protein. A major fraction of RT protein always stayed bound with the beads even after using imidazole concentration at 500 mM in elution buffer. The partially purified RT protein was dialyzed to remove imidazole and urea where final concentration of imidazole and urea was 20mM and 1 M respectively. The protein was assayed for reverse transcriptase activity by employing L1 element amplification protocol (LEAP) assay and the result showed that the 1M urea inhibited synthesis of cDNA on Alu RNA template in LEAP assay. It was surprised to see that very little amount of induced RT protein (less than 1%) was still in soluble fraction when analyzed by Western blotting; although most of the protein after induced expression in bacterial system misfolded and thus formed inclusion bodies. Although, purified RT protein from soluble fraction using Ni-agarose chromatography didn't show the bands in commassie stained SDS-PAGE gel, but it showed significantly high RT activity in reverse transcriptase assay. Previous studies mainly used poly(rA):oligo(dT) as template/primer combination followed by measuring of radioactive incorporation of α-^{32}PdNTPs (mainly dATP) to study the RT activity [Clements et al, 1998; Piskareva et al, 2003; Piskerva et al., 2006]. Here, we modified the assay as per LEAP protocol [Kulpa and Moran 2006]. We have used Alu RNA (ends with 10 As) as a template which is not present in bacterial genome from where the RT protein was purified. A RACE primer ends with 10T followed by 20 nucleotides sequence which does not have any sequence similarity with bacterial genome was used as primers for synthesis of cDNA using Alu RNA as template. The assay system used revealed significant good activity of recombinant human RT protein as compared with MMLV reverse transcriptase on the same template. In our RT activity assay, SV3 clone showed more activity than SV2 clone. The increase activity might be due to increased amount of soluble protein or 35 amino acids extra at the N-terminal region of SV3 clone compared to SV1 clone. It is important to mention that as per Malik et al [Malik et al., 1999] the human RT domain expand from 453-880 amino acids in human ORF2p; the SV1 and SV3 clone is 474-987 and 443-981 respectively.

4.2 Antibodies against L1 encoded proteins (ORF1p and ORF2p) are important tools to study the biology of L1 retrotransposon:

The antibodies against human L1 proteins (ORF1p and ORF2p) is very important tolls to study the biology of L1 retrotransposon, especially to find out its expression patterns in different parts of brain and in different types of cancers; sequence analysis showed in both instances copy number of L1 is higher compared to controls. Very few laboratories have made antibody against ORF2p and not a single one is commercially available [Ergün et al., 2004; Goodier et al., 2004; Sokolowski et al., 2014; De Luca et al, 2016]. Although, two peptide antibodies reported by Goodier et al [Goodier et al., 2004] 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 cell lines and cancer tissues. Here I have tried to make antibody from human ORF2 protein. The SDS-PAGE gel band purified ORF2p was injected to rabbit which didn't show any antibody response against ORF2p. Although, it is not known why the gel band purified human RT is not immunogenic in rabbit, the same experiment can be repeated in rat or mice to see if those animals show response again that antigen. Simultaneously purified human L1-ORF1p has also been tried to raise antibody against that protein. The bacterial induced His tag ORF1p showed multiple shorter bands after purification. Since, the purified ORF1p showed shorter bands after purification, antibody raising against ORF1p has been discontinued, as in parallel antibody against RRM domain of ORF1p showed beautiful reactivity against human ORF1p [Sur et al., 2017]. It is worthwhile to mention that several laboratories made antibody against human ORF1p [Raiz et al, 2012; Harris et al., 2010; Rodic et al., 2014; Chen et al.' 2012; Bratthauer et al.' 1992] and one monoclonal ORF1 antibody is commercially available from EMD Millipore [Rodic et al., 2014] .

4.3 ORF1p expression is common in OSCC:

It has long been hypothesized that L1 retrotransposition activity may contribute to the onset and progression of cancer. Only recently, with advances in DNA sequencing and

effective reagents (e.g. antibodies reactive against L1 proteins), have significant developments been made in our understanding of L1 biology in cancer. Indeed, several recent studies have demonstrated 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; Belancio et al., 2010]. Likewise, it has been shown that about half of common cancers express human L1 ORF1p [43]. Here, we build on these studies by characterizing L1 ORF1p 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 Cancr, World cancer report 2014; World health organization report http://www.who.int/oral_health/publications/]. Our analysis of 12 patient samples 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 [43].

Although localization studies have primarily observed ORF1p in the cytoplasm, occasionally limited number cells do show nuclear localization [Goodier et al., 2007; Doucel et al., 2010; Horn et al., 2014]. We observe similar localization in the samples which we carried out IHC for. Specifically, we see strong staining of nuclear ORF1p in three (C2, C3 and C4) out of the 12 samples tested (Figure R21C and Figure 22). 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].

4.4 L1 protein expression is associated with hypomethylation of the L1 promoter in OSCC:

A hallmark of cancer is an altered epigenetic landscape [Lund and van Lohuizen 2004; Jones et al., 2016]. A large number of frequently hypomethylated loci including the L1

promoter have been reported in many cancers [Daskalos et al., 2009; Saito et al., 2010; Rodic et al., 2014]. Our bisulfite analysis (Figure R19B) indicates that hypomethylation of the L1 5'-UTR also occurs in OSCC. Notably, of the eight paired normal cancer samples assayed here, four displayed both significant hypomethylation throughout the CG-rich L1 promoter and upregulation of L1ORF1p. Future studies will elucidate the impact of L1 hypomethylation on L1 protein expression and the number of new insertions events on cancers like OSCC.

4.5 Tumor suppressor p53 protein might have some role in L1 activation in OSCC:

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 and that this activity is evolutionarily conserved [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. Similarly, we observe elevated p53 expression in only ORF1p positive samples but not tissues where ORF1p was below the level of detection. Together, these data warrant further investigation into a potential role for mutant p53 might in L1 retrotransposition in OSCC. 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.

1. Agostini, F., Cirillo, D., Livi, C.M., DelliPonti, R. and Tartaglia, G.G. 2014. cc SOL omics: a webserver for solubility prediction of endogenous and heterologous expression in Escherichia coli. Bioinformatics. 30(20), pp.2975-2977.
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. Agrawal, P., Chen, Y.T., Schilling, B., Gibson, B.W. and Hughes, R.E. 2012. Ubiquitin-specific peptidase 9, X-linked (USP9X) modulates activity of mammalian target of rapamycin (mTOR). Journal of Biological Chemistry. 287(25), pp.21164-21175.
4. Ardeljan, D., Taylor, M.S., Ting, D.T. and Burns, K.H. 2017. The human long interspersed element-1 retrotransposon: an emerging biomarker of neoplasia. Clinical Chemistry. 63(4), pp.816-822.
5. Ariel, P. and Ryan, T.A. 2012. New insights into molecular players involved in neurotransmitter release. Physiology. 27(1), pp.15-24.
6. Arjan-Odedra, S., Swanson, C.M., Sherer, N.M., Wolinsky, S.M. and Malim, M.H. 2012. Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication of exogenous retroviruses. Retrovirology. 9(1), p.53.
7. 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.
8. Bao, W., Kojima, K.K. and Kohany, O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA. 6(1), p.11.
9. Belancio, V.P., Deininger, P.L. and Roy-Engel, A.M. 2009. LINE dancing in the human genome: transposable elements and disease. Genome medicine, 1(10), p.97.
10. Belancio, V.P., Roy-Engel, A.M., Pochampally, R.R. and Deininger, P., 2010. Somatic expression of LINE-1 elements in human tissues. Nucleic acids research, *38*(12), pp.3909-3922.

11. Bestor, T.H. and Bourc'his, D. 2004. Transposon silencing and imprint establishment in mammalian germ cells. Cold Spring Harbor symposia on quantitative biology. 69, pp. 381-388.
12. 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.
13. Bogerd, H.P., Wiegand, H.L., Hulme, A.E., Garcia-Perez, J.L., O'Shea, K.S., Moran, J.V. and Cullen, B.R. 2006. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proceedings of the National Academy of Sciences. 103(23), pp.8780-8785.
14. Bourc'his, D. and Bestor, T.H. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 431(7004), pp.96-99.
15. Bratthauer, G.L., Fanning, T.G.1992.Active LINE-1 retrotransposons in human testicular cancer. Oncogene. 7(3), pp.507-10.
16. Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V. and Kazazian, H.H. 2003. Hot L1s account for the bulk of retrotransposition in the human population. Proceedings of the National Academy of Sciences. 100(9), pp.5280-5285.
17. Burns, K.H. 2017. Transposable elements in cancer. Nature Reviews Cancer. 17(7), pp.415-424. doi: 10.1038/nrc.2017.35.
18. CDOE05_vol33_397_9/en/ (accessed on 26 Oct 2013). World Health Organization. Tobacco free initiative http://www.who.int/tobacco/research/cancer/en/(accessed on 26 Oct 2013).
19. Chang, C.C.H., Li, C., Webb, G.I., Tey, B., Song, J. and Ramanan, R.N., 2016. Periscope: quantitative prediction of soluble protein expression in the periplasm of Escherichia coli. Scientific reports, 6.
20. Chen, H., Lilley, C.E., Yu, Q., Lee, D.V., Chou, J., Narvaiza, I., Landau, N.R. and Weitzman, M.D. 2006. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. Current Biology. 16(5), pp.480-485.
21. 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.

22. 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.
23. 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.
24. Conti, A., Carnevali, D., Bollati, V., Fustinoni, S., Pellegrini, M. and Dieci, G. 2014. Identification of RNA polymerase III-transcribed Alu loci by computational screening of RNA-Seq data. Nucleic acids research. 43(2), pp.817-835.
25. Cordaux, R. and Batzer, M.A. 2009. The impact of retrotransposons on human genome evolution. Nature Reviews Genetics. 10(10), pp.691-703.
26. 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.
27. Craig, N.L. 2002. Mobile DNA: an introduction. Mobile DNA II , AMS, Washington, pp. 3-11.
28. Crow, Y.J., Hayward, B.E., Parmar, R., Robins, P., Leitch, A., Ali, M., Black, D.N., van Bokhoven, H., Brunner, H.G., Hamel, B.C. and Corry, P.C. 2006. Mutations in the gene encoding the 3′-5′ DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. Nature genetics. 38(8), pp.917-920.
29. Daskalos, A., Nikolaidis, G., Xinarianos, G., Savvari, P., Cassidy, A., Zakopoulou, R., Kotsinas, A., Gorgoulis, V., Field, J.K. and Liloglou, T. 2009. Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. International Journal of Cancer. 124(1), pp.81-87.
30. deKoning, A.J., Gu, W., Castoe, T.A., Batzer, M.A. and Pollock, D.D. 2011. Repetitive elements may comprise over two-thirds of the human genome. PLoS genetics. 7(12), p.e1002384.
31. De Luca, C., Guadagni, F., Sinibaldi-Vallebona, P., Sentinelli, S., Gallucci, M., Hoffmann, A., Schumann, G.G., Spadafora, C., Sciamanna, I. 2016. Enhanced expression of LINE-1-encoded ORF2 protein in early stages of colon and prostate transformation.Oncotarget. 7(4):4048-61. doi: 10.18632/oncotarget.6767.

32. 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.
33. Dewannieux, M., Esnault, C. and Heidmann, T. 2003. LINE-mediated retrotransposition of marked Alu sequences. Nature genetics. 35(1), pp.41-48.
34. Dewannieux, M. and Heidmann, T. 2005. L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. Journal of molecular biology. 349(2), pp.241-247.
35. 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.
36. 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.
37. Doucet-O'Hare, T.T., Rodić, N., Sharma, R., Darbari, I., Abril, G., Choi, J.A., Ahn, J.Y., Cheng, Y., Anders, R.A., Burns, K.H. and Meltzer, S.J. 2015. LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. Proceedings of the National Academy of Sciences. 112(35), pp.E4894-E4900.
38. Engels, W.R. and Preston, C.R. 1981. Identifying P factors in Drosophila by means of chromosome breakage hotspots. Cell. 26(3), pp.421-428.
39. Ergün, S., Buschmann, C., Heukeshoven, J., Dammann, K., Schnieders, F., Lauke. H., Chalajour, F., Kilic, N., Strätling, W.H., Schumann, G.G. 2004. Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult humantissues.J Biol Chem. 279(26) pp.27753-63.
40. Esnault, C., Maestre, J. and Heidmann, T. 2000. Human LINE retrotransposons generate processed pseudogenes. Nature genetics. 24(4), pp.363-367.
41. 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.

42. Faulkner, G.J. and Garcia-Perez, J.L. 2017. L1 Mosaicism in Mammals: Extent, Effects, and Evolution. Trends in Genetics. 33(11), pp.802-816.
43. Feng, Q., Moran, J.V., Kazazian, H.H. and Boeke, J.D. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell. 87(5), pp.905-916.
44. Fuhrman, S.A., Deininger, P.L., LaPorte, P., Friedmann, T. and Peter Geiduschek, E. 1981. Analysis of transcription of the human Alu family ubiquitous repeating element by eukaryotic RNA polymerase III. Nucleic acids research. 9(23), pp.6439-6456.
45. Garcia-Perez, J.L., Widmann, T.J. and Adams, I.R. 2016. The impact of transposable elements on mammalian development. Development. 143(22), pp.4101-4114.
46. Godbole, M., Chandrani, P, Gardi, N., Dhamne, H., Patel, K., Yadav, N., Gupta, S., Badwe, R., Dutt, A. 2017.miR-129-2 mediates down-regulation of progesterone receptor in response to progesterone in breast cancer cells. Cancer Biology & Therapy.
47. Gonçalves, I., Duret, L. and Mouchiroud, D. 2000. Nature and structure of human genes that generate retropseudogenes. Genome research. 10(5), pp.672-678.
48. Goodier, J.L. 2016. Restricting retrotransposons: a review. Mobile DNA. 7(1), p.16.
49. Goodier, J.L., Pereira, G.C., Cheung, L.E., Rose, R.J. and Kazazian Jr, H.H. 2015. The broad-spectrum antiviral protein ZAP restricts human retrotransposition. PLoS genetics. 11(5), p.e1005252.
50. 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.
51. Goodier, J.L., Cheung, L.E. and Kazazian Jr, H.H. 2012. MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. PLoS genetics. 8(10), p.e1002941.
52. Goodier, J.L. and Kazazian, H.H., 2008. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell. 135(1), pp.23-35.
53. 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.

54. Goodier, J.L., Ostertag, E.M., Engleka, K.A., Seleme, M.C., Kazazian Jr, H.H. 2004. A potential role for the nucleolus in L1 retrotransposition.13(10), pp.1041-1048.
55. Goodier, J.L., Ostertag, E.M., Du, K. and Kazazian, H.H. 2001. A novel active L1 retrotransposon subfamily in the mouse. Genome research. 11(10), pp.1677-1685.
56. Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L., Robertson, M. and Sargeant, L. 1991. Identification and characterization of the familial adenomatous polyposis coli gene. Cell. 66(3), pp.589-600.
57. 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.
58. Hamdorf, M., Idica, A., Zisoulis, D.G., Gamelin, L., Martin, C., Sanders, K.J. and Pedersen, I.M. 2015. miR-128 represses L1 retrotransposition by binding directly to L1 RNA. Nature structural & molecular biology. 22(10), pp.824-831.
59. Hancks, D.C. and Kazazian, H.H. 2016. Roles for retrotransposon insertions in human disease. Mobile DNA. 7(1), p.9.
60. Hancks, D.C., Goodier, J.L., Mandal, P.K., Cheung, L.E. and Kazazian Jr, H.H. 2011. Retrotransposition of marked SVA elements by human L1s in cultured cells. Human molecular genetics. 20(17), pp.3386-3400.
61. 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.
62. Hebditch, M., Carballo-Amador, M.A., Charonis, S., Curtis, R., Warwicker, J. 2017. Protein-Sol: a web tool for predicting protein solubility from sequence.Bioinformatics. 33(19), pp.3098-3100.
63. 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.
64. Herrero, R., Castellsagué, X., Pawlita, M., Lissowska, J., Kee, F., Balaram, P., Rajkumar, T., Sridhar, H., Rose, B., Pintos, J. and Fernández, L. 2003. Human papillomavirus and oral cancer: the International Agency for Research on Cancer multicenter study. Journal of the National Cancer Institute. 95(23), pp.1772-1783.

65. 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.
66. Horn, A.V., Klawitter, S., Held, U., Berger, A., JaguvaVasudevan, A.A., Bock, A., Hofmann, H., Hanschmann, K.M.O., Trösemeier, J.H., Flory, E. and Jabulowsky, R.A. 2013. Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity. Nucleic acids research. 42(1), pp.396-416.
67. Huang, C.R.L., Burns, K.H. and Boeke, J.D. 2012. Active transposition in genomes. Annual review of genetics. 46, pp.651-675.
68. Idica, A., Sevrioukov, E.A., Zisoulis, D.G., Hamdorf, M., Daugaard, I., Kadandale, P. and Pedersen, I.M. 2017. MicroRNA miR-128 represses LINE-1 (L1) retrotransposition by down-regulating the nuclear import factor TNPO1. Journal of Biological Chemistry. 292(50), pp.20494-20508.
69. Ikeda, T., Abd El Galil, K.H., Tokunaga, K., Maeda, K., Sata, T., Sakaguchi, N., Heidmann, T. and Koito, A. 2011. Intrinsic restriction activity by apolipoprotein B mRNA editing enzyme APOBEC1 against the mobility of autonomous retrotransposons. Nucleic acids research. 39(13), pp.5538-5554.
70. India Project Team of the International, I.P.T. and Cancer Genome Consortium. 2013. Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. Nature communications. 4, 2873.
71. International Agency for Research on Cancer. World cancer report 2014. https://shop.iarc.fr/products/wcr2014.
72. Iskow, R.C., McCabe, M.T., Mills, R.E., Torene, S., Pittard, W.S., Neuwald, A.F.et al. 2010. Natural mutagenesis of human genomes by endogenous retrotransposons.Cell.141,pp.1253-61.
73. Jacobs, F.M., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A.D., Katzman, S., Paten, B., Salama, S.R. and Haussler, D. 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature. 516(7530), pp.242-245.

74. Jacob, F. and Wollman, E.L. 1958. Episomes, a proposed term for added genetic elements. Comptesrendushebdomadaires des séances de l'Académie des sciences. 247(1), pp.154-156.
75. Jayalekshmi, P.A., Gangadharan, P., Akiba, S., Nair, R.R.K., Tsuji, M. and Rajan, B. 2009. Tobacco chewing and female oral cavity cancer risk in Karunagappally cohort, India. British journal of cancer. 100(5), pp.848-852.
76. Jones, P.A., Issa, J.P.J. and Baylin, S. 2016. Targeting the cancer epigenome for therapy. Nature reviews Genetics. 17(10), pp.630-641.
77. Kastenhuber, E.R. and Lowe, S.W. 2017. Putting p53 in context. Cell. 170(6), pp.1062-1078.
78. Kazazian Jr, H.H. and Moran, J.V. 2017. Mobile DNA in health and disease. New England Journal of Medicine. 377(4), pp.361-370.
79. 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.
80. Khazina, E. and Weichenrieder, O. 2009. Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. Proceedings of the National Academy of Sciences. 106(3), pp.731-736.
81. Kulpa, D.A. and Moran, J.V. 2006. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. Nature structural & molecular biology. 13(7), pp.655-660.
82. 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.
83. Lachenmeier, D.W., Przybylski, M.C. and Rehm, J. 2012. Comparative risk assessment of carcinogens in alcoholic beverages using the margin of exposure approach. International Journal of Cancer. 131(6).
84. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. and Funke, R. 2001. Initial sequencing and analysis of the human genome. Nature. 409(6822), pp.860-921.

85. 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.

86. Li, P., Du, J., Goodier, J.L., Hou, J., Kang, J., Kazazian Jr, H.H., Zhao, K. and Yu, X.F. 2017. Aicardi–Goutières syndrome protein TREX1 suppresses L1 and maintains genome integrity through exonuclease-independent ORF1p depletion. Nucleic acids research. 45(8), pp.4619-4631.

87. Li, X., Zhang, J., Jia, R., Cheng, V., Xu, X., Qiao, W.,Guo, F., Liang, C. and Cen, S. 2013. The MOV10 helicase inhibits LINE-1 mobility. Journal of Biological Chemistry. 288(29), pp.21148-21160.

88. Luan, D.D. and Eickbush, T.H. 1995. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. Molecular and cellular biology. 15(7), pp.3882-3891.

89. 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.

90. Lund, A.H. and van Lohuizen, M. 2004. Epigenetics and cancer. Genes & development. 18(19), pp.2315-2335.

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

92. 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.

93. Malone, C.D. and Hannon, G.J. 2009. Small RNAs as guardians of the genome. Cell. 136(4), pp.656-668.

94. 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.

95. Mandal, P.K. and Kazazian, H.H. 2008. SnapShot: vertebrate transposons. Cell, 135(1), pp.192-192.

96. 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

restriction enzyme-like LINE-encoded endonuclease. Eukaryotic cell. 3(1), pp.170-179.

97. Mandal, P.K., Rawal, K., Ramaswamy, R., Bhattacharya, A., Bhattacharya S. 2006. Identification of insertion hot spots for non-LTR retrotransposons: computational and biochemical application to Entamoeba histolytica. Nucleic Acids Research, 34: 5752-5763.
98. Martin, S.L. 2010. Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. RNA biology. 7(6), pp.706-711.
99. Martin, S.L., 2006. The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. BioMed Research International. 2006 (1), pp.45621.
100. 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.
101. McClintock, B. 1950. The origin and behavior of mutable loci in maize. Proceedings of the National Academy of Sciences. 36(6), pp.344-355.
102. 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.
103. Messing, J., Bharti, A.K., Karlowski, W.M., Gundlach, H., Kim, H.R., Yu, Y., Wei, F., Fuks, G., Soderlund, C.A., Mayer, K.F. and Wing, R.A. 2004. Sequence composition and genome organization of maize. Proceedings of the National Academy of Sciences of the United States of America. 101(40), pp.14349-14354.
104. 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.
105. Moldovan, J.B. and Moran, J.V. 2015. The zinc-finger antiviral protein ZAP inhibits LINE and Alu retrotransposition. PLoS genetics. 11(5), p.e1005121.
106. Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D. and Kazazian, H.H. 1996. High frequency retrotransposition in cultured mammalian cells. Cell. 87(5), pp.917-927.

107. Moran, J.V. and Gilbert, N. 2002. Mammalian LINE-1 retrotransposons and related elements. Mobile DNA II, pp. 836-869. American Society of Microbiology.

108. 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.

109. Muckenfuss, H., Hamdorf, M., Held, U., Perković, M., Löwer, J., Cichutek, K., Flory, E., Schumann, G.G. and Münk, C. 2006. APOBEC3 proteins inhibit human LINE-1 retrotransposition. Journal of Biological Chemistry. 281(31), pp.22161-22172.

110. 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), pp.443-446.

111. Murthy, V., Calcuttawala, A., Chadha, K., d'Cruz, A., Krishnamurthy, A., Mallick, I., Nair, S., Teni, T., Pawar, S., Talapatra, K., Patil, A., Bhatt, A., Chatterjee, S., Swain, M., Narayanan, P., Ghadyalpatil, N., Singhal, M., Kuriakose, M., Prabhash, K., Agarwal, J. and Parikh, P. 2017. Human papillomavirus in head and neck cancer in India: Current status and consensus recommendations. South Asian J Cancer. 6(3):93-98.

112. Muwonge, R., Ramadas, K., Sankila, R., Thara, S., Thomas, G., Vinoda, J. and Sankaranarayanan, R. 2008. Role of tobacco smoking, chewing and alcohol drinking in the risk of oral cancer in Trivandrum, India: a nested case-control design using incident cancer cases. Oral oncology. 44(5), pp.446-454.

113. Nagpal N, Kulshreshtha R. 2014.miR-191: an emerging player in disease biology. Front Genet, 5:99.

114. Niblock, W.J. 1902. Cancer in India. The Indian Medical Gazette, 37(5), pp.161.

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

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

117. Pandey, R., Mukerji. M. 2011. From 'JUNK' to just unexplored noncoding knowledge: the case of transcribed Alus.Brief Funct Genomics, 10(5):294-311.

118. 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.

119. Pei, B., Sisu, C., Frankish, A., Howald, C., Habegger, L., Mu, X.J., Harte, R., Balasubramanian, S., Tanzer, A., Diekhans, M. and Reymond, A. 2012. The GENCODE pseudogene resource. Genome biology. 13(9), p.R51.

120. 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.

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

122. 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.

123. 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.

124. Pizarro, J.G. and Cristofari, G. 2016. Post-transcriptional control of LINE-1 retrotransposition by cellular host factors in somatic cells. Frontiers in cell and developmental biology. 4.

125. 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.

126. Reibel, J. 2003. Tobacco and oral diseases. Medical Principles and Practice. 12(Suppl. 1), pp.22-32.

127. Rice, G.I., Bond, J., Asipu, A., Brunette, R.L., Manfield, I.W., Carr, I.M., Fuller, J.C., Jackson, R.M., Lamb, T., Briggs, T.A. and Ali, M. 2009. Mutations involved in Aicardi-Goutieres syndrome implicate SAMHD1 as regulator of the innate immune response. Nature genetics. 41(7), pp.829-832.

128. Richardson, S.R., Doucet, A.J., Kopera, H.C., Moldovan, J.B., Garcia-Pérez, J.L. and Moran, J.V. 2014. The influence of LINE-1 and SINE retrotransposons on mammalian genomes. Microbiology spectrum. 3(2).

129. 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.

130. Rodić, N., Steranka, J.P., Makohon-Moore, A., Moyer, A., Shen, P., Sharma, R., Kohutek, Z.A., Huang, C.R., Ahn, D., Mita, P. and Taylor, M.S. 2015. Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. Nature medicine. 21(9), pp.1060-1064.

131. Rodić, N. and Burns, K.H. 2013. Long interspersed element–1 (LINE-1): passenger or driver in human neoplasms?. PLoS genetics. 9(3), p.e1003402.

132. 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.

133. Schumann, G.G., Gogvadze, E.V., Osanai-Futahashi, M., Kuroki, A., Münk, C., Fujiwara, H., Ivics, Z. and Buzdin, A.A. 2010. Unique functions of repetitive transcriptomes. Int Rev Cell Mol Biol. 285, pp.115-188.

134. Schumann, G.G. 2007. APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition.BiochemSoc Trans. 35, pp.637–642.

135. Scott, E. C. et al. 2016.A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. Genome Res. 26, pp.745–755.

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

137. Shapiro, J.A.1969. Mutations caused by the insertion of genetic material into the galactose operon of Escherichia coli. J Mol Biol. 40(1), pp.93-105. No abstract available.

138. Sharma S, Nagpal N, Ghosh PC, Kulshreshtha R. 2017. P53-miR-191-SOX4 regulatory loop affects apoptosis in breast cancer. RNA. 23:1237-1246.

139. Sheehy, A.M., Gaddis, N.C., Choi, J.D., Malim, M.H. 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature. 418, pp.646–650.

140. Shukla, R., Upton, K.R., Muñoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., et al. 2013. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell. 153(1),pp.101-11.

141. 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.

142. Sokolowski, M., DeFreece, C.B., Servant, G., Kines, K.J., deHaro, D.L., Belancio, V.P.2014.Development of a monoclonal antibody specific to the endonuclease domain of the human LINE-1 ORF2 protein.5(1):29. doi: 10.1186/s13100-014-0029-x. eCollection 2014.

143. Smalheiser, N.R., Torvik, V.I. 2005. Mammalian microRNAs derived from genomic repeats. Trends Genet. 21(6), pp.322-326.

144. Solyom, S., Ewing, A.D., Rahrmann, E.P., Doucet, T., Nelson, H.H., Burns, M.B. et al. 2012. Extensive somatic L1 retrotransposition in colorectal tumors. Genome Res. 22, pp. 2328-38.

145. Speek, M.2001. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol. 21(6), pp.1973-1985.

146. Stetson, D.B., Ko, J.S.,Heidmann, T. and Medzhitov, R. 2008. Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell. 134(4), pp.587-598.

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

148. Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., and Hannon, G.J. 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature.453, pp.534–538.

149. 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.

150. Taylor al.1963. Bacteriophage-induced mutation in escherichia coli. Proc Natl AcadSci U S A.50, pp.1043-1051. No abstract available.

151. 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.

152. Tchénio, T., Casella, J.F., Heidmann, T.2000. Members of the SRY family regulate the human LINE retrotransposons. Nucleic Acids Res. 15;28(2), pp.411-5.

153. Ullu, E. and Tschudi, C. 1984. Alu sequences are processed 7SL RNA genes. Nature. 312(5990), pp.171-172.

154. Upadhyay, V., Singh, A. and Panda, A.K. 2016. Purification of recombinant ovalbumin from inclusion bodies of Escherichia coli. Protein expression and purification. 117, pp.52-58.

155. 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.

156. 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.

157. 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.

158. 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: cispreference versus trans complementation. Molecular and cellular biology. 21(4), pp.1429-1439.

159. Weiner, A.M. 1980. An abundant cytoplasmic 7S RNA is complementary to the dominant interspersed middle repetitive DNA sequence family in the human genome. Cell. 22(1), pp.209-218.

160. Wissing, S., Montano, M., Garcia-Perez, J.L., Moran, J.V. and Greene, W.C. 2011. Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. Journal of Biological Chemistry. 286(42), pp.36427-36437.

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

162. 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.

163. 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.

164. 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.

165. 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.

166. Yue, X., Zhao, Y., Xu, Y., Zheng, M., Feng, Z. and Hu, W. 2017. Mutant p53 in cancer: Accumulation, gain-of-function and therapy. Journal of Molecular Biology.429, pp.1595-1606.

167. 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.

168. Zhang, A., Dong, B., Doucet, A.J., Moldovan, J.B., Moran, J.V. and Silverman, R.H. 2013. RNase L restricts the mobility of engineered retrotransposons in cultured human cells. Nucleic acids research. 42(6), pp.3803-3820.

169. Zhao, K., Du, J., Han, X., Goodier, J.L., Li, P., Zhou, X., Wei, W., Evans, S.L., Li, L., Zhang, W. and Cheung, L.E. 2013. Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutieres syndrome-related SAMHD1. Cell reports. 4(6), pp.1108-1115.

170. Zheng, Y.H., Jeang, K.T. and Tokunaga, K. 2012. Host restriction factors in retroviral infection: promises in virus-host interaction. Retrovirology. 9(1), p.112.

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
```

Blue : EN domain: (1-239 aa)
Red: RT Domain: (453-880 aa)
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  LL  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
actaaaatcagagcagaactgaaggaaatagagacacaaaaaaccccttcaaaaaatcaat
 T  K  I  R  A  E  L  K  E  I  E  T  Q  K  T  L  Q  K  I  N
gaatccaggagctggtttttttgaaaggatcaacaaaattgatagaccgctagcaagacta
 E  S  R  S  W  F  F  E  R  I  N  K  I  D  R  P  L  A  R  L
ataaagaaaaaaagagagaagaatcaaatagacacaataaaaaatgataaaggggatatc
 I  K  KK  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
ctagccatatgtagaaagctgaaactggatccttccttacaccttatacaaaaatcaat
 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
acagcaaaagaaactaccatcagagtgaacaggcaacctacaacatgggagaaaattttt
 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  KK  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  SS  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  -
```

SV2RTSacFwd: 5'-ACTATT*GAGCTC***g**aaggaaatagagacacaa-3'

SV2RTNotRev: 5'-ACTATT*GCGGCCGC*tcagttttcttctagggt-3'

SV3RTSacFwd: 5'-ACT*GAGCTC***g**caggaagaagttgaatctctgaat-3'

ACTATT and ACT sequences at the begining for sitting restriction enzyme at the end of PCR amplified insert for proper digestion
*GAGCTC i*s SacI site
*GCGGCCGC*is NotI site
g an extra nucleotide added for making right frame
tca stop codon

6.1.3. pBS-hL1RT
Human RT domain (1422 –2961 nu) is cloned in pBS KS(-) EcoRI-BamHI sites

……..ttgatatc**gaattc**taccagaggtacaaggaggaactggtaccattccttctgaaactattc
caatcaatagaaaaagagggaatcctccctaactcattttatgaggccagcatcattctgatacca
aagccgggcagagacacaaccaaaaaagagaatttagaccaatatccttgatgaacattgatgca
aaaatcctcaataaaatactggcaaaccgaatccagcagcacatcaaaaagcttatccaccatgat
caagtgggcttcatccctgggatgcaaggctggttcaatatacgcaaatcaataaatgtaatccag
catataaacagagccaaagacaaaaaccacatgattatctcaatagatgcagaaaaagcctttgac
aaaattcaacaacccttcatgctaaaaactctcaataaattaggtattgatgggacgtatttcaaa
ataataagagctatctatgacaaacccacagccaatatcatactgaatgggcaaaaactggaagca
ttccctttgaaaaccggcacaagacagggatgccctctctcaccgctcctattcaacatagtgttg
gaagttctggccagggcaatcaggcaggagaaggaaataaagggtattcaattaggaaaagaggaa
gtcaaattgtccctgtttgcagacgacatgattgtttatctagaaaaccccatcgtctcagcccaa
aatctccttaagctgataagcaacttcagcaaagtctcaggatacaaaatcaatgtacaaaaatca
caagcattcttatacaccaacaacagacaaacagagagccaaatcatgggtgaactcccattcaca
attgcttcaaagagaataaaatacctaggaatccaacttacaagggatgtgaaggacctcttcaag
gagaactacaaaccactgctcaaggaaataaaagaggagacaaacaaatggaagaacattccatgc
tcatgggtaggaagaatcaatatcgtgaaaatggccatactgcccaaggtaatttacagattcaat
gccatccccatcaagctaccaatgactttcttcacagaattggaaaaaactactttaaagttcata
tggaaccaaaaaagagcccgcattgccaagtcaatcctaagccaaaagaacaaagctggaggcatc
acactacctgacttcaaactatactacaaggctacagtaaccaaaacagcatggtactggtaccaa
aacagagatatagatcaatggaacagaacagagccctcagaaataatgccgcatatctacaactat
ctgatctttgacaaacctgagaaaaacaagcaatggggaaaggattccctatttaataaatggtgc
tgggaaaactggctagccatatgtagaaagctgaaact**ggatcc**actagttctaga**gcggccgc**……
…

6.1.4:SV1: pET 28a-RT (Swapping of EcoRI- NotI fragment from pBS to pET28a)

atgggcagcagccatcatcatcatcacagcagcggcctggtgccgcgcggcagccatatggct
agcatgactggtggacagcaaatgggtcgcggatcc**gaattc**taccagaggtacaaggaggaactg
gtaccattccttctgaaactattccaatcaatagaaaaagagggaatcctccctaactcattttat
gaggccagcatcattctgataccaaagccgggcagagacacaaccaaaaaagagaatttagacca
atatccttgatgaacattgatgcaaaaatcctcaataaaatactggcaaaccgaatccagcagcac
atcaaaaagcttatccaccatgatcaagtgggcttcatccctgggatgcaaggctggttcaatata
cgcaaatcaataaatgtaatccagcatataaacagagccaaagacaaaaaccacatgattatctca
atagatgcagaaaaagcctttgacaaaattcaacaacccttcatgctaaaaactctcaataaatta
ggtattgatgggacgtatttcaaaataataagagctatctatgacaaacccacagccaatatcata
ctgaatgggcaaaaactggaagcattccctttgaaaaccggcacaagacagggatgccctctctca
ccgctcctattcaacatagtgttggaagttctggccagggcaatcaggcaggagaaggaaataaag
ggtattcaattaggaaaagaggaagtcaaattgtccctgtttgcagacgacatgattgtttatcta
gaaaaccccatcgtctcagcccaaaatctccttaagctgataagcaacttcagcaaagtctcagga
tacaaaatcaatgtacaaaaatcacaagcattcttatacaccaacaacagacaaacagagagccaa
atcatgggtgaactcccattcacaattgcttcaaagagaataaaatacctaggaatccaacttaca
agggatgtgaaggacctcttcaaggagaactacaaaccactgctcaaggaaataaaagaggagaca
aacaaatggaagaacattccatgctcatgggtaggaagaatcaatatcgtgaaaatggccatactg
cccaaggtaatttacagattcaatgccatccccatcaagctaccaatgactttcttcacagaattg
gaaaaaactactttaaagttcatatggaaccaaaaaagagcccgcattgccaagtcaatcctaagc
caaaagaacaaagctggaggcatcacactacctgacttcaaactatactacaaggctacagtaacc
aaaacagcatggtactggtaccaaaacagagatatagatcaatggaacagaacagagccctcagaa

ataatgccgcatatctacaactatctgatctttgacaaacctgagaaaaacaagcaatggggaaag
gattccctatttaataaatggtgctgggaaaactggctagccatatgtagaaagctgaaact**ggat**
ccactagttctaga**gcggccgc**actcgagcaccaccaccaccaccactgagatccggctgctaa

Red: pET 28a vectrorseq
Black: RT domain
Blue:pBS vector sequence

6.1.5. pET-hL1RT protein sequence

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

Red letters are extra amino acids coming from vector backbone.Black letters indicate RT protein sequence. There are extra 34 and 20 amino acids at the N-ter and C-ter respectively.

Total no. of amino acids translated including vector amino acids: 527
Molecular weight and PI of the protein is 61.2kDa and 9.1 respectively.

6.1.6. SV2: pET 30b-RT_{SV2} (PCR amplified fragment using primer pairs SV2RTSacFwd and SV2RTNot Rev as indicated L1 nucleotide sequence at the top cloned in SacI- NotI site of pET30b)

atgcaccatcatcatcattcttctggtctggtgccacgcggttctggtatgaaagaaaccgct
gctgctaaattcgaacgccagcacatggacagcccagatctgggtaccgacgacgacgacaaggcc
atggcgatatcggatccgaattc**gagctc**gaaggaaatagagacacaaaaaacccttcaaaaaatc
aatgaatccaggagctggtttttgaaaggatcaacaaaattgatagaccgctagcaagactaata
aagaaaaaaagagagaagaatcaaatagacacaataaaaaatgataaagggatatcaccaccgat
cccacagaaatacaaactaccatcagagaatactacaaacacctctacgcaaataaactagaaaat

ctagaagaaatggatacattcctcgacacatacactctcccaagactaaaccaggaagaagttgaa
tctctgaatcgaccaataacaggctctgaaattgtggcaataatcaatagtttaccaaccaaaaag
agtccaggaccagatggattcacagccgaattctaccagaggtacaaggaggaactggtaccattc
cttctgaaactattccaatcaatagaaaaagagggaatcctccctaactcattttatgaggccagc
atcattctgataccaaagccgggcagagacacaaccaaaaaagagaattttagaccaatatccttg
atgaacattgatgcaaaaatcctcaataaaatactggcaaaccgaatccagcagcacatcaaaaag
cttatccaccatgatcaagtgggcttcatccctgggatgcaaggctggttcaatatacgcaaatca
ataaatgtaatccagcatataaacagagccaaagacaaaaaccacatgattatctcaatagatgca
gaaaaagcctttgacaaaattcaacaacccttcatgctaaaaactctcaataaattaggtattgat
gggacgtatttcaaaataataagagctatctatgacaaacccacagccaatatcatactgaatggg
caaaaactggaagcattccctttgaaaaccggcacaagacagggatgccctctctcaccgctccta
ttcaacatagtgttggaagttctggccagggcaatcaggcaggagaaggaaataaagggtattcaa
ttaggaaaagaggaagtcaaattgtccctgtttgcagacgacatgattgtttatctagaaaacccc
atcgtctcagcccaaaatctccttaagctgataagcaacttcagcaaagtctcaggatacaaaatc
aatgtacaaaaatcacaagcattcttatacaccaacaacagacaaacagagagccaaatcatgggt
gaactcccattcacaattgcttcaaagagaataaaatacctaggaatccaacttacaagggatgtg
aaggacctcttcaaggagaactacaaaccactgctcaaggaaataaaagaggagacaaacaaatgg
aagaacattccatgctcatgggtaggaagaatcaatatcgtgaaaatggccatactgcccaaggta
atttacagattcaatgccatccccatcaagctaccaatgactttcttcacagaattggaaaaaact
actttaaagttcatatggaaccaaaaaagagcccgcattgccaagtcaatcctaagccaaaagaac
aaagctggaggcatcacactacctgacttcaaactatactacaaggctacagtaaccaaaacagca
tggtactggtaccaaaacagagatatagatcaatggaacagaacagagccctcagaaataatgccg
catatctacaactatctgatctttgacaaacctgagaaaaacaagcaatggggaaaggattcccta
tttaataaatggtgctgggaaaactggctagccatatgtagaaagctgaaactggatcccttcctt
acaccttatacaaaaatcaattcaagatggattaaagatttaaacgttaaacctaaaaccataaaa
accctagaagaaaac**gcggccgc**tga

6.1.7. pET 30b-RT$_{SV2}$protein sequence

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMAISDPNSSSKEIETQKTLQKI
NESRSWFFERINKIDRPLARLIKKKREKNQIDTIKNDKGDITTDPTEIQTTIREYYKHLYANKLEN
LEEMDTFLDTYTLPRLNQEEVESLNRPITGSEIVAIINSLPTKKSPGPDGFTAEFYQRYKEELVPF
LLKLFQSIEKEGILPNSFYEASIILIPKPGRDTTKKENFRPISLMNIDAKILNKILANRIQQHIKK
LIHHDQVGFIPGMQGWFNIRKSINVIQHINRAKDKNHMIISIDAEKAFDKIQQPFMLKTLNKLGID
GTYFKIIRAIYDKPTANIILNGQKLEAFPLKTGTRQGCPLSPLLFNIVLEVLARAIRQEKEIKGIQ
LGKEEVKLSLFADDMIVYLENPIVSAQNLLKLISNFSKVSGYKINVQKSQAFLYTNNRQTESQIMG
ELPFTIASKRIKYLGIQLTRDVKDLFKENYKPLLKEIKEETNKWKNIPCSWVGRINIVKMAILPKV
IYRFNAIPIKLPMTFFTELEKTTLKFIWNQKRARIAKSILSQKNKAGGITLPDFKLYYKATVTKTA
WYWYQNRDIDQWNRTEPSEIMPHIYNYLIFDKPEKNKQWGKDSLFNKWCWENWLAICRKLKLDPFL
TPYTKINSRWIKDLNVKPKTIKTLEENGR*

Red: pET 30b vector sequence
Black: RT domain
Total no. of amino acids translated including vector amino acids: 689
Molecular weight and PI of the protein is 80.2kDa and 9.5 respectively.

6.1.8. SV3: pET 30b-RT_{SV3} (PCR amplified fragment using primer pairs SV3RTSacFwd and SV2RTNot Rev as indicated L1 nucleotide sequence at the top cloned in SacI- NotI site of pET30b)

atgcaccatcatcatcatcattcttctggtctggtgccacgcggttctggtatgaaagaaaccgct
gctgctaaattcgaacgccagcacatggacagcccagatctgggtaccgacgacgacgacaaggcc
atggcgatatcggatccgaattc**gagctc**gcaggaagaagttgaatctctgaatcgaccaataaca
ggctctgaaattgtggcaataatcaatagtttaccaaccaaaaagagtccaggaccagatggattc
acagccgaattctaccagaggtacaaggaggaactggtaccattccttctgaaactattccaatca
atagaaaaagagggaatcctccctaactcattttatgaggccagcatcattctgataccaaagccg
ggcagagacacaaccaaaaaagagaattttagaccaatatccttgatgaacattgatgcaaaaatc
ctcaataaaatactggcaaaccgaatccagcagcacatcaaaaagcttatccaccatgatcaagtg
ggcttcatccctgggatgcaaggctggttcaatatacgcaaatcaataaatgtaatccagcatata
aacagagccaaagacaaaaaccacatgattatctcaatagatgcagaaaaagcctttgacaaaatt
caacaacccttcatgctaaaaactctcaataaattaggtattgatgggacgtatttcaaaataata
agagctatctatgacaaacccacagccaatatcatactgaatgggcaaaaactggaagcattccct
ttgaaaaccggcacaagacagggatgccctctctcaccgctcctattcaacatagtgttggaagtt
ctggccagggcaatcaggcaggagaaggaaataaagggtattcaattaggaaaagaggaagtcaaa
ttgtccctgtttgcagacgacatgattgtttatctagaaaaccccatcgtctcagcccaaaatctc
cttaagctgataagcaacttcagcaaagtctcaggatacaaaatcaatgtacaaaaatcacaagca
ttcttatacaccaacaacagacaaacagagagccaaatcatgggtgaactcccattcacaattgct
tcaaagagaataaaatacctaggaatccaacttacaagggatgtgaaggacctcttcaaggagaac
tacaaaccactgctcaaggaaataaaagaggagacaaacaaatggaagaacattccatgctcatgg
gtaggaagaatcaatatcgtgaaaatggccatactgcccaaggtaatttacagattcaatgccatc
cccatcaagctaccaatgactttcttcacagaattggaaaaaactactttaaagttcatatggaac
caaaaaagagcccgcattgccaagtcaatcctaagccaaaagaacaaagctggaggcatcacacta
cctgacttcaaactatactacaaggctacagtaaccaaaacagcatggtactggtaccaaaacaga
gatatagatcaatggaacagaacagagccctcagaaataatgccgcatatctacaactatctgatc
tttgacaaacctgagaaaaacaagcaatggggaaaggattccctatttaataaatggtgctgggaa
aactggctagccatatgtagaaagctgaaactggatcccttccttacaccttatacaaaaatcaat
tcaagatggattaaagatttaaacgttaaacctaaaaccataaaaaccctagaagaaaac**gcggcc**
gctga

6.1.9. pET 30b-RT_{SV3}protein sequence

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMAISDPNSSSQEEVESLNRPIT
GSEIVAIINSLPTKKSPGPDGFTAEFYQRYKEELVPFLLKLFQSIEKEGILPNSFYEASIILIPKP
GRDTTKKENFRPISLMNIDAKILNKILANRIQQHIKKLIHHDQVGFIPGMQGWFNIRKSINVIQHI
NRAKDKNHMIISIDAEKAFDKIQQPFMLKTLNKLGIDGTYFKIIRAIYDKPTANIILNGQKLEAFP
LKTGTRQGCPLSPLLFNIVLEVLARAIRQEKEIKGIQLGKEEVKLSLFADDMIVYLENPIVSAQNL
LKLISNFSKVSGYKINVQKSQAFLYTNNRQTESQIMGELPFTIASKRIKYLGIQLTRDVKDLFKEN
YKPLLKEIKEETNKWKNIPCSWVGRINIVKMAILPKVIYRFNAIPIKLPMTFFTELEKTTLKFIWN
QKRARIAKSILSQKNKAGGITLPDFKLYYKATVTKTAWYWYQNRDIDQWNRTEPSEIMPHIYNYLI
FDKPEKNKQWGKDSLFNKWCWENWLAICRKLKLDPFLTPYTKINSRWIKDLNVKPKTIKTLEENGR
*

Red: pET 30b vectrorseq
Black: RT domain

Total no. of amino acids translated including vector amino acids: 594
Molecular weight and PI of the protein is 68.7kDa and 9.6 respectively.

6.2. Appendix II

6.2.1. The details of patients used in this study are in the following table:

Serial no.	Age	Sex	Tissue type	ORF1p	p53	Smoking/tobacco Chewing status
1	45	Male	tongue	negative	ND	Yes
2	70	Male	buccal mucosa	positive	ND	Yes
3	29	Male	buccal mucosa	positive	positive	Yes
4	35	Male	buccal mucosa	positive	positive	Yes
5	53	Male	buccal mucosa	negative	ND	Yes
6	58	Male	buccal mucosa	positive	positive	Yes
7	46	Male	buccal mucosa	positive	ND	Yes
8	54	Male	tongue	negative	ND	Yes
9	67	Male	buccal mucosa	positive	ND	Yes
10	42	Male	tongue	negative	negative	Yes
11	45	Female	buccal mucosa	positive	negative	Yes
12	45	Male	buccal mucosa	negative	ND	Yes

6.3. Appendix III

6.3.1. >L1 Promoter (5'-UTR) sequence (L1.3, Accession # 19088.1)

GGGGGAGGAGCCAAGATGGCCGAATAGGAACAGCTCCGGTCTACAGCTCCCAGCGTGAGCGACGCAGAA
GACGGTGATTTCTGCATTTCCATCTGAGGTACCGGGTTCATCTCACTAGGGAGTGCCAGACAGTGGGCGC
AGGCCAGTGTGTGTGCGCACCGTGCGCGAGCCGAAGCAGGGCGAGGCATTGCCTCACCTGGGAAGCGCA
AGGGGTCAGGGAGTTCCCTTTCTGAGTCAAAGAAAGGGGTGACGGTCGCACCTGGAAAATCGGGTCACT
CCCACCCGAATATTGCGCTTTTCAGACCGGCTTAAGAAACGGCGCACCACGAGACTATATCCCACACCTGG
CTCGGAGGGTCCTACGCCCACGGAATCTCGCTGATTGCTAGCACAGCAGTCTGAGATCAAACTGCAAGGC
GGCAACGAGGCTGGGGGAGGGGCGCCCGCCATTGCCCAGGCTTGCTTAGGTAAACAAAGCAGCCGGGA
AGCTCGAACTGGGTGGAGCCCACCACAGCTCAAGGAGGCCTGCCTGCCTCTGTAGGCTCCACCTCTGGGG
GCAGGGCACAGACAAACAAAAAGACAGCAGTAACCTCTGCAGACTTAAGTGTCCCTGTCTGACAGCTTTG
AAGAGAGCAGTGGTTCTCCCAGCACGCAGCTGGAGATCTGAGAACGGGCAGACAGACTGCCTCCTCAAGT
GGGTCCCTGACTCCTGACCCCCGAGCAGCCTAACTGGGAGGCACCCCCCAGCAGGGGCACACTGACACCT
CACACGGCAGGGTATTCCAACAGACCTGCAGCTGAGGGTCCTGTCTGTTAGAAGGAAAACTAACAACCAG
AAAGGACATCTACACCGAAAACCCATCTGTACATCACCATCATCAAAGACCAAAAGTAGATAAAACCACAA
AG

Yellow sequence- analyzed for methylation state in this study.
Total CpGs=39 (red)
CpGs in amplified sequence=19

For both cancer and paired normal samples, the region of the L1 promoter (shaded yellow) amplified after bisulfite treatment and cloned. Three to five representative clones were sequenced from each cancer and paired normal samples and then blast using Rep Base [22] to find out in which L1 subfamily those clones are belonged to. The clones were then aligned using Clustal omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) to determine the number of unmethylated C in CpG dinucleotides.

As a representative, the alignment result for sample 1 [normal (N1)- 4 clones and cancer (C1) -4 clones] is shown below which was used to make Figure 1A. The similar kind of alignment was performed for other 7 samples but not included in the supplementary text.

C1-Cancer tissue from patient 1
N1-Paired normal tissue from patient 1
a,b,c,d- four representative clones sequenced

6.3.2. Alignment of consensus L1HS with four clones obtained from C1

```
L1HS      ----------------------------------------AAGGGGTCAGGGAGTTCCCTT
C1b       CCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTAAGGGGTTAGGGAGTTTTTTT
C1a       CCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTAAGGGGTTAGGGAGTTTTTTT
C1c       CCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTAAGGGGTTAGGGAGTTTTTTT
C1d       CCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTAAGGGGTTAGGGAGTTTTTTT
                                          ******* ********   **

L1HS      TCCGAGTCAAAGAAAGGGGTGACGGACG-CACCTGGAAAATCGGGTCACTCCCACCCGAA
C1b       TTCGAGTTAAAGAAAGGGGTGACGGACG-TATTTGGAAAATCGGGTTATTTATATTTGAA
C1a       TTCGAGTTAAAGAAAGGGGTGACGGACG-TATTTGGAAAATCGGGTTATTTTTATTCGAA
C1c       TTTGAGTTAAAGAAAGGGGTGATGGACGGTATTTGGAAAATCGGGTTATTTTTATTCGAA
C1d       TTTGAGTTAAAGAAAGGGGTGATGGACGGTATTTGGAAAATCGGGTTATTTTTATTTGAA
          *  **** **************  **   *  ************ * *   *    **

L1HS      TATTGCGCTTTTCAGACCGGCTTAAGAAACGGCGCACCACGAGACTATATCCCACACCTG
C1b       TATTGCGTTTTTTAGATTGGTTTAAAAAACGGCGTATTACGAGATTATATTTTATATTTG
C1a       TATTGCGTTTTTTCGACGGGTTTAAGAAACGGCGTATTACGAGATTATATTTTTTATTTG
C1c       TATTGCGTTTTTTCGACGGGTTTAAAAAATGGTGTATTATGAGATTATATTTTGTATTTG
C1d       TATTGCGTTTTTTCGATAGGTTTAAAAAAAGGCGTATTACGAGATTATATTTCGTATTTG
          ******* ****  **  ** **** *** ** * *  * **** *****     *  **

L1HS      GCTCGGAGGGTCCTACGCCCACGGAATCTCGCTGATTGCTAGCACAGCAGTCTGAGATCA
C1b       GTTCGGAGGGTTTTACGTTTACGGAATTTGGTTGATTGTTAGTATAGTAGTTTGAGATTA
C1a       GTTTGGAGGGTTTTATGTTTATGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATTA
C1c       GTTTGAAGGGTTTTATGTTTATGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATTA
C1d       GTTTGGAGGGTTTTATGTTTATGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATTA
          * *   *****  ** *   * *** * * * ****** *** * ** *** ****** *

L1HS      AACTGCAAGGCGGCAACGAGGCTGGGGGAGGGGCGCCCGCCATTGCCCAGGCTTGCTTAG
C1b       AATTGTAAGGTGGTAGCGAGGTTGGGGGAGGGGCGTTCGTTATTGTTTAGGTTTGTTTAG
C1a       AATTGTAAGGTGGTAGTGAGGTGGGGGGAGGGGTGTTTGTTATTGTTTAGGTTTGATTAG
C1c       AATTGTAAGGTGGTAGTGAGGTTGGGGGAGGGGTGTTTGTTATTGTTTAGGTTTGTTTAG
C1d       ATTTGTAAGGCGGTAGCGAGGTTGGGGGAGGGGCGTTTGTTATTGTTTAGGTTTGTTTAG
          *  ** **** ** *  ****  ********** *   *  ****   *** *** ****

L1HS      GTAAACAAAGCAGCCGGGAAGCTCGAACTGGGTGGAGCCCACCACAGCTCAAGGAGGCCT
C1b       GTAAATAAAGTGGTTGGGAAGTTCGAATTGGGTGGAGTTTATTATAGTTTAAGGAGGTTT
C1a       GTAAATAAAGTAGTTAGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTAAGGAGGTTT
C1c       GTAAATAAAGTAGTTTGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTAAGGAGGTTT
C1d       GTAAATAAAGTAGTTAGGAGGTTTGAATTGGGTGGAGTTTATTATAGTTTAAGGAGGTTT
          ***** ****  *   *** * *  ** *********   *  * ** * *******  *

L1HS      GCCTGCCTCTGTAGGCTCCACCTCTGGGGGCAGGGCACA--------------------
C1b       GTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACTAGTGCGGCCG
C1a       GTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACTAGTGCGGCCG
C1c       GTTTGTTTCTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACTAGTGCGGCCG
C1d       GTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACTAGTGCGGCCG
          *  **  * ****** *  *  * ****** **** * *
```

6.3.3. Alignment of consensus L1HS with four clones obtained from N1

```
L1HS      ----------------------------------------AAGGGGTCAGGGAGTTCCC
N1c       GCCCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTAAGGGGTTAGGGAGTTTTT
N1b       GCCCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTAAGGGGTTAGGGAGTTTTT
N1a       GCCCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTAAGGGGTTAGGGAGTTTTT
N1d       GCCCGACGTCGCATGCTCCCGGCCGCCATGGCCGCGGGATTAAGGGGTTAGGGAGTTTTT
                                                 ******* ********


L1HS      TTTCCGAGTCAAAGAAAGGGGTGACGGAC-GCACCTGGAAAATCGGGTCACTCCCACCCG
N1c       TTTTTGAGTTAAAGAAAGGGGTGACGGACGGTATTTGGAAAATCGGGTTATTTTTATTCG
N1b       TTTTCGAGTTAAAGAAAGGGGTGACGGAC-GTATTTGGAAAATCGGGTTATTTTTATTTA
N1a       TTTTTGAGTAAAAGAAAGGGGTGACGGAC-GTATTTGGAAAATCGGGTTATTTTTATTCG
N1d       TTTTTGAGTAAAAGAAAGGGGTGACGGAC-GTATTTGGAAAATCGGGTTATTTTTATTCG
          ***  **** ******************* * *  ************ * *   *


L1HS      AATATTGCGCTTTTCAGACCGGCTTAAGAAACGGCGCACCACGAGACTATATCCCACACC
N1c       AATATTGCGTTTTTTCGACGGGTTTAAAAAACGGCGTATTATGAGATTATATTTTGTATT
N1b       AATATTGCGTTTTTCGGATCGGTTTAAGAAACGGTGTATTACGAGATTATATTTCGTATT
N1a       AATATTGCGTTTTTCGGATCGGTTTAAAAAGCGGCGTATTACGAGATTATGTTTCGTATT
N1d       AATATTGCGTTTTTCGGATCGGTTTAAAAAGCGGCGTATTACGAGATTATGTTTCGTATT
          ********* ****  **  ** **** ** ***   *  * **** *** *     *


L1HS      TGGCTCGGAGGGTCCTACGCCCACGGAATCTCGCTGATTGCTAGCACAGCAGTCTGAGAT
N1c       TGGTTTGGAGGGTTTTACGTTTACGGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGAT
N1b       TGGTTTGGAGGGTTTTACGTTTATAGAGTTTCGTTGATTATTAGTATAGTAGTTTGAGAT
N1a       TGGTTCGTAGGGTTTTACGTTTATGGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGAT
N1d       TGGTTCGTAGGGTTTTACGTTTATGGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGAT
          *** * * *****  ****   *  ** * *** *****  *** * ** *** ******


L1HS      CAAACTGCAAGGCGGCAACGAGGCTGGGGGAGGGGCGCCCGCCATTGCCCAGGCTTGCTT
N1c       TAAATTGTAAAGCGGTAGTGAGGTTGGGGGAGGGGCGTTCGTTATTGTTTAGGTTTGATT
N1b       TAAATTGTAAGGCGGTAGCGAGGTTGGGGGAGGGGTGTTCGTTATTGTTTAGGTTTGTTT
N1a       TAAATTGTAAAGCGGTAGTGAGGTTGGGGGAGGGGTATTCGTTATTGTTTAGGTTTGATT
N1d       TAAATTGTAAAGCGGTAGTGAGGTTGGGGGAGGGGTATTCGTTATTGTTTAGGTTTGATT
           *** ** ** **** *  **** ***********    **  ****   *** *** **


L1HS      AGGTAAACAAAGCAGCCGGGAAGCTCGAACTGGGTGGAGCCCACCACAGCTCAAGGAGGC
N1c       AGGTAAATAAAGTAGATGGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTAAGGAGGT
N1b       ATTTAAATAAAGTAGTTGGGAAGTTTAAATTGGGTGGAGTTTATTATAGTTTAGGGAGGT
N1a       AGGTAAATAAAGTAGTTGGGAAGTTTGAAGTGGGTGGAGTTTATTACAGTTTAAGGAGGT
N1d       AGGTAAATAAAGTAGTTGGGAAGTTTGAAGTGGGTGGAGTTTATTACAGTTTAAGGAGGT
          *  **** **** **   ***** *  ** *********   *  * ** * * *****


L1HS      CTGCCTGCCTCTGTAGGCTCCACCTCTGGGGGCAGGGCACA------------------
N1c       TTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTTGGGTATAGATAAATCACTAGTGCGGC
N1b       TTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTTGGGTATAGATAAATCACTAGTGCGGC
N1a       TTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGGTAGGGTATAGATAAATCACTAGTGCGGC
N1d       TTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGGTAGGGTATAGATAAATCACTAGTGCGGC
           **  **  * ****** *  *  * ******  *** * *
```

6.3.4. **Sanger sequences of bisulfite treated L1 5'UTR clone sequences from 8 paired normal cancer tissues used for methylaion analysis of L15'UTR in OSCC.**

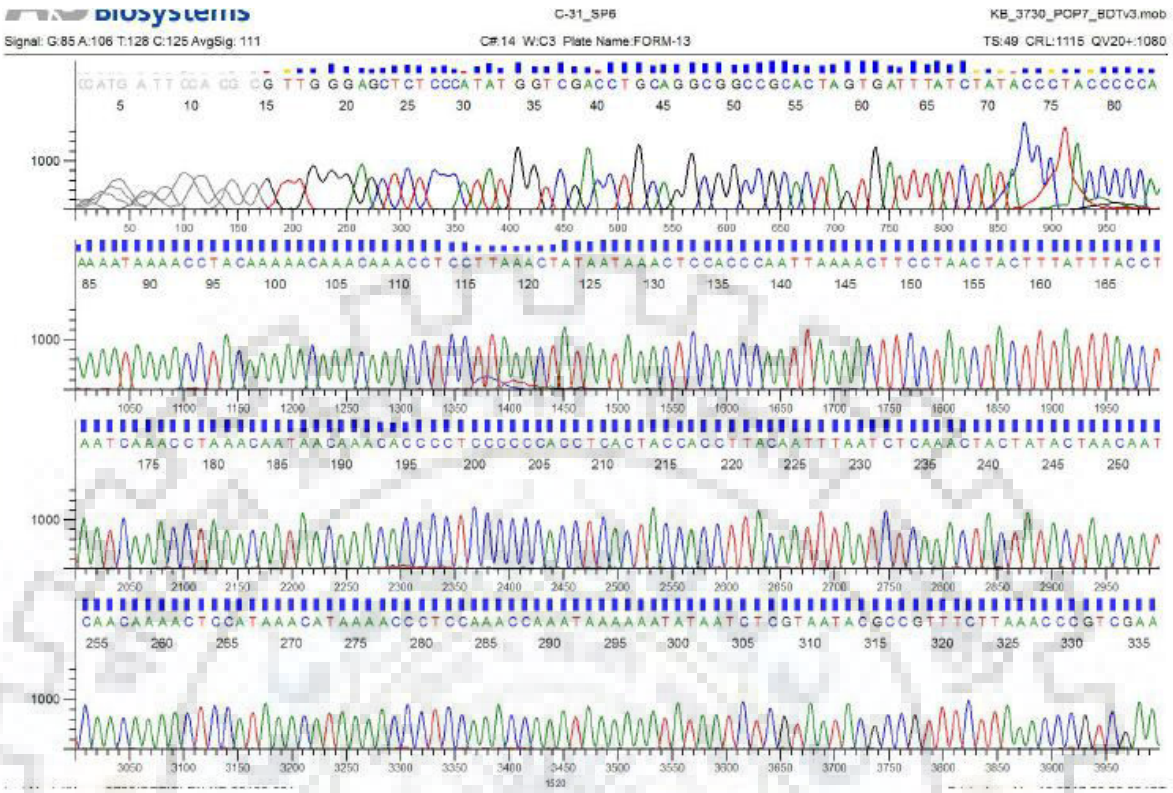
L1HS 5'UTR sequence nucleotide 209-571

AAGGGGTCAGGGAGTTCCCTTTCCGAGTCAAAGAAAGGGGTGACGGACGCACCTGGAAAATCGGGTCACTCCCACC
CGAATATTGCGCTTTTCAGACCGGCTTAAGAAACGGCGCACCACGAGACTATATCCCACACCTGGCTCGGAGGGTCC
TACGCCCACGGAATCTCGCTGATTGCTAGCACAGCAGTCTGAGATCAAACTGCAAGGCGGCAACGAGGCTGGGGGA
GGGGCGCCCGCCATTGCCCAGGCTTGCTTAGGTAAACAAAGCAGCAGGGAAGCTCGAACTGGGTGGAGCCCACCAC
AGCTCAAGGAGGCCTGCCTGCCTCTGTAGGCTCCACCTCTGGGGGCAGGGCACAGACA

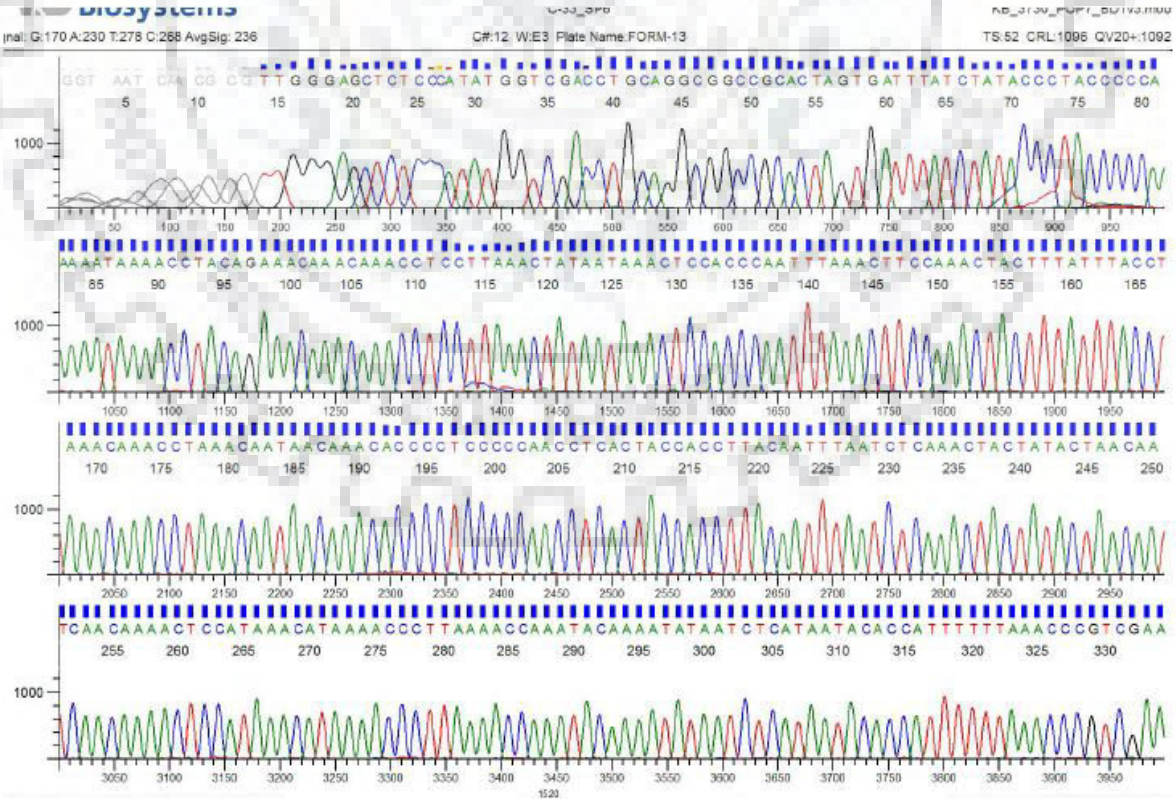
Paired cancer sample 1

>C1a
AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTCGACGGGTTTAAGAAACGGCGTATTACGAGATTATATTTTTTATTTGGTTTGGAGGGTTTTAT
GTTTATGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGTGAGGTGGGGGGAGGG
GTGTTTGTTATTGTTTAGGTTTGATTAGGTAAATAAAGTAGTTAGGAAGTTTTAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT
>C1b
AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGATGTATTTGGAAAATCGGGTTATTTATATTT
AAATATTGCGTTTTTTAGATTGGTTTAAAAAACGGCGTATTACGAGATTATATTTTATATTTGGTTCGGAGGGTTTTAC
GTTTACGGAATTTGGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGCGAGGTTGGGGGAGGG
GCGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTGGTTCGGAAGTTCGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT
>C1c
AAGGGGTTAGGGAGTTTTTTTTTTGAGTTAAAGAAAGGGGTGATGGACGGTATTTGGAAAATCGGGTTATTTTTATT
CGAATATTGCGTTTTTTCGACGGGTTTAAAAAATGGTGTATTATGAGATTATATTTTGTATTTGGTTTTAAGGGTTTTA
TGTTTATGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGTGAGGTTGGGGGAGGG
GTGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTTGGAAGTTTAAATTGGGTGGAGTTTATTATAGTTTA
AGGAGGTTTGTTTGTTTCTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT
>C1d
AAGGGGTTAGGGAGTTTTTTTTTTGAGTTAAAGAAAGGGGTGATAGACGGTATTTGGAAAATCGGGTTATTTTTATT
TGAATATTGCGTTTTTTCGATAGGTTTAAAAAAAGGCGTATTACGAGATTATATTTCGTATTTGGTTTGGAGGGTTTTA
TGTTTATGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAATTTGTAAGGCGGTAGCGAGGTTGGGGGAGGG
GCGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTAGGAGGTTTTAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

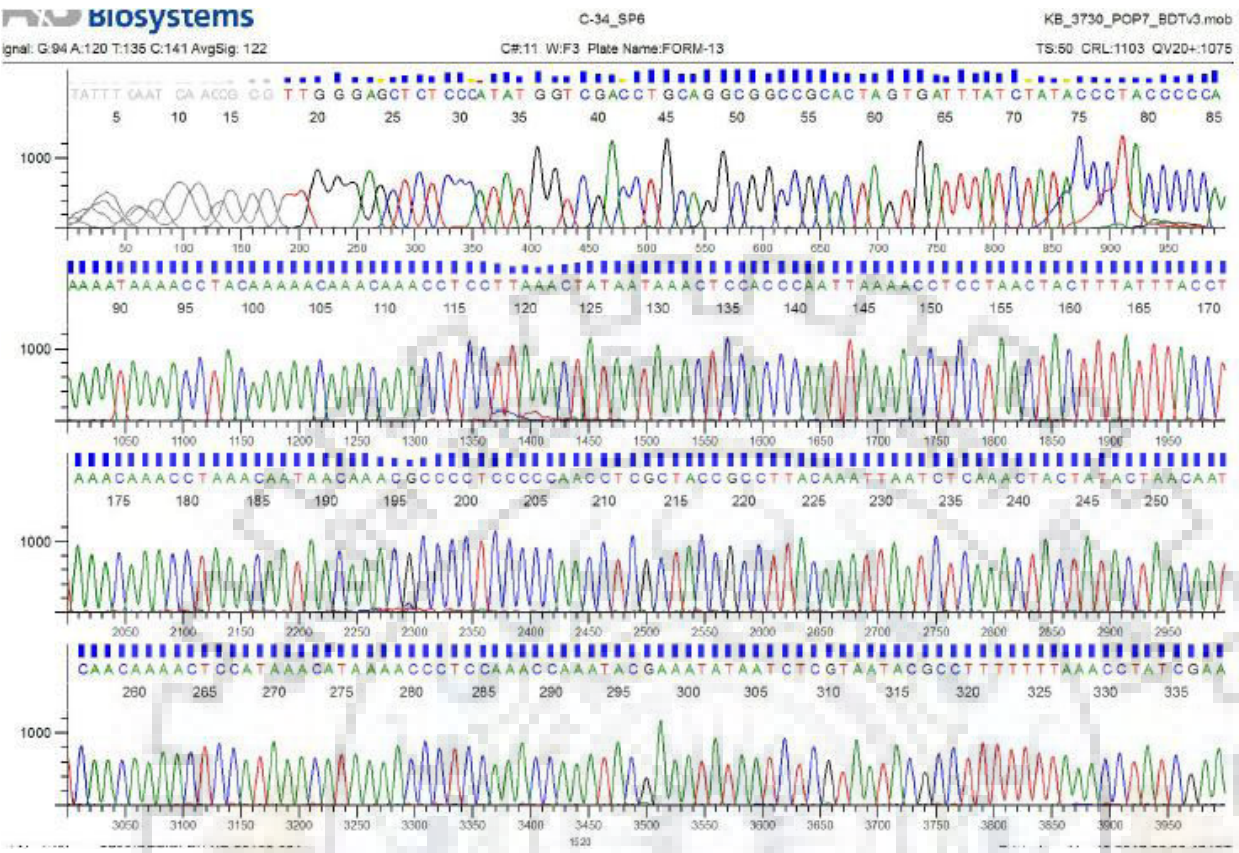
>C1a


C-31_SP6
KB_3730_POP7_BDTv3.mob
Signal: G:85 A:106 T:128 C:125 AvgSig: 111
C#:14 W:C3 Plate Name:FORM-13
TS:49 CRL:1115 QV20+:1080

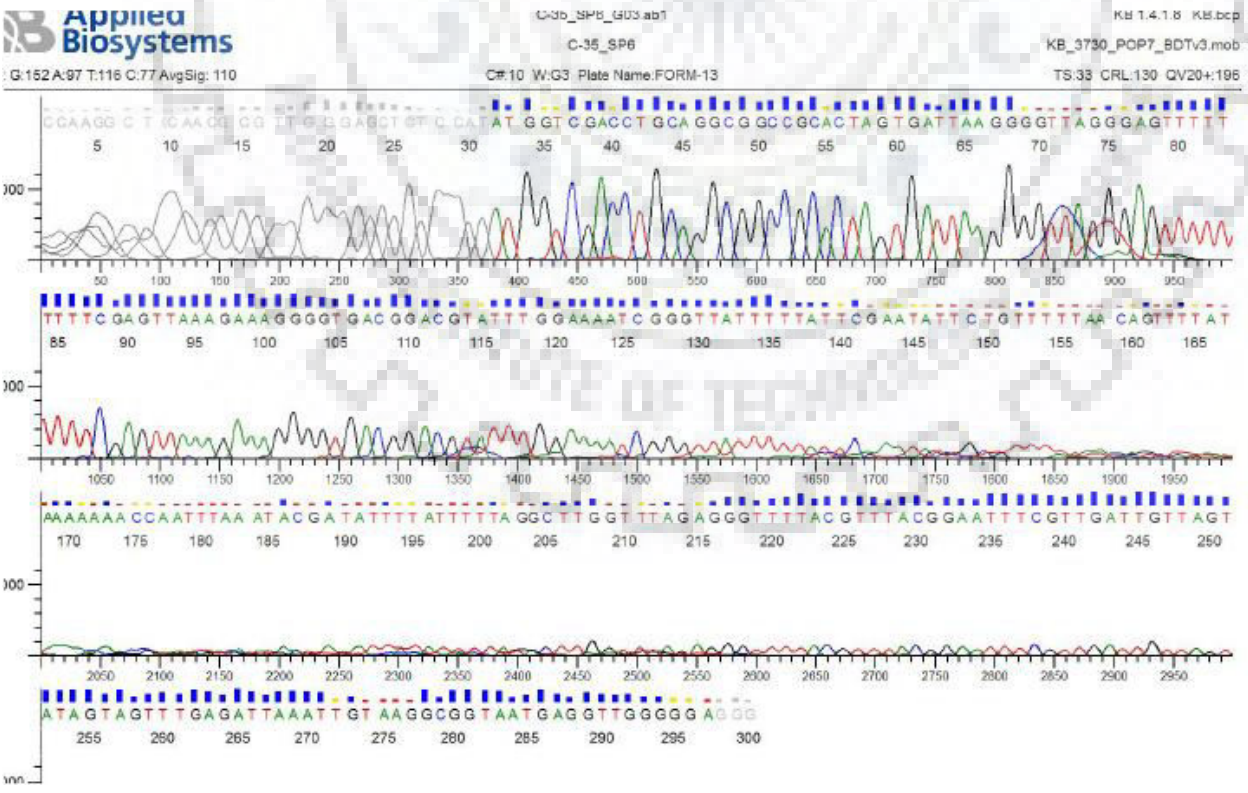
>C1b


Signal: G:170 A:230 T:278 C:268 AvgSig: 236
C#:12 W:E3 Plate Name:FORM-13
TS:52 CRL:1096 QV20+:1092

>C1c


Biosystems
C-34_SP6
KB_3730_POP7_BDTv3.mob
C#:11 W:F3 Plate Name:FORM-13
TS:60 CRL:1103 QV20+:1075

>C1d


Applied Biosystems
C-35_SP6
KB_3730_POP7_BDTv3.mob
G:152 A:97 T:116 C:77 AvgSig: 110
C#:10 W:G3 Plate Name:FORM-13
TS:33 CRL:130 QV20+:196

Paired normal sample 1

>N1a

AAGGGGTTAGGGAGTTTTTTTTTTGAGTAAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTCGGATCGGTTTAAAAAGCGGCGTATTACGAGATTATGTTTCGTATTTGGTTCGTAGGGTTTTA
CGTTTATGGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAAGCGGTAGTGAGGTTGGGGGAGGG
GTATTCGTTATTGTTTAGGTTTGATTAGGTAAATAAAGTAGTTAGGAAGTTTGAAGTGGGTGGAGTTTATTACAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>N1b

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTT
AAATATTGCGTTTTTCGGATCGGTTTAAGAAACGGTATATTACGAGATTATATTTCGTATTTGGTTTGGAGGGTTTTAC
GTTTATAGAGTTTCGTTGATTATTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGCGAGGTTGGGGGAGGG
GTGTTCGTTATTGTTTAGGTTTGTTTATTTAAATAAAGTAGTTAGGAAGTTTAAATTGGGTGGAGTTTATTATAGTTTA
GGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTTGGGTATAGATAAATCACT

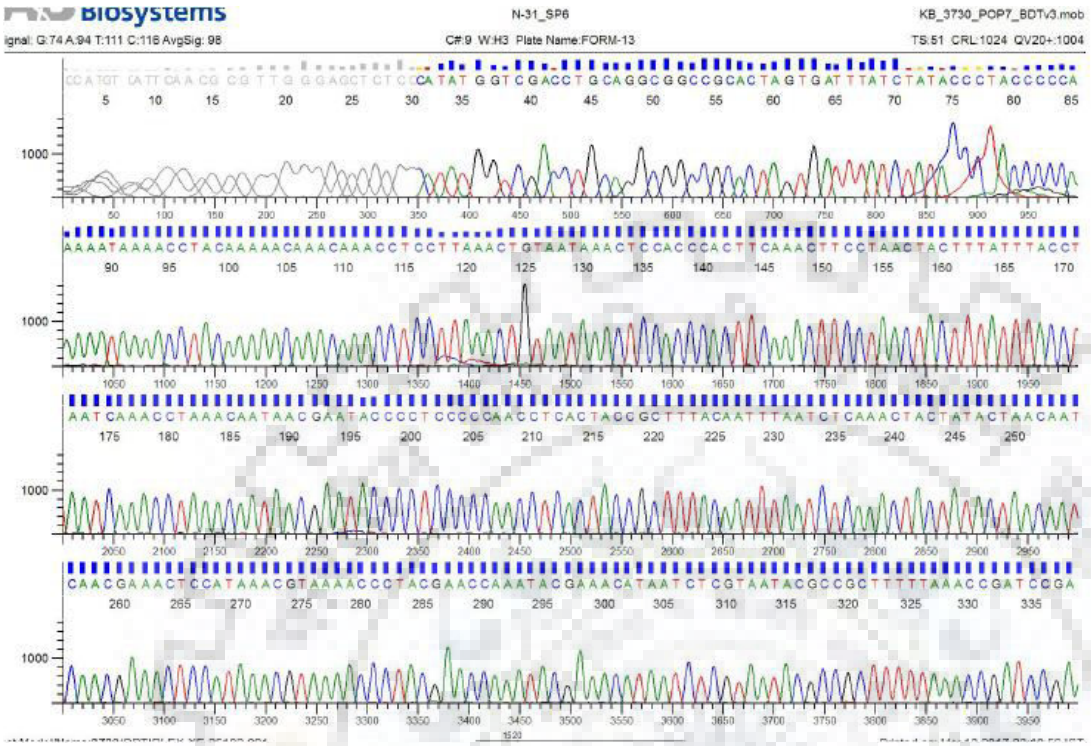
>N1c

AAGGGGTTAGGGAGTTTTTTTTTTGAGTTAAAGAAAGGGGTGACGGACGGTATTTGGAAAATCGGGTTATTTTTATT
CGAATATTGCGTTTTTTCGACGGGTTTAAAAAACGGCGTATTATGAGATTATATTTTGTATTTGGTTTGGAGGGTTTTA
CGTTTACGGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAAGCGGTAGTGAGGTTGGGGGAGGG
GCGTTCGTTATTGTTTAGGTTTGATTAGGTAAATAAAGTAGATAGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTTGGGTATAGATAAATCACT

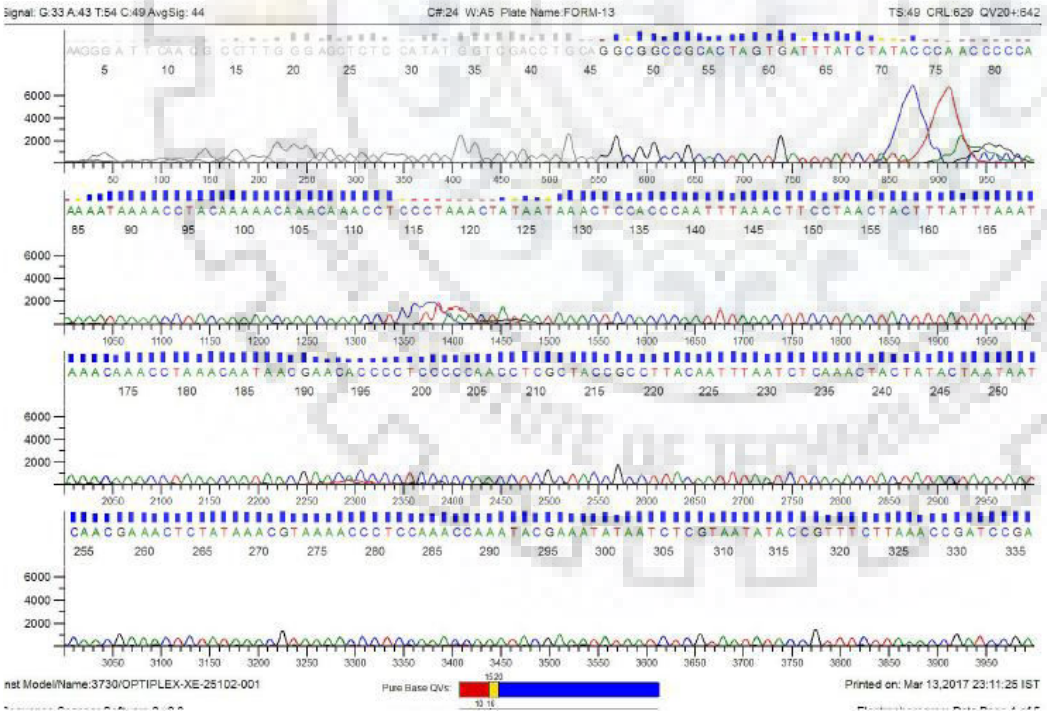
>N1d

AAGGGGTTAGGGAGTTTTTTTTTTGAGTAAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTCGGATCGGTTTAAAAAGCGGCGTATTACGAGATTATGTTTCGTATTTGGTTCGTAGGGTTTTA
CGTTTATGGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAAGCGGTAGTGAGGTTGGGGGAGGG
GTATTCGTTATTGTTTAGGTTTGATTAGGTAAATAAAGTAGTTAGGAAGTTTGAAGTGGGTGGAGTTTATTACAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

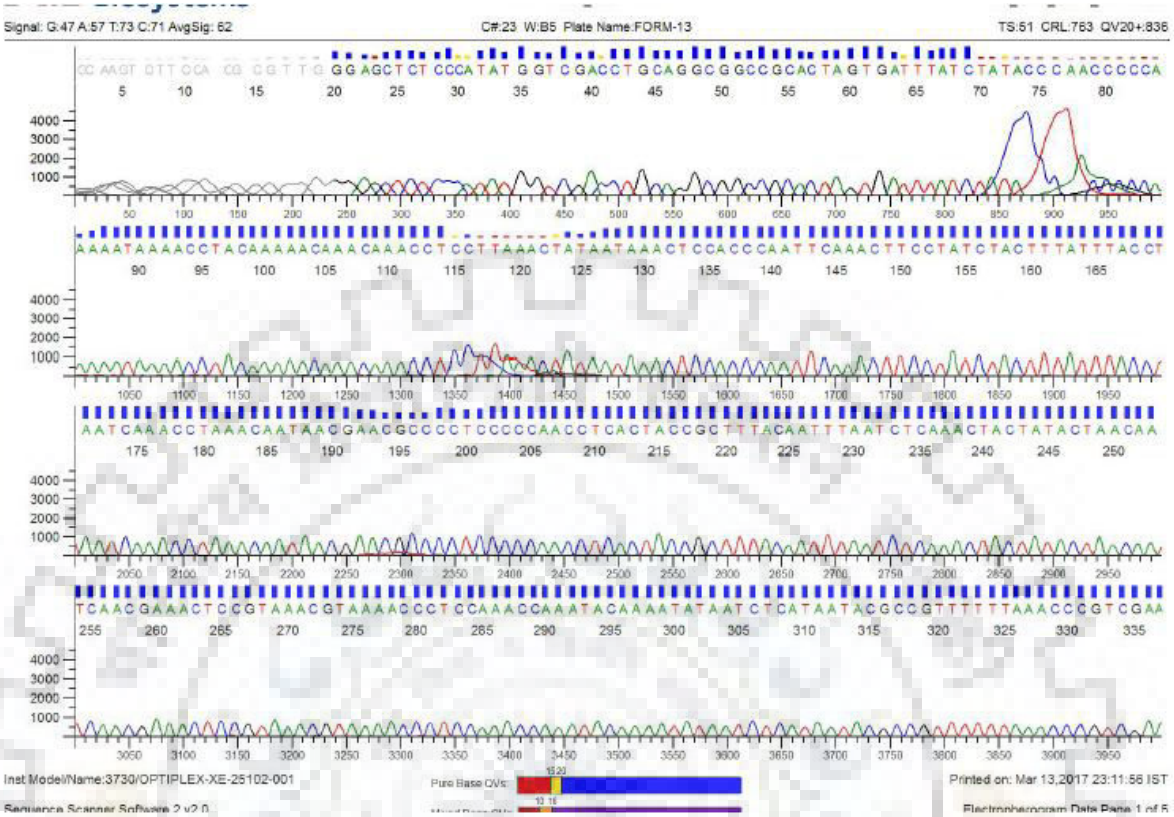
>N1a


Biosystems
N-31_SP6
KB_3730_POP7_BDTv3.mob
C#:9 W:H3 Plate Name:FORM-13
TS:51 CRL:1024 QV20+:1004

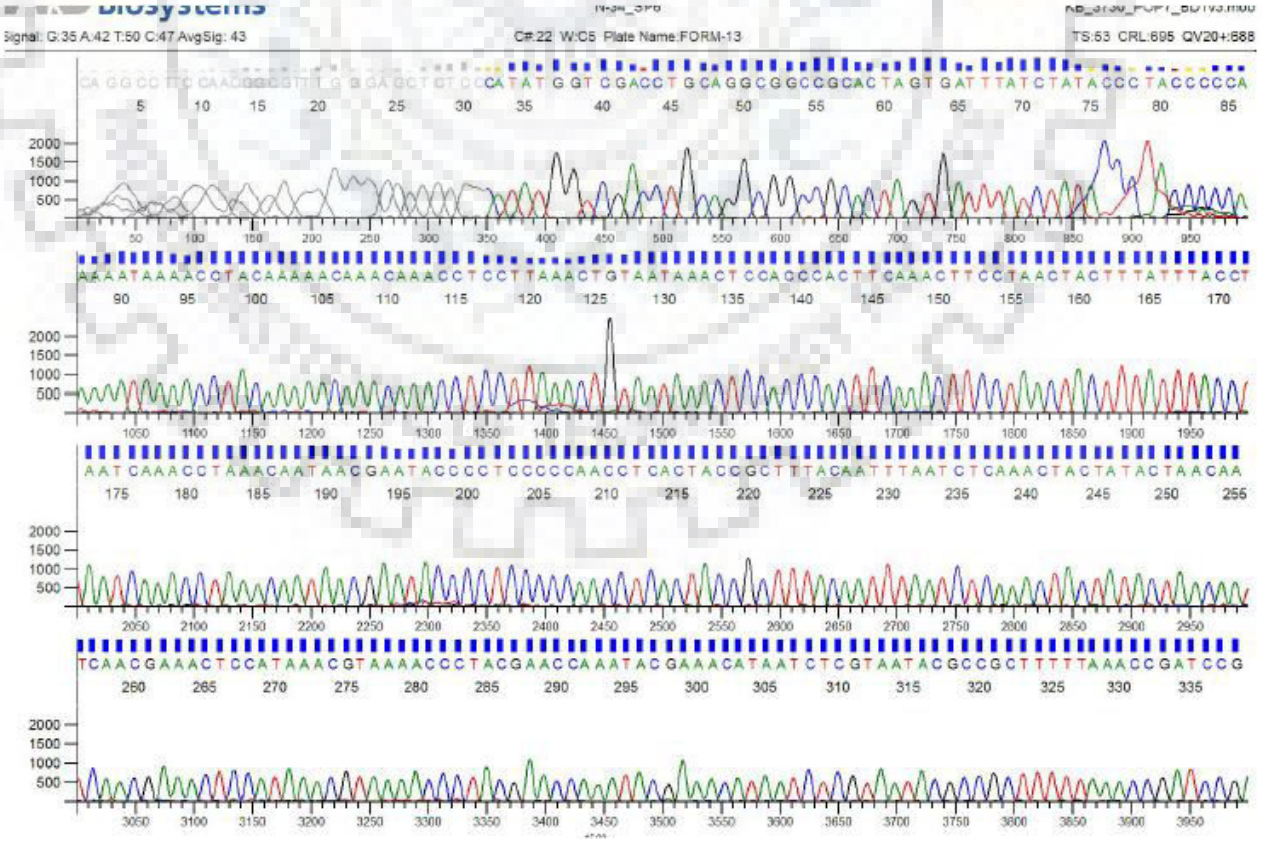
>N1b


C#:24 W:A5 Plate Name:FORM-13
TS:49 CRL:629 QV20+:642
Printed on: Mar 13,2017 23:11:25 IST

>N1c


Signal: G:47 A:57 T:73 C:71 AvgSig: 62
C#:23 W:B5 Plate Name:FORM-13
TS:51 CRL:763 QV20+:836
Inst Model/Name:3730/OPTIPLEX-XE-25102-001
Printed on: Mar 13,2017 23:11:56 IST

>N1d


Signal: G:35 A:42 T:50 C:47 AvgSig: 43
C#:22 W:C5 Plate Name:FORM-13
TS:53 CRL:695 QV20+:688

Paired cancer sample 2

>C2a
AGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTCG
AATATTGCGTTTTTTAGATCGGTTTAAGAAACAGCGTATTACGAGATTATATTTTATACTTGGTTTAGAGGGTTTTACG
TTTACGGAATTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAATGAGGTTGGGGGAGGGGT
GTTCGTTATAGTTTAGGTTTGTTTAGGTAAATAAAGTAGTCGGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTAA
GGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCCCGCGGCCATGGCGGCCGGAGC
ATGCGACTCGCCAAGT

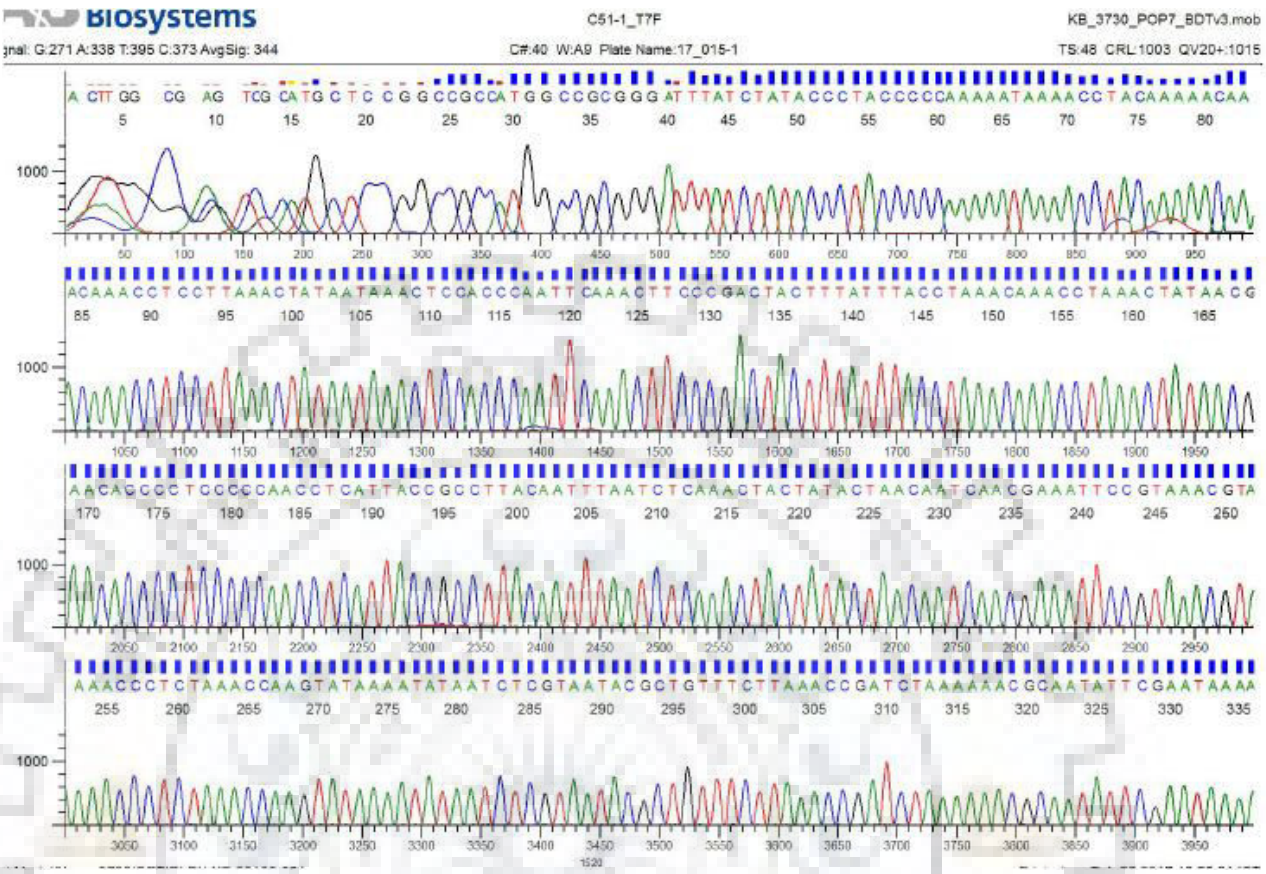
>C2b
AAGGGGTTAGGGAGTTTTTTTTTTGAGTTAAAGAAAGGGGTGATGGACGGTATTTGGAAAATCGGGTTATTTTTATT
CGAATATTGCGTTTTTTTCGACGGGTTTAAAAAATGGTGTATTATGAGATTATATTTTGTATTTGGTTTTAAGGGTTTTA
TGTTTATGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGTGAGGTTGGGGGAGGG
GTGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTTGGAAGTTTAAATTGGGTGGAGTTTATTATAGTTTA
AGGAGGTTTGTTTGTTTCTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCAC

>C2c
AAGGGGTTAGGGAGTTTTTTTTTTGAGTTAAAGAAAGGGGTGATAGACGGTATTTGGAAAATCGGGTTATTTTTATT
TGAATATTGCGTTTTTTTCGATAGGTTTAAAAAAAGGCGTATTACGAGATTATATTTCGTATTTGGTTTGGAGGGTTTTA
TGTTTATGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAATTTGTAAGGCGGTAGCGAGGTTGGGGGAGGG
GCGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTAGGAGGTTTTAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCAC

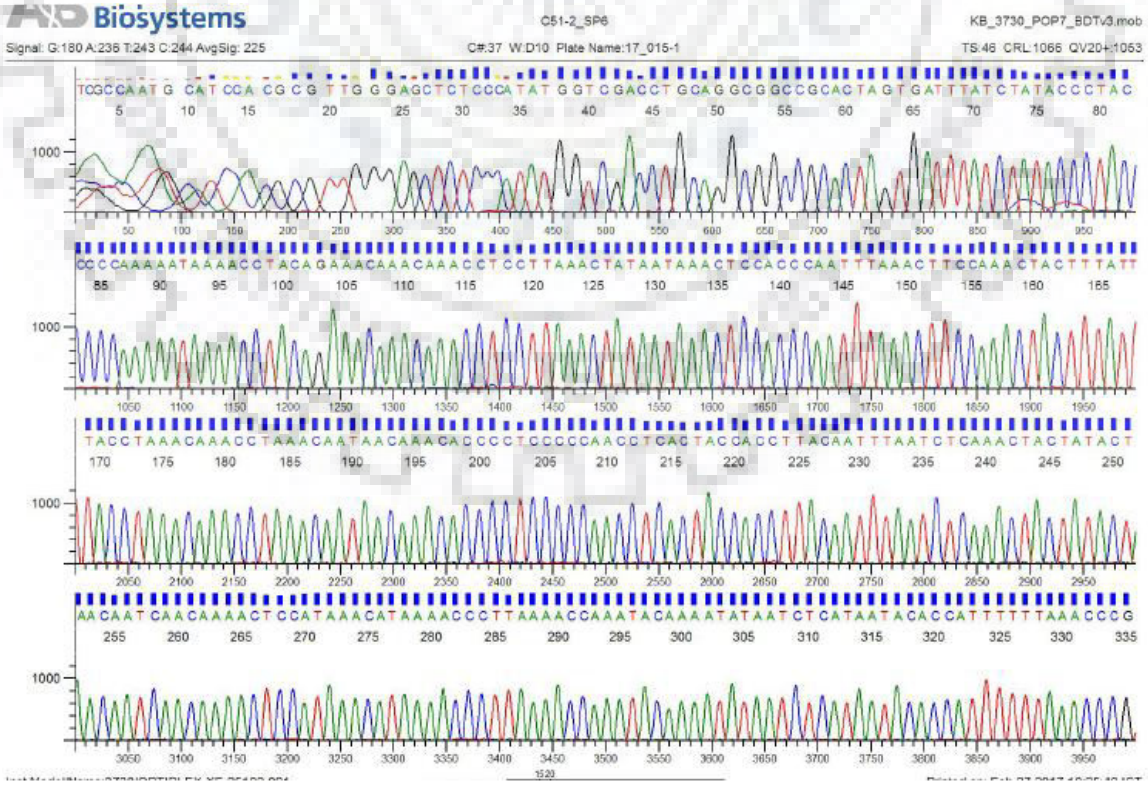
>C2d
AAGGGGTTAGGGAGTTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTTCGACGGGTTTAAGAAACGGCGTATTACGAGATTATATTTTTTATTTGGTTTGGAGGGTTTTAT
GTTTATGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGTGAGGTGGGGGGAGGG
GTGTTTGTTATTGTTTAGGTTTGATTAGGTAAATAAAGTAGTTAGGAAGTTTTAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>C2e
AAGGGGTTAGGGAGTTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGATGTATTTGGAAAATCGGGTTATTTATATTT
AAATATTGCGTTTTTTTAGATTGGTTTAAAAAACGGCGTATTACGAGATTATATTTTATATTTGGTTCGGAGGGTTTTAC
GTTTACGGAATTTGGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGCGAGGTTGGGGGAGGG
GCGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTGGTTCGGAAGTTCGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

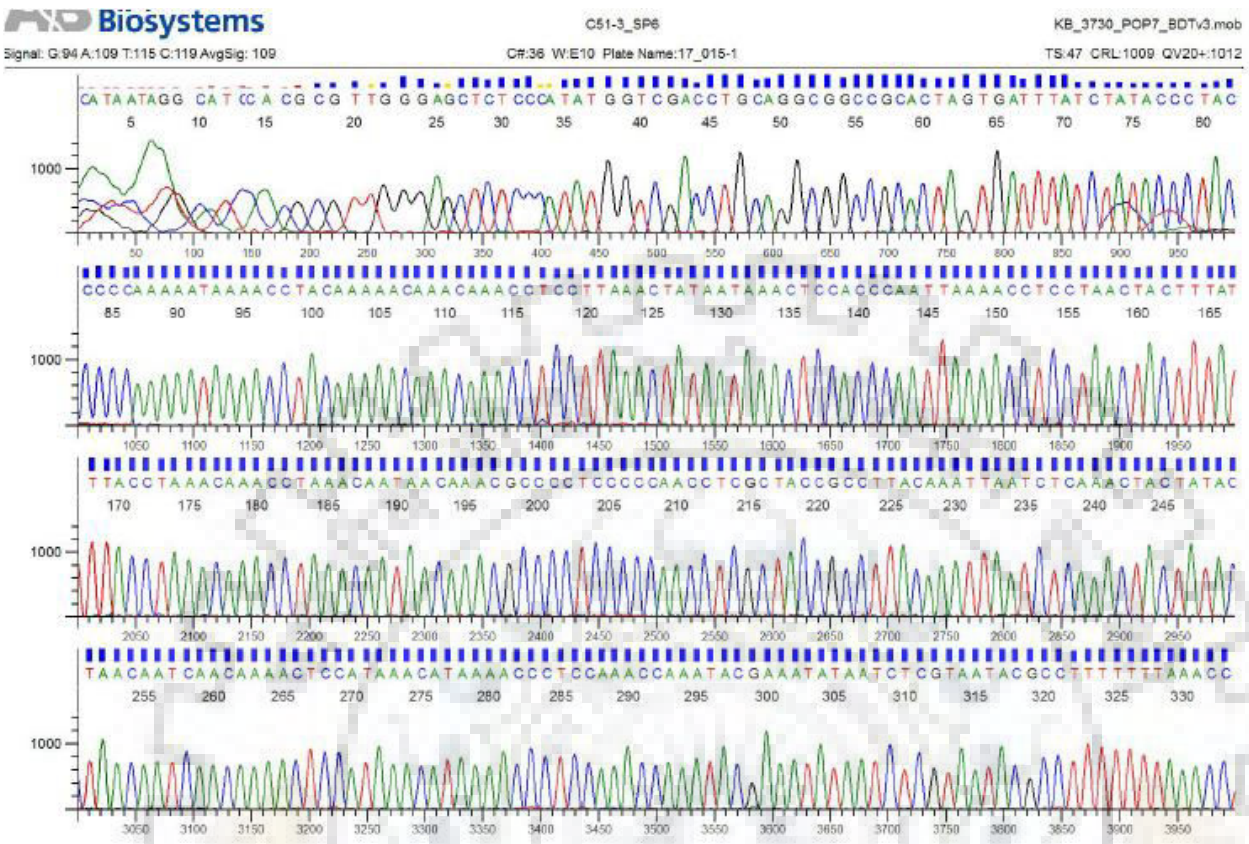
>C2a


Biosystems
C51-1_T7F
KB_3730_POP7_BDTv3.mob
gnal: G:271 A:338 T:395 C:373 AvgSig: 344
C#:40 W:A9 Plate Name:17_015-1
TS:48 CRL:1003 QV20+:1015

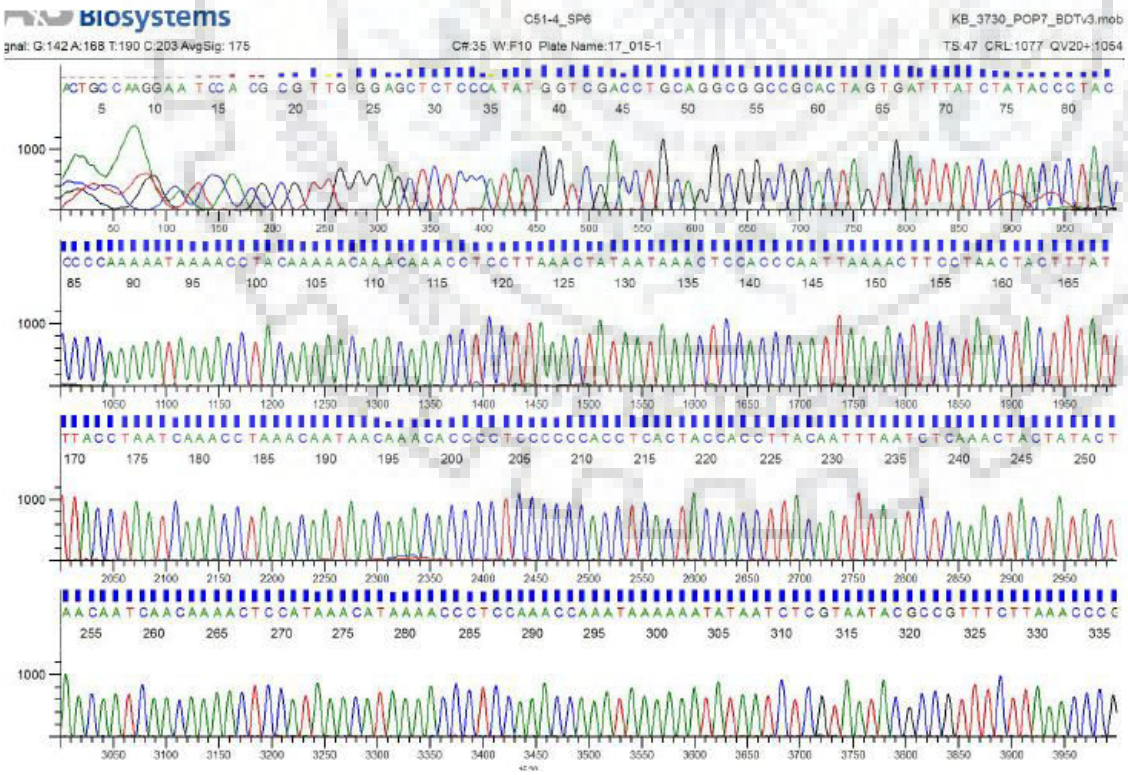
>C2b


Biosystems
C51-2_SP6
KB_3730_POP7_BDTv3.mob
Signal: G:180 A:236 T:243 C:244 AvgSig: 225
C#:37 W:D10 Plate Name:17_015-1
TS:48 CRL:1066 QV20+:1053

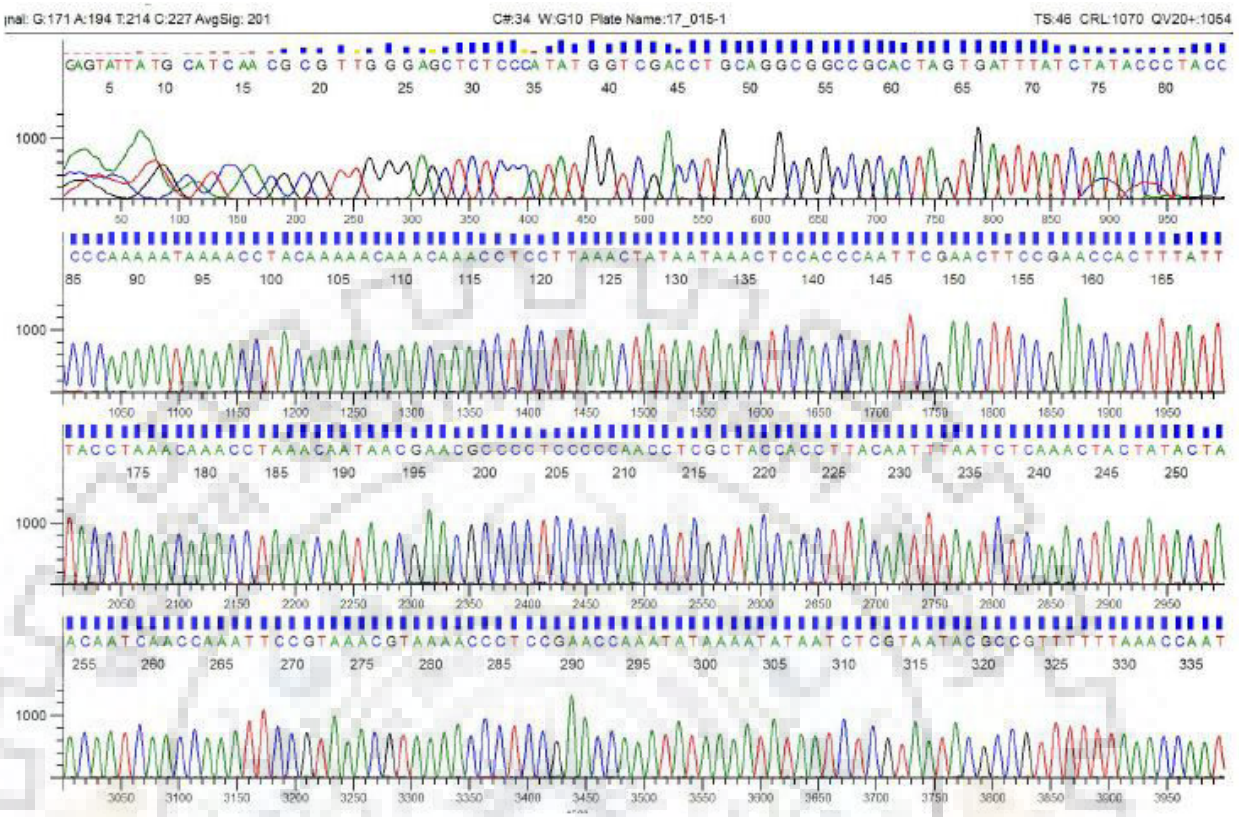
>C2c


Biosystems
C51-3_SP6
KB_3730_POP7_BDTv3.mob
Signal: G:94 A:109 T:115 C:119 AvgSig: 109
C#:36 W:E10 Plate Name:17_015-1
TS:47 CRL:1009 QV20+:1012

>C2d


Biosystems
C51-4_SP6
KB_3730_POP7_BDTv3.mob
C#:35 W:F10 Plate Name:17_015-1
TS:47 CRL:1077 QV20+:1054

>C2e



Paired normal sample 2

>N2a

AAGGGGTTAGGGAGTTTTTTTTTTGAGTAAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTCGGATCGGTTTAAAAAGCGGCGTATTACGAGATTATGTTTCGTATTTGGTTCGTAGGGTTTTA
CGTTTATGGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAAGCGGTAGTGAGGTTGGGGGAGGG
GTATTCGTTATTGTTTAGGTTTGATTAGGTAAATAAAGTAGTTAGGAAGTTTGAAGTGGGTGGAGTTTATTACAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>N2b

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTT
AAATATTGCGTTTTTCGGATCGGTTTAAGAAACGGTATATTACGAGATTATATTTCGTATTTGGTTTGGAGGGTTTTAC
GTTTATAGAGTTTCGTTGATTATTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGCGAGGTTGGGGGAGGG
GTGTTCGTTATTGTTTAGGTTTGTTTATTTAAATAAAGTAGTTAGGAAGTTTAAATTGGGTGGAGTTTATTATAGTTTA
AGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>N2c

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGATAGATGGTATTTGGAAAATTGGGTTATTTTTATT
TTAGTATTGTGTTTTTTTGACGGGTTTAAAAAACGGCGTATTAGGAGATTATATTTTATATTTGGTTCGGAGGGTTTTA
TGTTTACGGAGTTTTATCGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTAATAGTTAGGTTGGGGGAGGG

GCGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTCGGGAGGTTCGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCC

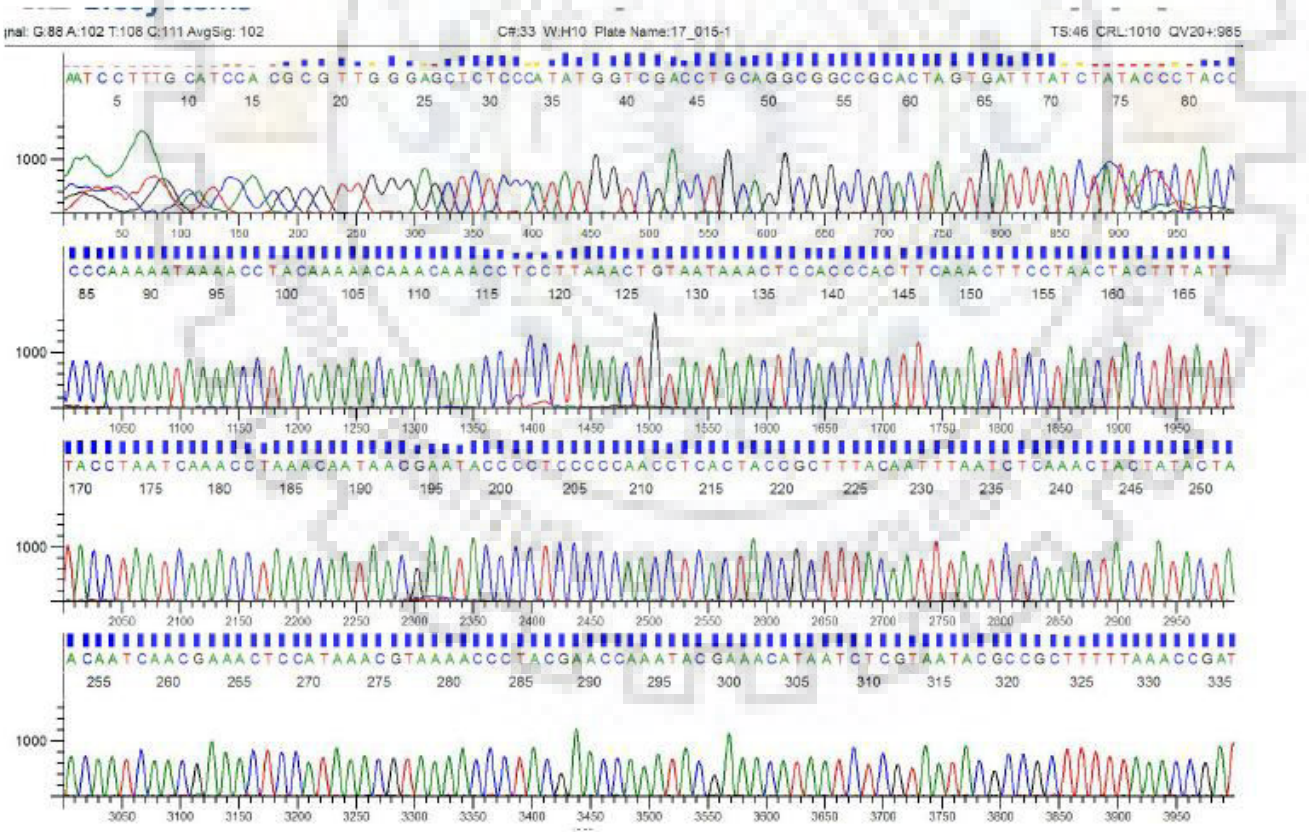
>N2d

AAGGGGTTAGGGAGTTTTTTTTTTGAGTAAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTCGGATCGGTTTAAAAAGCGGCGTATTACGAGATTATGTTTCGTATTTGGTTCGTAGGGTTTTA
CGTTTATGGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAAGCGGTAGTGAGGTTGGGGGAGGG
GTATTCGTTATTGTTTAGGTTTGATTAGGTAAATAAAGTAGTTAGGAAGTTTGAAGTGGGTGGAGTTTATTACAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

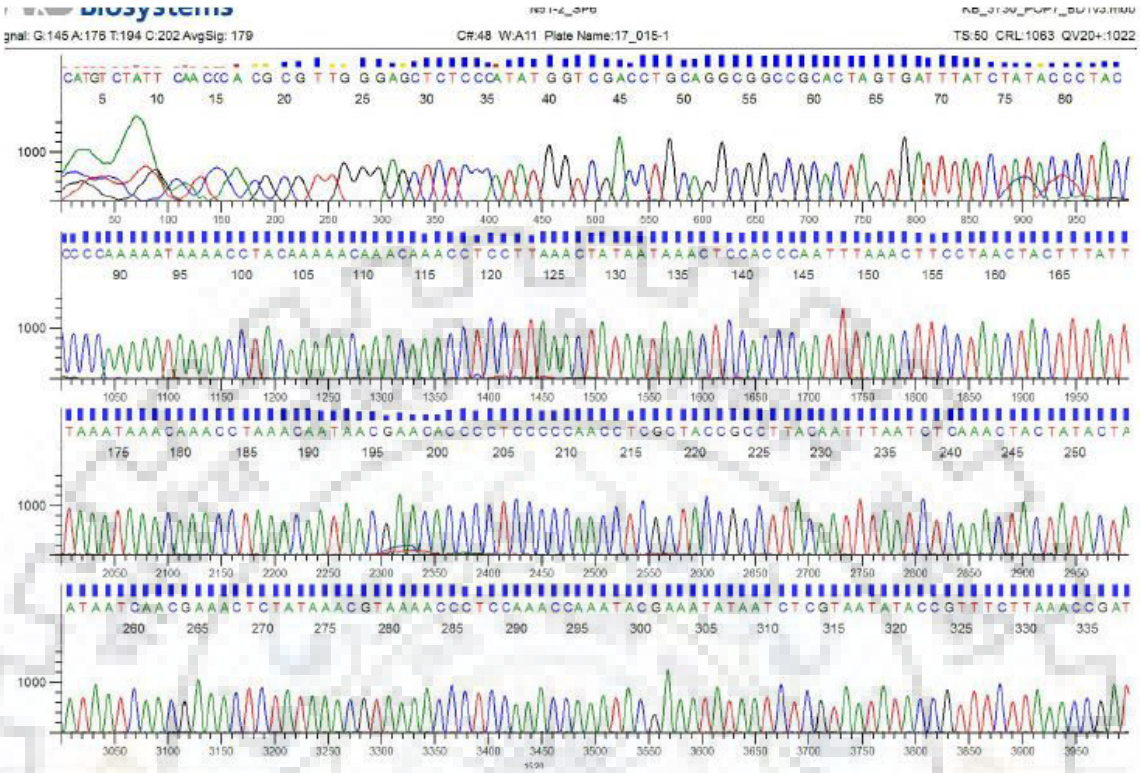
>N2e

AGGGGTTACGGAGTTTTTTTTTTGAGTTAAAGAAAGGGGTGCCGGACGGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTCGACGGGTTTAAAAAACGGCGTATTATGAGATTATATTTTGTATTTGGTTTGGAGGGTTTTAC
GTTTACGGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAAGCGGTAGTGAGGTTGGGGGAGGG
GCGTTCGTTATTGTTTAGGTTTGCTTAGGTAACTAAAGTAGATAGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

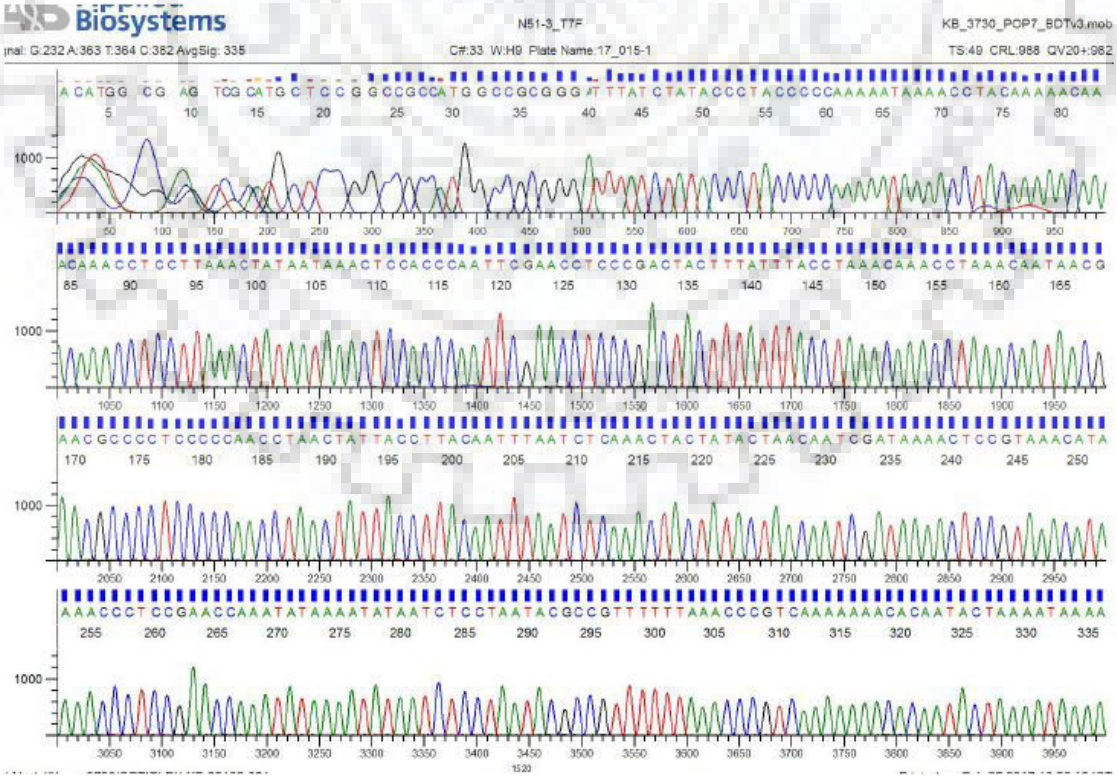
>N2a



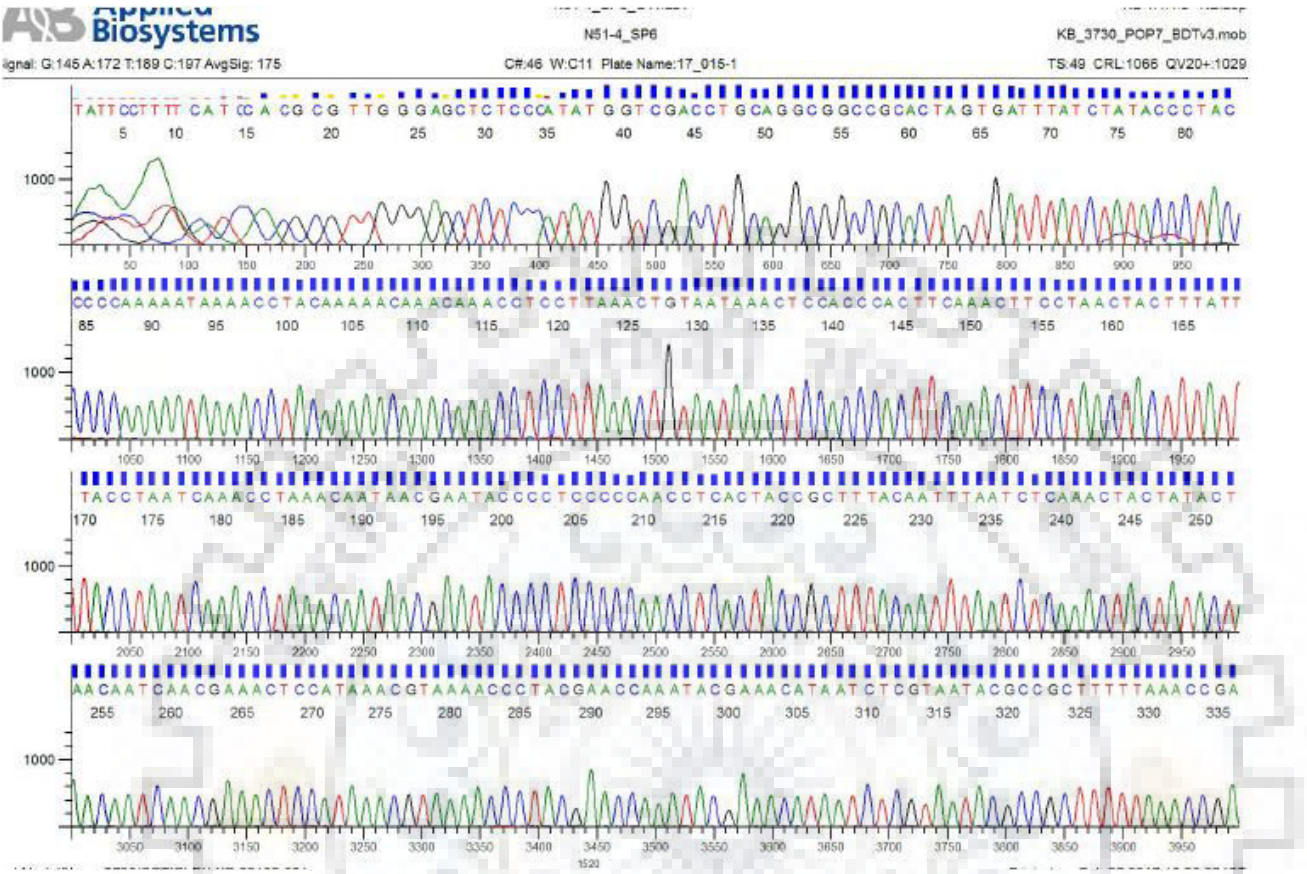
>N2b



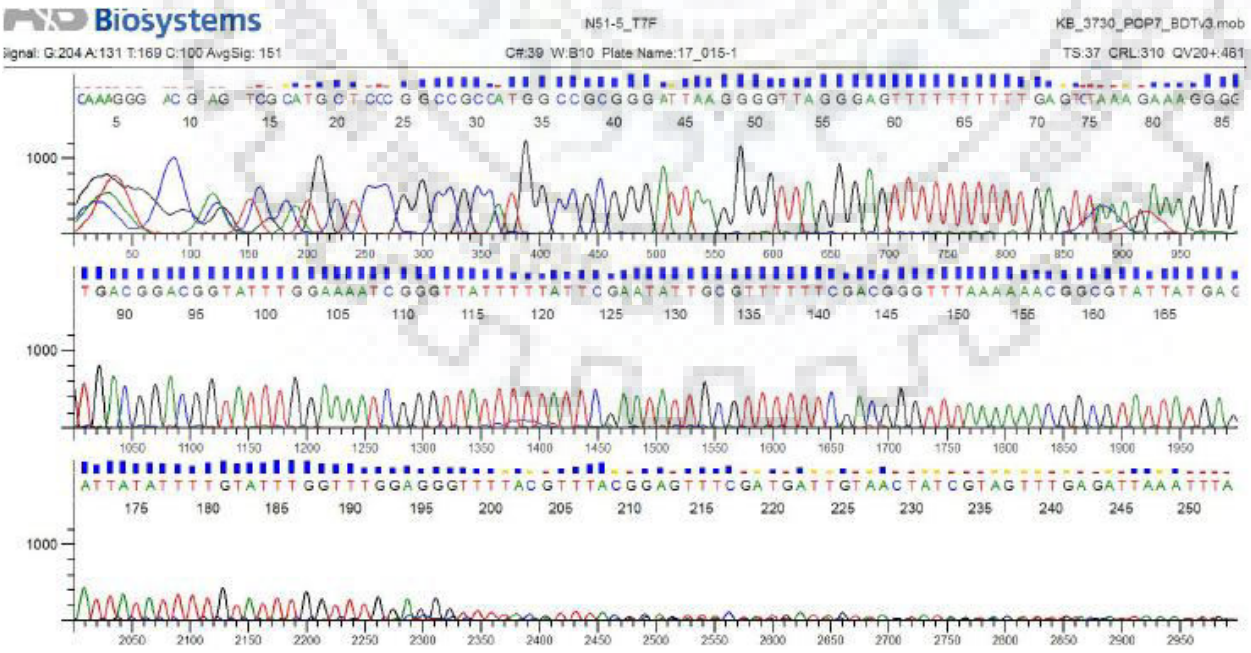
>N2c



>N2d


Applied Biosystems
N51-4_SP6
KB_3730_POP7_BDTv3.mob
Signal: G:145 A:172 T:189 C:197 AvgSig: 175
C#:46 W:C11 Plate Name:17_015-1
TS:49 CRL:1066 QV20+:1029

>N2e


Biosystems
N51-5_T7F
KB_3730_POP7_BDTv3.mob
Signal: G:204 A:131 T:169 C:100 AvgSig: 151
C#:39 W:B10 Plate Name:17_015-1
TS:37 CRL:310 QV20+:461

Paired cancer sample 3

>C3a

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTCTATTC
GAATATTGATTTTTTTAGATCGGTTTAAAAAATGGTGTATTATGAGATTATATTTTATATTTGGTTTAGAGGGTTTTAT
GTTTATGGAATTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGTGAGGTTGGGGGAGGGG
TGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTTGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTAA
GGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACTA

>C3b

AAGGGGTTAGGGAGTTTTTTTTTTTAGTTAAAGAAAGGGGTGATAGATGGTATTTGGAAAAGCGGGTTATTTTTATT
TTAATATAGTGTTTTTTAGATGGGTTTAAAAAATGGTATATTAGGAGATTATATTTCGTATTTGGTTCGGAGGGTTTTA
CGTTTATGGAGTTTTATTGAATGTTAGTATAGTAGTTTGAGATTAAACGGTAAGGTGGTAGTGAGGTTGGGGGAGGG
GGAGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTAGGAAGTTTGAATTGGGTGGAGTTTATTATAGTT
CAAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCAC

>C3c

AAGGGGTTAGGGAGTTTTTTTTTTAGTTAAAGAAAGGGGTGATAGACGGTATTTGGAAAATTGGGTTATTTTTATTTT
AATATTGTATAATTTTAATGGGTTTAATAAATGGTATATTAGGAGATTATATTTTATATATGGTTTGGAGGGTTTTACG
TTTACGGAGTTTTATTTGTTGTTAGTATAGTAGTTTGAGATAAAGTTGTAAGGCGGTAGTAAGGTTGGGGGAGGGGC
GTTCGTTATTGTTTAGGTTTGAGTTGGTAAATAAAGTGGTTAGGAAGTTTAAATTGGGTGGGGTTTATTAAAGATTAA
GGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACTA

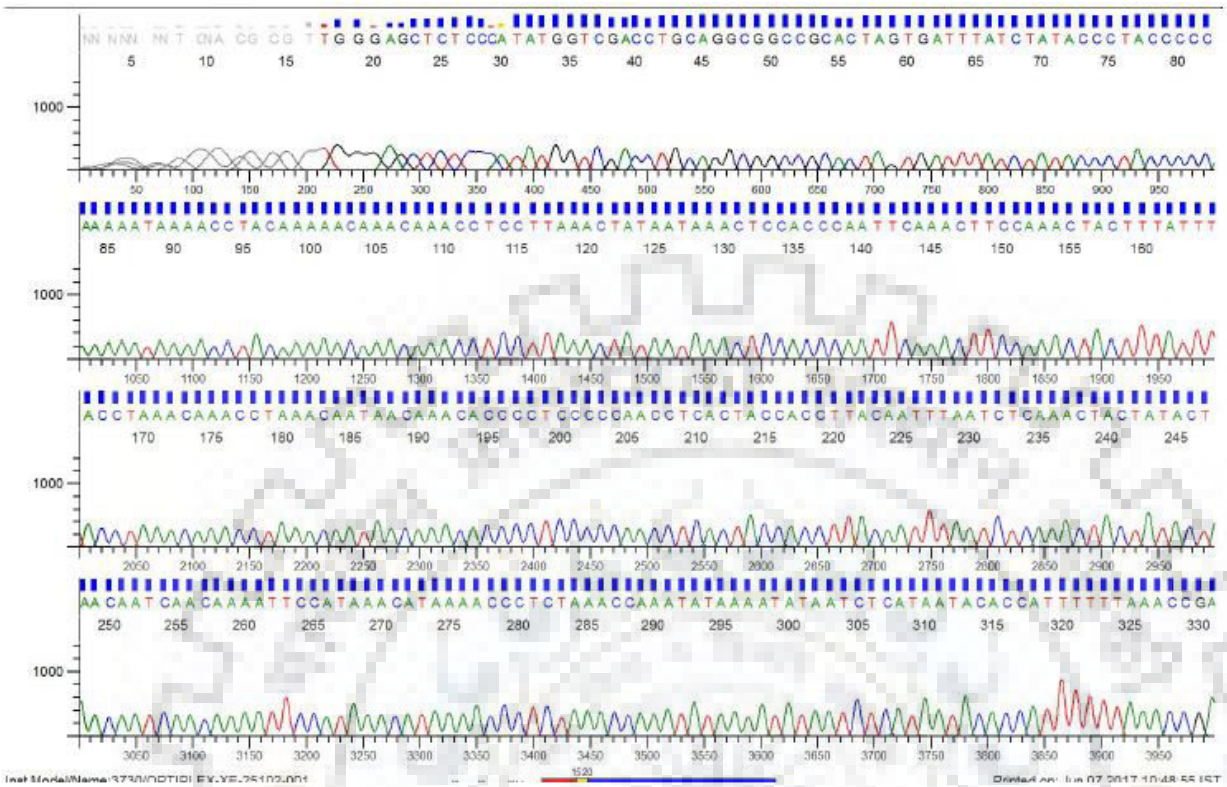
>C3d

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTCTATTC
GAATATTGATTTTTTTAGATCGGTTTAAAAAATGGTGTATTATGAGATTATATTTTATATTTGGTTTAGAGGGTTTTAT
GTTTATGGAATTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGTGAGGTTGGGGGAGGGG
TGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTTGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTAA
GGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

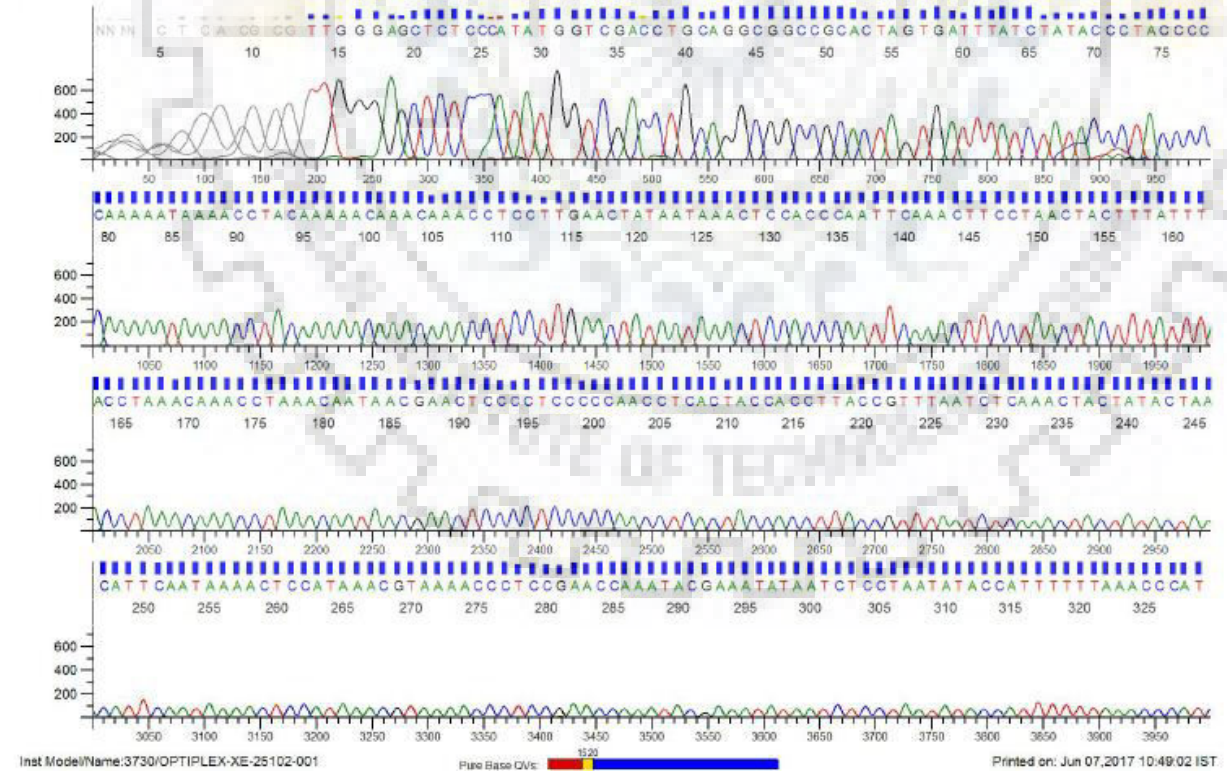
>C3e

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTCTATTC
GAATATTGATTTTTTTAGATCGGTTTAAAAAATGGTGTATTATGAGATTATATTTTATATTTGGTTTAGAGGGTTTTAT
GTTTATGGAATTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGTGAGGTTGGGGGAGGGG
TGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTTGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTAA
GGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

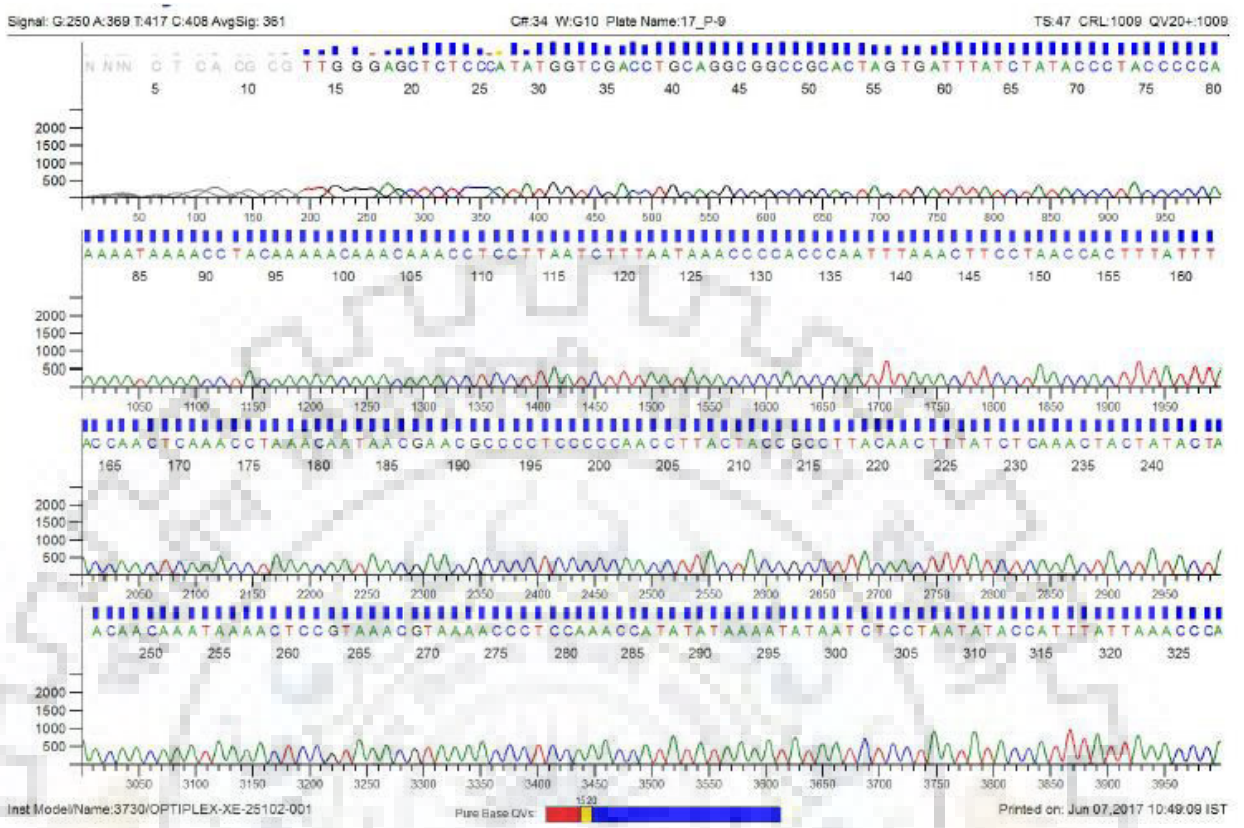
>C3a



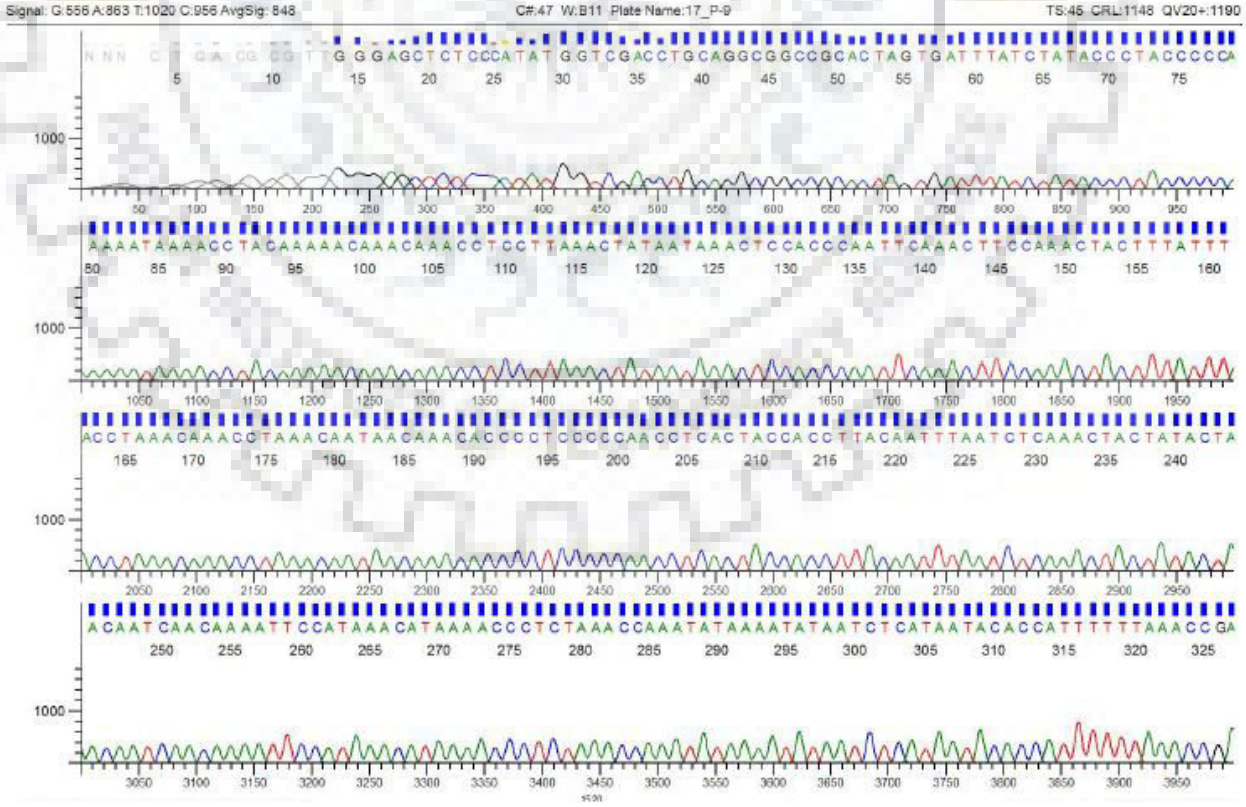
>C3b


Inst Model/Name:3730/OPTIPLEX-XE-25102-001
Pure Base QVs:
Printed on: Jun 07,2017 10:49:02 IST

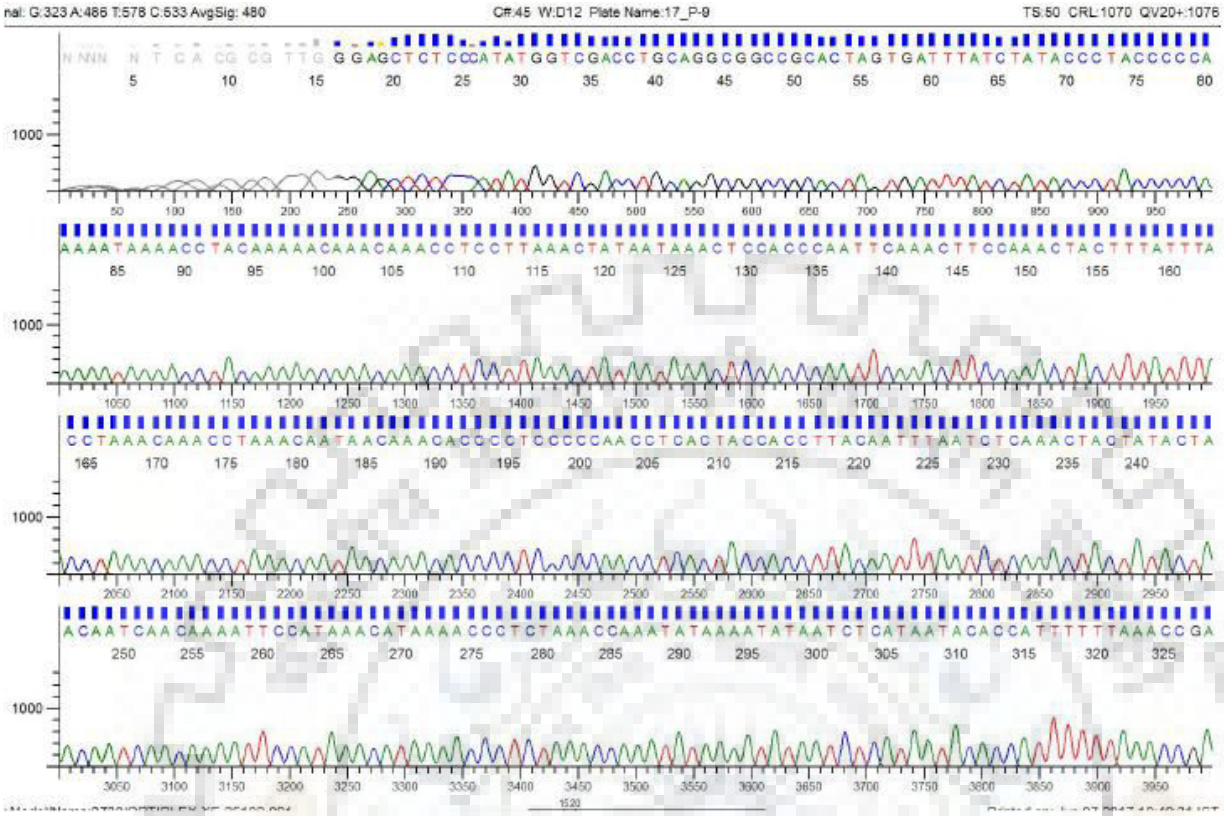
>C3c


Signal: G:250 A:369 T:417 C:408 AvgSig: 361
C#:34 W:G10 Plate Name:17_P-9
TS:47 CRL:1009 QV20+:1009
Inst Model/Name:3730/OPTIPLEX-XE-25102-001
Pure Base QVs:
Printed on: Jun 07,2017 10:49:09 IST

>C3d


Signal: G:556 A:863 T:1020 C:956 AvgSig: 848
C#:47 W:B11 Plate Name:17_P-9
TS:45 CRL:1148 QV20+:1190

>C3e



Paired normal sample 3

>N3a

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTAGATTGGTTTAAAAAACGGCGTATTACGAGATTATATTTTATATTTGGTTCGGAGGGTTTTAC
GTTTTCGGAATTTTGTTGATTGCTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGCGAGGTTGGGGGAGGGG
CGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTGGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTA
AGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>N3b

AAGGGGTTAGGGAGTTTTTTTTTTTAGTTAAAGAAAGGGGTGACGGATATATTTGGAAAATCGGGTTATTTTTATTCG
ATTATTGCGTTTTTTCGATTGGTTTAAAAAAAGGTATATTATGAGATTATATTTCGTATTTGGTTTAGAGGGTTTTACG
TTTAGGGAGTTTCGTTGATTGATAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGTTAGGTTGGGGGAGGGG
CGTTTATTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTAGGATGTTCGAATTGGGTGGAGTTTATTATAGTTTAA
GGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

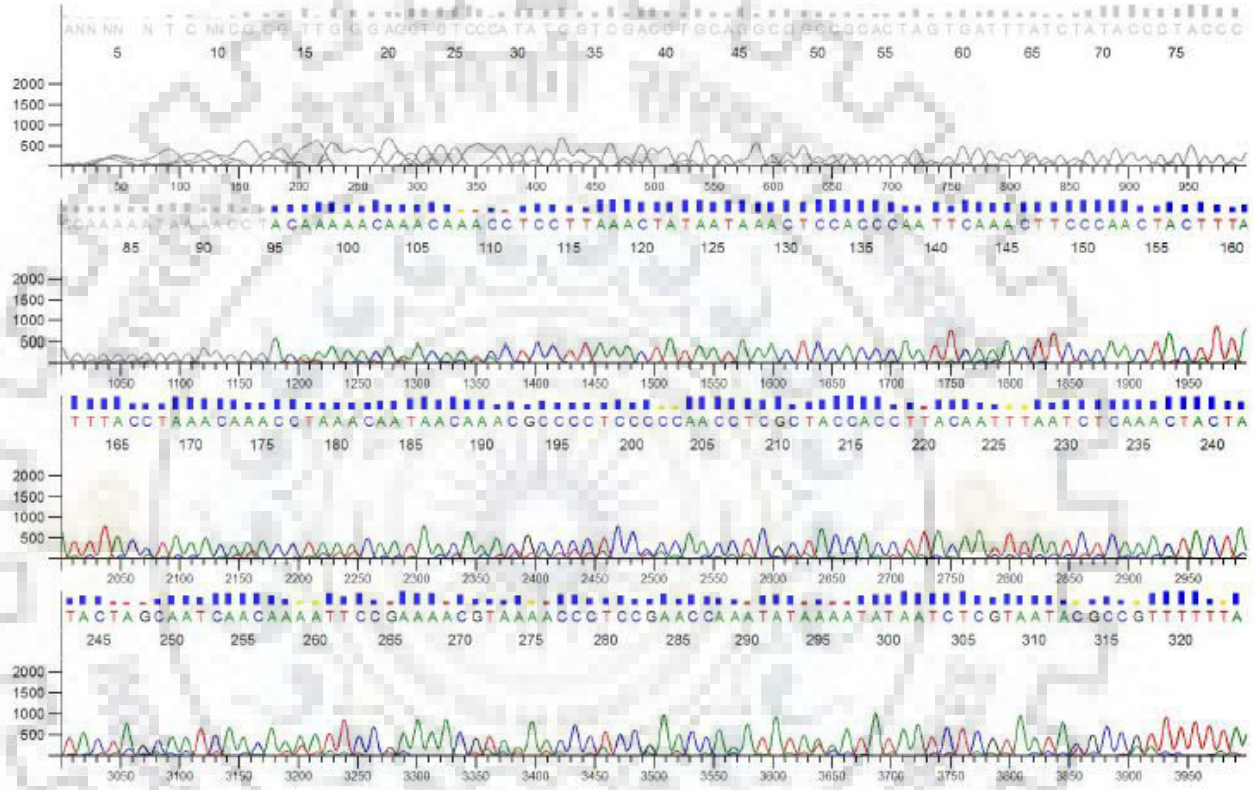
>N3c

AAGGGGTTAGGGAGTTTTTTTTTTTAGTTAAAGAAAGGGGTGACGGATGGTATTTGGAAAATTGGGTCGTTTTTATT
CGAATATTGCGTTTTTTTTGACGGGTTTAAAAAACGGCGTATTATAAGATTATATTTTGTATTTAGTTCGGAGGGTTTTA
TGTTTATGGAGTTTCGTTGATTGTTAGTATAGTAGATTGAGATTAAATTGTAAGGTGGTAGCGAGGTTGGGGGAGG
GGCGTTTATTATTGTTTAGGTGTGTTTAGGTAAATAAAGTAGTTGGGAAGTTTGAATTGGGTGGAGTTTATTATAGTT
TAAGGAGGTTTGCGTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

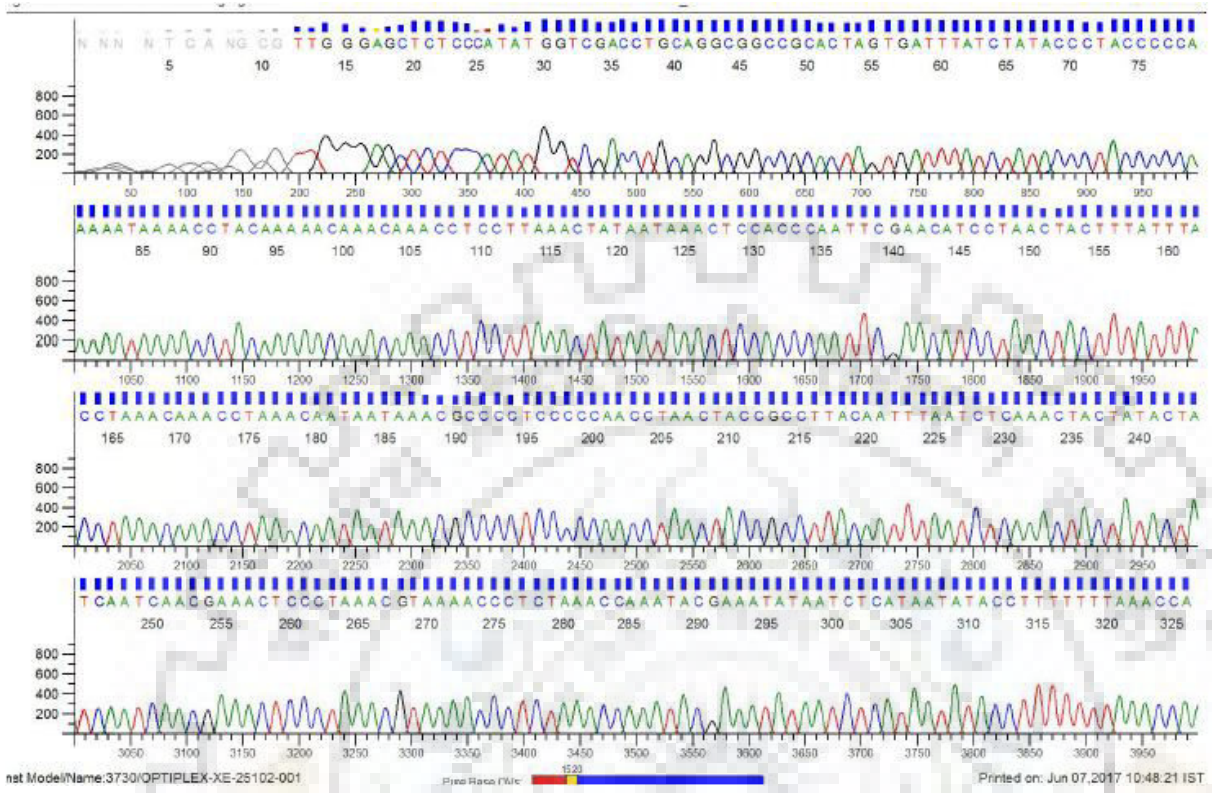
>N3d

AAGGGGTTAGGGAGTTTTTTTTTTTAGTTAAAGAAAGGGGTGACGGATGGTATTTGGAAAATTGGGTCGTTTTTATT
CGAATATTGCGTTTTTTTGACGGGTTTAAAAAACGGCGTATTATAAGATTATATTTTGTATTTAGTTCGGAGGGTTTTA
TGTTTATGGAGTTTCGTTGATTGTTAGTATAGTAGATTGAGATTAAATTGTAAGGTGGTAGCGAGGTTGGGGGAGG
GGCGTTTATTATTGTTTAGGTGTGTTTAGGTAAATAAAGTAGTTGGGAAGTTTGAATTGGGTGGAGTTTATTATAGTT
TAAGGAGGTTTGCGTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCAC

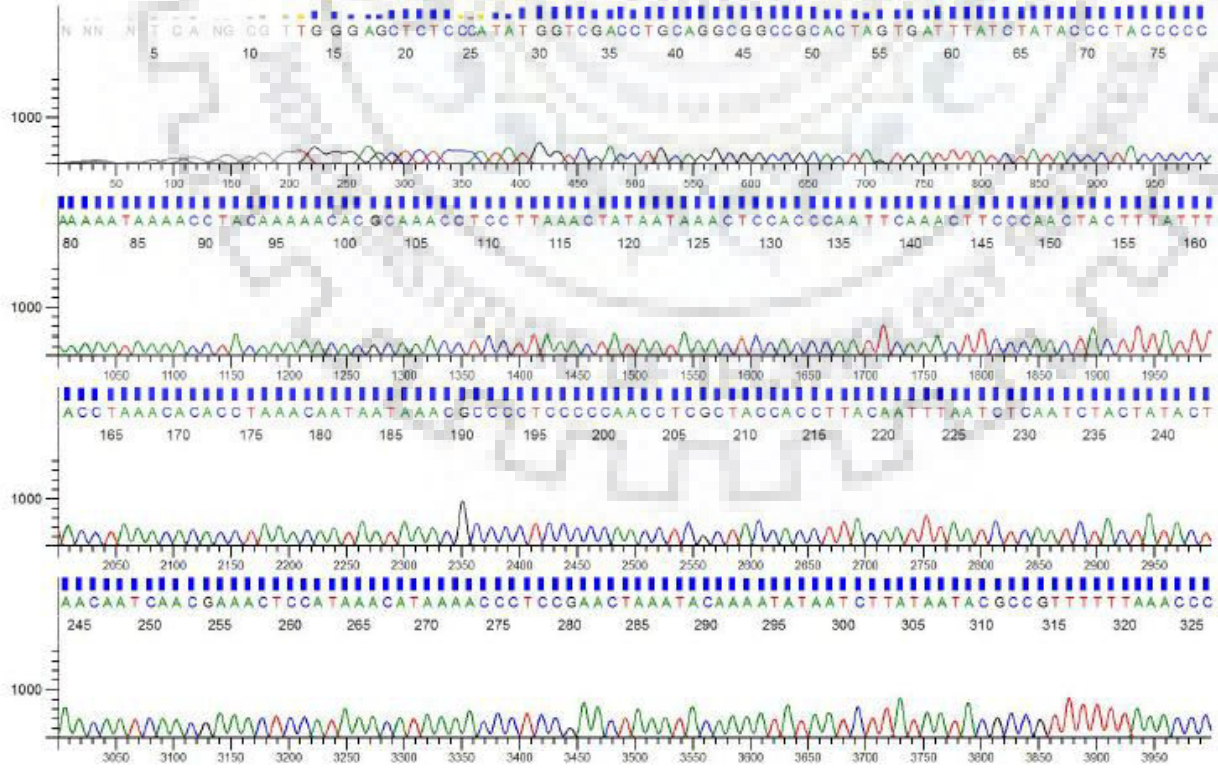
>N3a



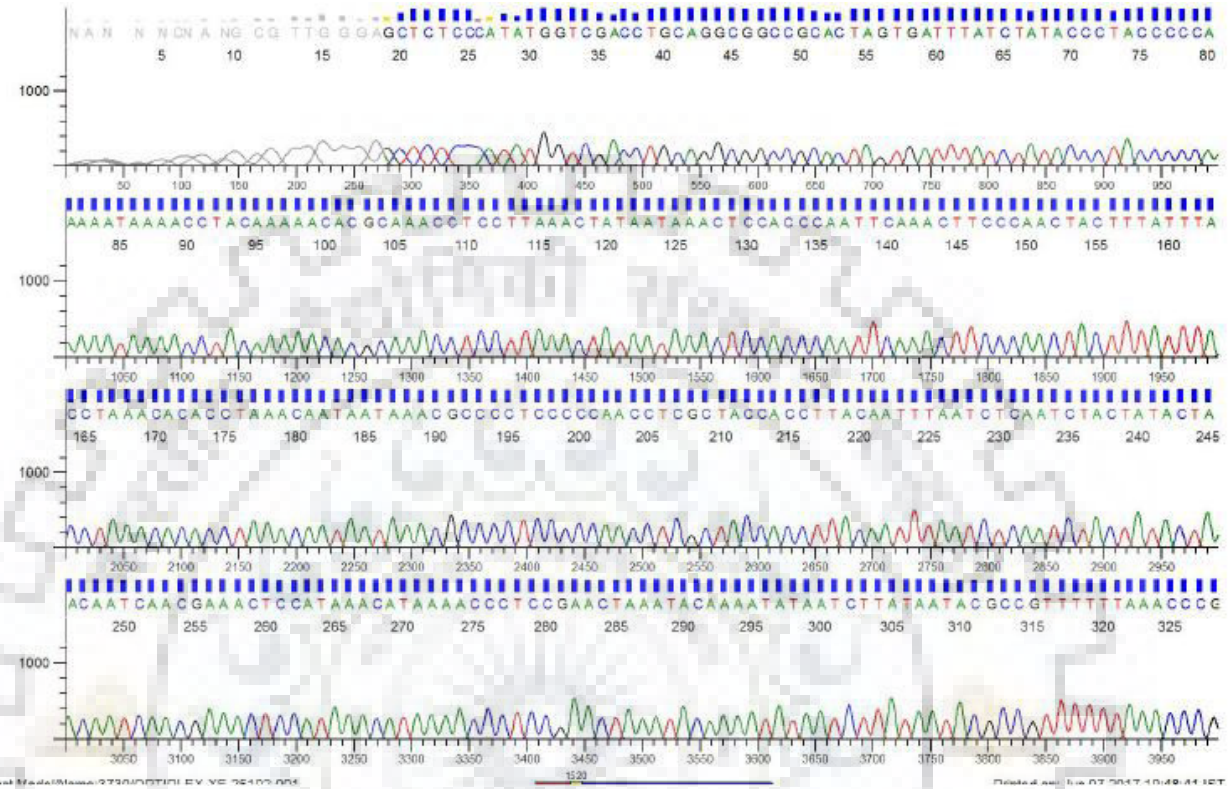
>N3b



>N3c



>N3d



Paired cancer sample 4

>C4a

AAGGGGTTAGGGAGTTTTTTTTTTTAGTTAAAGAAAGGGGTGATAGACGGTATTTGGAAAATTAGGTTATTTTTATTC
GAATATTGTATTTTTTCGATGGGTTTAAAAAACGGCGTATTAGGAGATTATATTTCGTATTTGGTTTAAGGGGTTTTAC
GTTTACGGAGTTTTATTGATTGTTAGTATAGTAGTTTGAGGTTAAATTGTAAGGCGGTAGCGAGGTTGGGGGAGGGG
GCGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTCGGGAAGTTCGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGGTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCAC

>C4b

AAGGGGTTAGGGAGTTTTTTTTTTTAGTTAAAGAAAGGGGTGATAGATTGTATTTGGAAAATTGGGTTATTTTTATTT
TAATATTGGGTTTTTTTAATGGGTTTAATAAACGGTATATCAGGAGATTATATTTCGTATCTGGTTTGGAGGGTTTTAT
ATTTACGGAGTTTTGTTTATTGTTAGTTTAGTAGTTTGAGATTAAATTGTAAGGTGGTAGTGAGGTTGGGGGAGGGG
TGTTTATTATTGTTTAGTTTTGAGTAGGTAAATAAAGTGGTTAGGAAGTTTGAATTAGGTGGAGTTTATTATAGTTTA
AGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTGGGGGTAGGGTATAGATAAATCACTA

>C4c

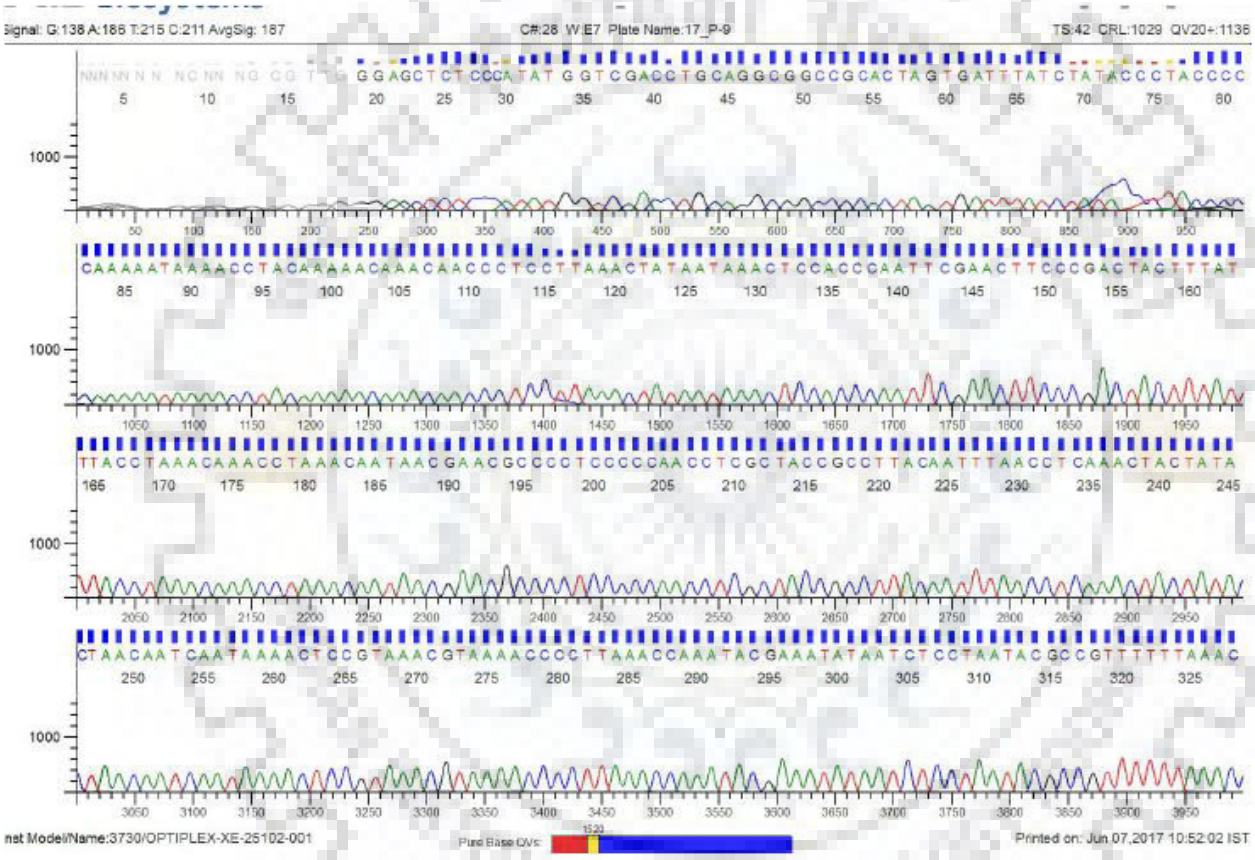
AAGGGGTTAGGGAGTTTTTTTTTTTTAGTTAAAGAAAGGGGTGACGGACGGTATTTGGAAAATCGGGTTATTTTTATT
AGAATATTGCGTTTTTTCGATAGGTTTAAAAAACGGTGTATTATTAGATTATATTTTATATTTGGTTTGGAGGGTTTTA
TGTTTATGGAATTTTTTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGCGAGGTTGGGGGAGGG

GCGTTCGTTATTGTTTAGGTTTGATTAGGTAAATAAAGTAGTTGGGAAGTTGGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCAC

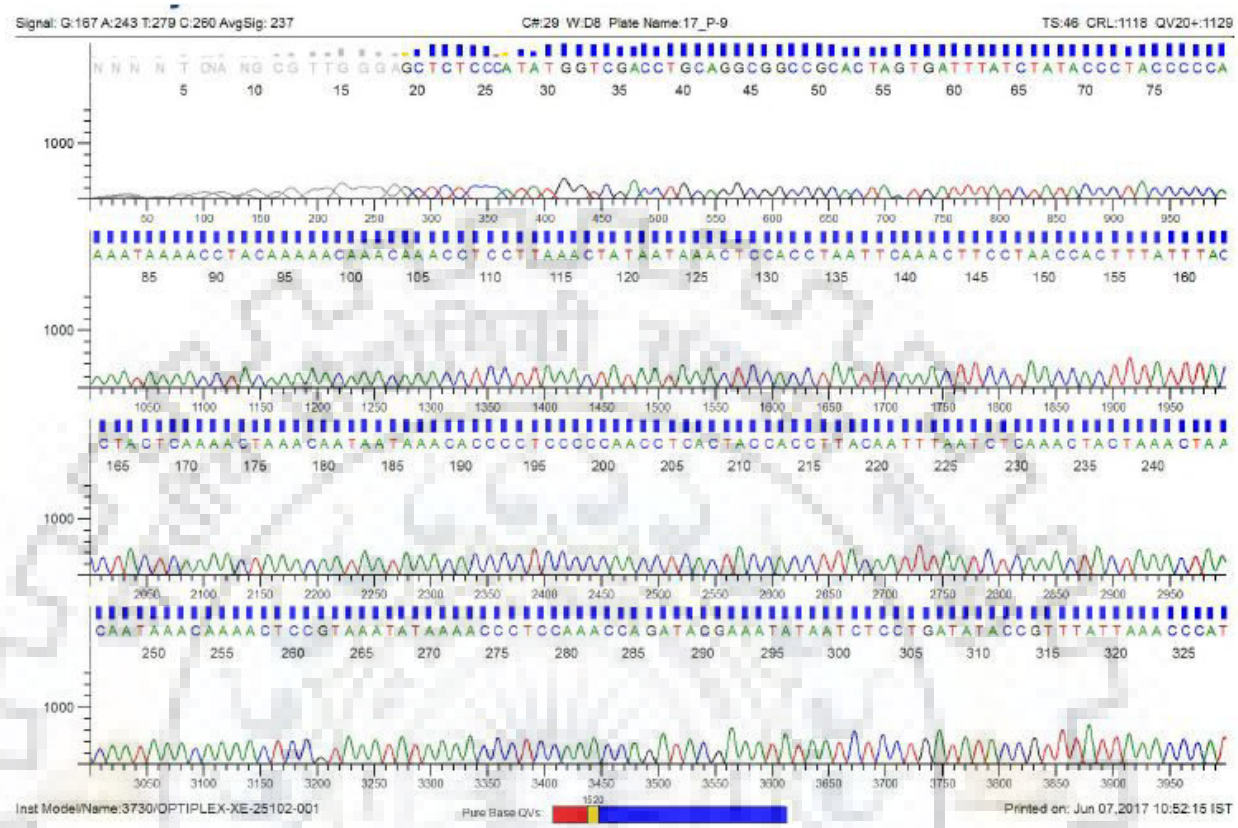
>C4d

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGATAGTATTTGGAAAATCGGGTTATTTTTATT
CGAATATTGCGTTTTTTCGGCGGGTTTAAAAAAACGGCGTATTACGAGATTATATTCGGTATTTGGTTCGGAGGGTTT
TACGTTTACGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGCGAGGTTGGGGGAG
GGGTGTTTGTTATTGTTTAGGTTTGATTAGGTAAATAAAGTAGTTAGGAAGTTTGGATTGGGTGGAGCTTATTATAGT
TTAAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCA

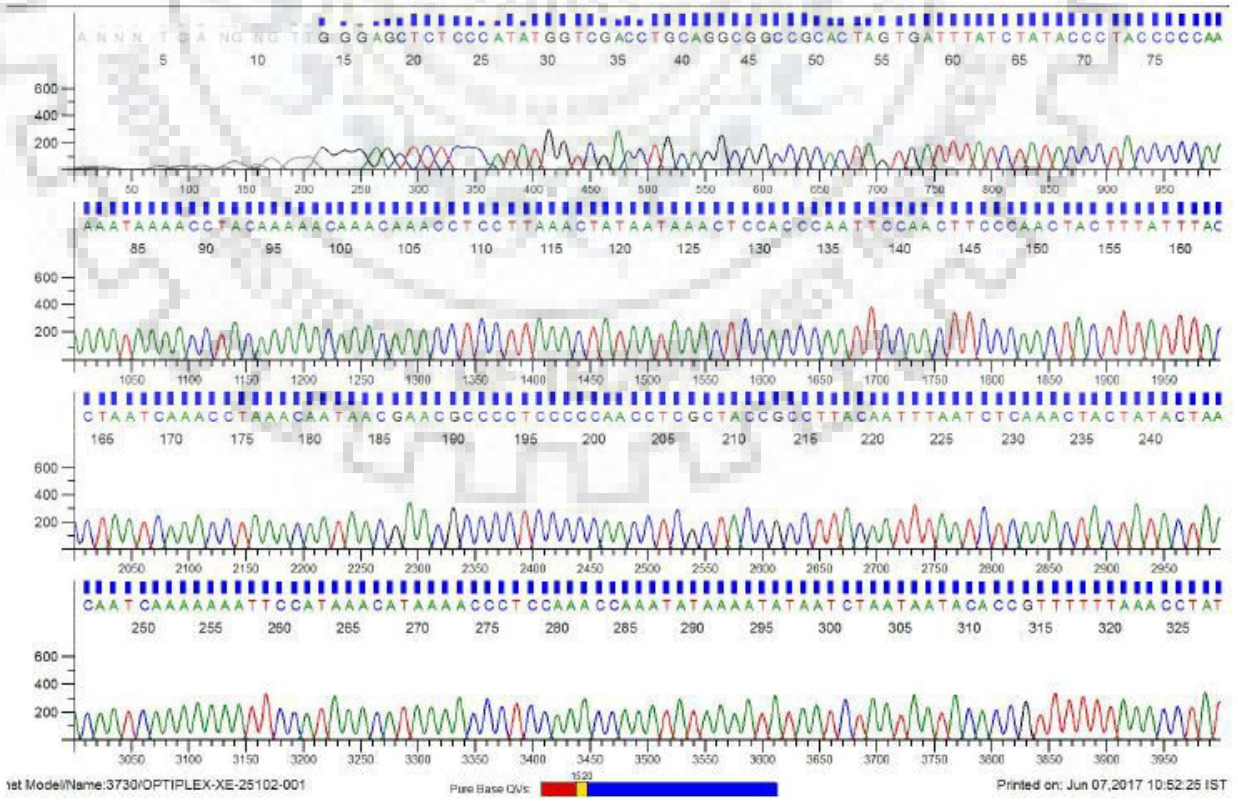
>C4a



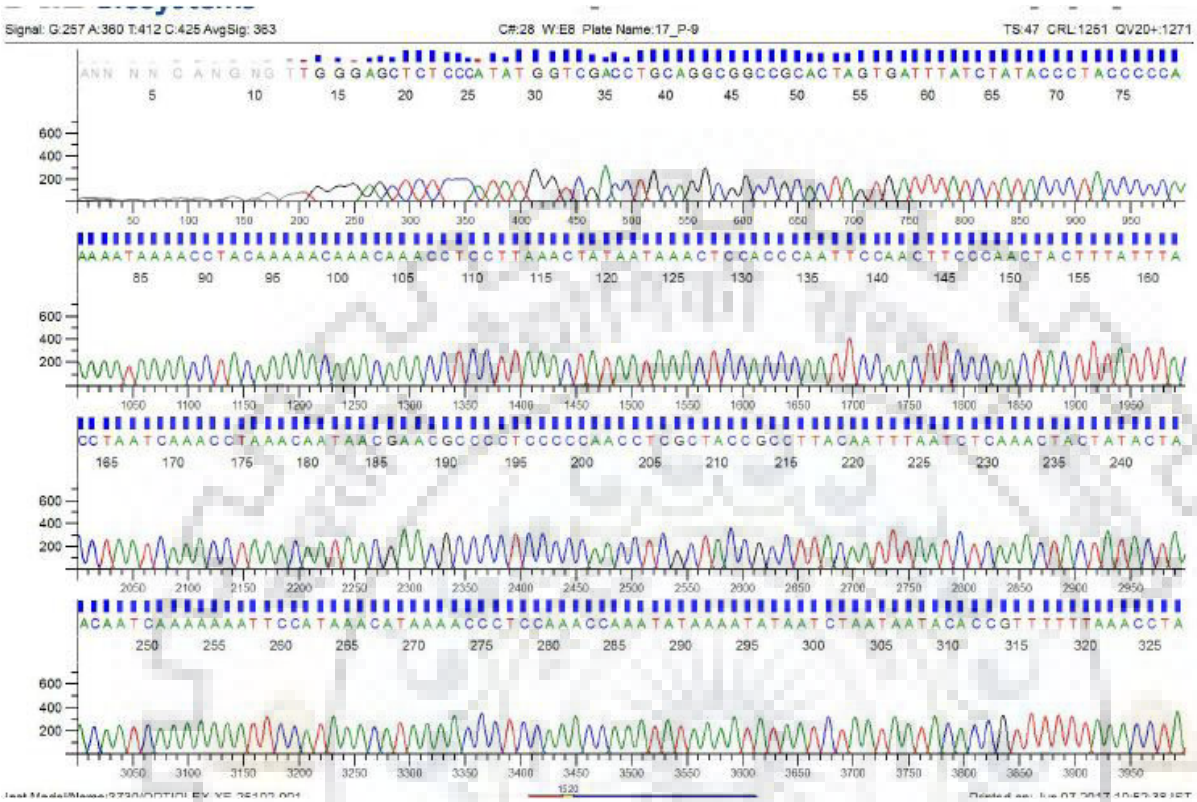
>C4b


Signal: G:167 A:243 T:279 C:260 AvgSig: 237
C#:29 W:D8 Plate Name:17_P-9
TS:46 CRL:1118 QV20+:1129
Inst Model/Name:3730/OPTIPLEX-XE-25102-001
Pure Base QVs:
Printed on: Jun 07,2017 10:52:15 IST

>C4c


Inst Model/Name:3730/OPTIPLEX-XE-25102-001
Pure Base QVs:
Printed on: Jun 07,2017 10:52:25 IST

>C4d



Paired Normal sample 4

>N4a

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGATTTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGTGTTTTTCGGATCGGTTTAAAAAACGGGAGATTATATTTTCGAGATTATATTCGGTATTTGGCGTGGAGG
GTTTTACGTTTACGGAGTTTCGTTGGTTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGCGAGGTTGGG
GGAGGGGCGTTCGTTATTGTTTAGGTTTGATTAGGTAAATAAAGTAGTTGGGAAGTTCGAATTGGGTGGAGTTTATT
ATAGTTTAAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAA

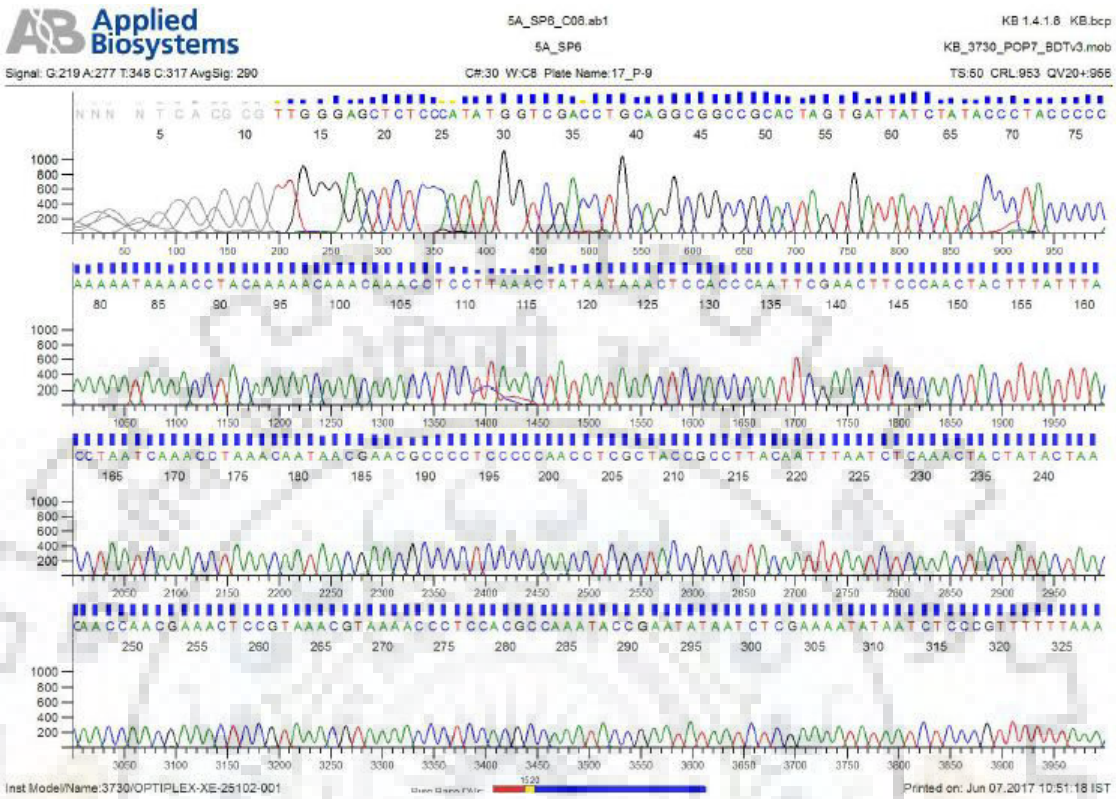
>N4b

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGACGGTATTTGGAAAATCGGGTTATTTTTATT
CGAATATTGCGTATTTTCGACGGGTTTAAAAAACGGCGAATTACGAGATTATATTTTATATTTGGCTTGGAGGGTTTT
ACGTTTACGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATCAAATTGTAAGGTGGTAGCGAGGTTGGGGGAGG
GGTATTCGTTATTGTTTAGGTTTGTTTAGATAAATAAAGTAGTTTGGAAGTTTGAATTGGGTGAATTGGGCGGAGTTT
ATTATAGTATAAGGAGGTCTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATA

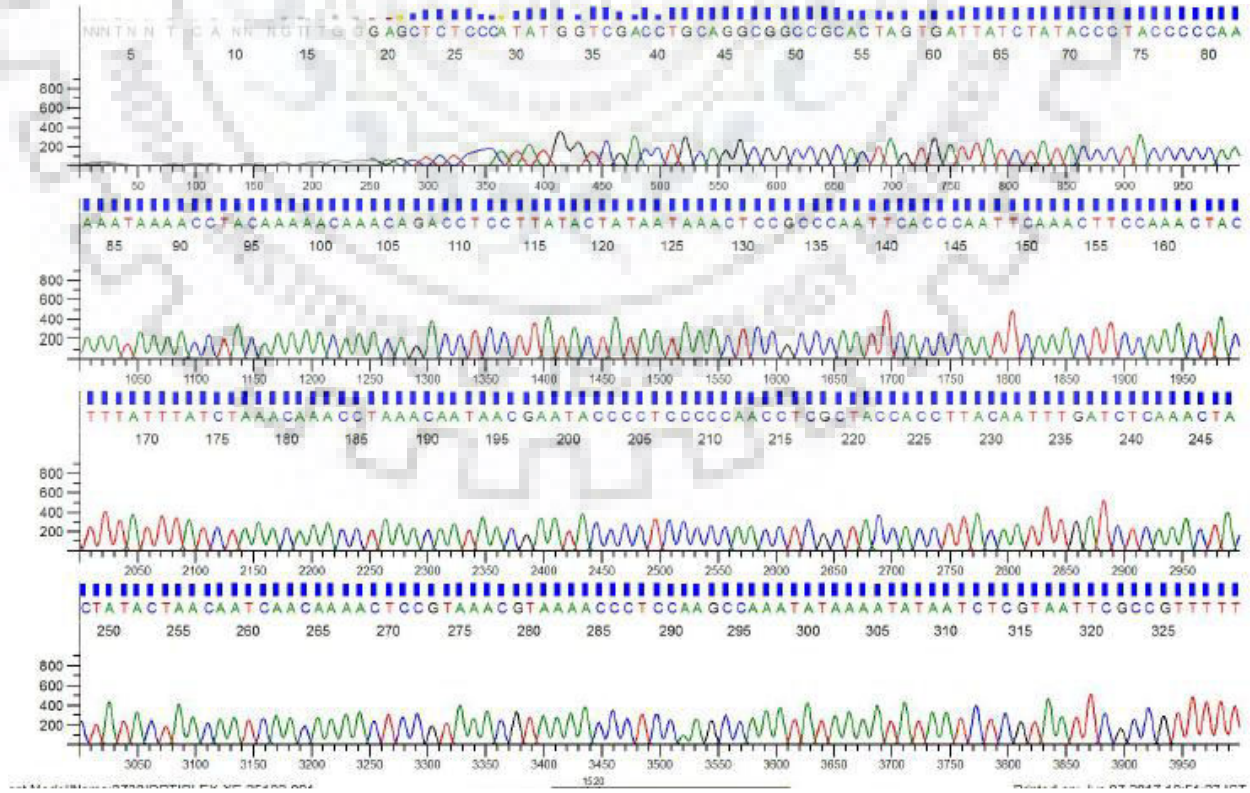
>N4c

AAGGGGTTAGGGAGTTTTTTTTTTTTAGTTAAGGGAAGTTATGATAGTGGTATTTGGAAAAATGGAATATTTGCGTTT
AAATATTGCGTTTTTTTAAAGGTTTTAGTAAATAGTATATTAGGAGATTATATTTCGTGTTTGGTTTGGTGGGTTTTAT
TTTTATTAAGTTTTGTTTATTGTTAGTGTAGTAGTTTGAGATTAGTTTGTAAGGTAGTAGTTTGGTAGGGGGAGGGGT
ATTTATTATTGCTGAGATTTGAGTAGGTAAATAAAGTAGTTTAGGAAGTTCAAATTGGGTAGAGTTTATCGTAGTTTA
GTTAGGTTTGTTGTTTTTGTAGATTTTATTTTGGGGGTAGGGTATAGATAAATCACTA

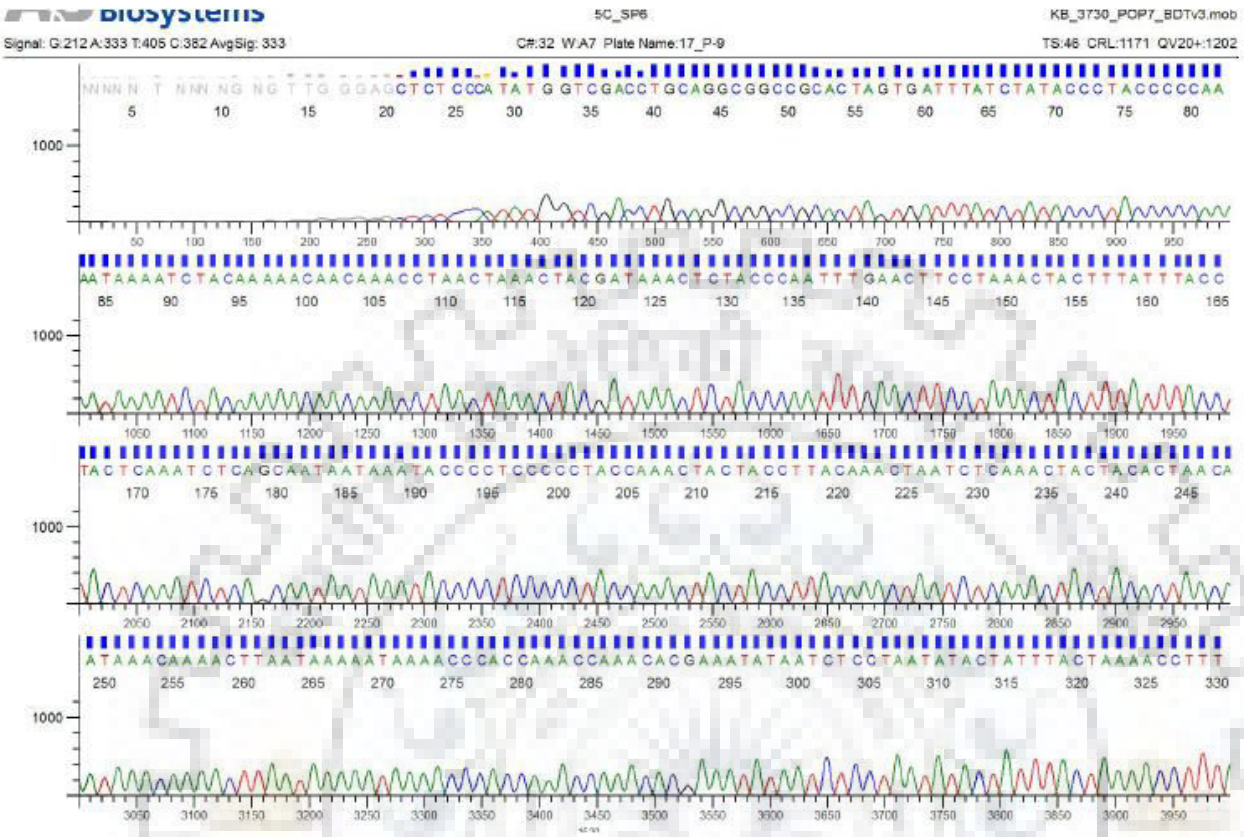
>N4a


Applied Biosystems
5A_SP6_C08.ab1
5A_SP6
KB 1.4.1.6 KB.bcp
KB_3730_POP7_BDTv3.mob
Signal: G:219 A:277 T:348 C:317 AvgSig: 290
C#:30 W:C8 Plate Name:17_P-9
TS:60 CRL:953 QV20+:966
Inst Model/Name:3730/OPTIPLEX-XE-25102-001
Printed on: Jun 07,2017 10:51:18 IST

>N4b



>N4c



Paired cancer sample 8

>C8a

AAGGGGTTAGGGAGTTTTTTTTTTAAGTTAAAGAACGGAGTGACGGACGGTATTTGGAAAATCGGGTTATTTTTATT
CGAATATTGCGTTTTTTCGGCGGGTTTAAAAAAACGGCGTATTACGAGATTATATTTCGTATTTGGTTCGGAGGGTTT
TATATTTACGGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGCGAGGTTGGGGGAG
GGGTATTTGTTATTGTTTAGGTTTGATTAGGTAAATAAAGTAGTCGGGAAGTTCGAATTGGGTGGAGTTTATTATAGT
TTAAGGAGGTTTGTTTGTATTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>C8b

AAGGGGTTAGGGAGTTTTTTTTTTAAGTTAAAGAACGGAGTGACGGACGGTATTTGGAAAATCGGGTTATTTTTATT
CGAATATTGCGTTTTTTCGGCGGGTTTAAAAAAACGGCGTATTACGAGATTATATTTCGTATTTGGTTCGGAGGGTTT
TATATTTACGGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGCGAGGTTGGGGGAG
GGGTATTTGTTATTGTTTAGGTTTGATTAGGTAAATAAAGTAGTCGGGAAGTTCGAATTGGGTGGAGTTTATTATAGT
TTAAGGAGGTTTGTTTGTATTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>C8c

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGATGGTATTTGGAAAATCGGGTTATTTTTATT
TAAATATTGCGTTTTTTCGACGGGTTTAAAAAACGGCGGATTACGAGATTATATTTCGTATTTGGTTGGGAGGGTTTT
ACGTTTACGGAGTTTCGTTGGTTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGCGAGGTTGGGGGAG

GGGTGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTAGGGAAGTTTGAATTGGGTGGAGTTTATTATAGT
TTAAGGAGGTTTGTTTGTTTTTGTAGATTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

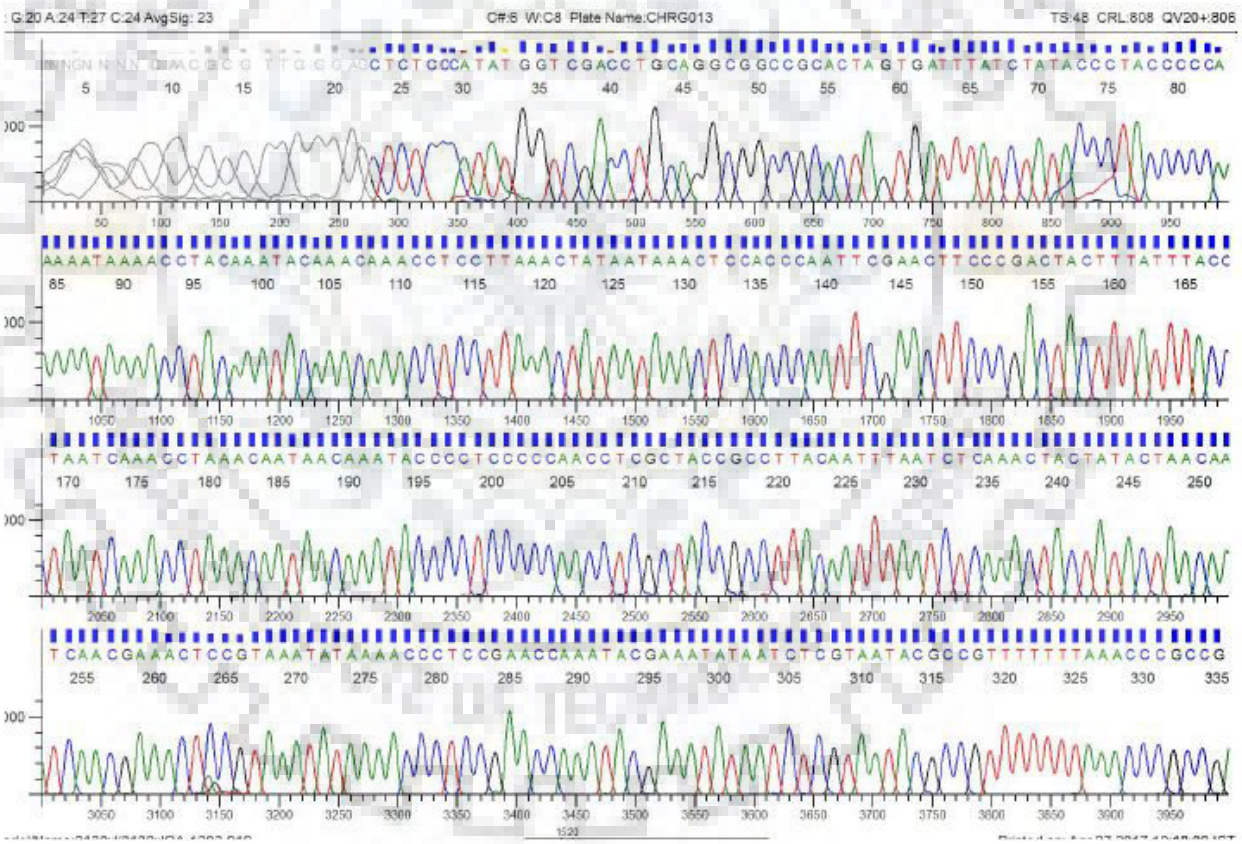
>C8d

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAGGAAAGGGGTGACGAATGTATTTGGAAAATCGGGTTATTTTTATTT
GAATATTGCGTTTTTTAGATCGGTTTAAAAAACGGTGTATTATGAGATTATATTTTATATTTGGTTGAGAGGGTTTTAC
GTTTACGGAATTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGGGGTAATGAGGTTGGGGGAGGG
GCGTTCGTTATTGTTTAGGTTTGTTTGGGTAAATAAAGGAGTTAGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

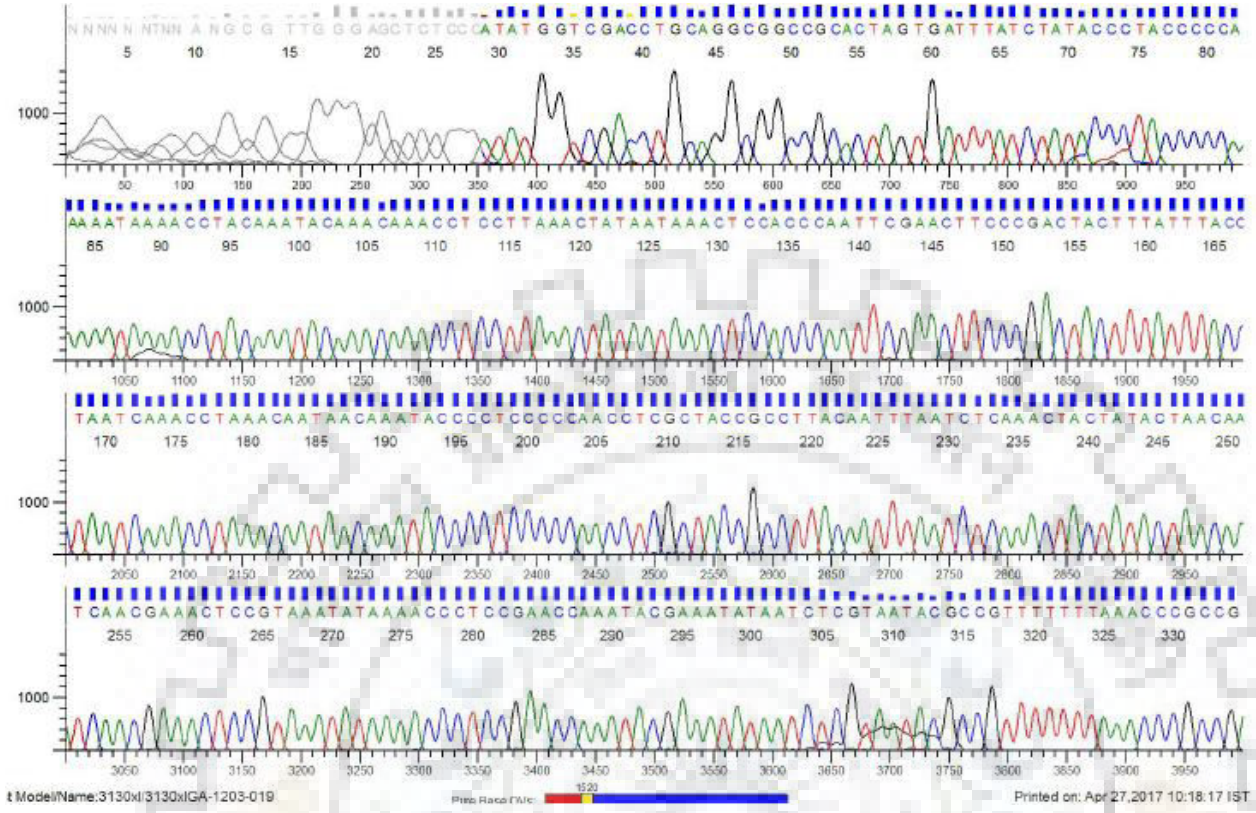
>C8e

AAGGGGTTAGGGAGTTTTTTTTTTAAGTTAAAGAACGGAGTGACGGACGGTATTTGGAAAATCGGGTTATTTTTATT
CGAATATTGCGTTTTTTCGGCGGGTTTAAAAAAACGGCGTATTACGAGATTATATTTCGTATTTGGTTCGGAGGGTTT
TATATTTACGGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGCGAGGTTGGGGGAG
GGGTATTTGTTATTGTTTAGGTTTGATTAGGTAAATAAAGTAGTCGGGAAGTTCGAATTGGGTGGAGTTTATTATAGT
TTAAGGAGGTTTGTTTGTATTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

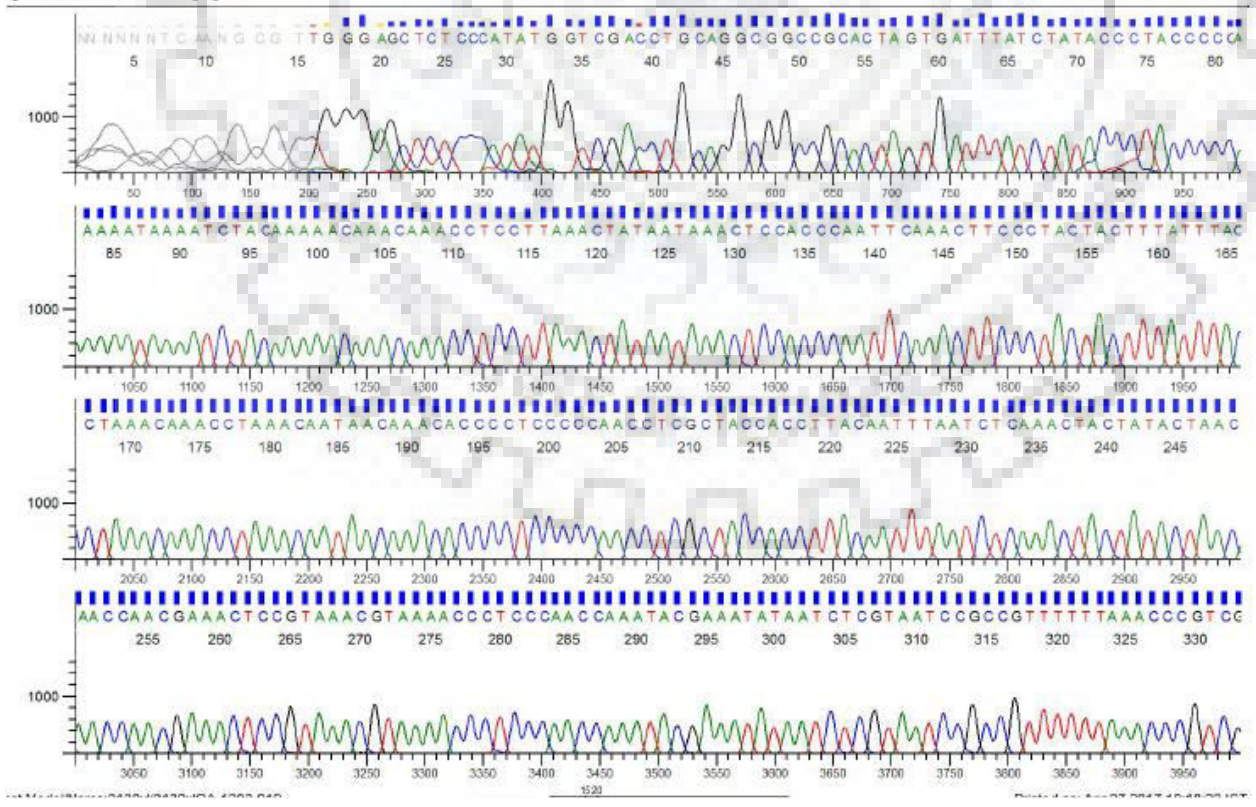
>C8a



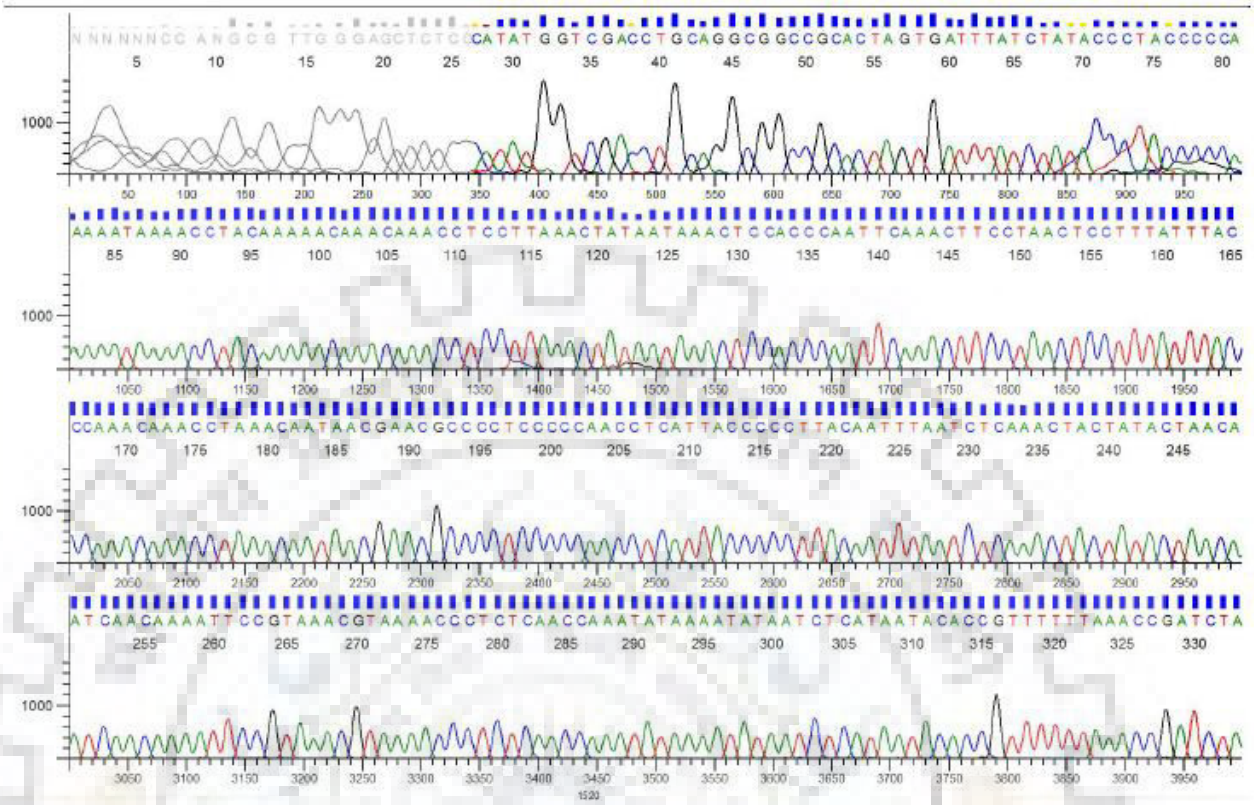
>C8b



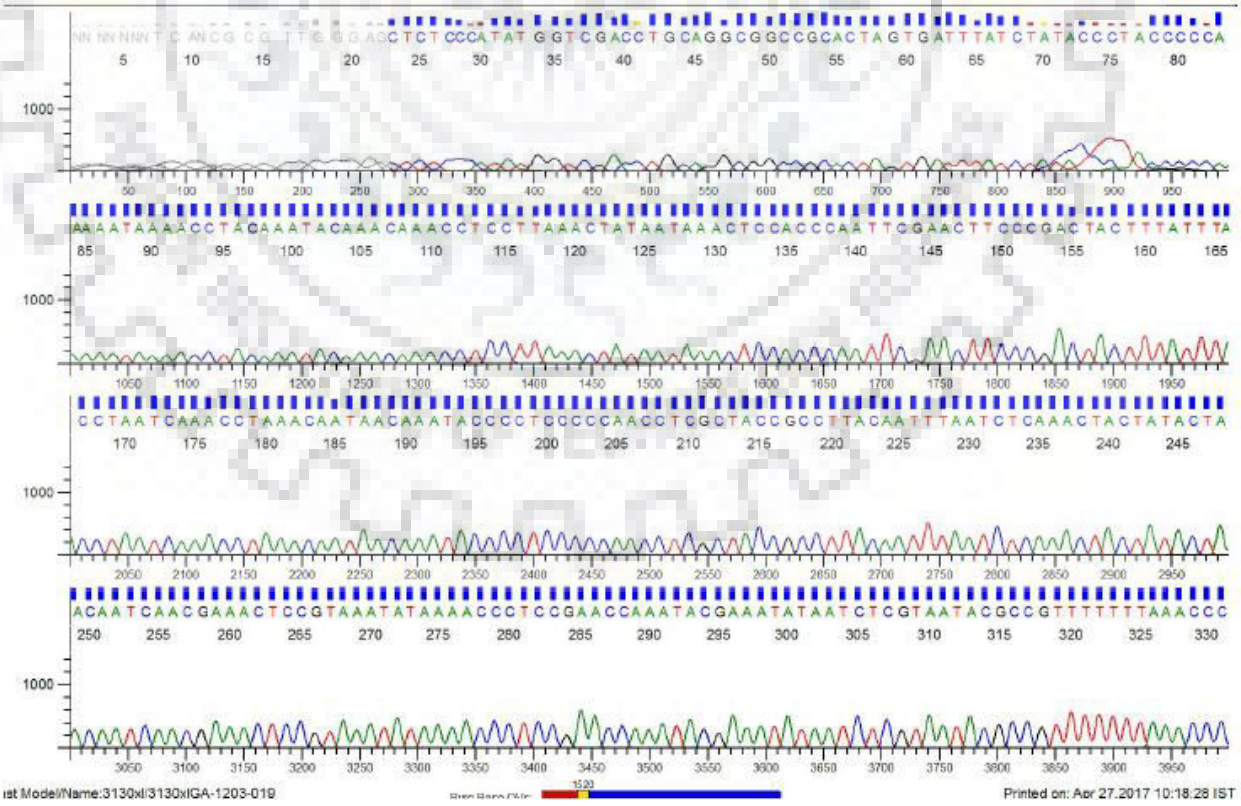
>C8c



>C8d



>C8e


ist Model\Name:3130xl/3130xlGA-1203-019
Printed on: Apr 27,2017 10:18:28 IST

Paired normal sample 8

>N8a

AAGGGGTTAGGGAGTTTTTTTTTTTAGTTAAAGAAAGGGGTGATAGATGGTATTTGGAAAATCGGGTTATTTTTATTT
TAATATTGCGTTTTTTCGATAGGTTTAAAAAACGGCGTATTAGGAGATTATATTTCGTATTTGGTTTAGAGGGTTTTAC
GTTTACGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGTTAGGTTGGGGGAGGGG
CGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTAGGAGGTTCGAATTGGGTGGAGTTTATTATAGTTTA
AGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>N8b

AAGGGGTTAGGGAGTTTTTTTTTTTAGTTAAAGAAAGGGGTGATAGACGGTATTTGGAAAATTAGGTTATTTTTATTC
GAATATTGTATTTTTTCGATGGGTTTAAAAAACGGCGTATTAGGGGATTATATTTCGTATTTGGTTTAAAGGGTTTTAC
GTTTATGGAGTTTTATTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGCGAGGTTGGGGGAGGG
GCGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTCGGGAAGTTCGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGGTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCTACT

>N8c

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTAGATCGGTTTAAGAAACGGCGTATTACGAGATTATATTTTATATTTGGTTCGGAGGGTTTTAC
GTTTACGGAATTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAACGAGGTTGGGGGAGGG
GCGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTCGGGAAGTTCGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

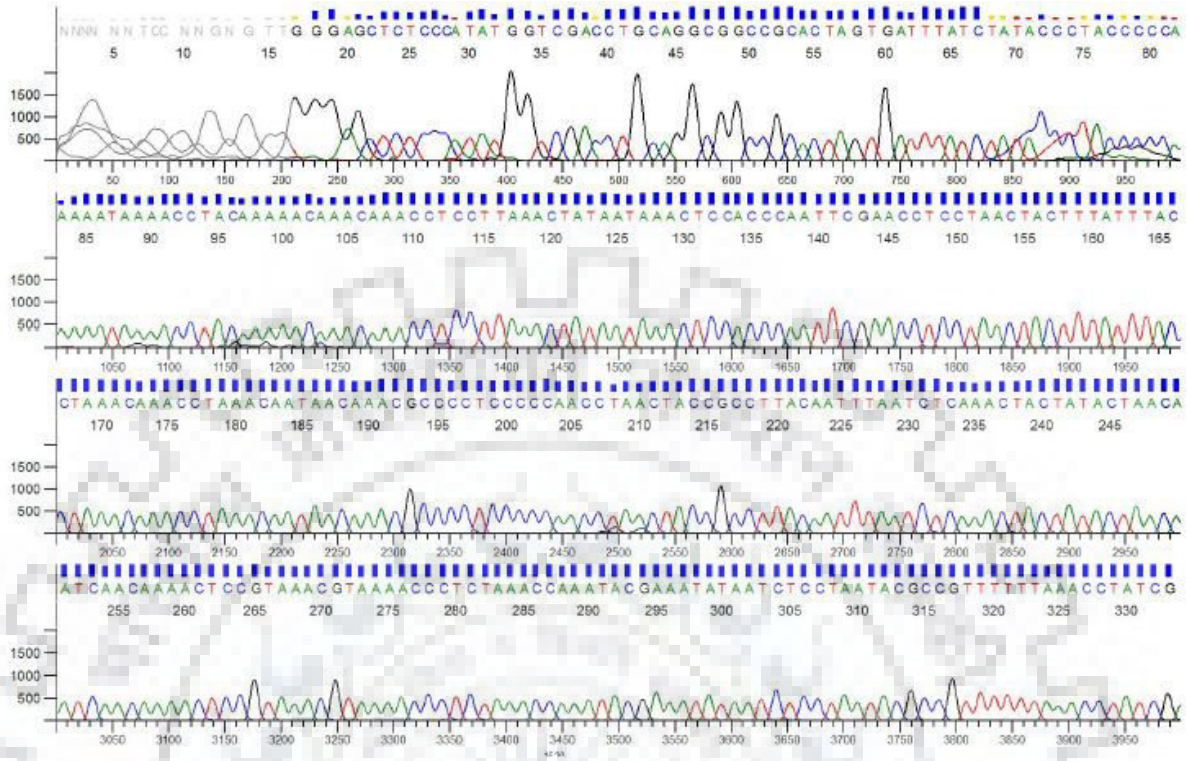
>N8d

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGATAGAGGGTATATGGAAAATCGGGTTATTTTTATT
CGAATATTGTATTTTTCGGATCGGTTTAAAAAATGGTGTATTATAAGATTATATTTTATATTTGGTTGGGAGGGTTTTA
CGTTTATGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTAGTAGCGAGGTTGGGGGAGGG
GTATTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTTGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTA
AGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

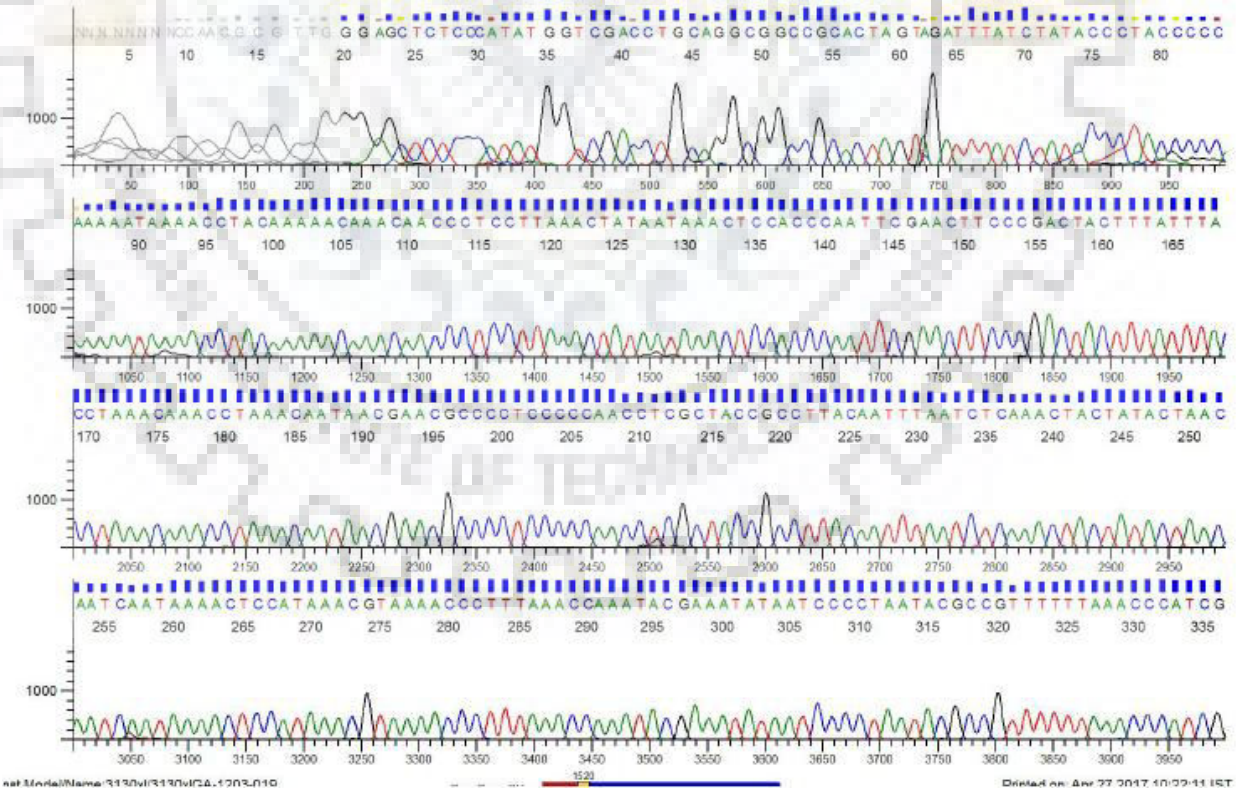
>N8e

AAGGGGTTAGGGAGTTTTTTTTTTTAATTAAAGAAAGGGGTGACGGACGGTATTTGGAAAATCGGGTTATTTTTATT
CGAATATTGCGTTTTTTCGAAGGGTTTAAAAAATGGCGTATTACGAGATTATATTTCGCATTTGGTTCGGAGGGTTTT
ATTTTATCGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGCGAGGTTGGGGGAGGG
GCGTTTATTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTAGGAAGTTCGAATTGGGTGGAGTTTATTATAGCTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

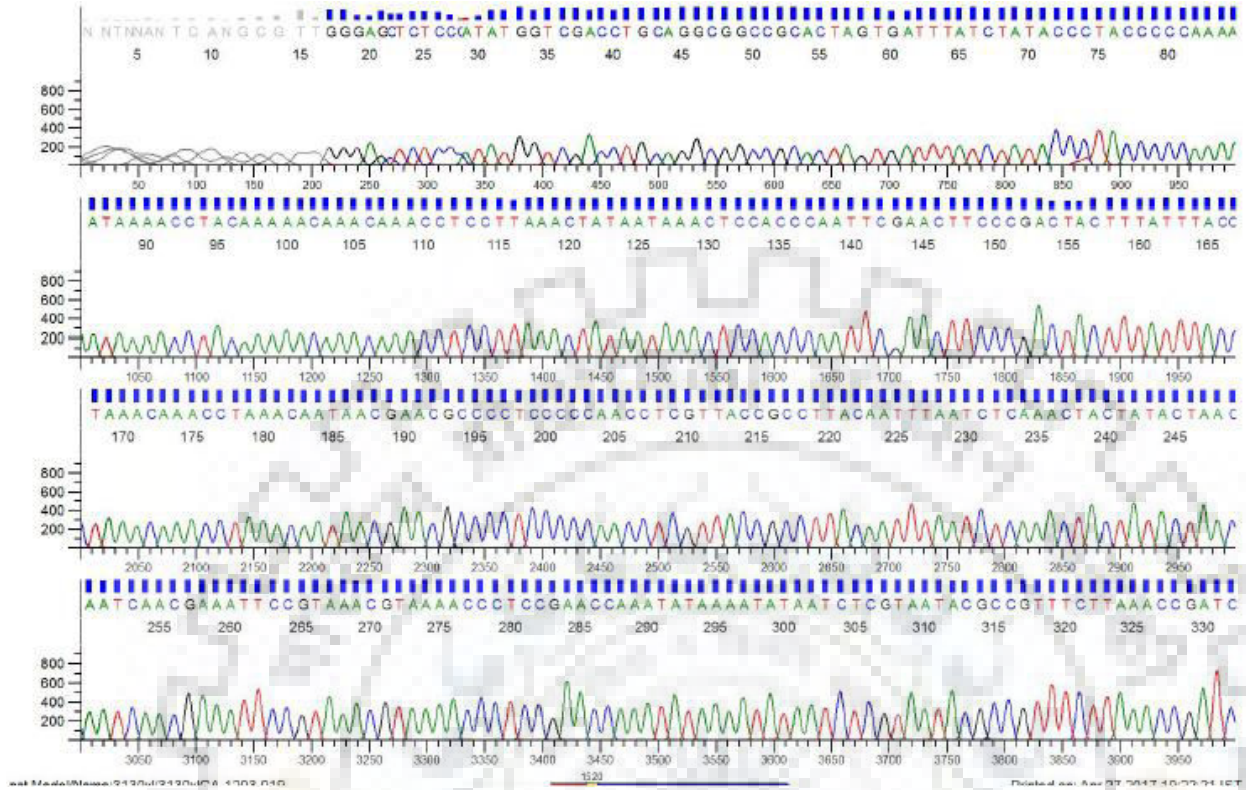
>N8a



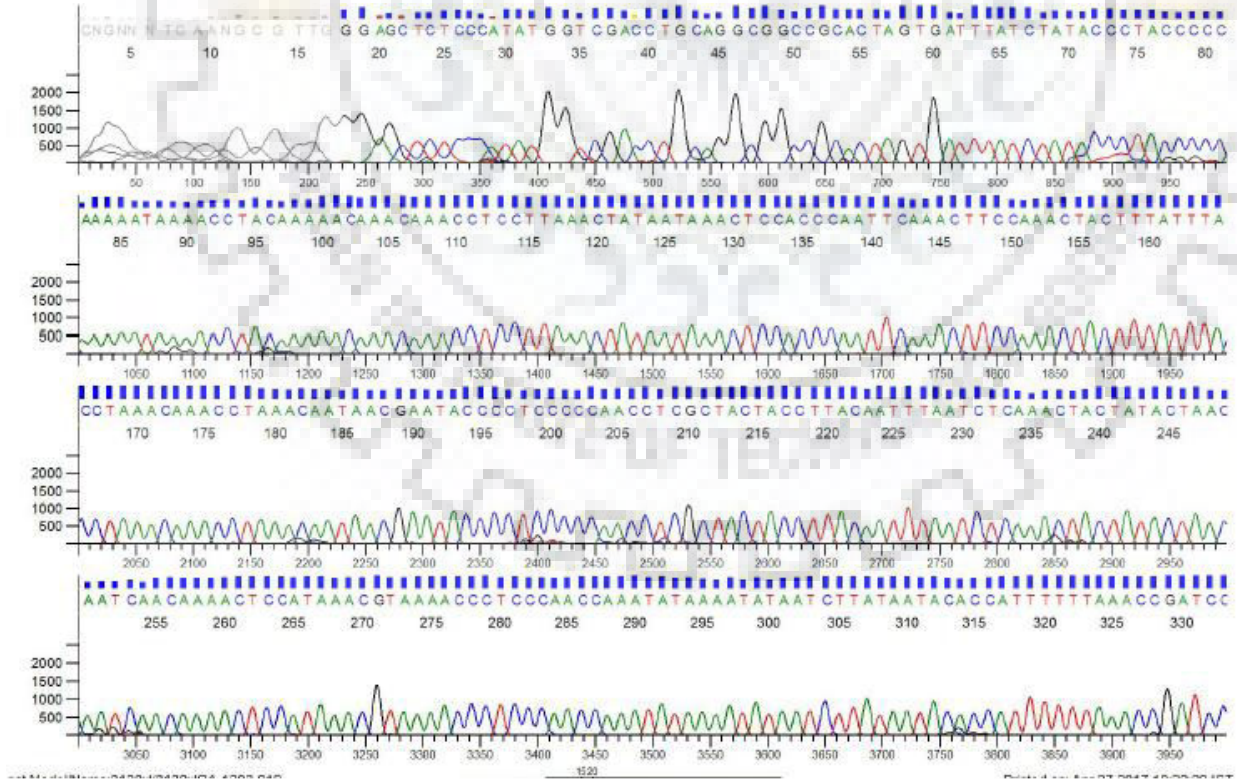
>N8b



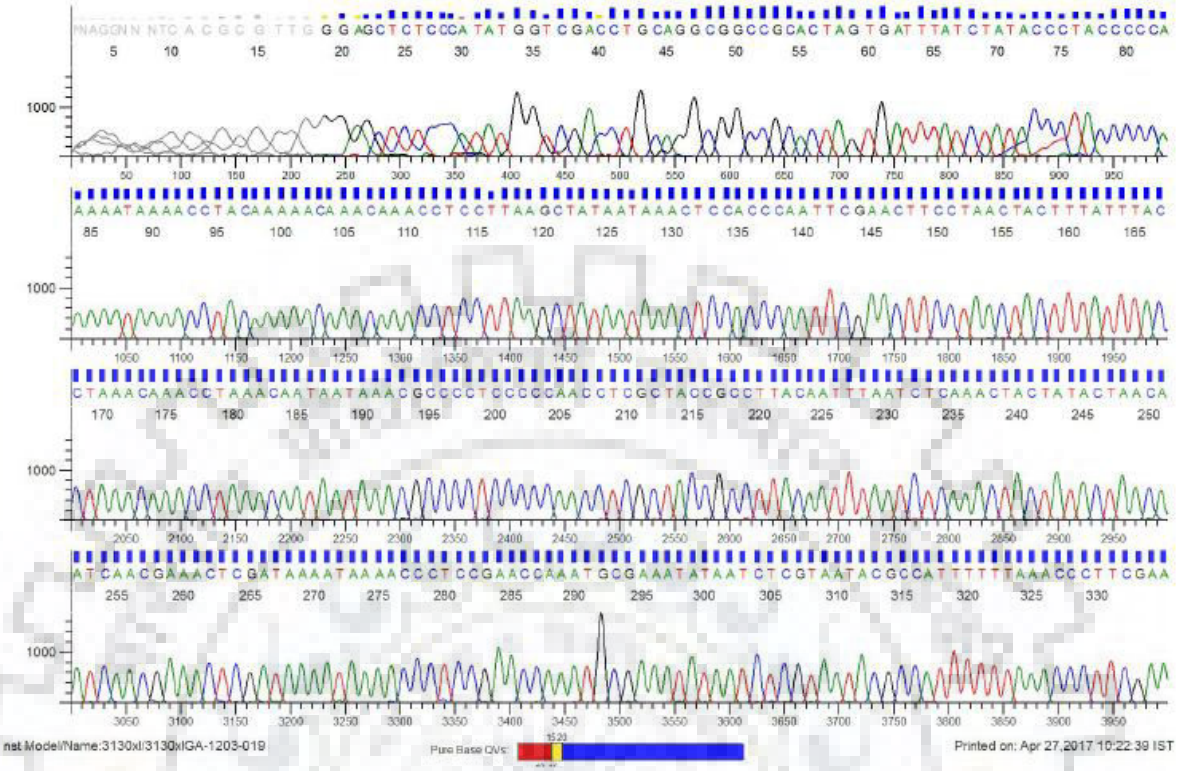
>N8c



>N8d



>N8e



Paired cancer sample 13

>C13a

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGAGGGACGTATTTGGAAAATCGGGTTATTTTTATTT
AAATATTGGGTTTTTTCGACGGGTTTAAAAAATGGTGTATTACGAGATTATATTTCGTATTTGGATCGGAGGGTTTTA
CGTTTATGGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGTGAGGTTGGGGGAGGG
GTGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAATAGGGAAGCTTGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>C13b

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTAGATCGGTTTAAGAAATGGCGTATTACGAGATTATATTTTATATTTGGTTCGGAGGGTTTTAC
GTTTACGGAATTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAATGAGGTTGGGGGAGGGG
CGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTGGGAAGTTCGAATTGGGTGGAGTTTATTATAGTTTA
AGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>C13c

AAGGGGTTAGGGAGTTTTTTTTTTGAGTTAAAGAAAGGGGTGATGGACGTATTTGGAAAATCGGATTATTTTTATTC
GAATAATGCGTTTTTTTGATTAGTTTAAAAAAAGGGTATATTATGAGATTATATTTCGTATTTGGTTTAGAGGGTTTTA
TGTTTACGGAGTTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGTTGGGTTGGGGGAGGG
GCGTTTATTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTTGGAAGTTCGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

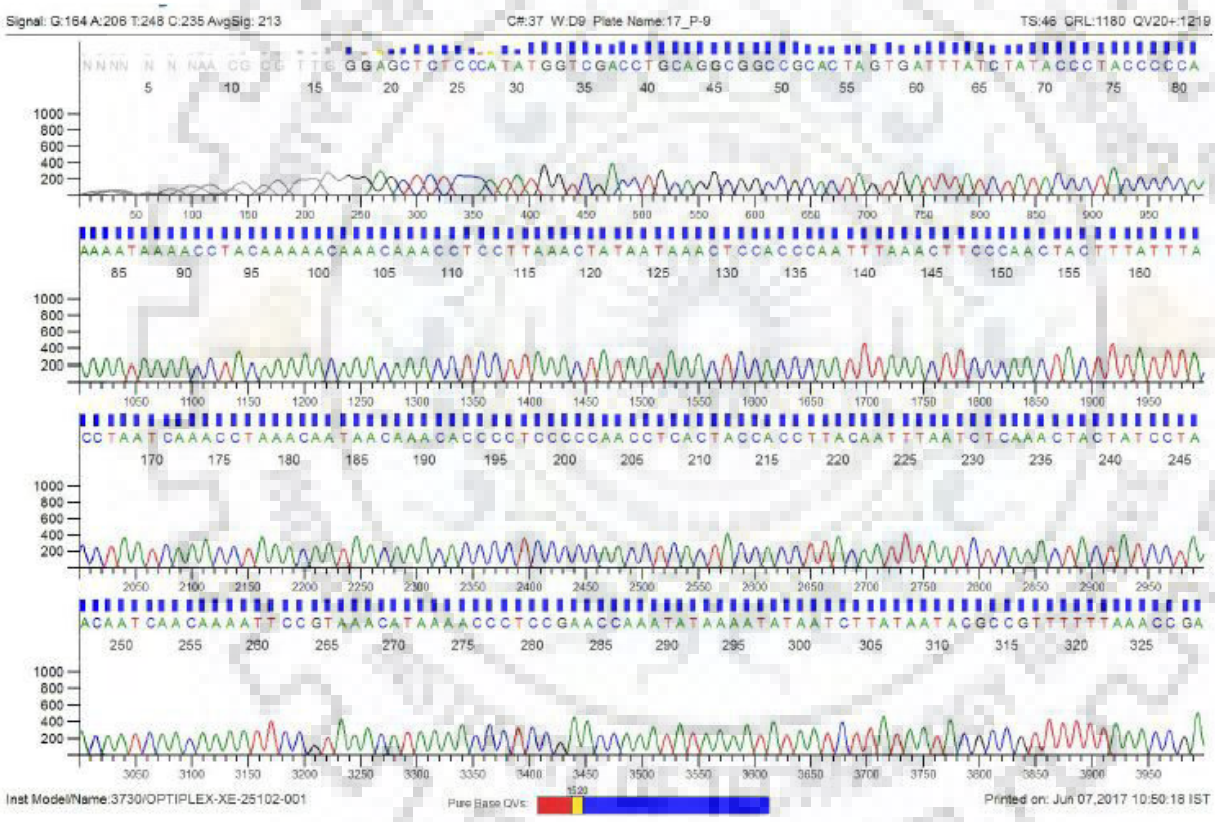
>C13d

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGATGGACGTATTTGGAAAATTGGGTTATTTTTATTT
GAATATTGCGTTTTTTAGATCGGTTTAAGAAACGGCGTATTACGAGATTATATTTTATATTTGGTTTAGAGGGTTTTAC
GTTTATGGAATTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAACGAGGTTGGGGGAAGGG
CGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTCGGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTA
AGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

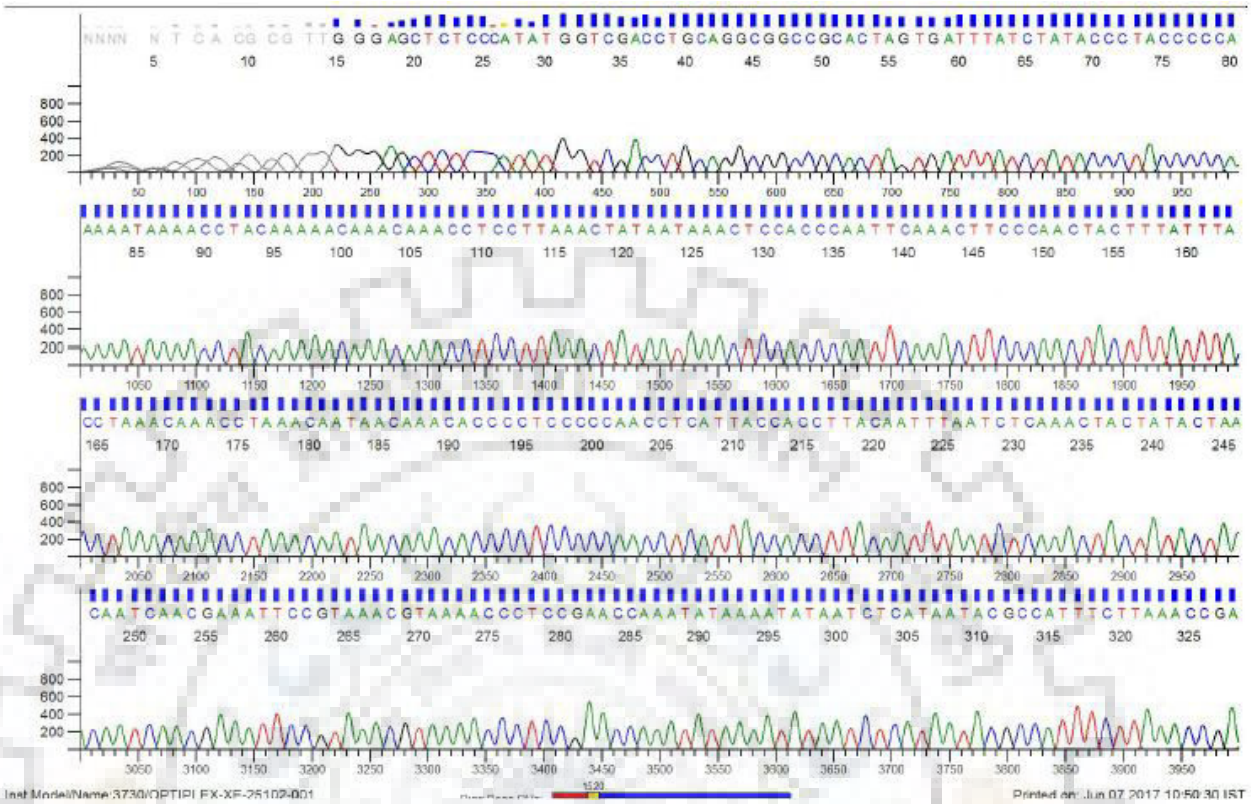
>C13e

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGATGGATGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTAGATCGGTTTAAAAAATGGCGTATTATGAGATTATATTTTATATTTGGTTTAGAGGGTTTTAT
ATTTACGGAATTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAATGAGGTTGGGGGAGGGG
TGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTAGGAAGTTTTAATTGGGTGGAGTTTATTATAGTTTAA
GGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

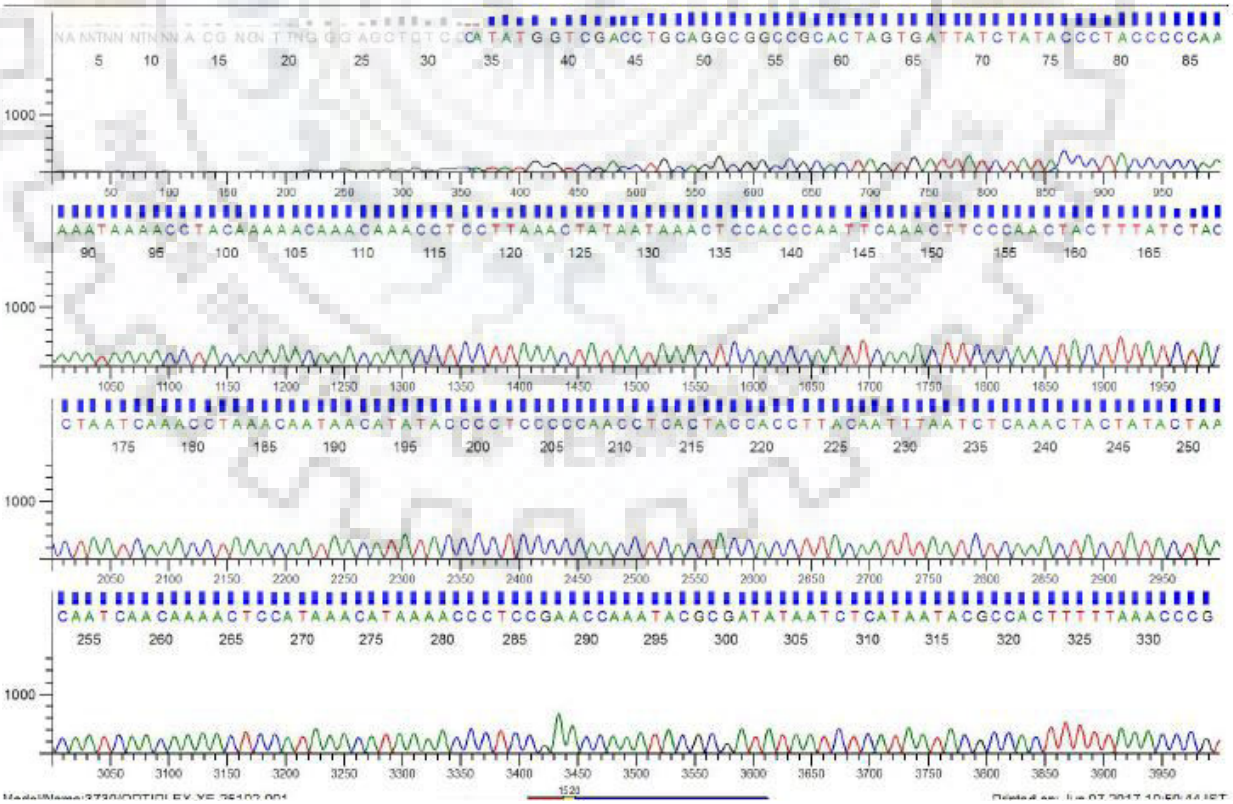
>C13a



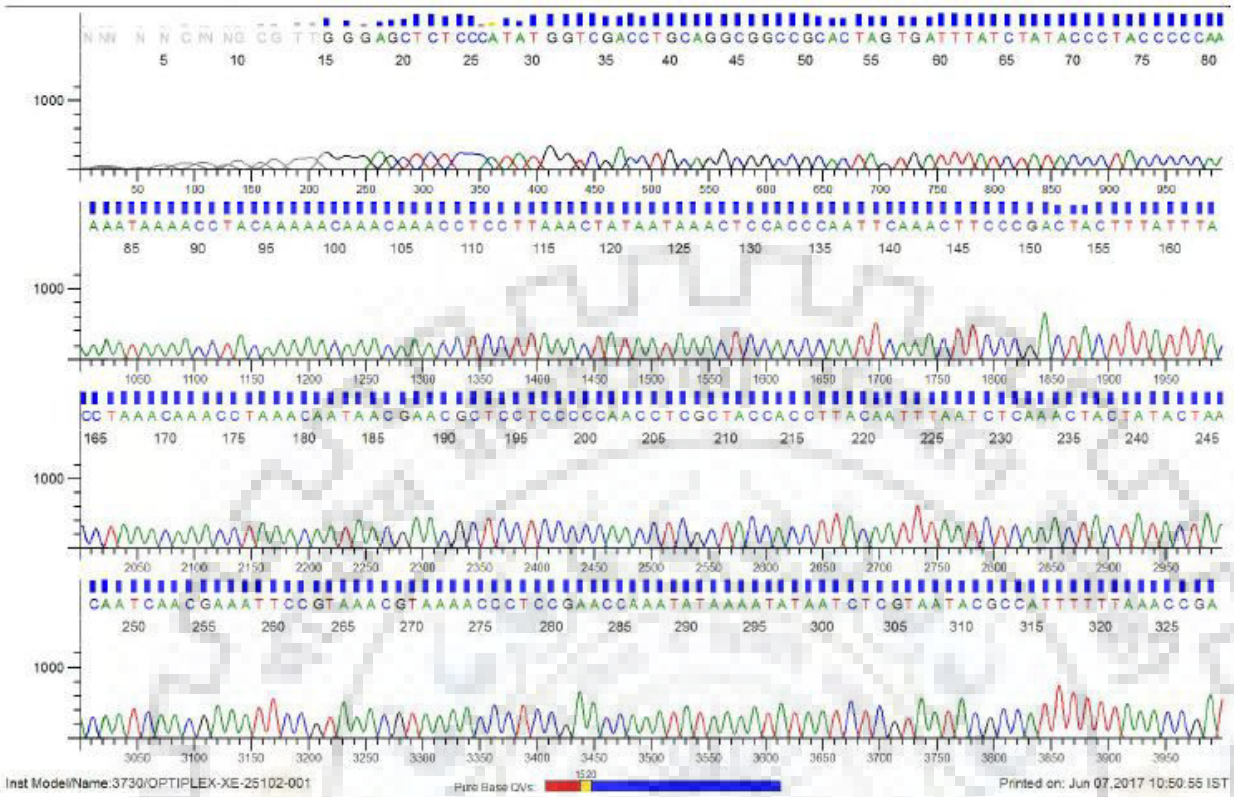
>C13b



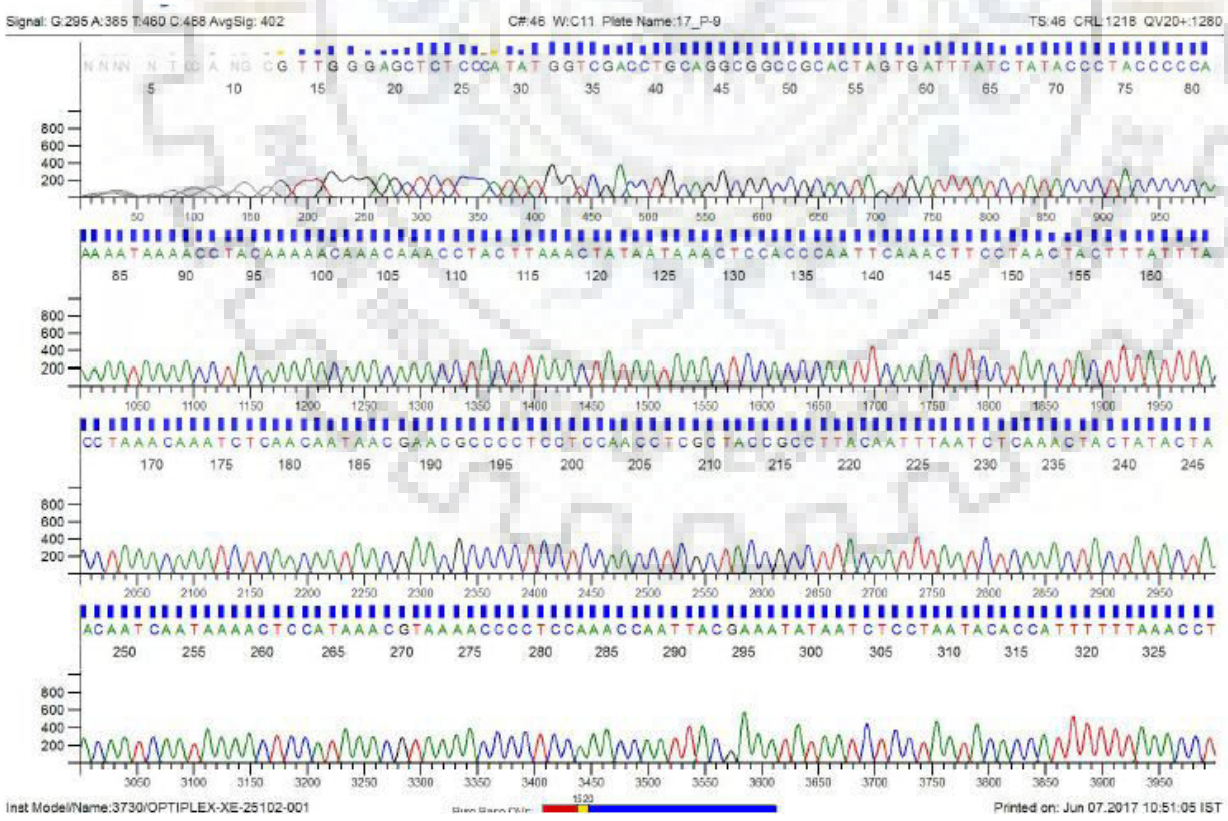
>C13c



>C13d


Inst Model/Name:3730/OPTIPLEX-XE-25102-001
Printed on: Jun 07,2017 10:50:55 IST

>C13e


Signal: G:295 A:385 T:460 C:488 AvgSig: 402
C#:46 W:C11 Plate Name:17_P-9
TS:46 CRL:1218 QV20+:1280
Inst Model/Name:3730/OPTIPLEX-XE-25102-001
Printed on: Jun 07,2017 10:51:05 IST

Paired normal sample 13

>N13a

AAGGGGTTAGGGAGTTTTTTTTTTGAGTTAAAGAAAGGGGTGACGGACGTATTTGGAAAATTGGGTTATTTTTATTT
GAATATTGCGTTTTTTAGATCGGTTTAAAAAACGGCGTATTATAAGATTATATTTTATATTTGGTTCGGAGGGTTTTAT
GTTTACGGAATTTTGTTGATTGTTAGGATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGTGAGGTTGGGGGAGGG
GTGTTTGTTATTGTTTAGGTTTGATTAGGTAAATAAAGTAGTTGGGAAGTTTAAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>N13b

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGATGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTAGATCGGTTTAAGAAATGGCGTATTATGAGATTATATTTTATATTTGGTTCGGAGGGTTTTAC
GTTTACGGAATTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAATGAGGTTGGGGGAGGGG
TGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTGGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTA
AGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>N13c

AAGGGGTTAGGGAGTTTTTTTTTTTTAGTTAAAGAAAGGGGTGACGGATGGTATTTGGAAAATTGGGTTATTTTTATTT
GAATATTGCGTTTTTTTCGACGGGTTTAAAAAGTGGCGTATTATGAGATTATATCGCGTATTTGGTTCGGAGGGTTTTA
TGTTTATGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGTGAGGTTGGGGGAGGGG
GTATATGTTATTGTTTAGGTTTGATTAGGTAGATAAAGTAGTTGGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAATCACT

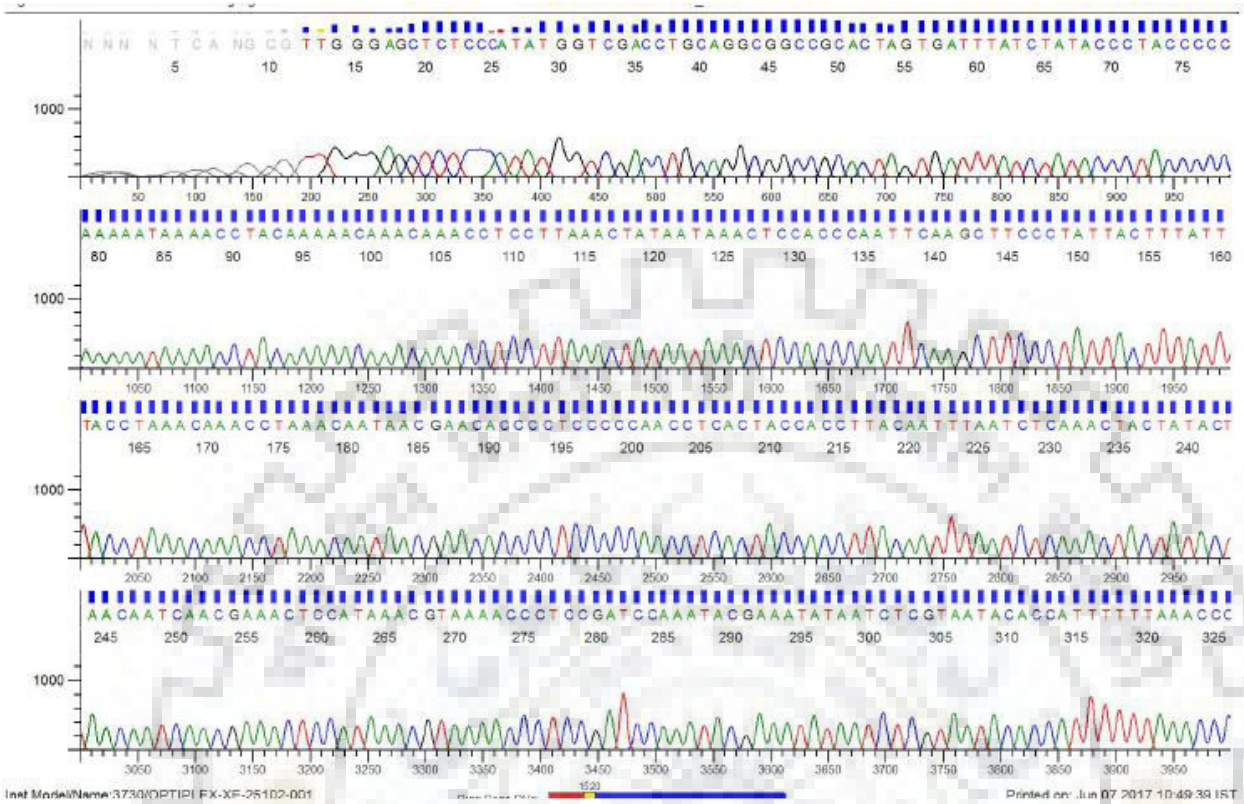
>N13d

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGATGTATTTGGAAAATTAGGTTATTTTTATTC
GAATATTGCGTTTTTTAGATCGGTTTAAAAAATGGCGTATTACGAGATTATATTTTATATTTGGTTCGGAGGGTTTTAC
GTTTACGGAATTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGCGAGGTTGGGGGAGGA
GCGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTCGGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

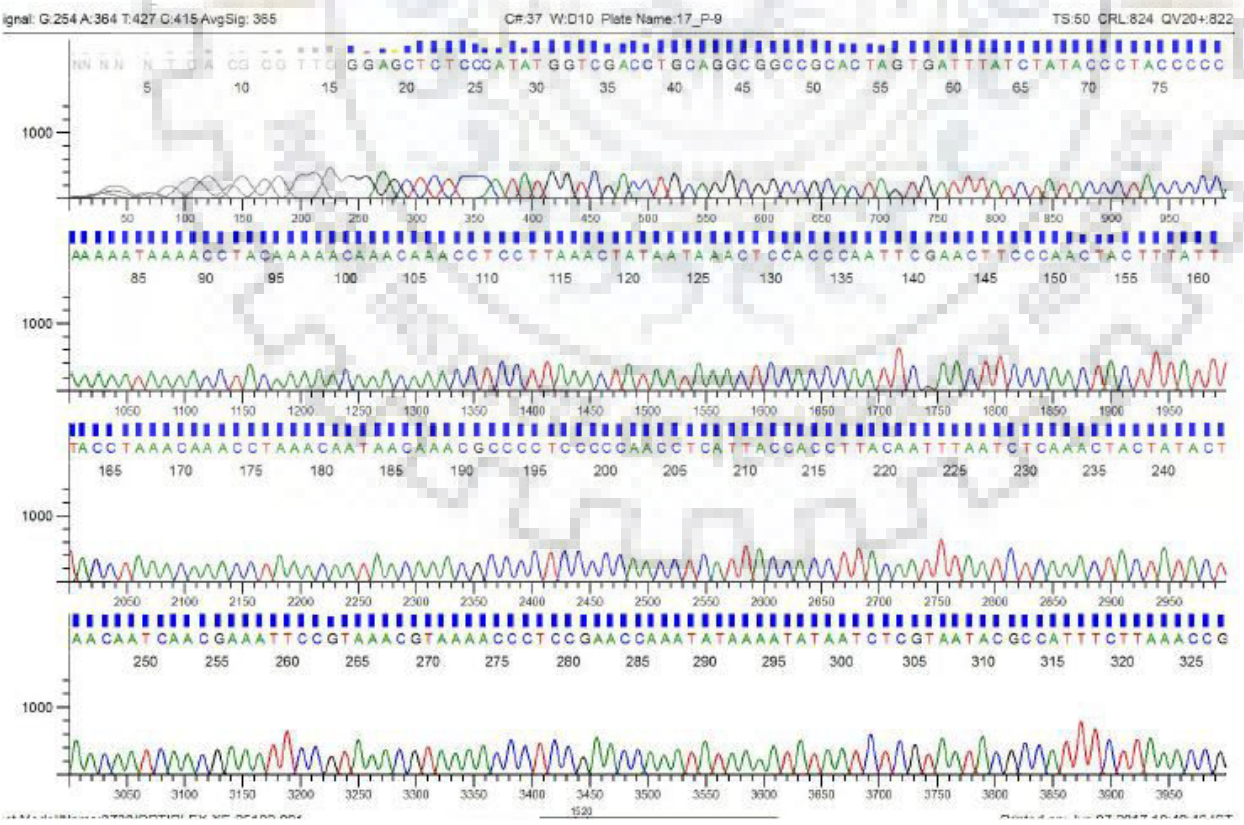
>N13e

AAGGGGTTAGGGAGTTTTTTTTTAGTTAAAGAAAGGGGTGATAGATGGTATTTGGAAAATCGGGTTATTTTTATTCG
AATATTGCGTTTTTTCGATAGGTTTAAAAAATGGTGTATTAGGAGATTATATTTCGTAATTGGTTTGGAGGGGTTTTA
CGTTTATGGAGTTTTATTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGCGAGGTTGGAGGAGGG
GCGTTCGTTATTGTTGAGATTTGTTTAGGTAAATAAAGTAGTTAGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTT
AAGTAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

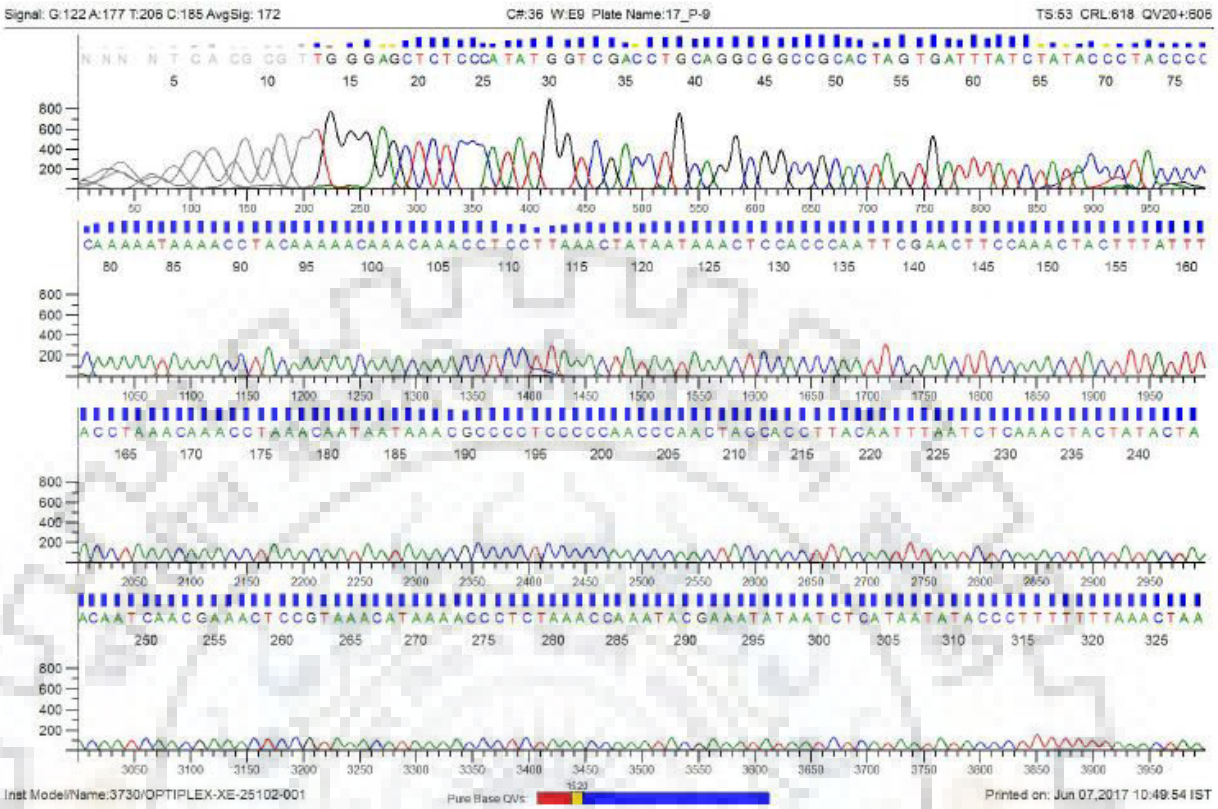
>N13a



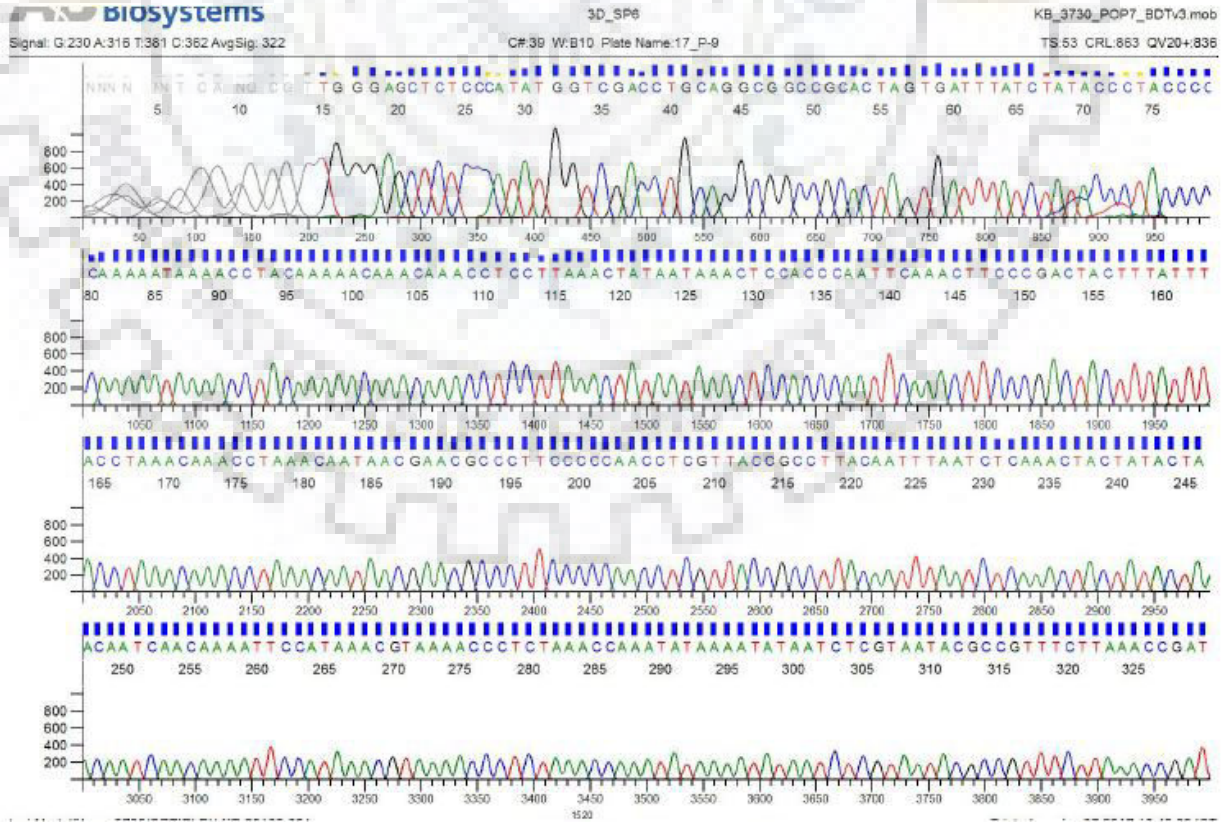
>N13b



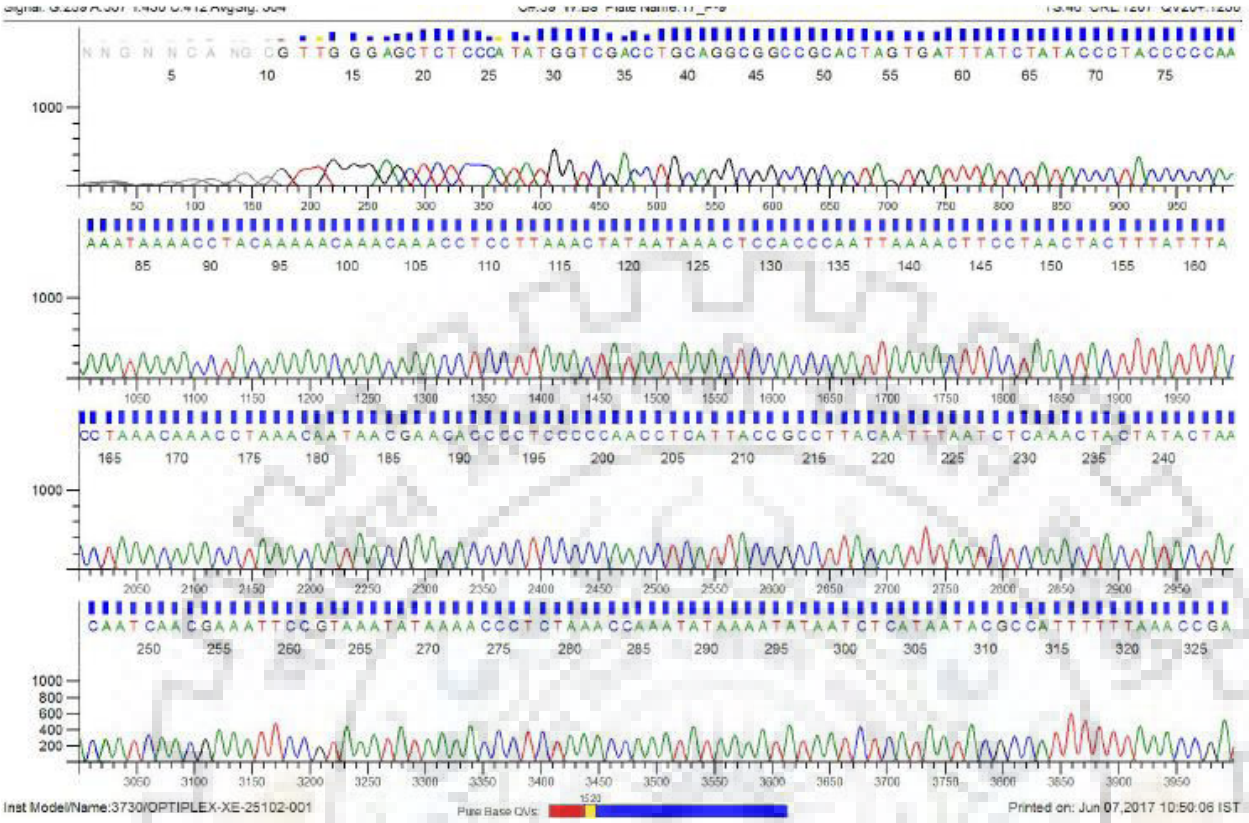
>N13c


Signal: G:122 A:177 T:206 C:185 AvgSig: 172
C#:36 W:E9 Plate Name:17_P-9
TS:53 CRL:618 QV20+:606
Inst Model/Name:3730/OPTIPLEX-XE-25102-001
Pure Base QVs:
Printed on: Jun 07,2017 10:49:54 IST

>N13d


Biosystems
3D_SP6
KB_3730_POP7_BDTv3.mob
Signal: G:230 A:316 T:361 C:362 AvgSig: 322
C#:39 W:B10 Plate Name:17_P-9
TS:53 CRL:863 QV20+:836

>N13e



Paired cancer sample 14

>C14a

AAGGGGTTAGGGAGTTTTTTTTTTTAGTTAAAGAAAGGGGTGATAAATAGTATTTGGAAAATCGGGCTATTTTTATTT
TAATATTGCGTTTTTTTAACGGGTTTTATAAATGGTATATTAGGAGATTATATTCGGTATTTAGTTCGGAGGGTTTTAT
GTTTACGGAGTTTCGTTTATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGCGAGGTTGGTGGAGGGG
CGTTTATTATTGTTTAGGTTTGAGTAGGTAAATAAGGTGGTTAGTTGGGAAGTTGGAATTGGGTGGAGTTTATTATA
GTTTAAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>C14b

AAGGGGTTAGGGAGTTTTTTTTTTTAGTTAAGGGAAGAGGGGATAAACGGTATTTGGAAGATCGGGTTATTTTTATT
TTAATATTGCGTTTTTTAGACGGTTTTAGTAAACGGTATATTAGGAGATTATATTTCGTGTTTGGTTTAGAGGGTTTTA
TGTTTATAGAGTTTTATTTATTGTTAGTATAGTAGTTTGAGATTAAATTGTAATGTAGTAGTTAGGTTGGGGGAGGGG
CGTTTGTTATTGTTGAGGTTTGAGTAGGTAAATAAAGTTTTTAGGAAGGGAAGTTTGAATTGGGTGGAGTTTATCGT
AGTTTAAGGAGGTTTGTTTGTTTTTGTAGATTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

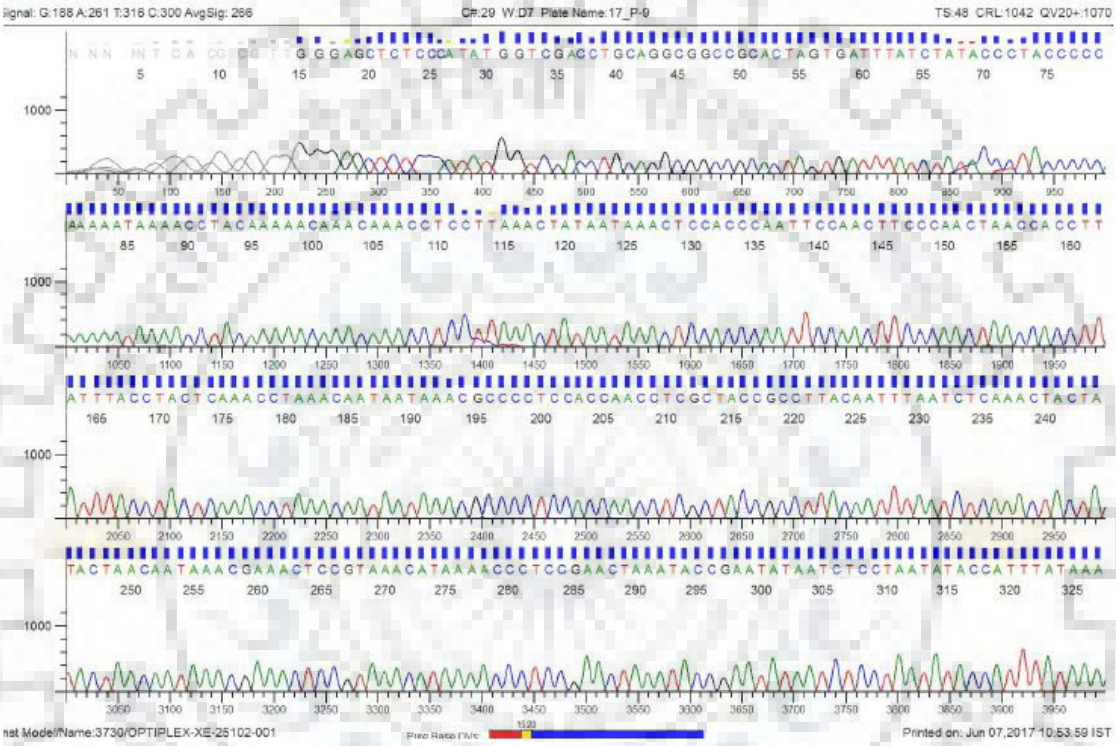
>C14c

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTAGATCGGTTTAAGAAACGGCGTACTATGAGATTATATTTTATATTTGGTTTGGAGGGTTTTAC
GTTTACGGAATTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAATGAGGTTGGGGGAGGGG
TGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTGGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTA
AGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

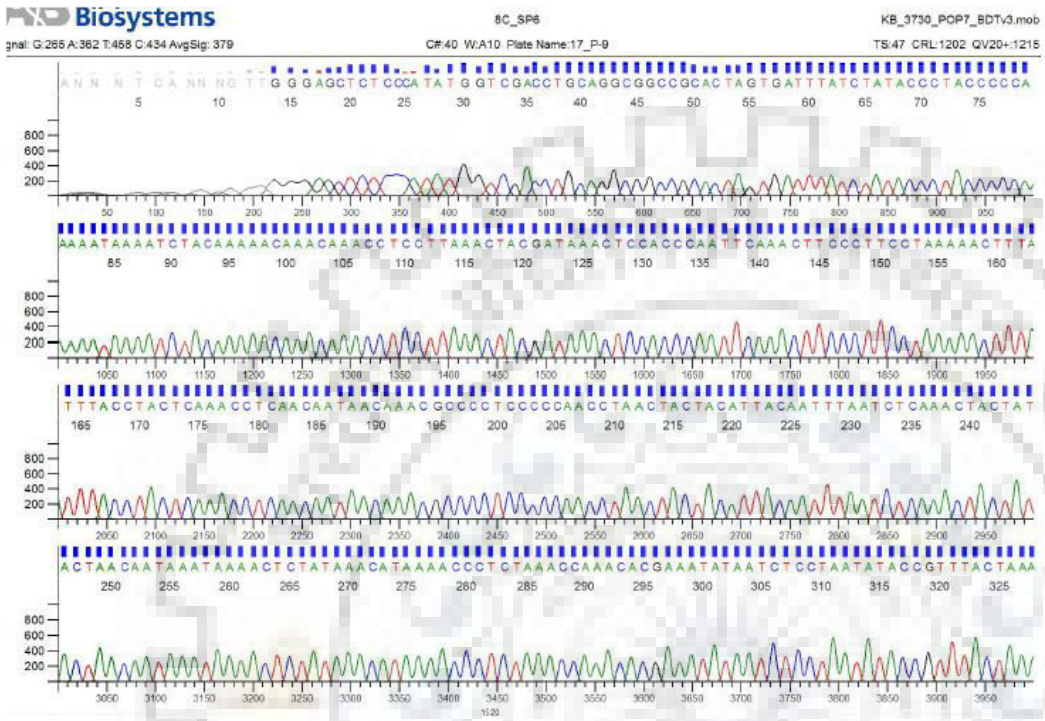
>C14d

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGACGTATTTGGAAAATCGGGTTATTATTAGTT
GAATATTGCGTTTTTTGGACCGGTTTAAAAAACGGCGTATTACGATATTATATTTCGTACGTGGTTTCGAGGGTTTTA
CGTTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGCAGCGAGGTTGGGGGAGGGGCGTTCGTT
ATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTAGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTAAGGAGGTT
TGTTTTTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACTAGTGCGGCCGC

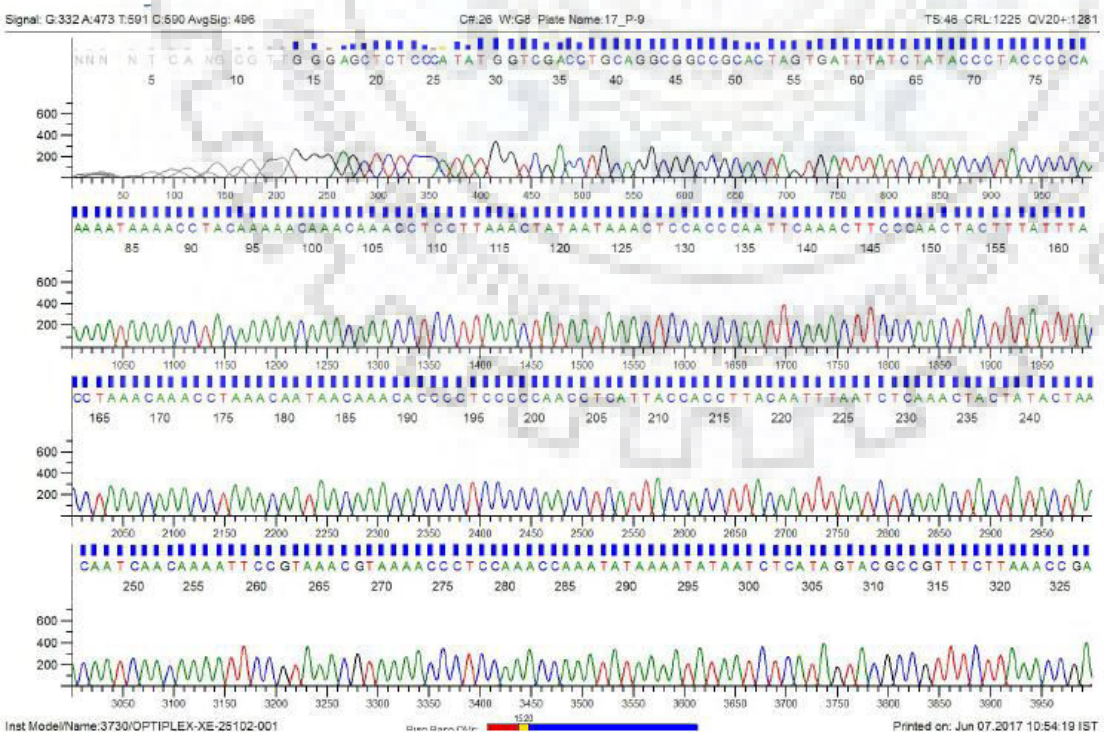
>C14a



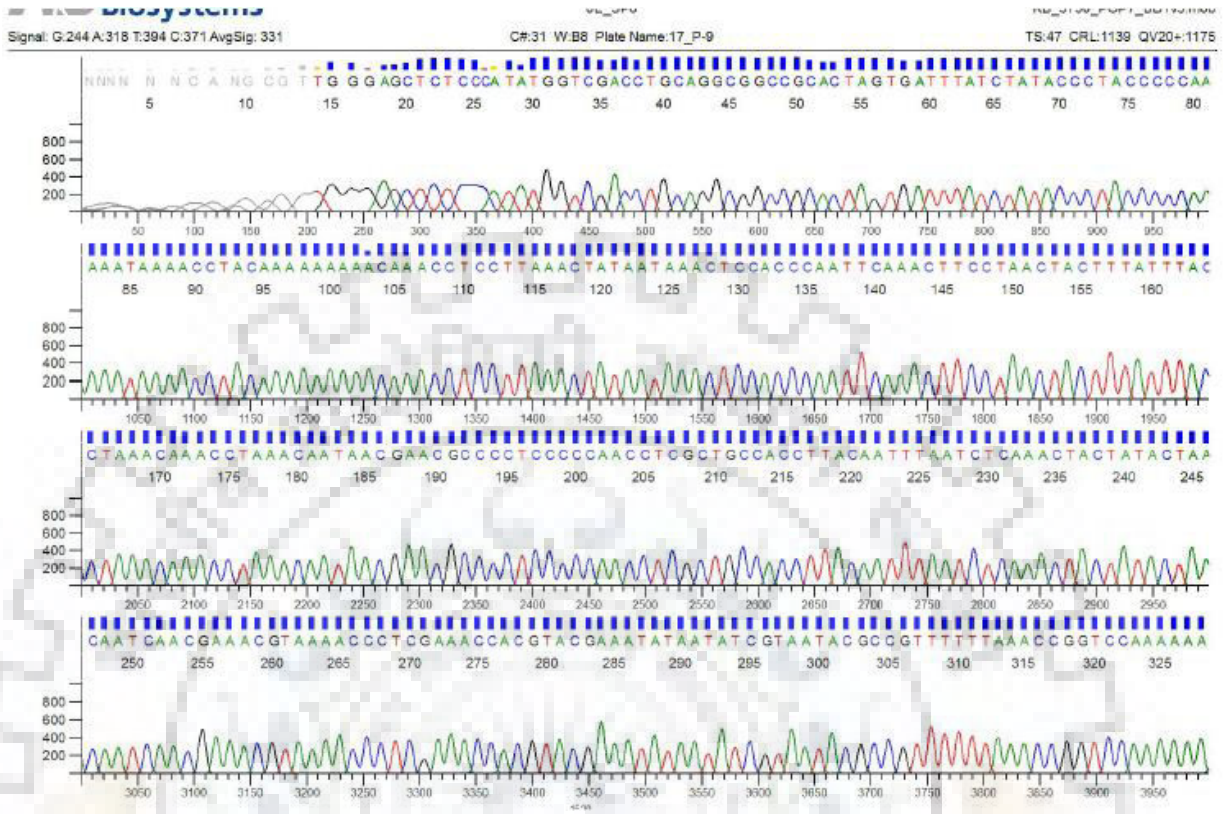
>C14b


Biosystems
8C_SP6
KB_3730_POP7_BDTv3.mob
C#:40 W:A10 Plate Name:17_P-9
TS:47 CRL:1202 QV20+:1215

>C14c


C#:26 W:G8 Plate Name:17_P-9
TS:48 CRL:1225 QV20+:1281
Printed on: Jun 07,2017 10:54:19 IST

>C14d



Paired normal sample 14

>N14a

AAGGGGTTAGGGAGTTTTTTTTTTTAGTTAAAGAAAGAGGTGACGATGGTATTTGGAAAAAAGGGTTATTTTTATCG
TAATATTGTGTTTTTTTAATTGTTTTAGTAAACGGTATATTACGAGATTATATTTCGTATTTGGTTTGGAGGGTTTTATA
TTTATGGAGTTTTATTTATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGCGAGGCTGGGGGAGGGGT
TTTTATTATTGTTTAGGTTTGAGTAGGTAAATAAAGTGGTTAGGAAGTTTAAATTGGGTGGAGTTTATTAAAGTTTAA
GGAGGTCGGTTTGTTTTTGTAGATTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>N14b

AAGGGGTTAGGGAGTTTTTTTTTGAGTTAAAGAAAGGGGTGATAGATGGTATTTGGAAAATCGGGTTATTTTTATTT
TAATATTGCGTTTTTTCGATAGGTTTAAAAAACGGTATATTACGAGATTATATTTCGTATTTGGTTTAGAGGGTTTTAC
GTTTACGGAGTTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGTTAGGTTGGGGGAGGGG
CGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTAGGAGGTTCGAATTGGGTGGAGTTTATTATAGTTTA
AGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>N14c

AAGGGGTTAGGGAGTTTTTTTTTTCGAGTTAAAGAAAGGGGTGATAGATGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTTAGATCGGTTTAAAAAACGGCGTATTACGAGATTATATTTTATATTTGGTTCGGAGGGTTTTAC
GTTTATTGATTGTGTTAGTAATTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGCGAGGTT
GGGGGAGGGGCGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTTGGAAGTTTGAATTGGGTGGAGTTT
ATTATAGTTTAAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

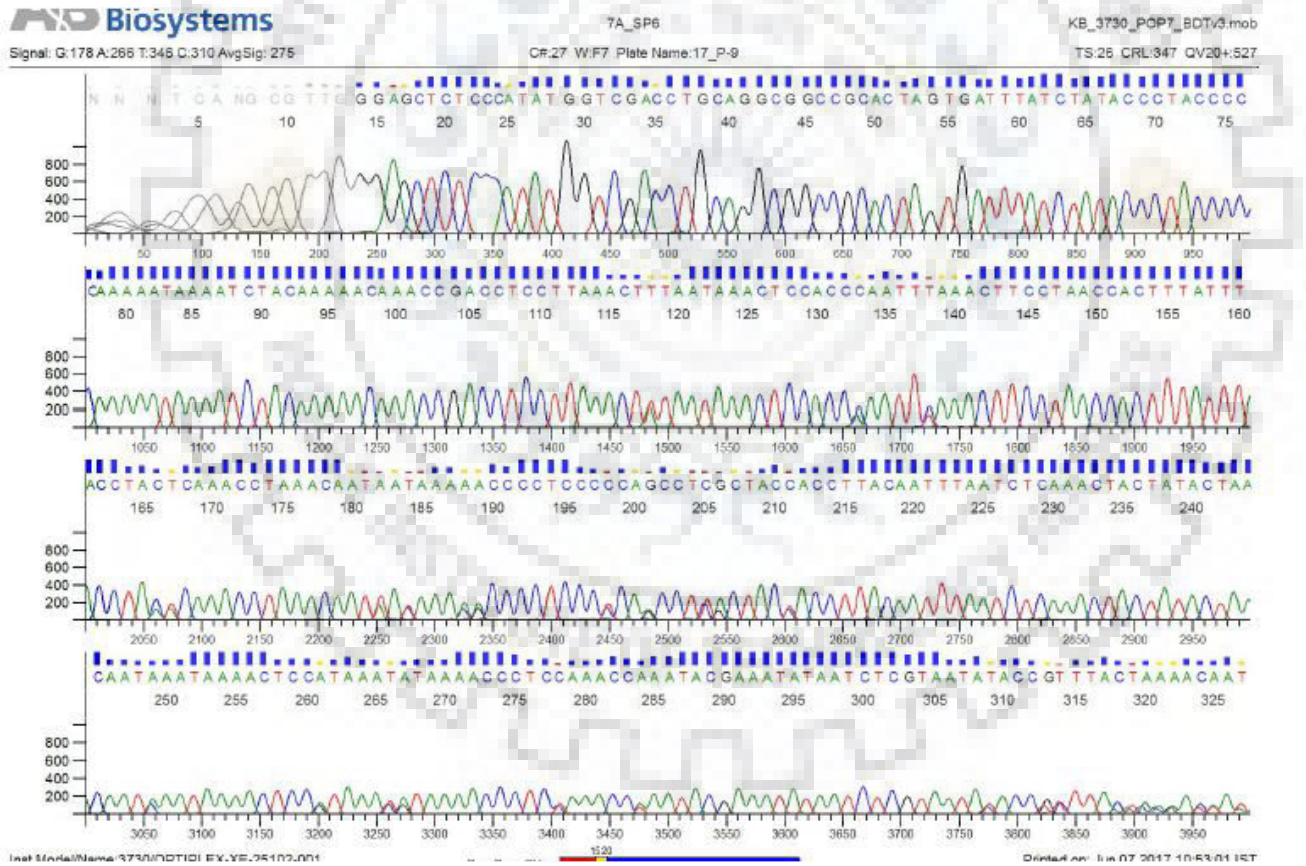
>N14d

AAGGGGTTAGGGAGTTTTTTTTTTTAGTTATGGGAAATCGTGATAGACGGTATTTGGGAAATCGGGAAATTTTTATTT
TAATATTGCGTTTTTTTAATGGTTTTAGTAAATGGTATATTAGGAGATTATATTTCGTATTTGGTTTAGAGGGATTTAT
GTTTATAGAGTTTGGTTTATTGTTAGTATAGTAGTTTGAGATCGAATTGTAAGGTAGTAGTGAGGTTGGGGGAGGGG
TATTCGTAATTGTTGAGGTTTGAGAAGGTAAATAAAGTAGTTTGAAAGTTCGAATTGGGTGGAGATTATCGTAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGATTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

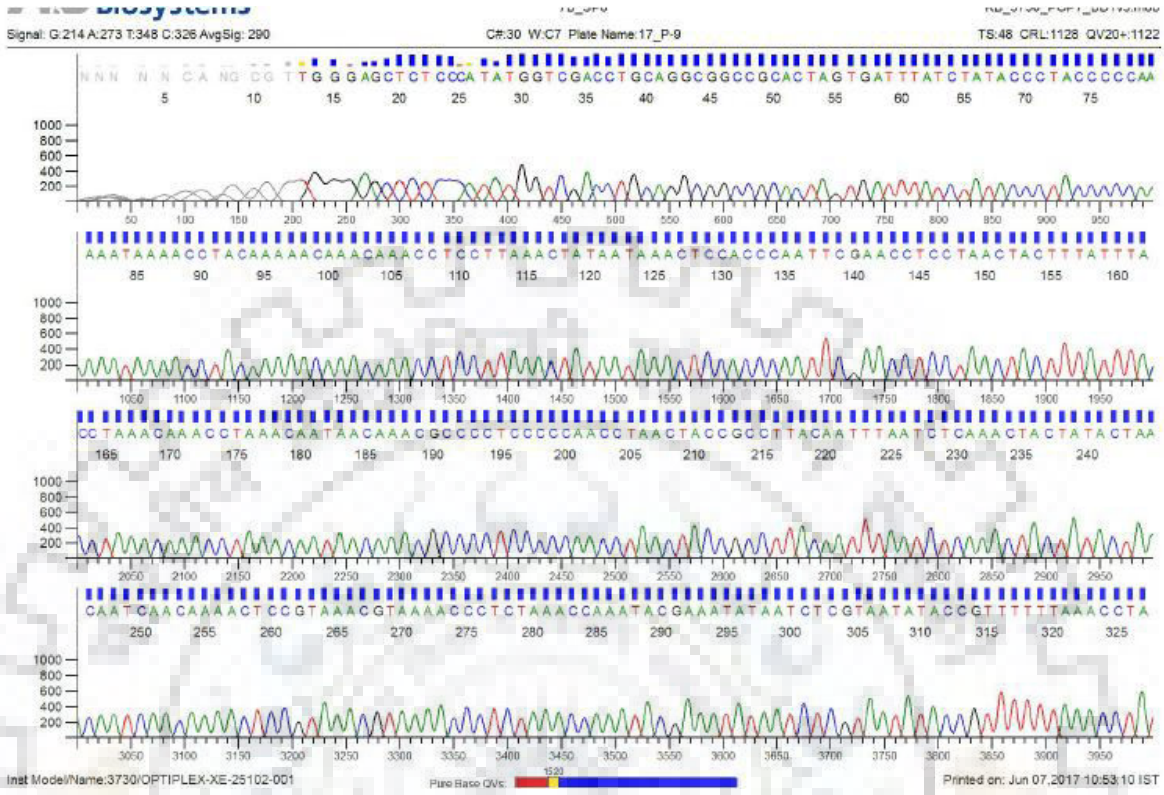
>N14e

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGGATAGTATTTGGAAAATCGGGTTATTTTTATT
CGAATATTGTTTTTTTTAGATCGGTTTAAAAAACGGCGTATTAGGAGATTATATTTTATATTTGGTTTAGAGGGTTTTA
TATTTATGGAATTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAACGAGGTTGGGGGAGGG
GCGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTAGGAAGTTCGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

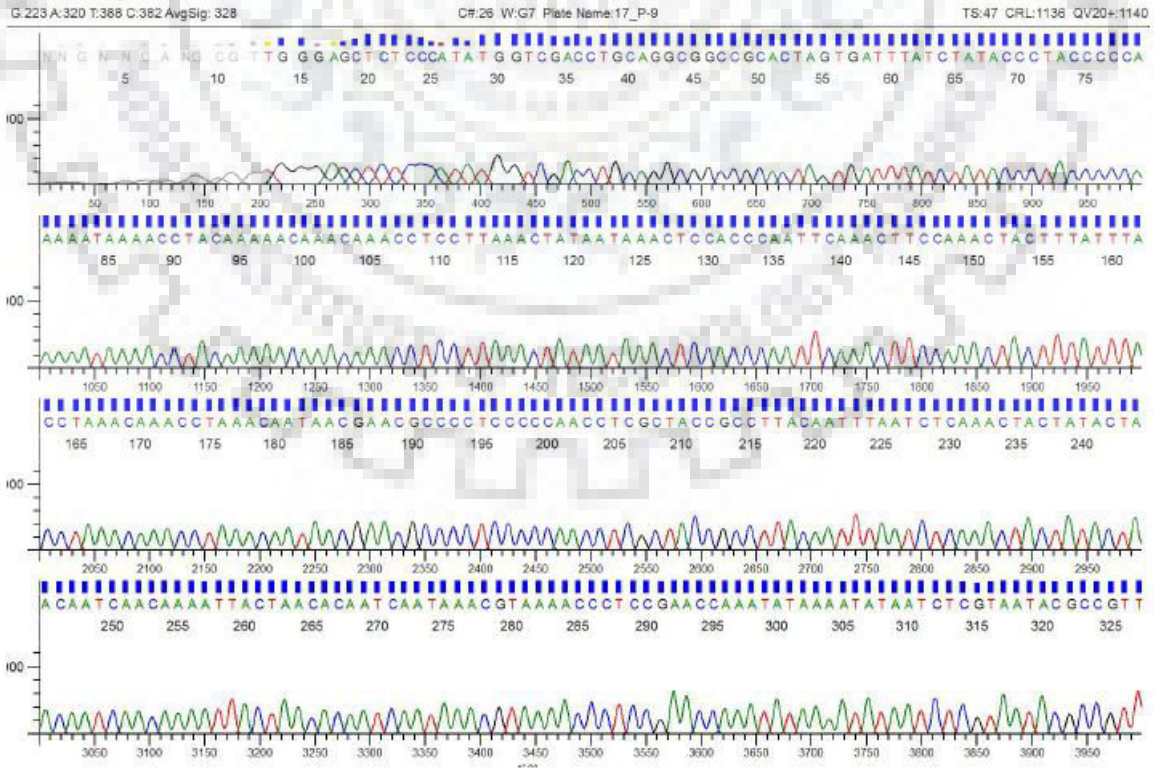
>N14a



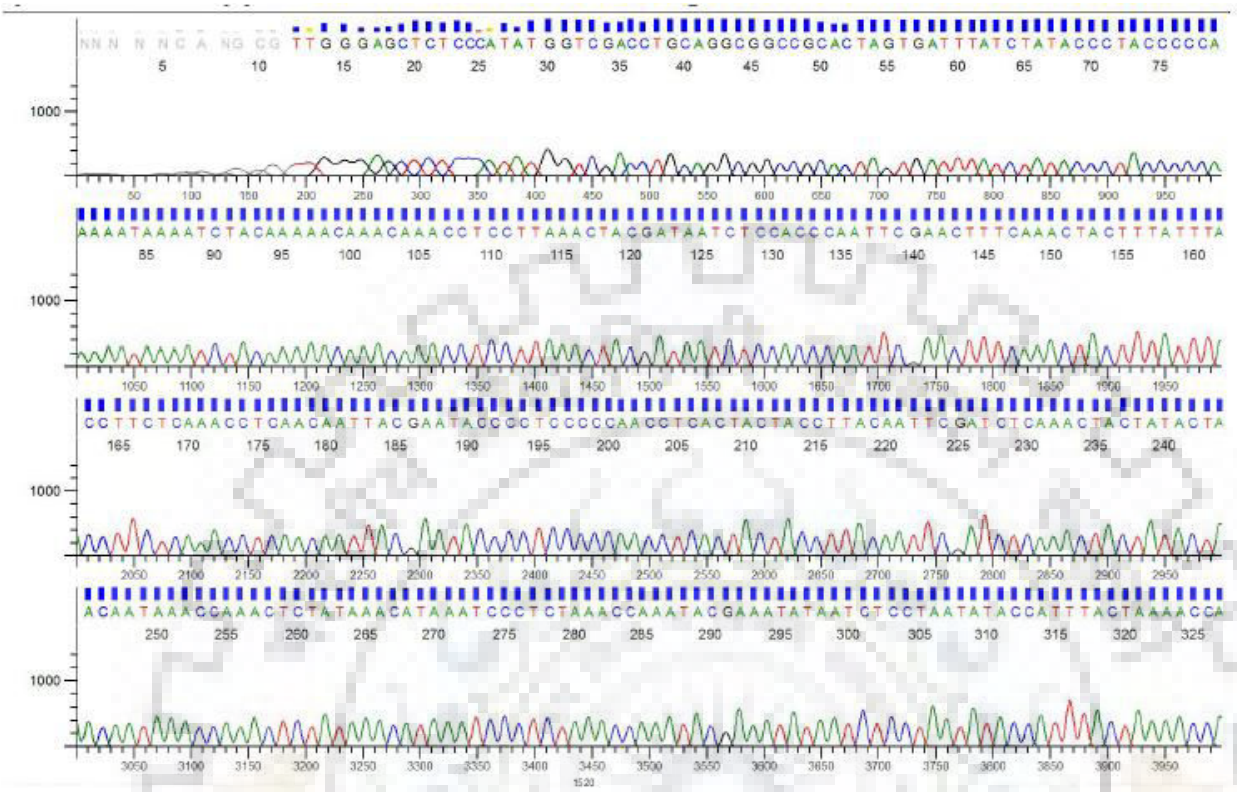
>N14b


Signal: G:214 A:273 T:348 C:326 AvgSig: 290
C#:30 W:C7 Plate Name:17_P-9
TS:48 CRL:1128 QV20+:1122
Inst Model/Name:3730/OPTIPLEX-XE-25102-001
Pure Base QVs:
Printed on: Jun 07,2017 10:53:10 IST

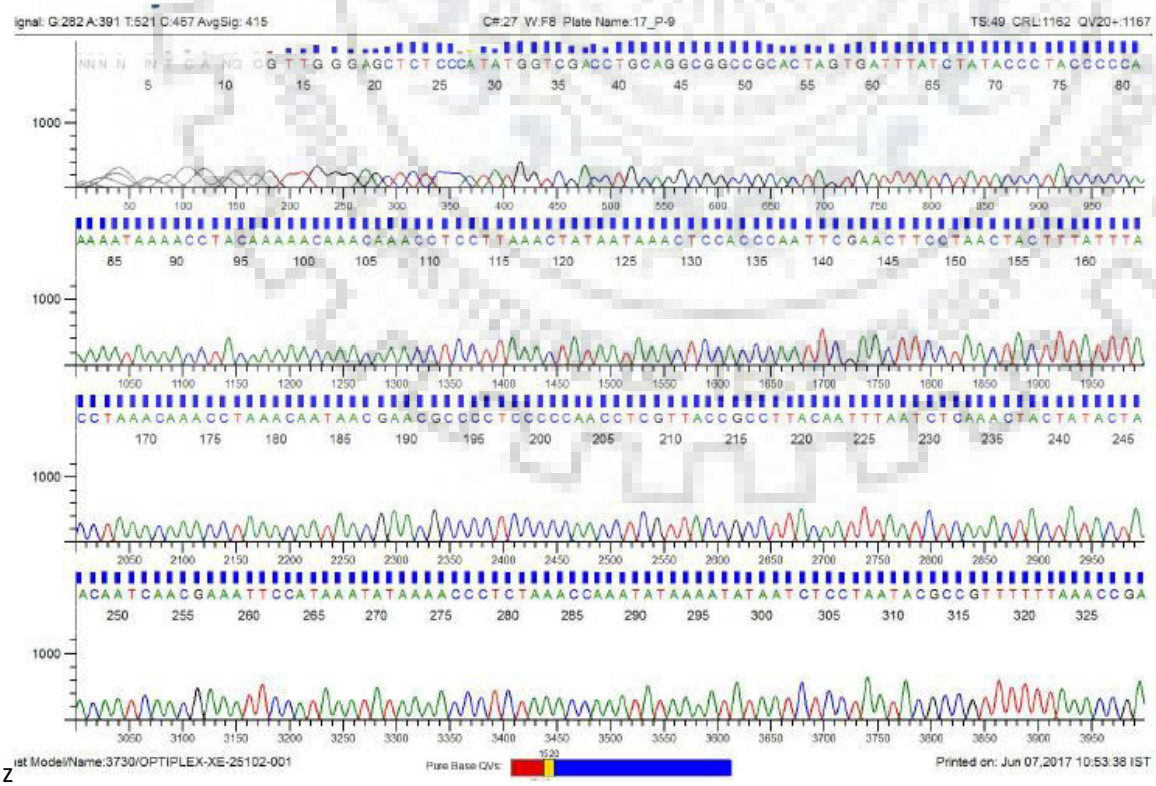
>N14c


G:223 A:320 T:388 C:382 AvgSig: 328
C#:26 W:G7 Plate Name:17_P-9
TS:47 CRL:1136 QV20+:1140

>N14d



>N14e



z

Paired cancer sample 15

>C15a

AAGGGGTTAGGGAGTTTTTTTTTTAGTTAAAGAAAGGGGTGACGGACGGTATTTGGAAAATCGGGTTATTTTTATTT
GAATATTGCGTTTTTTCGACGGGTTTAAAAAATAGCGTATTACGAGATTATATTTTGTATTTGGTTCGGAGGGTTTTAC
GTTTATGGAGATTCGTTGATTGTTAGTATAGTAGTTTGATATTAAATTGTAAGGTAGTAGCGAGGTTGGGGGAGGGG
CGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTGGGAAGTTCGAATTGGGTGGAGTTTATTACAGTTAA
AGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>C15b

AAGGGGTTAGGGAGTTTTTTTTTTGAGTTAAAGAAAGGGGTGATGGATGTATTTGGAAAATTGGGTTATTTTTATTT
GAATATTGTGTTTTTTAGATTAGTTTAAAAAATGGTGTATTATGAGATTATATTTTATATTTGGTTTGGAGGGTGTTAT
GTTTATGGAATTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTGGTAGTGAGGTTGGGGGAGGGG
AGTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTTAGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTA
AGGAGGTTTGTTTGTTCTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>C15c

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGACGAACGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTAGATCGGTTTAAGAAACGGCGTATTACGAGATTATATTTTATATTTGGTTTAGAGGGTTTTAC
GTTTACGGAATTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAACGAGGTTGGGGGAGGG
GCGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTCGGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

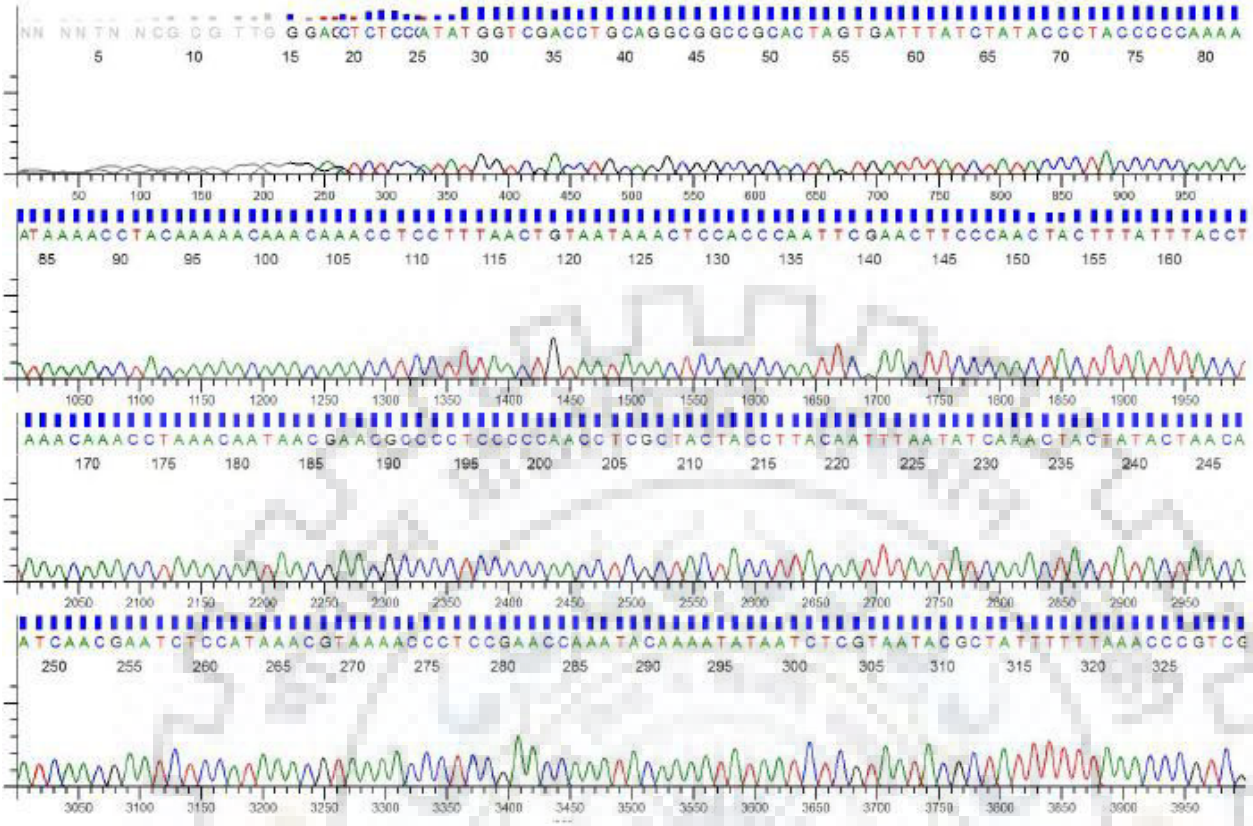
>C15d

AAGGGGTTAGGGAGTTTTTTTTTTGAGTTAAAGAAAGGGGTGATGGATGTATTTAGAAAATTGGGTTATTTTTATTTG
AATATTGTGTTTTTTAGATTGGTTTAAAAAATGGTGTATTATGAGATTATATTTTATATTTGGTTTGGAGGGTTTTACG
TTTATGGAATTTTGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGTAGTAGTGAGGTTGGGGGAGGGGA
GTTTGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTGGGGAAGTTTGAATTGGGTGGAGTTTATTATAGTTTAA
GGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

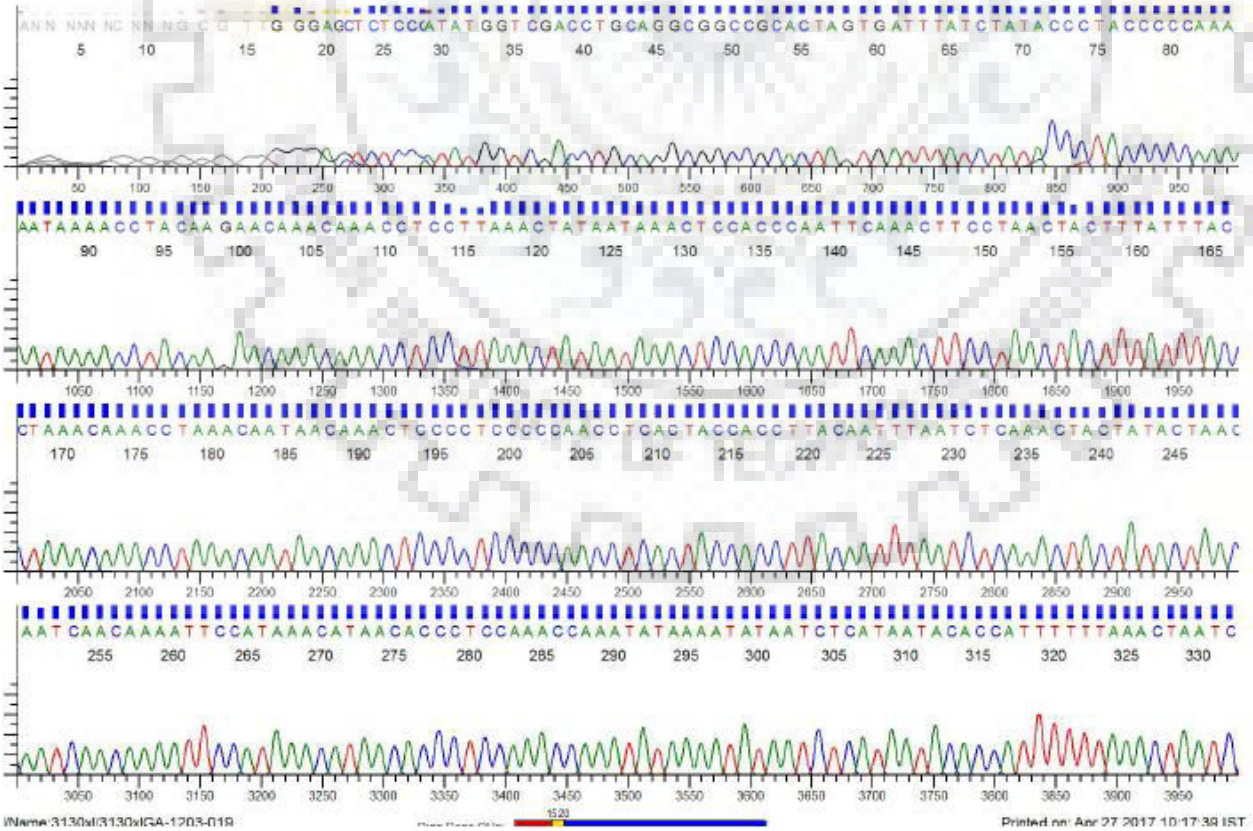
>C15e

AAGGGGTTAGGGAGTTTTTTTTTAGTTAAAGAAAGGGGTGATAGACCGTATTTGGGAAATTGGGTTATTTTTATTTTA
ATATTGCTTTTTTTCGACCGGTTTAAAAAACGCTCCCTTAGCAGATTATATTTCGTATTTGGTTCGGAGGGTTTTACGT
TTACGGGGTTTTGTTGATTGTTAGTATAGTATTTTGAGATTAAATTGTAAGGTGGTAGCGAGGTTGGGGGAGGGGC
GTTCGTTATTGTTTAGGTTCGTTTAGGGAAATAAAGTAGTTAGGAAGTTTGAATCGGGTGGAGTTTATTATAGTTTAA
GGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

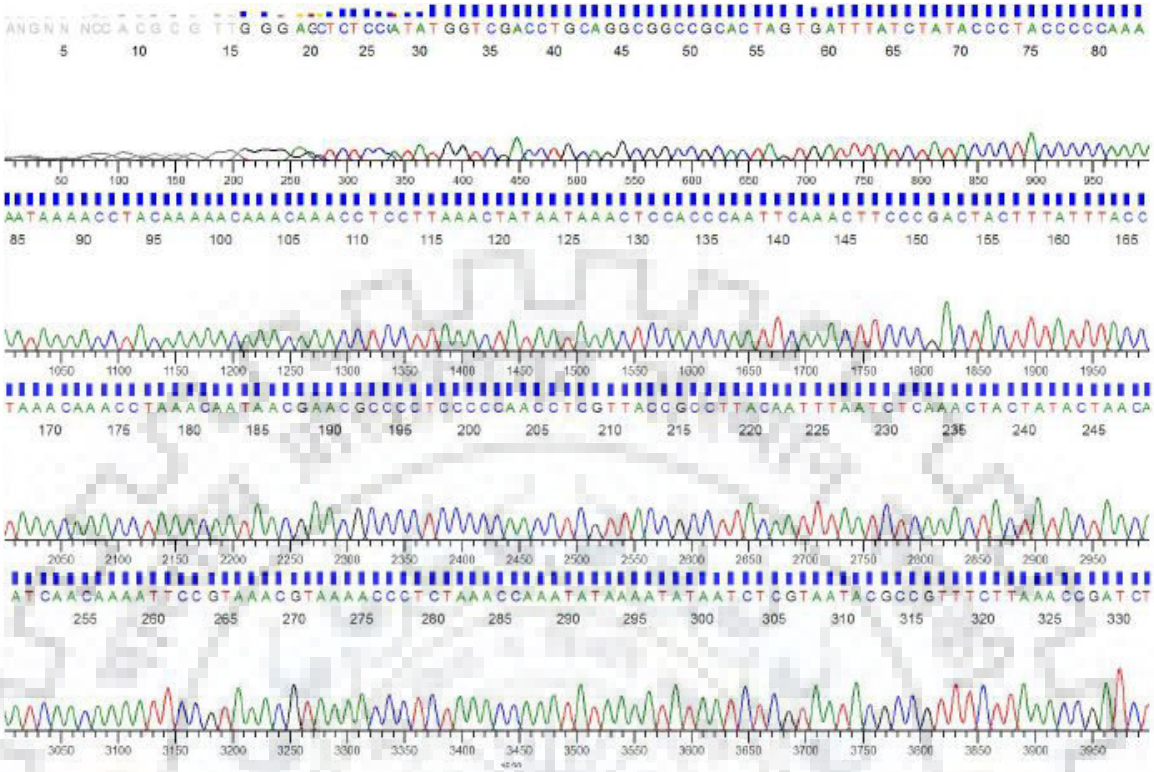
>C15a



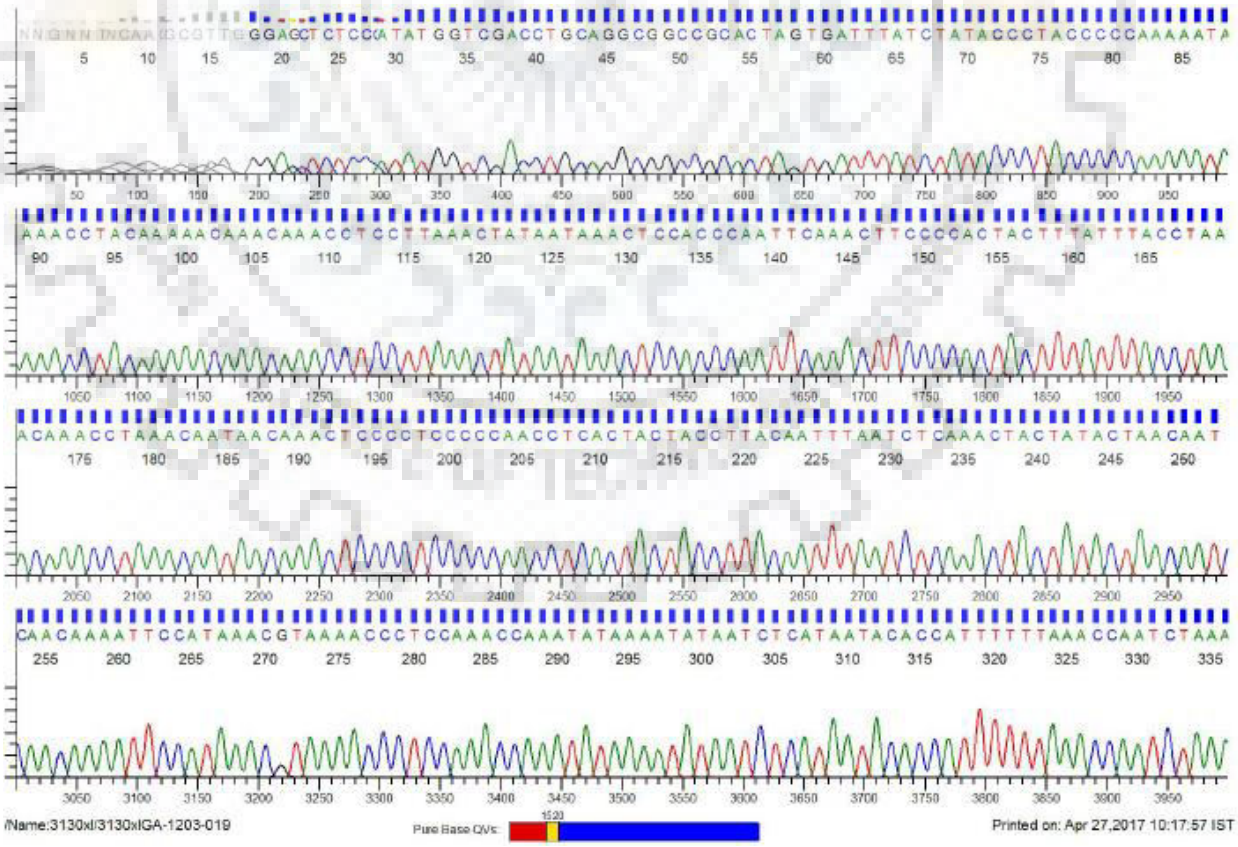
>C15b



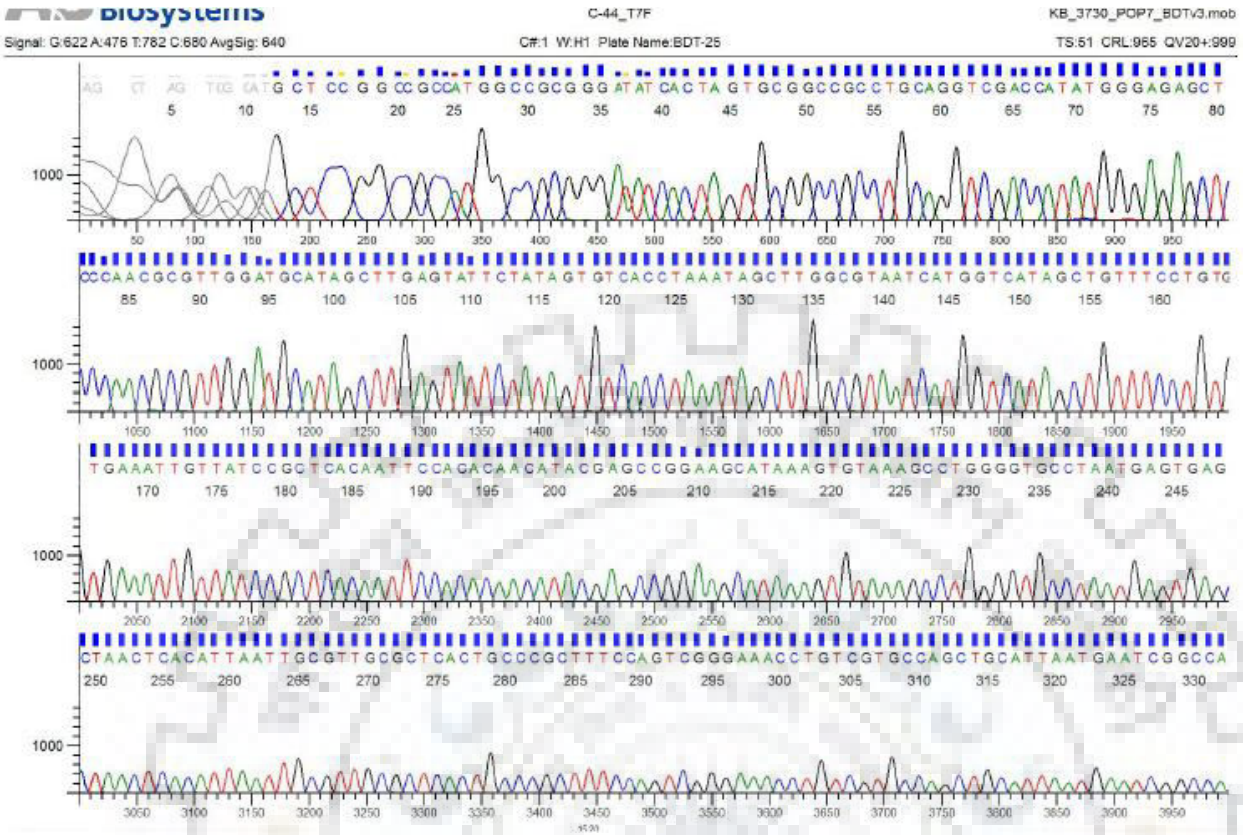
>C15c



>C15d



>C15e



Paired normal sample 15

>N15a

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGATGGATTTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTAGATCGGTTTAAAAAACGGTGTATTACGAGATTATATTTTATATTTGGTTTAGAGGGTTTTAC
GTTTACGGAATTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGCGAGGTTGGGGGAGGG
GTTTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAGGTAGTTAGGAAGTTCGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>N15b

AAGGGGTTAGGGAGTTTTTTTTTTAGTTAAAGAAAGGGGTGATAGACGGTATTTGGAAAATCGGGTTATTTTTATTTT
AATATTGCGTTTTTTTAATGGGTTTAATAAAGGGTATATTAGGAGATTATATTTTGTATTTGGTTCGGAGGGTTTTATA
TTTATGGAGTTTCGTTTATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAGCGAGGTTGGGGGAGGGGT
ATTTGTTATTGTTATGGTTTGAGTAGGTAAAAAAAGCGGTCGGGAAGTTCGAATTGGGTGGAGTTTATTATAGTTTA
AGGAGGTTTGTTTGTTTTTGTAGTTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

>N15c

AAGGGGTTAGGGAGTTTTTTTTTTCGAGTTAAAGAAGGGGGTGATAGACGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTAGATCGGTTTAAGAAACGGCGTATTATGAGATTATATTTTATATTTGGTTTAGAGGGTTTTAC
GTTTACGGAATTTTGTTGATTGTTAGTATAGTAGTTTGAGGTTAAATTGTAAGGCGGTAACGAGGTTGGGGGAGGG
GTGGGGTTTGTTTGTTCTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACTAGTGCGGCCGCCTGCAGGT
CGACCATATGGGAGAGCTCCCAACGCNTTGG

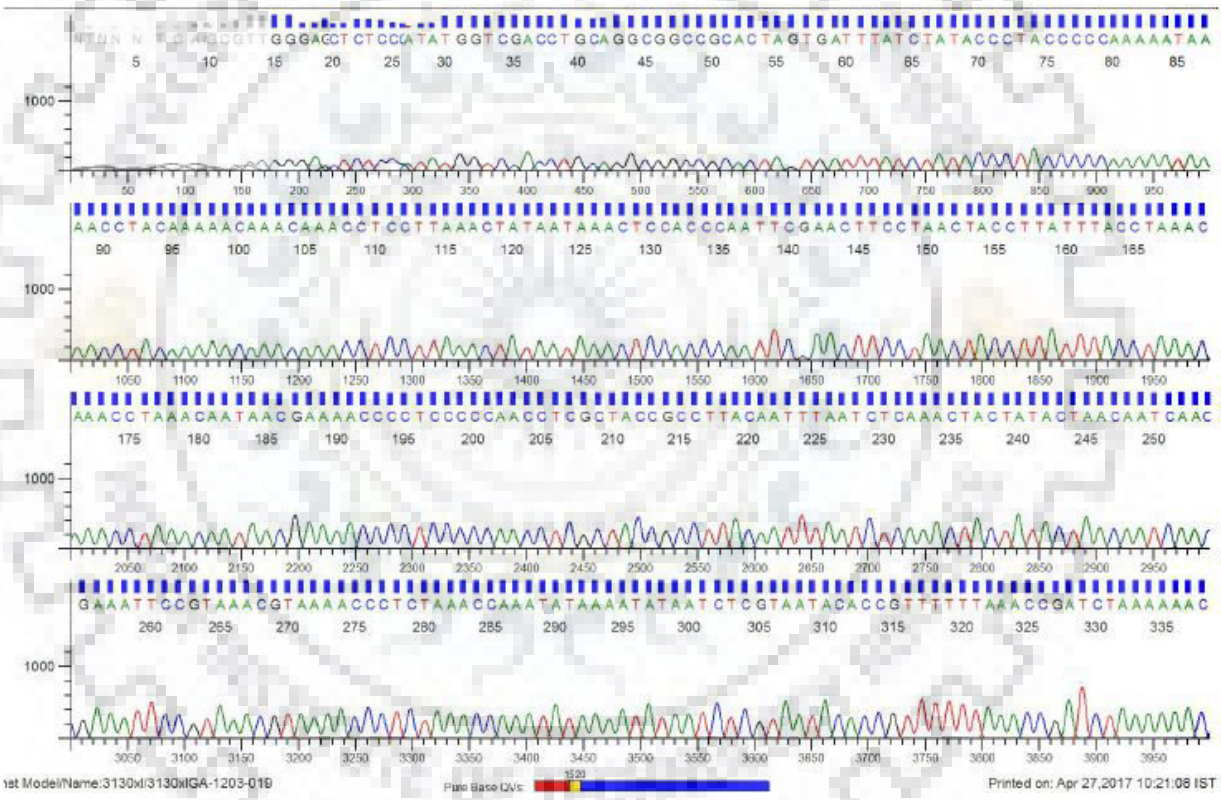
>N15d

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAAGGGGTGATGGATGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTAGATCGGTTTAAGAAACGGCGTATTACGAGATTATATTTTATATTTGGTTTAGAGGGTTTTAC
GTTTACGGAATTTCGTTGATTGTTAGTATAGTAGTTTGAGATTAAATTGTAAGGCGGTAACGAGGTTGGGGGAGGG
GCGTTCGTTATTGTTTAGGTTTGTTTAGGTAAATAAAGTAGTCGGGAAGTTCGAATTGGGTGGAGTTTATTATAGTTT
AAGGAGGTTTGTTTGTTTTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACT

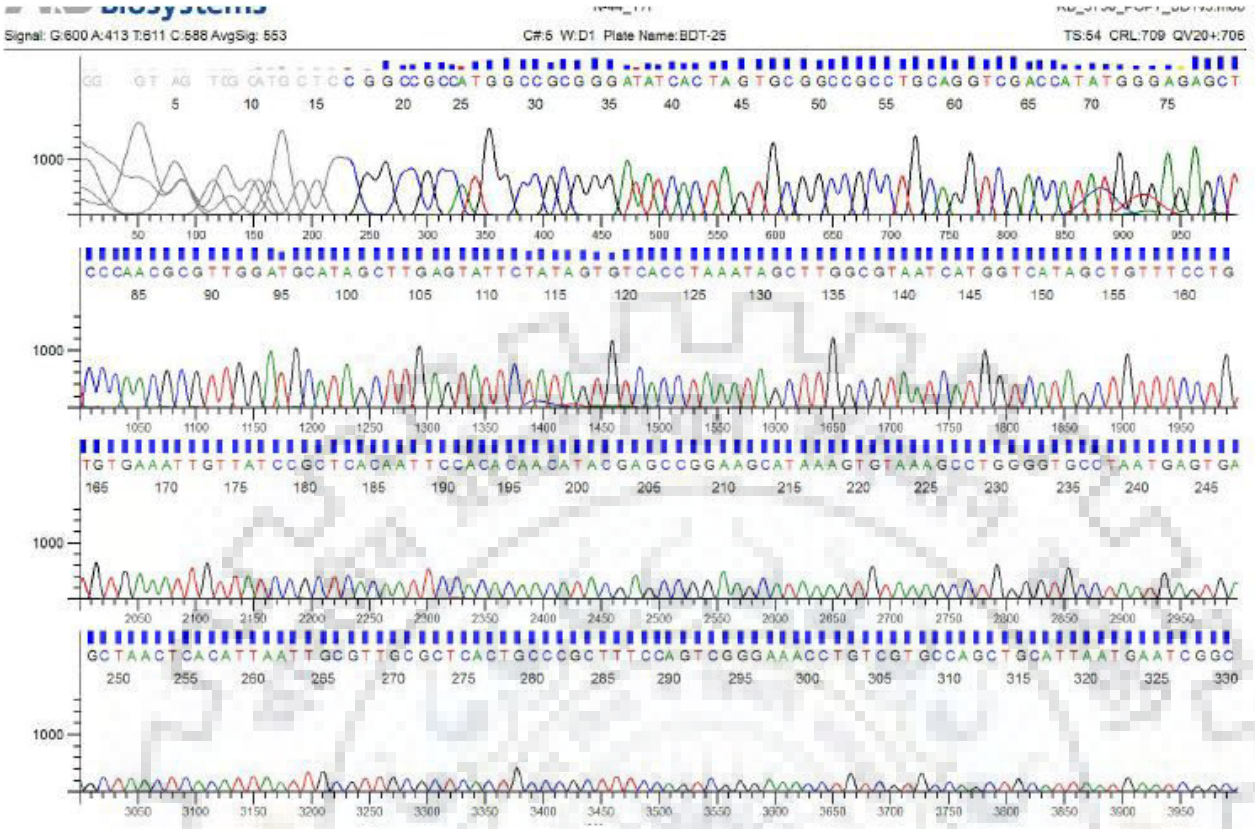
>N15e

AAGGGGTTAGGGAGTTTTTTTTTCGAGTTAAAGAAGGGGGTGATAGACGTATTTGGAAAATCGGGTTATTTTTATTC
GAATATTGCGTTTTTTAGATCGGTTTAAGAAACGGCGTATTATGAGATTATATTTTATATTTGGTTTAGAGGGTTTTAC
GTTTACGGAATTTTGTTGATTGTTAGTATAGTAGTTTGAGGTTAAATTGTAAGGCGGTAACGAGGTTGGGGGAGGG
GTGGGGTTTGTTTGTTCTTGTAGGTTTTATTTTTGGGGGTAGGGTATAGATAAATCACTAGTGCGCCCGCCTGCAGGT
CGAAAAAAAACG

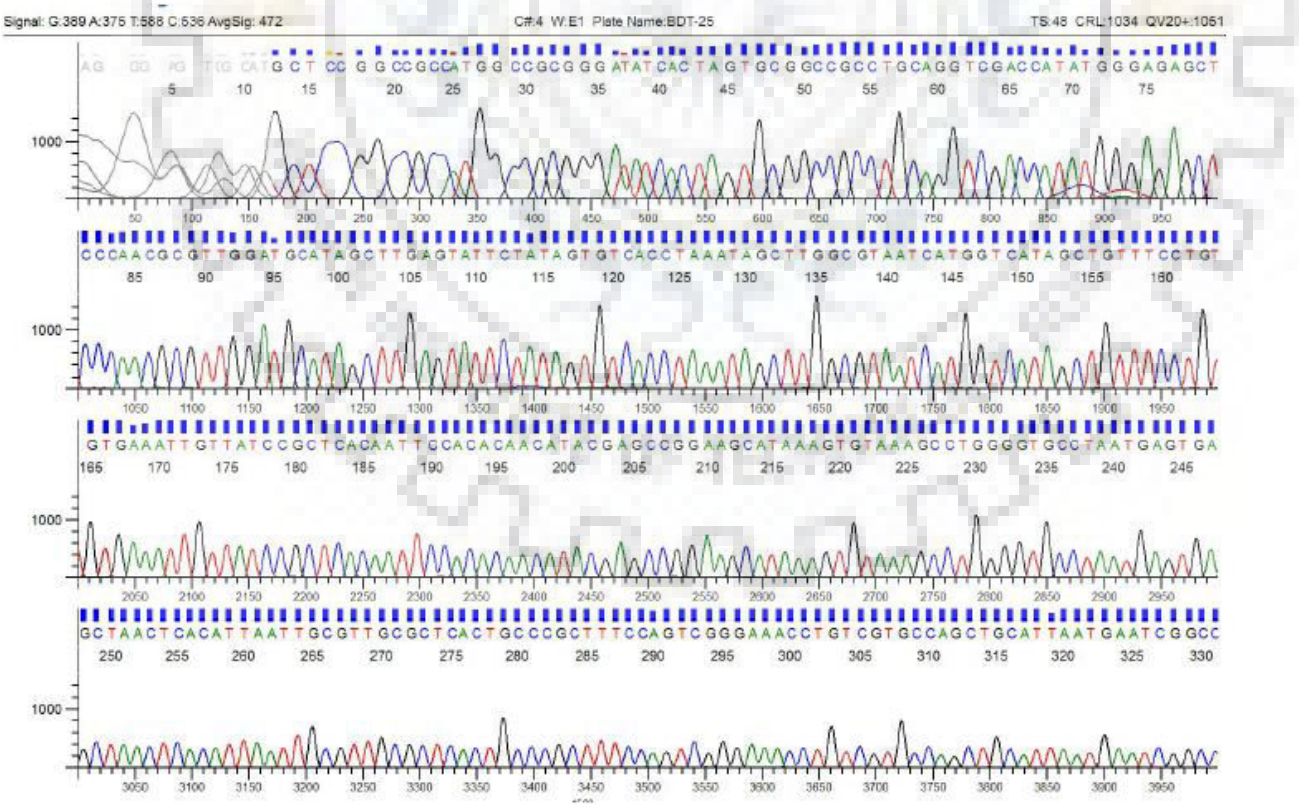
>N15a



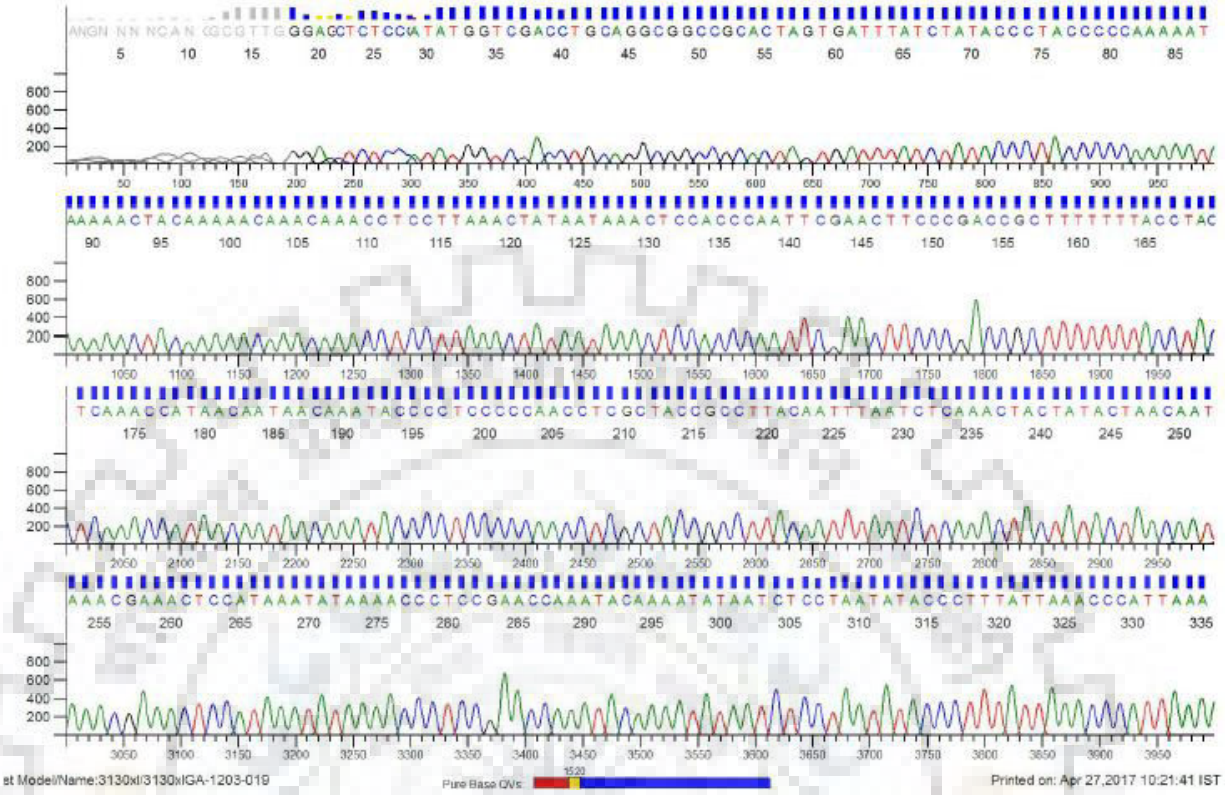
>N15b


Signal: G:600 A:413 T:611 C:588 AvgSig: 553
C#:6 W:D1 Plate Name:BDT-25
TS:54 CRL:709 QV20+:706

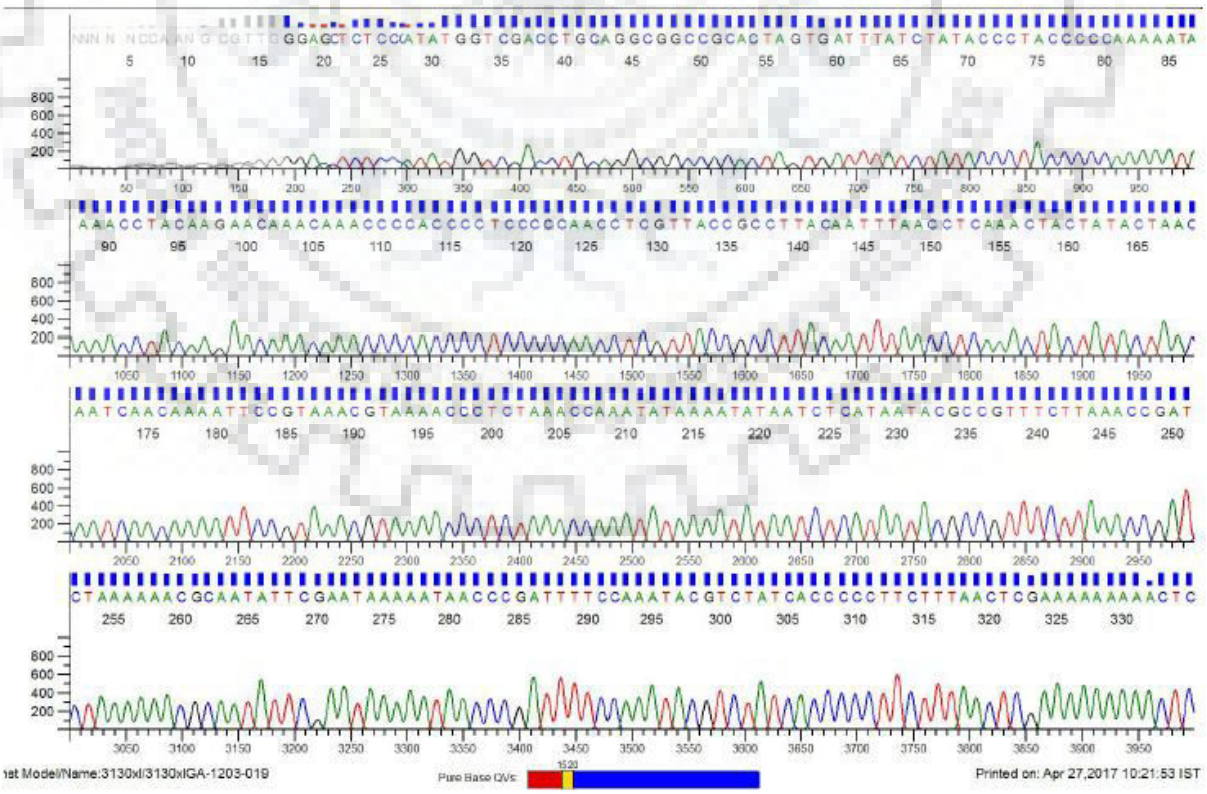
>N15c


Signal: G:389 A:375 T:588 C:636 AvgSig: 472
C#:4 W:E1 Plate Name:BDT-25
TS:48 CRL:1034 QV20+:1051

>N15d


Pure Base QVs
Printed on: Apr 27,2017 10:21:41 IST

>N15e


Pure Base QVs
Printed on: Apr 27,2017 10:21:53 IST

6.3.5. Calculations for plot R20A:

% Methylation was calculated as per following formula

∑M/∑(M+hM) X 100

M= Methylated CpG; hM = hypomethylated; CpG; N = Normal ; T = Tumor

N	67.9	65.5	54.6	71.1	64.5	58.3	58.1	78.3
T	46.6	52.6	32.2	45.7	71.7	48.3	54	44.6

The above value was plotted in R20A to depict % of methylation of L1 promoter in 8 paired normal cancer OSCC samples.

Calculation for plot R20C:

Mean % of methylation for normal (N=8) 64.7875

Mean % of methylation for tumor (T=8) 49.4625

Paired t-test: Monday, May 14, 2018, 10:01:50 AM

Data source: Data 1 in Notebook1

Normality Test (Shapiro-Wilk): Passed (P = 0.945)

Treatment Name	N	Missing	Mean	Std Dev	SEM
Col 1	8	0	64.787	7.776	2.749
Col 2	8	0	49.462	11.150	3.942
Difference	8	0	15.325	13.052	4.615

t = 3.321 with 7 degrees of freedom.

95 percent two-tailed confidence interval for difference of means: 4.413 to 26.237

Two-tailed P-value = 0.0127

The change that occurred with the treatment is greater than would be expected by chance; there is a statistically significant change (P = 0.013)

One-tailed P-value = 0.00637

The sample mean of treatment Col 1 exceeds the sample mean of treatment Col 2 by an amount that is greater than would be expected by chance, rejecting the hypothesis that the population mean of treatment Col 2 is greater than or equal to the population mean of treatment Col 1. (P = 0.013)

Power of performed two-tailed test with alpha = 0.050: 0.812

Power of performed one-tailed test with alpha = 0.050: 0.909

Calculations for plot R20B: % of methylation for each of the 20 CpG dinucleotide positions was calculated as per following formula

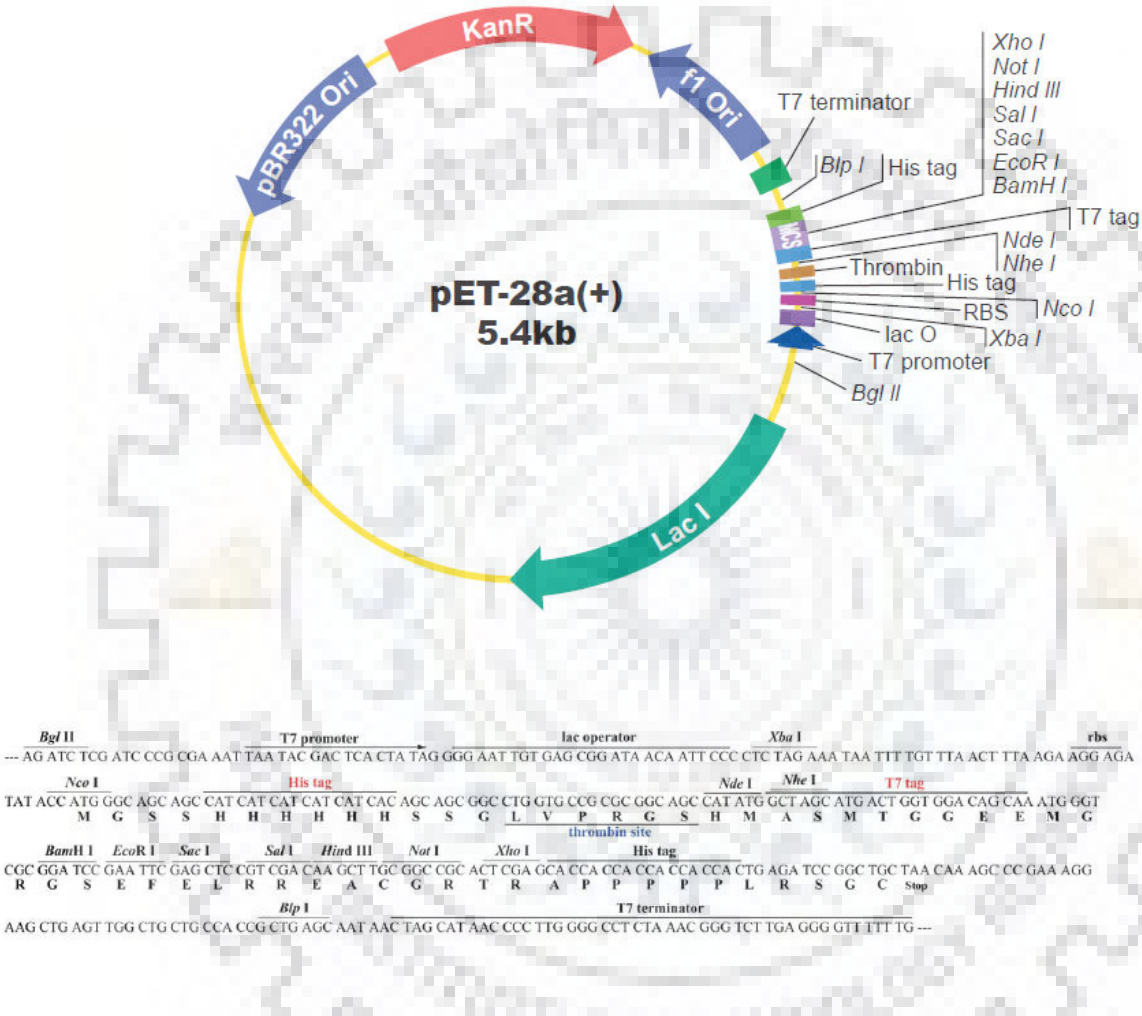
∑M/∑(M+hM) X 100

N	55.8	65.6	60	85.2	75.7	82.8	56.7	84.8	75.7	73.3	47	76.4
	51.5	65.7	68.5	60	62.8	65.7	20.5	42.8				
T	42.8	62.1	64.7	72.9	51.4	85.2	70	57.1	56.7	51.7	33.3	50
	48.5	30.5	33.3	48.6	50	40.5	20.5	23.5				

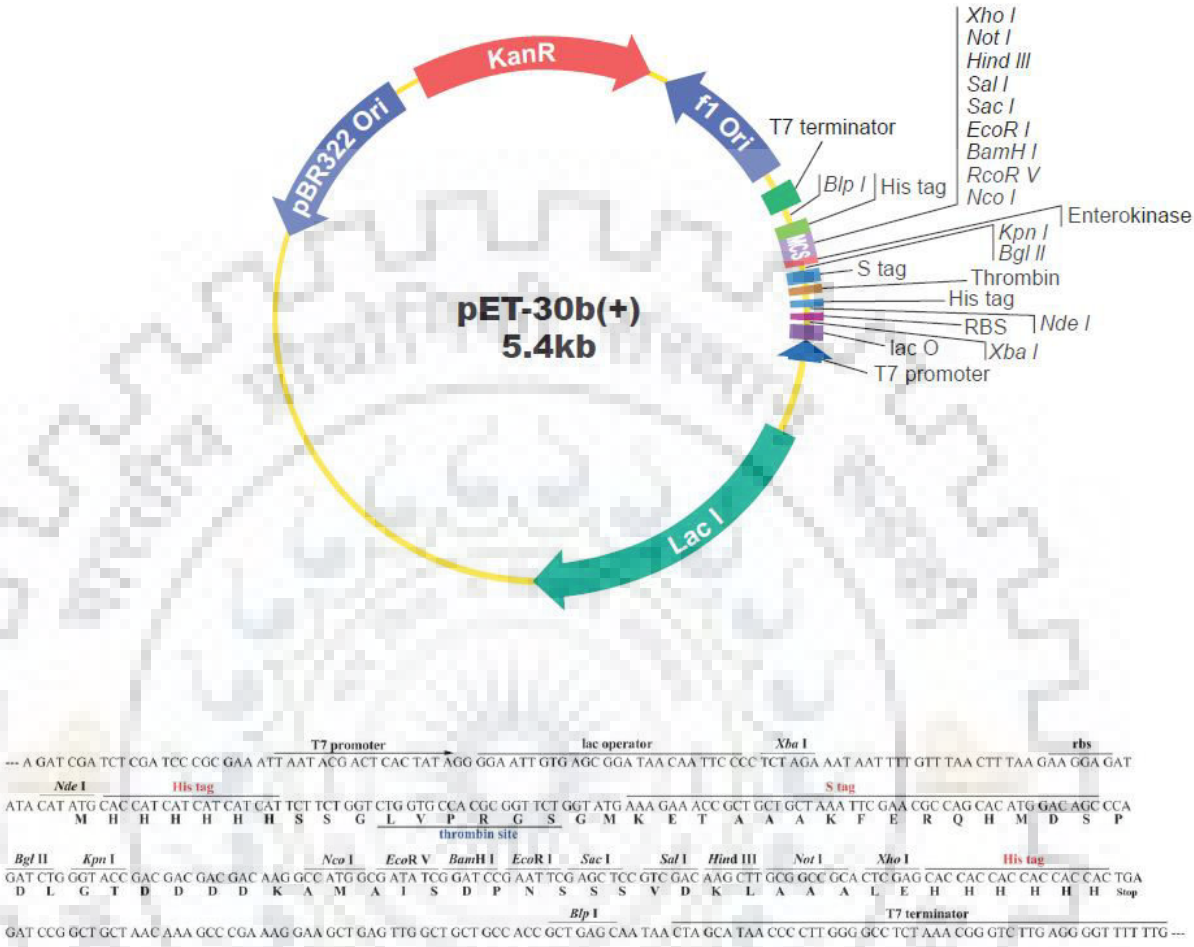
M= Methylated CpG; hM = hypomethylated; CpG; N = Normal ; T = Tumor

6.4.Appendix IV

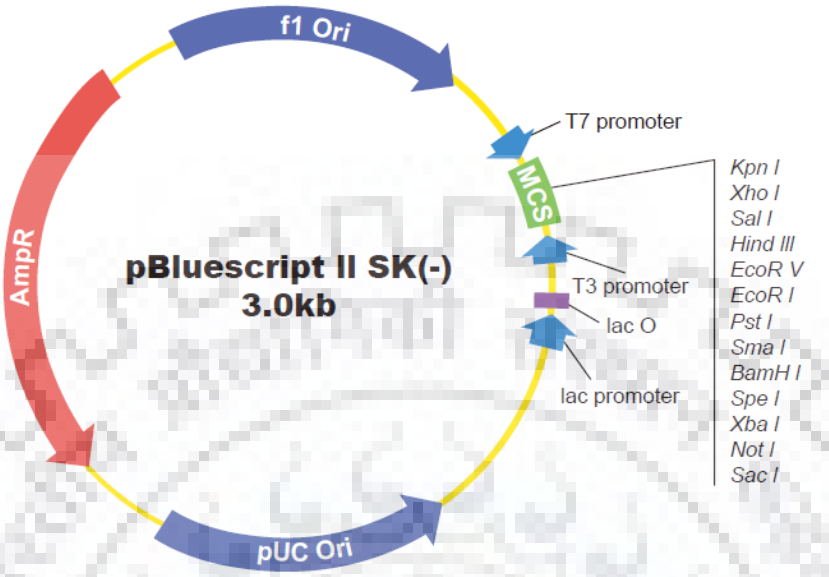
6.4.1. Plasmid maps used in this study:



Plasmid map of pET-28a cloning and expression vector



Plasmid map of pET-30b cloning and expression vector



T7 promoter

--- TAA TAC GAC TCA CTA TAG GGC GAA TTG GGT ACC GGG CCC CCC CTC GAG GTC GAC GGT ATC GAT AAG CTT GAT ATC GAA TTC CTG CAG CCC GGG GGA

BamH I Spe I Xba I Not I Sac I

TCC ACT AGT TCT AGA GCG GCC GCC ACC GCG GTG GAG CTC CAG CTT TTG TTC CCT TTA GTG AGG GTT AAT TGC GCG CTT GGC GTA ATC ATG GTC ATA ---

T3 promoter

Plasmid map of pBS cloning vector

6.5. Appendix V

6.5.1. Poster Abstract I

5th Annual International Conference on Advances in Biotechnology (BIOTECH 2015), Indian Institute of Technology Kanpur, Kanpur India March 2015

Title: Characterization of human LINE1 encoded reverse transcriptase

Savita Budania, Ramasare Prasad and Prabhat K Mandal

ABSTRACT

Long INterpersed Element 1 (LINE-1 or L1), a class of mobile genetic element belongs to the non-LTR retrotransposon category. L1 is actively jumping in our genome thus responsible for more than half of the human genome .Its ongoing activity increasing the genome size in every generation and also creating insertional mutaions. An active L1 is 6.0 kb in length, encodes two proteins designated as ORF1p (40kD) and ORF2p (150kD) which is required for jumping the L1mRNA. ORF1p is a single stranded nucleic acid binding proteins where as ORF2p has distinct reverse transcriptase (RT) and endonuclese (EN) activity. It is hypothesized that L1 RNA jumps by coupled reverse transcription and integration mechanism where L1mRNA reverse transcribed by ORF2p mediated RT activity at the genomic target site. RT activity is the most crucial step in the process of human L1 retrotransposition. Here we are demonstrating the human RT activity by cloning and purifying the RT domain only.

Key words: LINE1, genome, ORF1p, reverse transcriptase.

6.5.2. Poster Abstract II

Conference: "Biotech Day 2017" at Department of Biotechnology Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, March 2016

Title: L1 retrotransposon activity in head and neck cancer patients

Savita Budania, Debpali Sur, Jitendra Nangal and Prabhat K Mandal

ABSTRACT

Retrotransposons are the major structural and functional modulator in any genome as these sequences jump from one place to another; thus can change the size of the genome as well as regulate the expression of genes close to the site of insertion. Human genome is the major hub for a non LTR retrotransposon called Long INterpersed Elements (LINEs or L1). There are almost 500,000 copies of L1 occupied around 30% of the human genome. Although most of them are inactive around 150-200 copies are actively jumping in human genome particularly in brain, germ cells and cancers. Jumping of L1 has several consequences and most detrimental when it insert within an essential gene thus can completely abolish the function of that gene. An active L1 is 6 kb in length and encodes two proteins (ORF1 and ORF2). Both proteins convert L1 mRNA to cDNA in a very complicated mechanism. The L1 mRNA is transcribed from its own promoter which is situated in L1 5'UTR. Here we are showing the L1 activity in head and neck cancer, the 3rd highest in globe and commonest malignancies in Indian Males. We collected samples from a Government hospital located at Bikaner Rajasthan, the highest hit region across India. We detected significant expression of L1 ORF1p in cancer tissues by employing Immunohistochemistry (IHC). We have also seen very high level of L1 promoter hypomethylation in cancer tissues compared to normal counterpart. In general L1 do not transcribe in normal cells due to heavy methylation at L1 promoter and thus no retrotransposition. This suggests that due to hypomethylation of promoter L1s are highly transcribing in cancer tissues and might be jumping frequently in cancer genome; and making the cancer more aggressive.

6.5.3. Poster Abstract III

The Mobile Genome: Genetic and Physiological Impact of Transposable Element EMBO EMBL Symposium, Heidelberg, Germany, October 2017

Title: Extensive LINE-1 promoter hypomethylation and ORF1p expressionin Oral squamous cell carcinoma

Savita Budania, Debpali Sur, Shilpi Saxena,Jitendra Nangal, Manash Biswas, Ramasare Prasad, Vijay Yadav, Dustin C. Hancks, and Prabhat K. Mandal

ABSTRACT

Oral squamous cell carcinoma (OSCC) ranked eighth among all common cancer in the world with a very high mortality rate. The occurrence of OSCC is quite high in Indian population especially among the youth because of excessive use of tobacco. Here we investigated Long INterpersed Element 1 (LINE-1 or L1) retrotransposon activity in OSCC samples in the same population. There are almost 500,000 copies of L1 occupied around 30% of the human genome. Although most of them are inactive, around 150-200 copies are actively jumping in human genome. L1 encodes two proteins designated as ORF1p and ORF2p and expression of both proteins are critical for the process of retrotransposition. Here we have analysed L1 activity by looking the presence of ORF1p in 12 paired cancer-normal tissues. Immunohistochemistry (IHC) with human ORF1 antibody showed the presence of ORF1p in more than 50% cancer samples (7 out of 12). Investigating L1 promoter methylation status, showed significant hypomethylation of L1 promoter in cancer tissues compared to its normal counterpart. Our data shows very high L1 retrotransposon activity in OSCC which might have some significant role in the onset and progression of this particular type of cancer.