IDENTIFICATION OF DELETERIOUS nsSNPs IN CELL CYCLE REGULATORY GENE E2F1 AND THEIR ROLE IN SUSCEPTIBILITY TO COMMON CANCERS IN HIMACHAL PRADESH POPULATION

A thesis submitted in fulfillment for the requirements of the degree of

Doctor of Philosophy

by

SANJAY SINGH

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology

Waknaghat, Solan, HP - 173234, India

May, 2019

`

TABLE OF CONTENTS

`

`

`

`

DECLARATION

I certify that:

a. The work contained in this thesis is original and has been done by me under the guidance of my supervisor.
b. The work has not been submitted to any other organization for any degree or diploma.
c. Wherever, I have used materials (data, analysis, figures or text), I have given due credit by citing them in the text of the thesis.

Sanjay Singh **Date:** _____/_____/2019

Enrollment No. 136556

Department of Biotechnology & Bioinformatics

Jaypee University of Information Technology

Waknaghat, Solan, India

`

CERTIFICATE

This is to certify that the thesis entitled, **"Identification of deleterious nsSNPs in cell cycle regulatory gene *E2F1* and their role in susceptibility to common cancers in Himachal Pradesh population"** which is being submitted by **Sanjay Singh (Enrollment No. 136556)** in fulfillment for the award of degree of **Doctor of Philosophy** in **Biotechnology** at Jaypee **University of Information Technology, Waknaghat, India** is the record of candidate's own work carried out by him under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Dr. Harish Changotra

Associate Professor

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology

Waknaghat, India

`

ACKNOWLEDGEMENT

The list of the people I need to thank will not fit to a single Acknowledgement section. I just mention some people whose contribution is obvious. It is my privilege and honor to express my gratitude to the following without whose guidance and support I would not have been able to complete my Ph.D thesis.

My first debt of gratitude must go to my supervisor, **Dr. Harish Changotra**. The insuperable contribution and peer support rendered by my supervisor is incomprehensible. He patiently provided the vision, encouragement and advice necessary for me to proceed through the doctorial program and complete my dissertation. He has been a strong and supportive adviser to me throughout my research

I wish to thank my parents, **Uday Pratap Singh** and **Rita Singh** and **my brother Ajay Singh**. Their love, support and motivation provided me inspiration and encouragement, and they were the driving force to successfully complete my work. I owe them everything and wish I could show them just how much I love and appreciate them.

I emphatically express my loyal and venerable thanks to JUIT administration, **Prof (Dr) Vinod Kumar (Vice Chancellor), Prof (Dr) S.D. Gupta (Director & Dean, Academics & Research), Maj. Gen (Retd.) Rakesh Bassi (Registrar), Prof (Dr) Y. Medury (Former COO, JES), Prof (Dr) S.K. Kak (Former vice chancellor), Brig (Retd.) K.K. Marwah (Former Registrar), Brig (Retd.) Balbir Singh (Former Registrar), Prof (Dr) T.S. Lamba (Former Dean, Academic & Research), Prof (Dr) R.S. Chauhan (Former Head, BT & BI), Prof (Dr.) Sudhir Kumar (Head, BT & BI)** for providing opportunity to pursue a Doctorate Degree, teaching assistantship and advanced lab infrastructure to accomplish this scientific venture of my life.

I am also highly beholden to **Prof (Dr) Manish Gupta, Prof (Dr) Rajeev Kumar Seam, Indira Gandhi Medical College, Shimla** for their guidance, valuable suggestions and tremendous cooperation.

I am also thankful to **Prof (Dr) Karanjeet Singh, Dr. Jitendraa Vashistt, Dr Tirthraj Singh** for their guidance and valuable suggestions throughout my research work. I wish to convey my sincere thanks to all the faculty members of Department of Biotechnology and Bioinformatics, for their help and guidance at the various stages of this study.

`

I am also thankful to all the members of technical and non-technical staff of the department, especially **Mr. Baleshwar, Mrs. Mamta, Mrs. Somlata Sharma, Mrs. Sonika, and Mr. Ismail** for their assistance and valuable contributions.

I am fortunate to have friends who have always stood beside me. I extend my heartfelt thanks to all my friends for their sustained support and ever needed cooperation. It is my pleasure to express my gratitude to all research scholars of the Biotechnology & Bioinformatics Department for keeping me blessed with best wishes.

I would like to express my heartfelt gratitude to all those who have contributed directly or indirectly towards obtaining my doctorate degree and apologize if have missed out anyone.

Last, but not the least, I thank the one above all of us, omnipresent God, for answering my prayers, for giving me the strength to plod on during each and every phase of my life.

All may not be mentioned, but no one is forgotten.

Sanjay Singh

`

ABSTRACT

Deregulated expression most of the E2Fs members have been observed in many human cancers. In this study, we examined the association damaging single nucleotide polymorphisms (SNPs) of *E2F1* gene with risk of lung, cervical, head and neck cancer in 395 patients and 230 cancer-free-controls. Before conducting case-control association study, we prioritized SNP on the basis of computational analysis. In the computational analysis, six amino-acid variants (Cys227Phe, Arg252His, Val295Asp, Cys298Tyr, Arg56Trp, and Tyr59Cys) of E2F1 protein were most deleterious mutations. The analysis was on the basis of the effects of mutation on protein structure, function, and amino-acid conservation. Molecular dynamics simulations performed to investigate the deep insight about structural changes. In the statistical analysis of genotypes, the mutant genotype TT of rs3213172 (Arg252His) was associated with the elevated risk for lung cancer (LC) with an (odd ratio (OR) = 4.8; 95% confidence interval (CI) = 2.479-9.560, P <0.001) in the homozygous model. This nsSNP was significantly associated with HNC and cervical cancer risk in genetic association analysis. The rs3213173 (Val276Met) polymorphism was associated with lung and cervical cancer risk in genotypic as well as *allelic* model. Current study revealed significant association of rs3213176 (Gly 393Ser) polymorphism with LC and HNC risk. The rs2071054 (C/T) polymorphism was also associated with HNC cancer risk in *allelic* model. Haplotypes of nsSNPs were significantly associated LC, HNC and cervical cancer risk. These results suggest that the rs3213172 (C/T), rs3213173 (C/T), rs3213176 (G/A) and rs2071054 (C/T) polymorphisms of *E2F1* gene could be used as an effective biomarker for genetic susceptibility to LC, HNC and cervical cancer. Larger studies on different ethnic groups will provide better insight about these polymorphisms.

`

LIST OF ABBREVIATIONS

%	Percent
˚C	Degree Celsius
µg	Micro gram
aa	Amino acid
bp	Base pair
DNA	Deoxyribonucleic acid
Kb	Kilo base
UTR	Untranslated region
COSMIC	Catalogue of Somatic Mutation in Cancer
NCBI	National Centre for Biotechnology Information
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SASA	Solvent Accessible Surface Area
Rg	Radius of Gyration
WHO	World Health Organization
mL	Milliliter
mM	Millimoler
Min	Minutes
E2F	Early gene factor-2
ER	Endoplasmic Reticulum
EtBr	Ethidium bromide
GWAS	Genome-wide association studies
SNP	Single nucleotide polymorphisms
PCR	Polymerase chain reaction
RFLP	Restriction Fragment Length Polymorphism
PTEN	phosphatase and tensin homolog
BRCA1	BReast Cancer 1
BRCA2	BReast Cancer 2
CDKs	Cyclin-dependent kinases
TF	Transcription factor
CHIP	chromatin immunoprecipitation
DHFR	Dihydrofolate reductase
TK	Tyrosine kinase
FAS	Fetal Alcohol Syndrome
PFKB,	Phophofructokinase B
Sirt6	silent information regulator 6
PDK	Phosphoinositide-Dependent Kinase
TOP1MT	Mitochondrial Topoisomerase I
NANOG	Nanog Homeobox protein
v-ATPase	Vacuolar-type ATPase
ATG1	Autophagy related 1
DRAM1	DNA Damage Regulated Autophagy Modulator 1
MAP1LC3	Microtubule-associated proteins 1A/1B light chain 3B
APAF1	Apoptotic protease activating factor 1
PUMA	p53 upregulated modulator of apoptosis

`

TDP1	Tyrosyl-DNA Phosphodiesterase 1
TDP2	Tyrosyl-DNA Phosphodiesterase 2
NIEHS	National Institute of Environmental Health Sciences SNPs program
MAF	Macrophage-activating factor
LC	Lung cancer
HNC	Head and neck cancer
CaCx	Carcinoma of the Cervix
PAHs	Polycyclic aromatic hydrocarbons
IARC	International Agency for Research on Cancer
TSNAs	Tobacco-specific nitrosamines
GST	Glutathione-S-transferase
ROS	Reactive dangerous oxygen species
ADH	Alcohol dehydrogenase
EBV	Epstein–Barr virus
HPV	Human papillomavirus
OSCC	Oral squamous cell carcinoma
SCC	Squamous cell carcinoma
HIV	Immunodeficiency virus
SIL	Squamous intraepithelial lesions
CIN	Cervical intraepithelial lesions
CYP1A1	Cytochrome P4501A1
GST	Glutathione S-transferase
GSTM1	Glutathione S-transferase M1
GSTT1	Glutathione S-transferase T1
NSCLC	Non-small cell lung cancer
NAT2	N-acetyltransferase 2
ERCC1	Excision Repair Cross-Complementation Group 1
XPD	X-Ray Photoelectron Diffraction
OGG1	8-Oxoguanine glycosylase 1
MDM2	Mouse double minute 2
FAS	Fetal Alcohol Syndrome
FASL	Fetal Alcohol Syndrome ligand
NK	Natural killer
IL-2	Interleukins
IFN-Y	Interferon gamma
BER	Base excision repair
ARF	Acute Renal Failure
DP	Dimerization partners
HCC	Human hepatocellular carcinoma
PDB	Protein Data Bank
GVGD	Grantham Variation and Grantham Deviation
SIFT	Sorting Intolerant from Tolerant
PhD-SNP	Predictor of Human Deleterious-SNP
PROVEAN	Protein Variation Effect Analyzer
SNAP	Screening of Non Acceptable Polymorphism
HMMs	Hidden markov models
MUSCLE	Multiple Sequence Comparison by Log-Expectation
ANOLEA	Atomic Non-Local Energy Assessment

`

SDM	Site-directed mutator
GROMACS	GROningen MAchine for Chemical Simulations
MD simulation	Molecular dianamic simulation
IGMC	Indira Gandhi Medical College
RBC	Red Blood cell
WBC	White blood cell
RT	Room temperature
TAE buffer	Tris-acetate-EDTA buffer
NEB	New England Biolabs
OSSE	Online Sample Size Estimator
TNM	Tumor Lymph node Metastasis
OR	Odd ratio
CI	Confidence Interval

`

LIST OF FIGURES

`

LIST OF TABLES

INTRODUCTION

Cancer is a group of diseases which is characterized by uncontrolled cellular proliferation and it spreads to any part of the body. It differs from benign tumors, which do not spread to other parts of the body [1]. Common signs and symptoms of most the cancers include a lump, unusual bleeding, prolonged cough, change in bowel movements, and unusual weight loss. These symptoms are not specific for any cancer and they may have other causes [2, 3].

Globally, cancer is responsible for the second highest pool of deaths following cardiovascular diseases. According to the GLOBOCAN 2018 report, 18.1 million cancer incidences and 9.6 million cancer mortality was reported in 2017 worldwide [3, 4]. Worldwide, lung cancer has highest incidences and mortality [3, 5].

In India, breast, oral cavity, lung, cervical, and stomach cancers contribute around 46% cancer incidences and these cancers contribute to 46.5% of cancer mortality [3, 4]. This may be due to high consumption of tobacco in any form by individuals, drinking alcohol, lower awareness about this deadly disease, and socioeconomic status of the population.

Genetic dissimilarity is a key contributor of cancer risk. Genetic association of different cancers have been elucidated by genome-wide association studies (GWAS). BRCA1 and BRCA2 are the well known genes whose mutations are associated with breast cancer susceptibility and these genes are used to check breast and ovarian cancer susceptibility in genetic counselling of these cancer [6]. Uncontrolled proliferation of cancerous cells is mainly governed by cell-cycle regulatory genes like, p53, pRB, PTEN, BRCA1 and BRCA2 [6, 7]. In many human cancers, deregulated expression of Early Gene Factor 2 (E2F) family members has also been detected [8].

E2F transcription factor 1 (E2F1) is a transcription factor (TF) that regulates the expression of many genes involved in cell growth and differentiation [9-11]. When pRB is bound to the E2F1, it inhibits the transcriptional activity of E2F1–Dimerization Partner (DP) complex. This interaction also prevents the binding of transcription co-regulators to the promoters of their target genes [12]. During cell-cycle progression, E2F1 is detached from E2F1-pRB complex due to cyclin-dependent kinases (CDKs) mediated phosphorylation of pRB. E2F1 then binds to the promoters of the target genes, increases their expression and these are involved in the synthesis of DNA and mitosis [13-15] . E2F1 has been discovered 30 years ago and sufficient data is

available now that suggests an important role of this transcription factor in cell cycle progression. Moreover, it is also involved in various other processes of the cell like apoptosis induction [16], senescence [17] and DNA-damage response [18]. "Genome-wide location studies revealed other important functions of this transcription factor like it also binds to the promoter of various genes and may alter expression of these genes which are involved in several cellular pathways [13-15]".

Initially, it was believed that E2Fs play role in the transcriptional activation of genes that are involved in G1/S transition and DNA replication [9-11, 19]. Recent molecular biology techniques such as DNA microarray and chromatin immunoprecipitation (ChIP) assay provided new insights about E2F target genes and new target genes were found which play roles in DNA damage response, factors associated in chromatin assembly and condensation, segregation of chromosome, mitotic spindle checkpoint, cell differentiation and development apoptosis induction [20-23]. Mutations in E2F1 gene and genes targeted by E2F1 protein are in the list of most commonly altered genes in the human cancers [24]. Deregulated expression or activities of most of the E2Fs family have been observed in various cancers [8, 25-27]. It has been clearly seen that E2F1 deficient mice consistently exhibited a defect in thymocyte apoptosis and these E2F1deficient mice have more risk for developing tumor [28, 29]. It has been observed that the apoptotic activity was highest in E2F1 among E2Fs family it induces apoptosis by p53 dependent as well as independent mechanisms [30].

Some important genes such as p53 and Myc play an important role to regulate cancer cell metabolism [31, 32]. E2F1 is also associated with metabolic reprogramming of cancer cells. It regulates the expression genes like, *DHFR, TK, FAS, PFKB, Sirt6, PDK TOP1MT, EVOVL2, NANOG, v-ATPase, ATG1, DRAM1, MAP1LC3* etc. that have an impact on cancer metabolism. This metabolic reprogramming is required to accomplish the need of cancerous cells, which is considered as hallmarks of cancer [33, 34].

According to a recent study, E2F1 determine cell fate by its expression level. Low levels of this protein are sufficient to increase the expression of genes which require to G1 to S transition. Intermediate levels increase the expression of growth arrest genes such as p18, p19 and p27 while, E2F1 mediated apoptosis is determine by its high levels which increase the expression of key apoptotic genes including *APAF1, PUMA, HRK* and *BIM* [35-38].

E2F1 is the extensively investigated member of E2Fs family. The *E2F1* gene is located on chromosome 20q (length of 10.71 kb). It contains 6 introns and 7 exons, which encodes E2F1 protein (437 residues). The E2F1 functional protein has four domains DNA binding (127-192), coiled-coil (177-217 residues) Heptad repeat (201-245), marked-box domain (252-367) and transactivation domain (368- 347) [39, 40]. N-terminal DNA-binding domain (127-192 residues) interacts with TDP1 and TDP2 and stimulates E2F1-dependent transcription. C-terminal acidic amino acid transactivation domain of E2F1 is involved in regulation of DNA synthesis, cell-cycle progression, cellular proliferation and apoptosis [40, 41]. Marked box domain (of E2F1) and its flanking region interact with Jab1 and this region of E2F1 is associated with apoptosis induction [42].

The genetic makeup of an individual has an important contribution in disease susceptibility. One of the well known susceptibility factors is single nucleotide polymorphisms (SNPs). It may influence the expression pattern of a gene, protein function and predisposition to different diseases [43, 44]. Different genetic epidemiological studies have shown the roles of SNPs in many genes with cancer susceptibility. Hence, determination of SNPs, which act as genetic susceptibility markers for tumors, has become an important current research interest. Such as SNPs in DNA repair pathway genes [45], apoptotic pathways genes [46], genes associated alcohol metabolism could affect the risk of cancer [47].

SNPs play a major role in understanding the genetic basis of many complex diseases. More than 10 million SNPs have been reported in the human genome [48]. Still, a major challenge for the researcher is to categorize the functional SNPs which are associated with the disease [49]. Non-synonymous SNPs (nsSNPs) is responsible for changes in the amino acid residues and are key factors associated with a functional diversity of the encoded proteins [50]. Various epidemiological studies demonstrated the role of promoter SNPs with risk of head and neck cancer [51].

SNPs also affect the expression pattern of the gene such as SNP of the promoter region of *E2F1* at position -897 predominantly associated with the reduction in expression of this gene in HeLa cell line [52]. The intronic variant of the gene also linked in gene regulation, e.g. in a study, an

intronic variant 13964 G>C of *TP53* gene was found to be associated in hereditary breast cancer patients in North-America [53].

However, the roles of non-synonymous SNPs (nsSNP) and intronic variants of *E2F1*gene with the susceptibility of cancer have not been described so far. In this study, we specifically focus on nsSNPs and an intronic variant of *E2F1* gene. We hypothesized that nsSNP in this gene might alter their functions which may predispose individuals to cancer risk. 143 nsSNPs are reported in *E2F1* gene in different databases like, Ensembl [54], dbSNP [55], UniProt [56], and COSMIC [57]. However, there is no integrated *in silico* as well as genetic association study has been undertaken regarding *E2F1* gene. Therefore, in this study we used an integrated computational approach to prioritize SNP for genetic association study. The *in silico* most damaging germ-line mutations and two other nsSNPs of *E2F1* gene which were predicted as disease risk SNP of *E2F1* gene by National Institute of Environmental Health Sciences SNPs program (NIEHS SNPs Program) were further analyzed by conducting case-control association study. We also included an intronic variant of *E2F1* gene for case-control association.

To the best of my knowledge a few studies are available regarding *E2F1* genetic variants with cancer risk and there is no *in silico* approach used to identify deleterious nsSNP in this gene. Therefore in this study, we used computational approach to identify deleterious nsSNPs and also checked the association of genetic variants with cancer (lung, head and neck and cervical) risk. We hope that this approach will provide important clues regarding E2F1 genetic variants with cancer risk.

In this research work, we used the following objective.

1. To identify (*in silico*) deleterious nsSNPs in *E2F1* gene.
2. To develop and optimize genotyping methods for selected SNPs.
3. To analyze the association of E2F1 genetic variants with the cancer risk (Lung Cancer, Head and Neck Cancer and Cervical Cancer).

REVIEW OF LITERATURE

2.1. Cancer

According to the WHO, Cancer is a group of diseases characterized by uncontrolled cellular proliferation beyond their usual boundaries that can invade to the neighbouring tissues and in advance stage, it spread to other organs. Neoplasm is another term which is used for a malignant tumor. Cancer can affect any part of the body and has various anatomic and molecular subtypes that each requires precise executive strategies. According to the 2018 WHO report, Cancer is the second leading cause of death in the World and it accounted 9.6 million deaths in 2017, most commonly diagnosed cancers worldwide are; lung, breast, colorectal, prostate and liver cancer [1, 4].

2.1.1. Global and Indian Epidemiology

Cancer is a critical public health problem worldwide. 18.1 million new cancer cases and 9.6 million cancer related deaths occurred in 2017 worldwide. According to the recent report, lung cancer remains the deadliest cancer, globally. 11.6% (2.094 million) cancer incidences and 18.4% (1.8 million) cancer mortality occurred due to lung cancer worldwide in 2017 [4]. Breast cancer is the second most frequently diagnosed cancer (2.089 million new cases, 11.6%). However, in terms of mortality, it ranks 5th position (627,000, 6.6%) because of the relatively favorable prognosis [1, 4]. Incidence and mortality of five most common cancers are given in Fig 2.1.

Breast cancer is one of the most frequently diagnosed cancers in India. 15.46% (1, 62,468) of all new cases and 12.11% (87,090) of all cancer deaths occur due to breast cancer in India [4]. Oral cavity cancer is the second most commonly diagnosed cancer in India. 11.42% of all cancers incidence (1, 19,982) and 10.09% (72,616) of all cancer related deaths occur due to oral cavity cancer in India. Common cancer incidence and mortality in India are given in Fig 2.2 [1, 4].

Common cancers incidence and mortality status worldwide

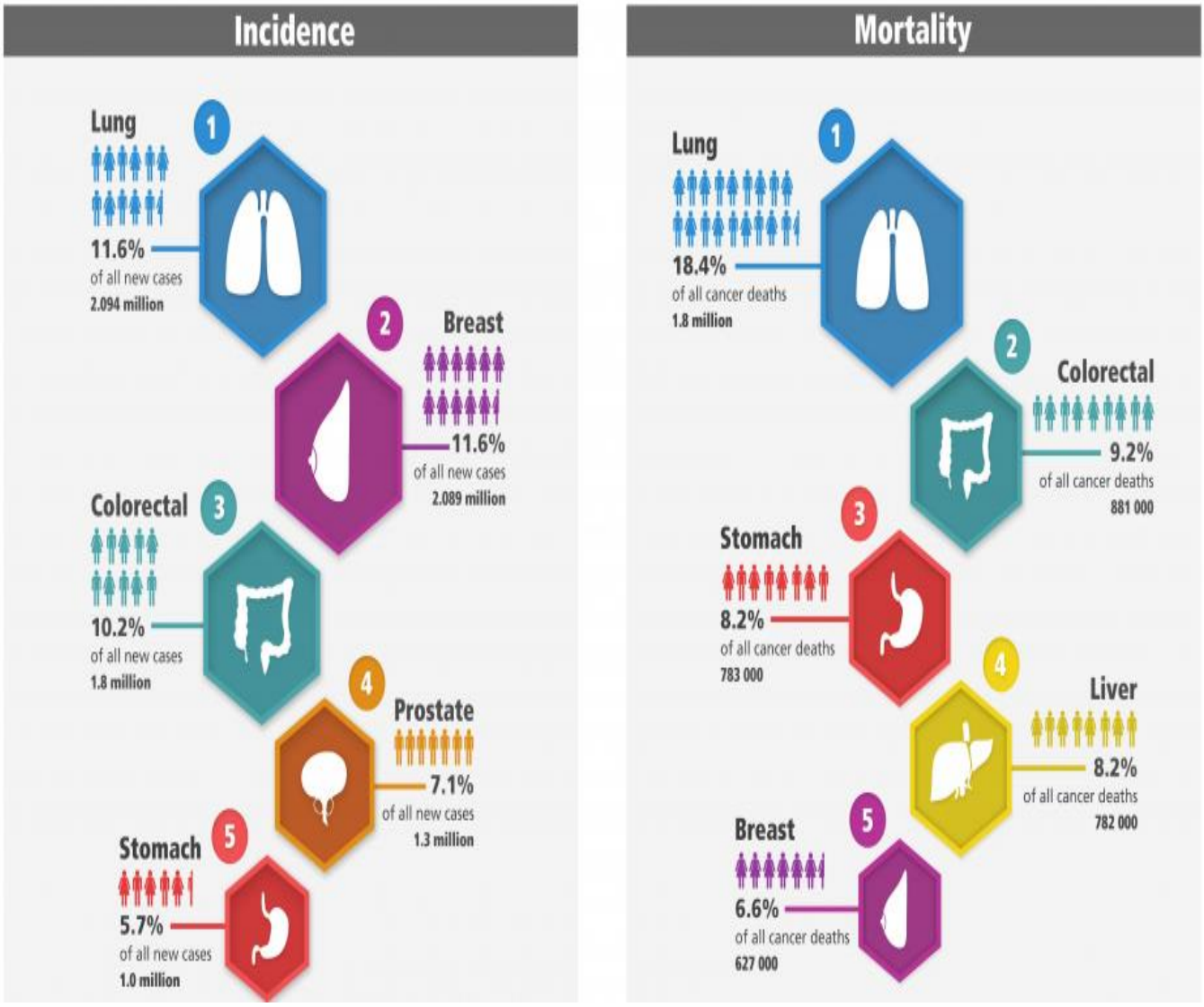


Fig 2.1 Incidence and mortality of common cancers in worldwide scenario [4].

Common cancers incidence and mortality status in India

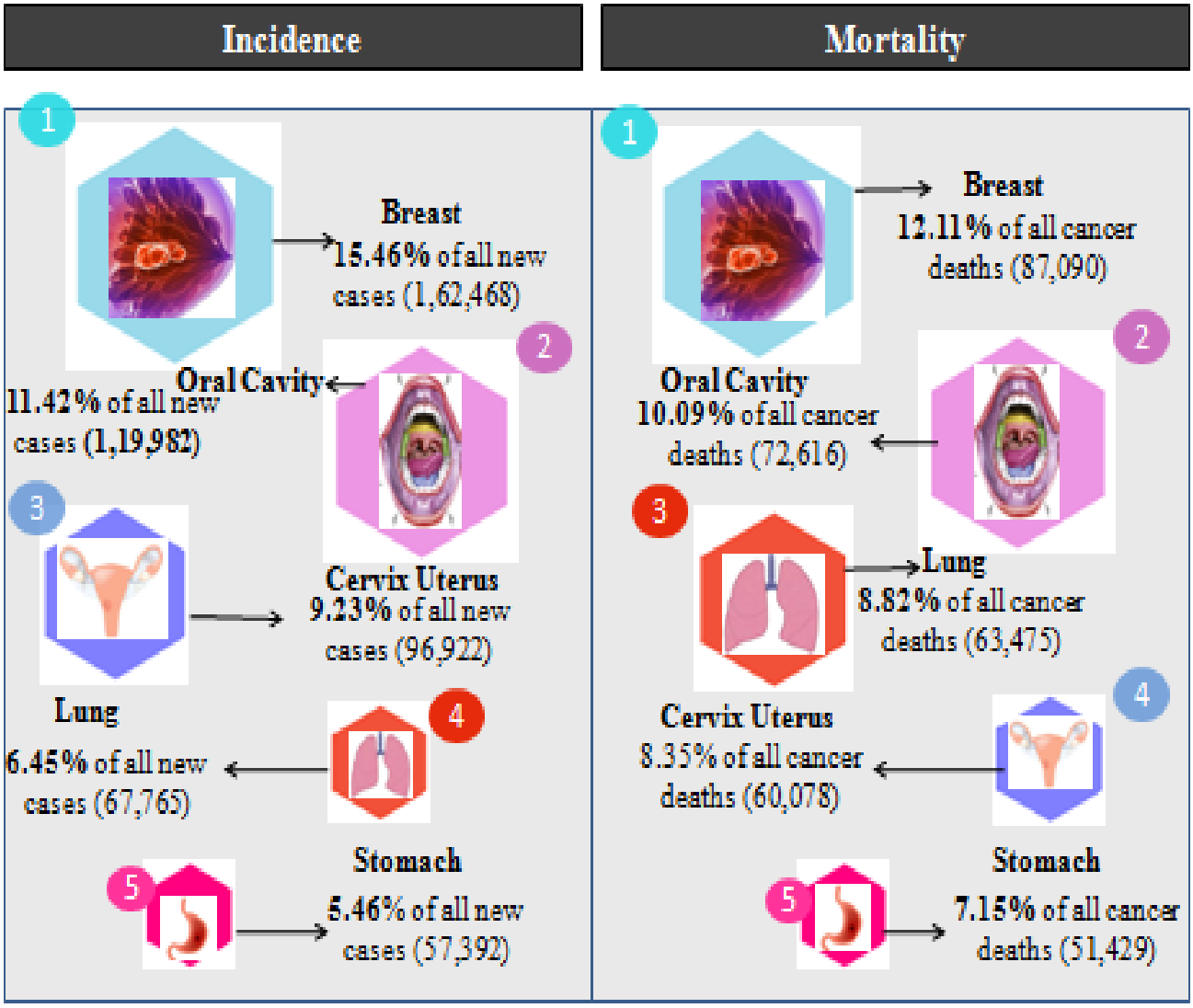


Fig 2.2 Incidence and mortality of common cancers in India [4].

2.1.2. Etiological agents of cancer

Cancers are mainly associated with exposure to environmental carcinogen, behavioral exposures, as well as lifestyle. The common environmental factors that contribute to cancer death, include tobacco (it contributes to 25-30% cancer related death), obesity (30–35%), infections (15–20%), radiation ionizing and non-ionizing (more than 10%), lack of physical activity, and environmental pollutants [58]. In the developing countries, almost 20% of cancers are due to infections such as hepatitis B virus, hepatitis C virus, and human papilloma virus [59]. Reproductive factors like multiple child, birth late child birth and hormonal therapy are also associated with cancer risk [60]. These factors are associated with endometrial cancer, ovarian cancer and breast cancer susceptibility in females [61, 62]

Cancers (30-50%) could be avoided by reducing risk factors: 22% cancerous deaths are associated with tobacco use, in the developing countries 20% cancers arise due to the various infections [58]. Additional 10% are due to obesity, a poor diet, physical inactivity, and drinking alcohol [2, 63]. These factors change the function of gene partially or completely [2, 63]. Usually, 4-7 events are required for formation of pre-malignant cell from the normal cell. Inherited mutations also play an important role in cancer susceptibility; around 5–10% of cancers arise due to the genetic factors [64]. Some well known cancer risk factors are given in Fig 2.3

Risk factors for cancer

Fig 2.3 Representations of some well known risk factors for cancer

2.1.3 Etiology of lung cancer

According to the recent WHO report, lung cancer the mortality rate will continue to rise, due to increased use of tobacco globally, especially in Asia [4]. Use of tobacco in any form is not safe and it is the main risk factor for lung cancer and any types of lung cancer are mainly associated with cigarette smoking [65]. Human lung and lung cancer representation are given in Fig 2.4.

2.1.3.1 Tobacco smoking

Regular and excessive exposure of tobacco mediated carcinogen in the cells are main factor of lung cancer initiation [5]. Lung cancer mainly arises in smokers and it is rare in non-smoker. Cigarette smoking is a major risk factor of lung cancer. Gaseous and particulate mixture are present in cigarette smoke [66]. The 95% part of mainstream smoke arises due to 400-500 gaseous compounds and 5% weight comes from 3500 particulate components [67]. Various effective cancer causing agents occur in mainstream smoke, which included polycyclic aromatic hydrocarbons (PAHs), aromatic amines and other inorganic and organic compounds, such as arsenic, chromium benzene, and vinyl chloride, The PAHs and *N*-nitrosamines become carcinogenic after metabolic activation. Cellular homeostasis determines the activation and detoxification of carcinogenic compound which probably affects individual to cancer risk. Radioactive materials such as radon, bismuth, and polonium are also present in tobacco smoke. Around 50 potential carcinogens have been identified by the International Agency for Research on Cancer (IARC) in tobacco smoke [68, 69]. The *N*-nitrosamines is associated with lung carcinoma [70].

4-(methylnitrosamino)-1(3-pyridyl)-1-butanone (NNK), induce adenocarcinoma of the lung in experimental animals model. Other Tobacco-specific nitrosamines (TSNAs) have been associated with the esophagus, bladder, pancreas, oral cavity, and larynx cancer [71]. Tobacco carcinogen like, NNK can bind to the DNA and from DNA adducts which are responsible for mismatching of nucleotides and resulting in mutations in DNA [72].

The DNA repair mechanisms remove adduct from DNA and restore normal DNA. If cells retain damaged DNA, it may lead to apoptosis. Failure of the normal DNA repair process is associated with mutations [72].

NNKs can mediate an array of downstream signaling that includes alteration in critical oncogenes and tumor suppressor genes that ultimately can result in unrestricted, cellular growth and tumorigenesis [73].

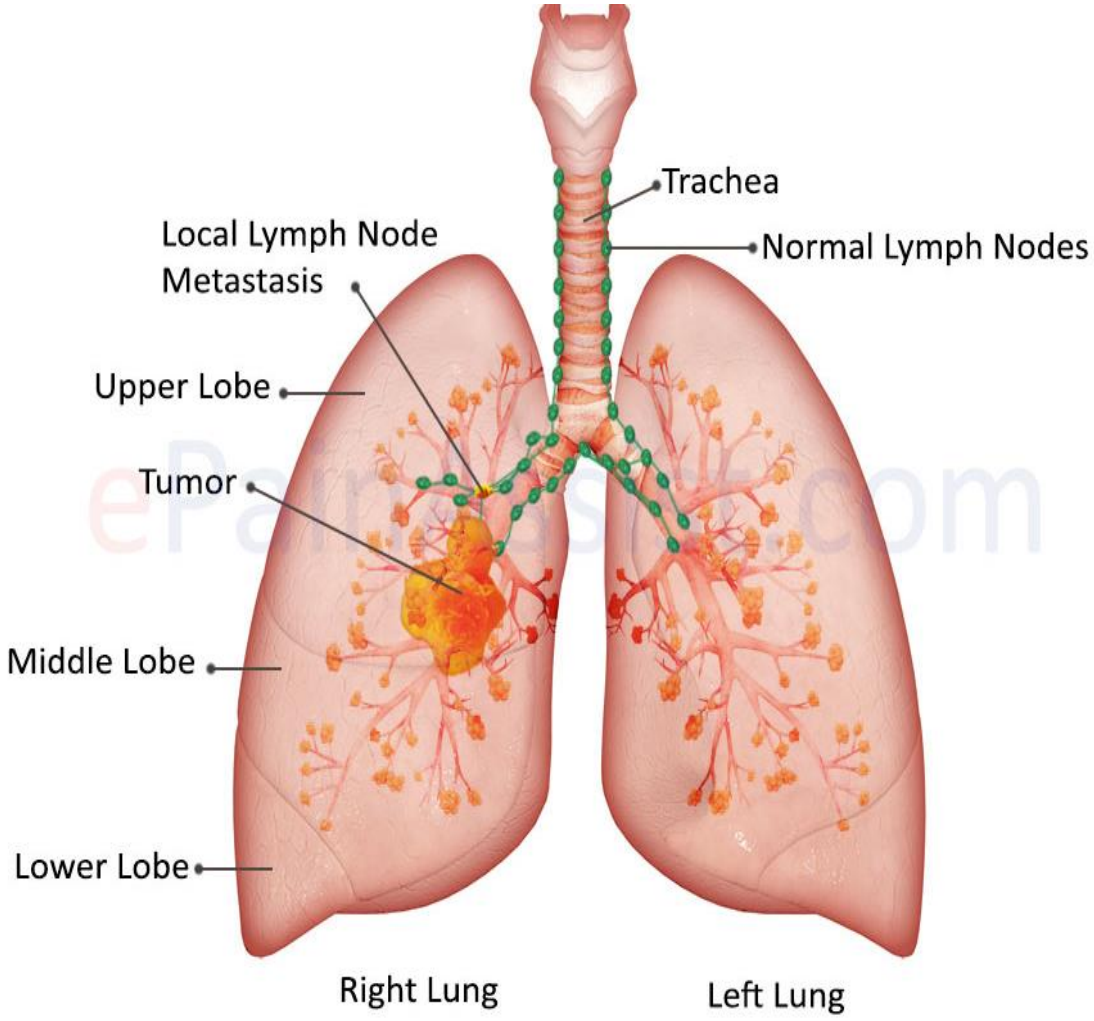


Fig 2.4 Human lung and lung cancer [72].

2.1.3.2 Different types of smoking

Different forms of tobacco smoking like cigar smoking and pipe smoking has been also linked with increased risk of lung cancer. These risk factors are weaker in causing cancer than the cigarette smoking [74]. A cohort study included >130,000 men who smoked (cigar smokers) continuously for more than 12 years and men who were non-smokers showed 5.1 times risk of lung cancer in smokers compared to non-smokers [74]. Another study also showed 2.1 times higher risk of lung cancer in cigar smoker than non-smokers [75]. Pipe and cigar smokers also have a comparatively higher risk of LC than non-smokers or former smokers [76].

2.1.5 Never Smokers

15% of lung cancer (LC) in males and more 53% of LC in females are not associated with smoking. Never smokers contribute to 25% of all LC cases worldwide. In South Asian countries, more than 80% of women with lung cancer are never smokers [77]. One of the studies conducted in the United States (US) revealed that 19% of lung cancer in females and 9% in males occurs in never smokers [78]. The incidence of lung cancer in never smokers varies geographically [79]. A study, conducted in Europe which included data from 1950 to 1990, they could not find major changes in the percentage of never smoker in male LC patients [80]. Females are more susceptible to LC than men and its prevalence is higher in a certain part of the world such as Asia [5].

2.1.6 Etiology of head and neck cancer (HNC)

Genetic and epigenetic factors are two main factors which influence squamous cell carcinoma (SCC) of HNC mainly arises by the influence of tobacco, alcohol, diet, viruses, radiation, ancestral and genetic predisposition, immunosuppression, occupational risks, and way of life. Use of tobacco is one the most important risk factors of cancers (including HNC, esophagus, LC, urinary bladder, renal, pelvis, and pancreas cancers). The association between smoking and HNC risk has been established by epidemiological studies [81]. The major risk factors of tobacco smoke are the PAHs, benzpyrene and nitrosamines which form adduct with DNA [73]. These adducts are responsible for mutagenesis of DNA during DNA replication [82]. Cyt-P450 and glutathione-S-transferase (GST) are two important enzyme which involved in oxygenation and

conjugation of these compound during its metabolism [83]. Marijuana is also known as ganja or bhang. It is smoked as cigarettes. The potent carcinogen like benzpyrene, phenols, terpenes and phytosterols, acids, are released during cannabinoids burn. Marijuana smoking is not only an independent risk factor for oral cancer development but also it is associated with oral cancer carcinogenesis with tobacco and alcohol exposure [83]. The utilization of smokeless tobacco has become prevalent globally; consumption of smokeless tobacco is mainly associated with oral cancer [84]. Structure of oral cavity and oral cavity cancer are presented in Fig 2.5.

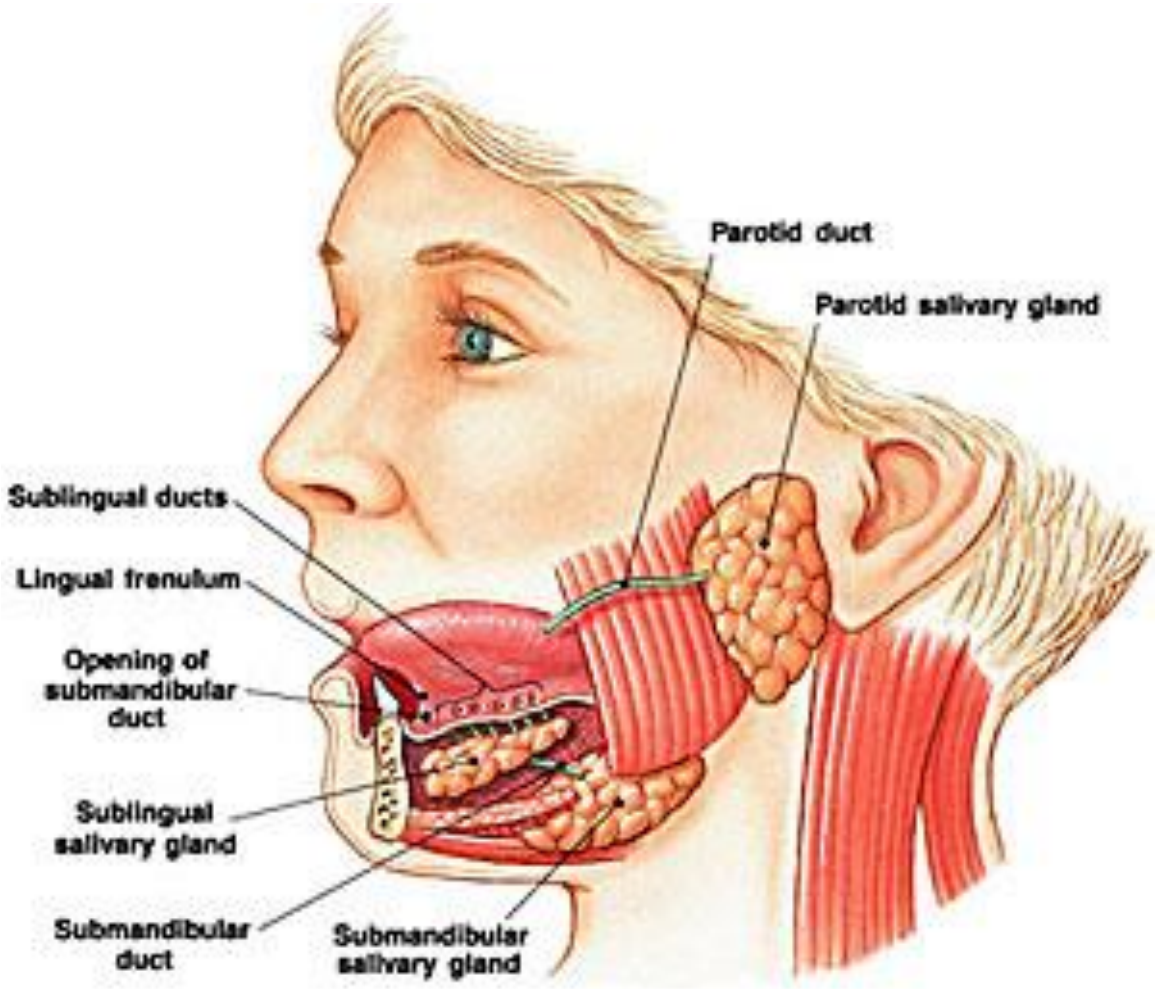


Fig 2.5 Representations of oral cavity and oral cavity cancer [83].

2.1.7 Betel quid

Chewing of betel quid in different forms is one of the most frequent habits in Southeast Asia, especially in India. Betel quid are usually mixture of betel leaf, areca nut, slaked lime, and tobacco. Most commonly used form of these mixtures are khaini, zarda, gadakhu, and mawa which consumed in different parts of India; nass, naswar/niswar in Central Asia and Middle East; shammah in Saudi Arabia, and toombak in Sudan. Different studies have shown the role of these tobacco products in HNC development [85].These tobacco products are also associated with precancerous stage like leukoplakia, erythroplakia, and oral submucous fibrosis [85].

Various *in-vitro* studies on oral mucosal fibroblasts have shown that some essential ingredients of betel quid are genotoxic and cytotoxic and induce cellular proliferation. By these studies, it is also clear that dangerous reactive oxygen species (ROS), methylating agent and metabolic intermediates generated during betel quid consumption cause DNA damage of various types [86].

2.1.7 Alcohol

Role of alcohol in carcinogenesis of oral cancer has been established in human but in the animal study, the carcinogenic property of ethanol has not been confirmed [9]. Alcohol plays a synergistic effect in oral cancer development; individuals who are a smoker as well as drinker have more risk for oral cancer than drinker only [9]. In one of the epidemiological studies conducted in India, it has also been found that alcohol act as an independent risk factor for oral leukoplakia [87]. Keller et.al (1969) could not demonstrated association of minor salivary gland tumors with a heavy smoker or heavy drinker [88]. The role of alcohol as an independent risk factor for oral cancer carcinogenesis is still uncertain. Alcohol is shown to increase the permeability of carcinogens into the oral mucosa that alters the morphology of mucosal epithelial cell which is characterized by epithelial atrophy [11].

Acetaldehyde is the main metabolite generated during alcohol metabolism which is mainly controlled by enzyme alcohol dehydrogenase (ADH). Acetaldehyde further oxidizes to acetate by aldehyde dehydrogenase (ALDH). In cultured mammalian cells, it has been shown that acetaldehyde causes DNA damage. It is associated with exchange of sister chromatids and

specific gene mutations [89]. The injuries caused by alkylating agents are repaired by enzyme 6-methylguanitransferase acetaldehyde inhibits the activity of this enzyme. These mentioned consequences of acetaldehyde initiate or promote tumor formation.

Increased accumulation of acetaldehyde in the body either by increase of its production or decrease of its elimination, is considered deleterious. Accumulation of acetaldehyde (CH_3CHO) could occur due to increased activity of ADH enzyme which is present in oral microflora and in the oral mucosa [89]. ADH type-3 genotype is associated with fast oxidation of C_2H_5OH to CH_3CHO and these are more susceptible to oral cancer [90]. In epidemiological study, it has been reported that the genetic variants of two enzymes, ADH and ALDH are related to the increased risk of alcohol-related cancers [90].

2.1.8 Diet and nutrition

Different epidemiological and laboratory analysis has established the role of diet and nutrition to the risk of cancer development [91]. Regular use of fruits and vegetables in diet mostly carrots, fresh tomatoes, and green peppers reduce the risk of oral and pharyngeal cancer [91]. Some food groups are associated with higher risk of oral cancer such as processed meats, soups, red meat, salted meat, cheese, pulses, polenta, pasta or rice, millet, and cornbread [92].

Certain studies have shown that the role of macronutrients and trace elements present in the food groups i.e. vitamins and their analogs (13-cis-retinoic acid and β-D-glucopyranosyl ascorbic acid (AA) shows protective association against cancer. The some of the vitamins and trace elements like vitamin A (retinol), E (α-tocopherol); carotenoids (β-carotene); potassium; and selenium are associated with cancer protection (38–43). Vitamin A, vitamin C, and vitamin E are essential for reducing free radical that causes mutations in DNA, lipid peroxidation of cell membranes and change the activity of enzymes [91]. Even though all the anti-carcinogenic mechanisms of carotenoids are not known, but these agents act as anti-oxidants, pro-oxidants, immune modulator, inhibits mutagenesis, reduce the nuclear damage, protect from various neoplastic events, and protect against photo-induced tissue damage [93].

A number of treatment trials with β-carotene have been completed for oral cancer and have revealed significant success rates [94]. It has been proven that Vitamin E inhibits tumor development in mice model and this has been attributed to effective stimulation of immune

response by vitamin E and role of vitamin E as a potent anti-oxidant has also been shown which is caused by hydroxyl radicals. Clinical trials with α-tocopherol revealed that this vitamin is a non-toxic anti-oxidant like β-carotene. But vitamin E has also been shown to associate with promoting skin tumor formation when its higher concentration is higher (80μmol) [94]. There is no such study available which reports the sole use of Amino Acids in the treatment of head and neck cancer. The diet in which low intake of fruits and vegetables and usually rich in vitamin C predisposed to increased risk of oral cancer [21]. It has been reported that the serum level of β-carotene (anti-oxidant) is lower in cigarette smokers than non-smokers. The quid chewing habit also reduces the serum levels of vitamins A, C, and B12; folate; and β-carotene than non-quid-chewers [95].

2.1.9 Environmental factors

The latent or chronic infections of Epstein–Barr virus (EBV), human papillomavirus (HPV), and herpes simplex virus are mainly responsible for inducing malignant transformation by hijacking host cell cycle regulatory machinery. Some of the viral genes have proto-oncogenic property and these genes become oncogenes when inserted into host DNA and the ultimate result of this insertion is malignant transformation [96]. EBV is associated with oral hairy leukoplakia and lympho proliferative disease. The causal correlation of squamous cell carcinoma (SCC) of oral cancer with EBV with is still indistinct [97]. HPVs are the most frequent viruses which associated with oral cancer carcinogenesis. Certain HPV types (HPVs 16, 18, 31, 33, 35, and 39) referred to as 'high-risk' types are associated with OSCC and oral premalignant lesions. Two major HPVs gene products E6 and E7 bind with p53 and Rb tumor suppressor protein and disrupt their activity. It disrupts cell cycle control, DNA replication, DNA repair, and apoptosis [98]. HPV 16 is found to be associated with most of the genital cancers it is the also frequently diagnosed in oral cancers, which clearly determine the possible role of HPV infection in the oral cavity carcinogenesis [99].

2.1.10 Immunosuppression

Individuals with weak immune system are more susceptible to develop oral cancers. During organ transplantation immunosuppressive drug was given and studies have shown increased chances of getting lip cancer in organ transplanted patients. This is also observed in patients having exposure to solar radiation and if the patients are regularly exposed to other risk factors.

However, the direct role of immunosuppression with lip cancer development was not demonstrated in the study (an observational study) [30].

2.1.11 Occupational Risks

Occupational exposure of excessive solar radiation (UV-rays) is associated lip cancers. Exposure of SO_2, asbestos, pesticides, mists from strong inorganic acids and burning of fossil fuels have also been known to cause cancers of posterior mouth, pharynx and larynx [100]. Certain occupations have been reported to place people at increased risk for the development of salivary gland carcinomas; these include manufacturing of rubber products, plumbing (exposure of metals), and woodworking in an automobile industry [101].

2.1.12 Etiology of cervical cancer (CaCx)

The incidence and mortality of cervical cancer are highest among women of developing countries [102]. The persistent infection of HPV is the main etiological agent of cervical cancer. It contributes to 5.2% of all cancers worldwide and it is also associated with anogenital and a subset of head and neck cancers [103, 104].

The incidence and mortality of cervical cancer in Indian women is 22.9% and 20.7%, respectively. 432.2 million females in India belong to 15 years or older than 15 years. Females belonging to this age (15-44) group have a higher chance of HPVs infection. Cervical cancer is the second most common cancer in women aged 15–44 years [1]

More than 200 types of HPVs are identified in which 40 HPV types can be easily dispersed through sexual contact or by the mucous membranes of infected people to the mucous membranes or skin of their partners [104]. HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 66 are ‘defined as carcinogens’ due to their oncogenic potential [105-107]. Whereas low-risk HPVs types are related with benign lesions like condyloma accuminata, high-risk HPV types are linked with the progression of high-grade lesions and malignant tumors. Persistent infection with HPV has been proven to be a main etiological agent which associated with squamous intraepithelial lesions (SIL) [108].

The HPV-16 and HPV-18 contributes to about 62.6% and 15.7% of cervical cancers, respectively [109]. Moreover, among all the HPV-associated cancers, the HPV types 16 and 18

contribute about 63 to 80% of penile cancers, 80 to 86% of vulva/vaginal cancers, 93% of anal cancers, and 89 to 95% of oropharyngeal cancers [110]. Therefore, HPV-16 and HPV-18 have been the primary focus of HPV vaccine development. Representative image of cervical cancer is given in Fig 2.6.

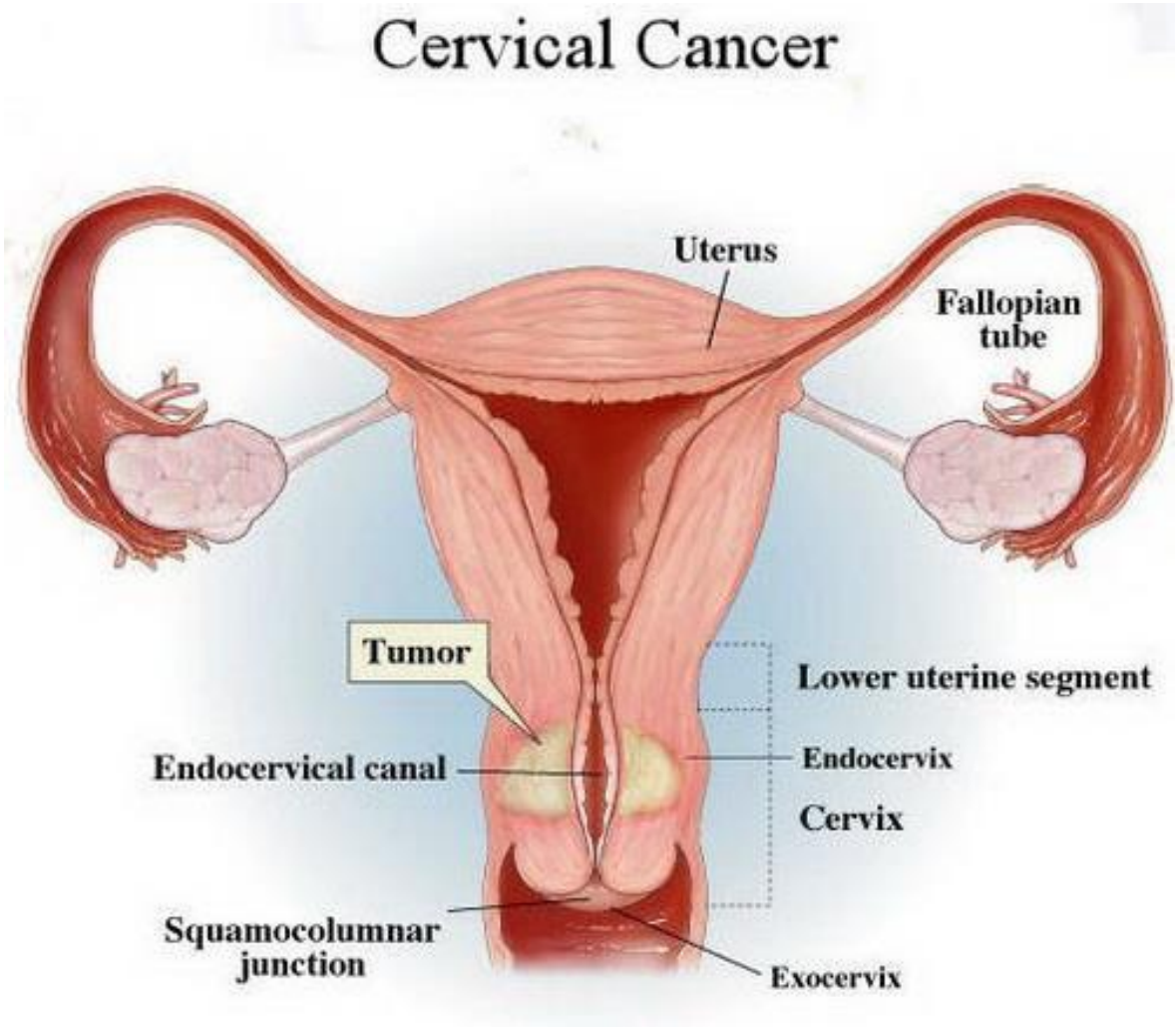


Fig. 2.6 Representation of female genital organ and cervical cancer [110].

2.1.13 Human papilloma virus types in India

Persistence infection of HPV is the main cause of cervical cancer. More than 200 HPV types have been reported in which HPV 16 and 18 are mainly categorize as high risk HPV type while rest HPVs are low risk type for cervical cancer [23]. The incidence of HPV varies from 87.8% - 96.67% among cervical cancer patients in India [111-114]. The epidemiological studies revealed that two most frequent HPV types HPV-16 and 18 are associated with invasive cervical cancer in which HPV-16 prevalence has been found more frequent [115].

The incidences of other high risk HPV types are not common. The genital warts and benign cervical lesions are associated with low-risk HPV types like HPV-6 and HPV-11 [29]. HPV has also been observed general population as well as women with benign cervical cytology. A meta-analysis conducted by Bhatla et al (2008) [115], revealed that there was no considerable variation in the incidence of HPV infection in North and South India. However, the prevalence of HPV-16 and 45 appeared to be higher in North India, while the prevalence of HPV-35 is higher in South India.

The human papilloma virus persistence was higher for HR-HPV type and this persistence was found to be highest for HPV-16, 45, 67, 31, 51, 59 [35]. The average period of persistence of HPV-16 was 12.5 months [116, 117]. The incidence of HPV-18 was higher than that of HPV-16 among adolescents in the US, while HPV-16 was associated commonly with high-grade SIL cytology [118].

2.1.14 The medical need for vaccination

Cervical cancer is the third leading cause of cancer-related deaths in women worldwide and second leading cause of cancer-related mortality in Indian women [1, 119-121]. The commercial availability of defensive HPV vaccines, Gardasil and Cervarix, has been introduced to prevent cervical cancer. The incidence of cervical cancer has decreased in high-income countries by introducing organized gynecological screening programs and appropriate treatment of the detected pre-cancerous lesions. By implication of newly commercial available HPV vaccines reduce the rate of cervical cancer in developed countries. However, cervical cancer incidence and mortality is higher in underserved populations where these vaccines are less effectively

implemented [5]. In spite of defensive vaccines and Pap smears, the prevalence of cervical cancer is rising continuously in the developing countries due to high cost and complicated infrastructure crucial to execute efficient screening and vaccination programs. For this reason, the 83% cervical cancer cases occur in developing countries, and this load of cervical cancer is expected to rise to 90% in developing countries by 2020 [122]. In addition, the existing preventive vaccines are useful in defending against the high-risk oncogenic HPV types 16 and 18, however, the incidence HPV16 and 18 differ across populations worldwide [109]. Thus, wide-range of cross-protection against oncogenic HPV types is extremely desirable in the second generation of HPV vaccine candidates. An additional limitation of existing commercially available vaccines in that they do not have therapeutic effects [123]. There is no effective way to eradicate a persistent infection of HPV before the development of pre-cancerous lesions. Worldwide, there is a huge burden of HPV infections and HPV-associated diseases which are undetected, untreated, and slowly progressing toward malignant transformation. The available vaccine is not effective against all kind of high-risk HPV viruses hence, there is a great medical need for the improvement in existing vaccine and also a novel preventive and therapeutic vaccine is needed to overcome the HPV-associated disease.

2.1.15 Genetics and cancer susceptibility

The incidence, prevalence, and mortality of various cancers differ in the different ethnic groups [124-126]. Like, prostate cancer which frequently occur in African-American population while the incidence of this cancer lowest in Japanese population [127]. The various causes of these differences are intrinsic variation, i.e. genetic variation, and extrinsic variation, which include diversity in social, economic, and geographical environments. Understanding these variations in cancer risk and the underlying mechanisms of these differences is important for future research as well as healthcare practices that can extent ethnic boundaries. Genetic differences are the key players in cancer risk, and in the recent genome-wide association studies (GWAS) a number of various genes variants have been elucidated that play roles disease susceptibility. The genetic variants of BRCA1 and BRCA2 genes are the most popular genetic risk factors which are associated with breast and ovarian cancer risk in females [128-131].

The genetic makeup of an individual is also associated with the pathogenesis of lung cancer, whether it is associated with certain type of lung cancer risk, with or without exposure to cigarette smoke or an individual's response to the treatment. Spitz et al (2007) developed a

prediction analysis for lung cancer risk by incorporating different variables, i.e. smoking history, exposure to passive tobacco smoke, occupational exposures of dust and asbestos and family history of cancer [132, 133].

A meta-analysis including 32 studies revealed that the persons with a family history of lung cancer have a 2-fold increased risk for lung cancer [134]. In a study in which family linkage approach was used, it was reported that the first relationship of familial LC to the region on chromosome 6q23–25 (146cM– 164cM). *CYP1A1* gene polymorphisms and their association with LC risks have been contradictory in various studies [135]. A meta-analysis including 16 studies showed no significant association with the Ile462Val genetic variant of this gene; when all the smokers and non-smokers were compared. However, in the analysis, showed increased the risk of SCC in whites, especially in female and non-smokers [136].

GST gene products help to metabolize the wide-range of carcinogen. Benhamou et al (2002) [137], conducted a meta-analysis that showed a 17% increased risk of LC in individuals who had *GSTM1* null mutation. Another meta-analysis conducted in 53,000 case-controls revealed 18% increased risk of LC among individuals who were *GSTM1* null [138]. Recent, meta-analysis revealed the association of GSTM1 and GSTT1 gene polymorphisms with lung cancer risk in Caucasians, Asians as well as in Indian population [139].

Amos et al (2008) performed a genome-wide association study (GWAS) to scan tagged SNPs for LC risk in histologically confirmed non-small cell lung cancer (NSCLC). They identified a risk locus for LC at chromosome 15q25.1, a region which contains the nicotinic acetylcholine receptor genes [140]. Interaction analysis of gene variants (NAT2 and mEH) revealed that NAT2 polymorphisms with certain acetylation phenotype and mEH polymorphisms were associated with certain activity level with LC risk. Patients who had NAT2 slow-acetylation and mEH high-activity genotype are more susceptible to LC. However, in non-smokers, 50% decreased the risk of lung cancer was observed among individuals carrying combined NAT2 slow-acetylation and mEH high-activity genotype [141].

Genotype-phenotype studies conducted regarding DNA repair pathway gene such as nucleotide excision repair (ERCC1, XPD, and XPA), base excision repair (XRCC1 and OGG1), and double-strand break repair (XRCC3), and mismatch repair pathways have also been studied and polymorphisms in these gene are associated with LC risk. A number of cell-cycle regulatory

genes have been associated in LC susceptibility, like tumor suppressor genes *p53* and *p73*, *MDM2* and the apoptosis genes encoding FAS and FASL [142].

Mutagen sensitivity is also associated with an increased risk of LC [143]. Combined risk of LC was higher in individuals with mutagen sensitivity [142]. Despite of various genetic association studies, the all genes responsible for the increased risk of LC remain unknown [144]. Genetic Susceptibility to Environmental Carcinogens (GSEC) and the International Lung Cancer Consortium (ILCCO) are making collaborative efforts on lager number of samples to find additional pool of genes that might play role in the development of LC.

LC risk can be determined by host genetic makeup. Individuals who are genetically susceptible have a higher risk of lung cancer if they smoke tobacco. With the advent of new and advanced technology, it could be possible to characterize group of individuals as a genetically high risk for lung cancer [144].

The genetic makeup of individual has been shown to be an important risk factor associated with SCC of HNC. A followed up study conducted by Copper e*t al.* revealed the association of genetics with HNC susceptibility. They found that 31 individual developed cancers of the respiratory tract and upper aero-digestive tract out of 105 relatives of head and neck cancer patients [145]. But, a few epidemiological studies determined the role of genetic factors in oral cancer susceptibility with coexisting risk factors like smoking and alcohol. Genetic variants of genes encoding P450 enzymes and XMEs enzymes are associated with predisposition to tobacco-induced head and neck cancers [146].

The etiological factor of HNC was associated with tobacco, alcohol, HPV, and some molecular drivers. According to the recent studies, micro-RNA (miRNAs) mediated transcriptional regulation of gene is also associated with cancer risk. miRNAs are small fragment RNAs and are about 20±22 nucleotides in length, that can negatively control the efficiency and stability of mRNAs translational by targeting specific mRNAs [147]. It has been established that the expression of approximately 10~30% of all human genes can be regulated by mature miRNAs [148]. Abnormal expression of miRNA-146a is associated to many cancers [149]. Zhang et al conducted a meta-analysis in 2485 cases and 11034 controls and found that the miRNA-146a (genetic variant) is associated with increased risk of SCC of HNC in Chinese patients [150].

Alcohol itself is not a carcinogen, but its metabolite acetaldehyde which is released during alcohol metabolism has carcinogenic properties. Acetaldehyde is eliminated by ALDH2 enzymes, thus this determines the concentration of acetaldehyde in blood acetaldehyde after drinking. A genetic polymorphism in *ALDH2* gene (ALDH2 Glu487Lys) is associated with its reduced enzyme activity. Individuals carrying Lys allele have a high concentration of acetaldehyde in blood than individuals who carrying Glu allele, thus Lys allele enhances the risk for esophageal cancer [151, 152]. Transforming growth factor beta 1 (TGF-β1) is a member of cytokines superfamily. It regulates both immune system as well as cellular functions i.e. cell differentiation and proliferation, extracellular matrix production, apoptosis, and angiogenesis [153].

Genetic variant T869C of TGF-β1 gene was associated with SCC of HNC susceptibility in Caucasian but not in Asian population [154]. CYP2E1RsaI/PstI polymorphism was associated increased risk of SCC of HNC among Asian population [155]. Cervical cancer is the third most frequently diagnosed malignancy and the fourth leading cause of cancer-related death in females globally [1]. Infection of HPV, use of oral contraceptive pills, immunosuppression, hormone therapy, and cigarette smoking might be associated with cervical cancer risk [156, 157].

But, cervical cancer arises in fraction of women if they are exposed to the risk factors, implying that along with these factors genetic makeup of individuals also contribute significantly in cervical cancer susceptibility. p53 is one of the master controller of the cell-cycle, its impaired activity has been observed in many cancers. MDM2 is a negative regulator of p53, it inhibits the activity of p53 that leads to carcinogenesis [158]. SNP at the 309^{th} position in the first intron (rs2279744 T/G), increase the affinity for stimulatory protein 1 binding which increase MDM2 expression [159]. The elevated levels of MDM2 suppress the activity of p53 and the damaged cells escape the cell cycle checkpoint [160]. A meta-analysis revealed the significant association of T309G polymorphism of MDM2 gene with cervical cancer risk in Asian population [161].

HPV is associated with malignancy of cervical cancer. This virus interacts with p53 and promotes cancer development. The p53 act as a tumor suppressor and it arrests the cell cycle in G1 phase. p53 activates DNA repair pathway if DNA damage occurs [162]. HR-HPV proteins E6 and E7 are responsible for proteasomal degradation of p53 [163].

nsSNP in exon 4 (rs1042522; Arg72Pro) of TP53 gene is associated with cervical cancer susceptibility. Binding affinity analysis revealed that Arg variant binds to the HR-HPV E6 protein with higher efficiency than the Pro variant [164]. The Arg variant of p53 is associated with the development of SIL to cervical cancer [165]. Various studies have indicated that certain cytokines might play important roles in the processes of inflammatory cell infiltration and malignant cell transformation [14, 15]. Interleukin-10 (IL-10) is a multitasking cytokine which is mostly secreted by Th2 cells, monocytes and macrophages, keratinocytes, tumor cells as well as Th1 cells [16, 17]. IL-10 is involved in various biological effects i.e. it stimulates mast cells maturation, increases the B cells proliferation and differentiation, and restrains type 1 immune responses by inhibiting the release of IL-2, IFN-γ and additional cytokines [18]. IL-10 has both activities tumor-promoting (immunosuppressive) and tumor-inhibiting (anti-angiogenic) [19].

It has been well established that cervical cancer is mostly initiated by HPV infection, and tumor TNF-α is an inflammatory cytokine which play an important role in the immune response of cervical cancer malignancy [166, 167]. Currently, there are a number SNPs identified in the TNF-α gene which can control the transcription and production of TNF-α, i.e. TNF- α rs1800629 and TNF-α rs361525 [168, 169]. TNF-α rs1800629 (G/A) is the most studied polymorphism, which occurs at the 308 position of TNF-α gene and it is associated with the expression of TNF-α gene [170, 171]. A recent meta-analysis indicates that rs1800629 (G/A) polymorphism of TNF-α gene is associated with increased risk of cervical cancer, especially in Caucasians population [172].

Different types of external and internal factors are associated with different types of DNA damages. If the damage DNA is not repaired it can cause genetic instability, mutations and cancer. The repairing of different types of DNA damages is important for maintaining genomic integrity [173]. Among different DNA repair mechanisms operating in humans, the base excision repair (BER) is the primary defends to cell against lesions generated by ionizing radiation, alkylating agents, and lesions created by endogenous DNA-damaging agents like viruses [9]. Different genetic association studies revealed the association of *XRCC1* gene polymorphisms with cancer risk [174, 175]. Meta-analysis on this polymorphism suggests that Arg194Trp variant of XRCC1 gene is associated with cervical cancer risk and another polymorphism (Arg399Gln) of this gene might be a low-penetrant risk for cervical cancer in Asians [176].

A recent meta-analysis on *XRCC3* gene polymorphism (Thr241Met) is associated with cervical cancer risk in Asian population under all genetic models [177]. Genetic association and etiological studies have revealed that different cancers are associated with different etiological factors. Along with these etiological factors, genetic makeup of individuals also play an important role in cancer susceptibility [178]. Cancer mainly arises due to the mutations in cell cycle regulatory genes because this machinery controls the cell proliferation [131, 178].

2.1.16 Cellular homeostasis and cancer

The transcription factor p53 is a master controller of cellular homeostasis which is activated during cancer associated stress responses including DNA damage responses, cellular checkpoint responses and oncogenes. In normal cellular conditions, levels of p53 protein are low and its levels are controlled by rapid degradation by ubiquitin-dependent E3 ligase MDM2 [179, 180]. Various stress responses inhibit the MDM2-mediated p53 degradation and induce pathways that activate the p53 mediated cellular responses i.e. growth arrest, senescence and apoptosis [179].

Mutations in TP53 gene have been reported to associate significantly in different cancers; around 50% of all human cancer arises due to the mutations in this gene. The functions of p53 could be determined by overexpression of MDM2 or by the inactivating ARF [180]. An additional family of transcription factors that have an effect on cell fate and cancer development is the E2F family. This family of transcription factors are downstream effectors of pRB. Initially, it was believed that E2Fs are essential for G1 to S transition. However, latest microarray analysis revealed that E2Fs are not only involved in G1 to S transition but also are involved in other biological processes, including DNA replication, mitosis, DNA damage response, checkpoints, DNA repair, differentiation and autophagy by transactivate and repress gene expression [181, 182]. In E2Fs family, E2F1 is able to induce apoptosis significantly [183]. E2F1 can induced both p53-dependent and p53-independent apoptosis [182].

2.1.17 *E2F1* gene and cancer

E2F1 gene is located on chromosome 20 which contains 7 exons and 6 introns and codes 437 AA residues E2F1 protein (Fig 2.7). The E2Fs is a family of transcription factors and it consists of 8 members, generally, it classified into 3 classes: E2F1–E2F3a (typical activators), E2F3b–E2F6 (canonical repressors) and E2F7-E2F8 (atypical repressors) [184-186]. All activator E2Fs require dimerization partners (DP) for its binding to the DNA at promoter regions of target genes and activates the downstream signaling of cell cycle progression and DNA repair [187-189]. Activators E2Fs represent structural similarity with E2F1 and all activators required dimerization partner (DP1/DP2) for effective binding with DNA. Recently, a third DP protein has been identified which is an inhibitory binding partner and it reduces the E2F1 mediated transcription [190]. E2F1 to E2F6 are associated with dimerization partner (DP1 or DP2) and form heterodimeric complexes that bind to DNA in a sequence-specific (TTTC/GG/CCGC). E2F7 and E2F8 have two DBD and do not require dimerization partner proteins (DP) for binding to DNA. E2F7 and E2F8 make homodimer or E2F7/E2F8 heterodimer to bind to the target [191]. E2F mediated cell-cycle progression is given in Fig 2.8

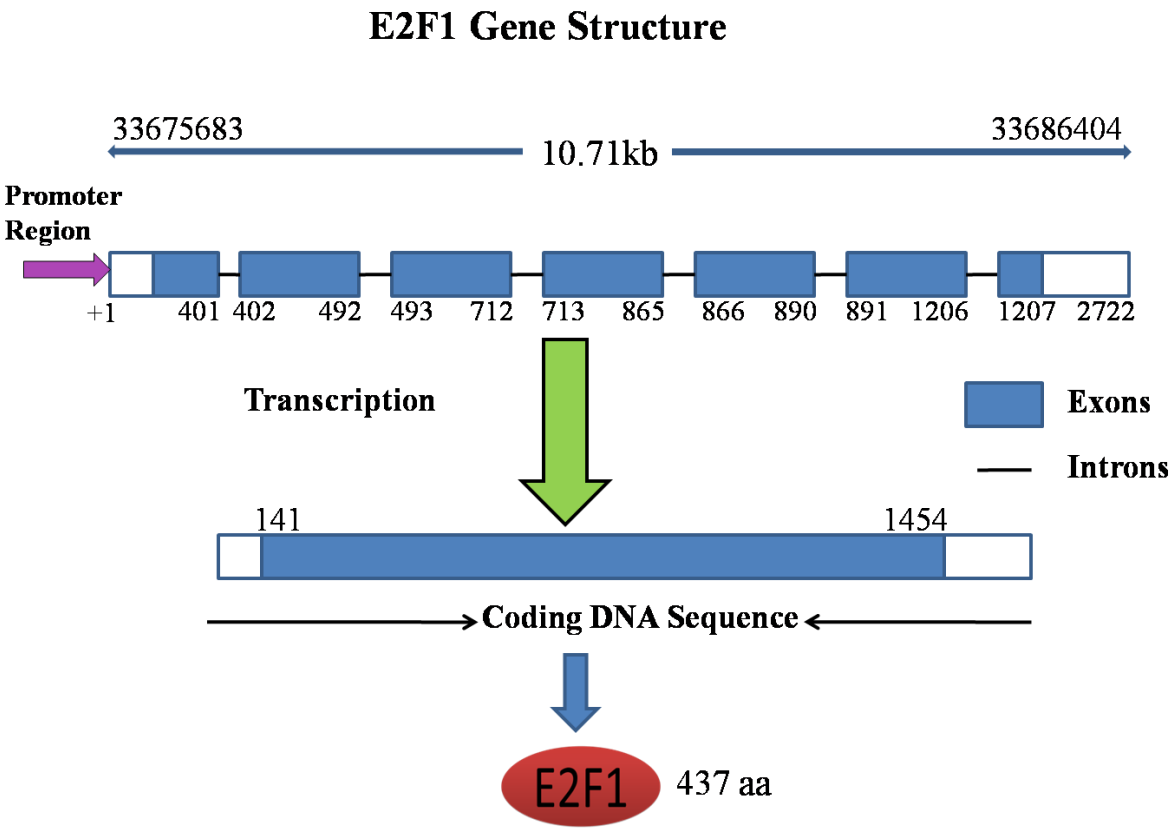


Fig 2.7 Systematic representation of E2F1 gene structure.

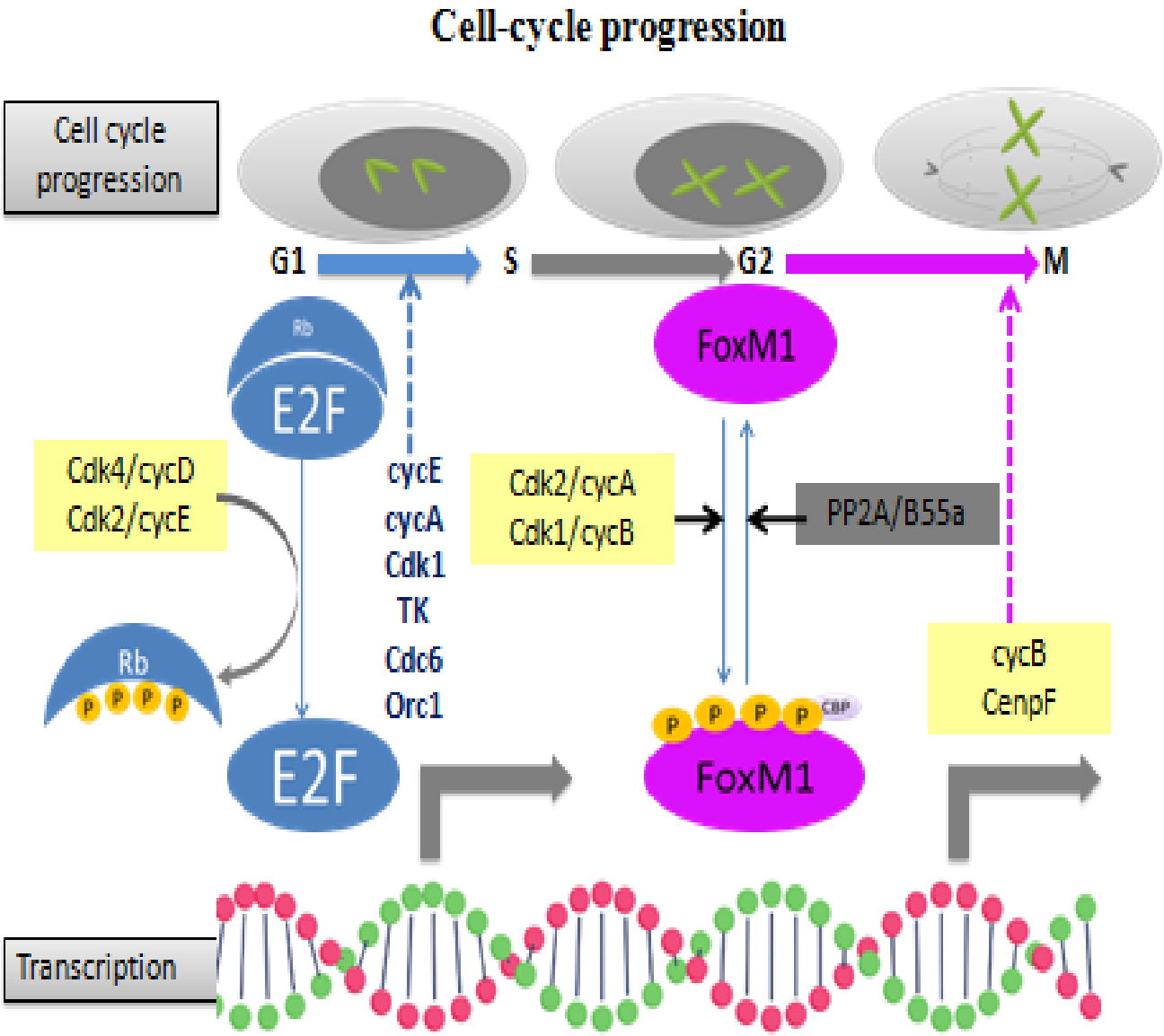


Fig 4.8 E2F mediated cell-cycle progression [12].

Generally transcriptional activity of E2F1 is highest among activator E2Fs in a context to pro-apoptotic gene activation. Overexpression of E2F1 in mouse models predisposes the animals to some type of cancer like vaginal, skin, and forestomach epithelium due to loss of cell-cycle control. On the other hand, increased activity of E2F1 also increases the transcription of p14ARF and therefore activation p53 pathway which is particularly common in tumors that have pRB inactivation. Hence, tumor formation is increased when efficient p53 is absent [192-194]. On the other hand, E2F1 deficient mice also develop a spontaneous tumor in specific tissues. The most frequent malignancies are histiosarcomas, hemiangiosarcomas, hepatocarcinomas, and LC. This tissue-specific carcinogenic transformation of the cell is possibly might be due to loss of an E2F1 checkpoint and may be decreased the expression of pro-apoptotic target genes, thus these finding supporting the role of E2F1 protein as a tumor suppressor [29, 195]. Somatic mutations of DP1 partner also reduce the E2F1–DP1 mediated transcriptional activity which decreases the levels of apoptosis after DNA damage [196]. The different cell-lines studies revealed the role of E2F1 in tumor suppression i.e. esophageal cancer cell lines induces tumor cell apoptosis *via* E2F1/ARF/MDM2/p53 pathway [197]. E2F1 mediated cellular homeostasis presented in Fig 2.9.

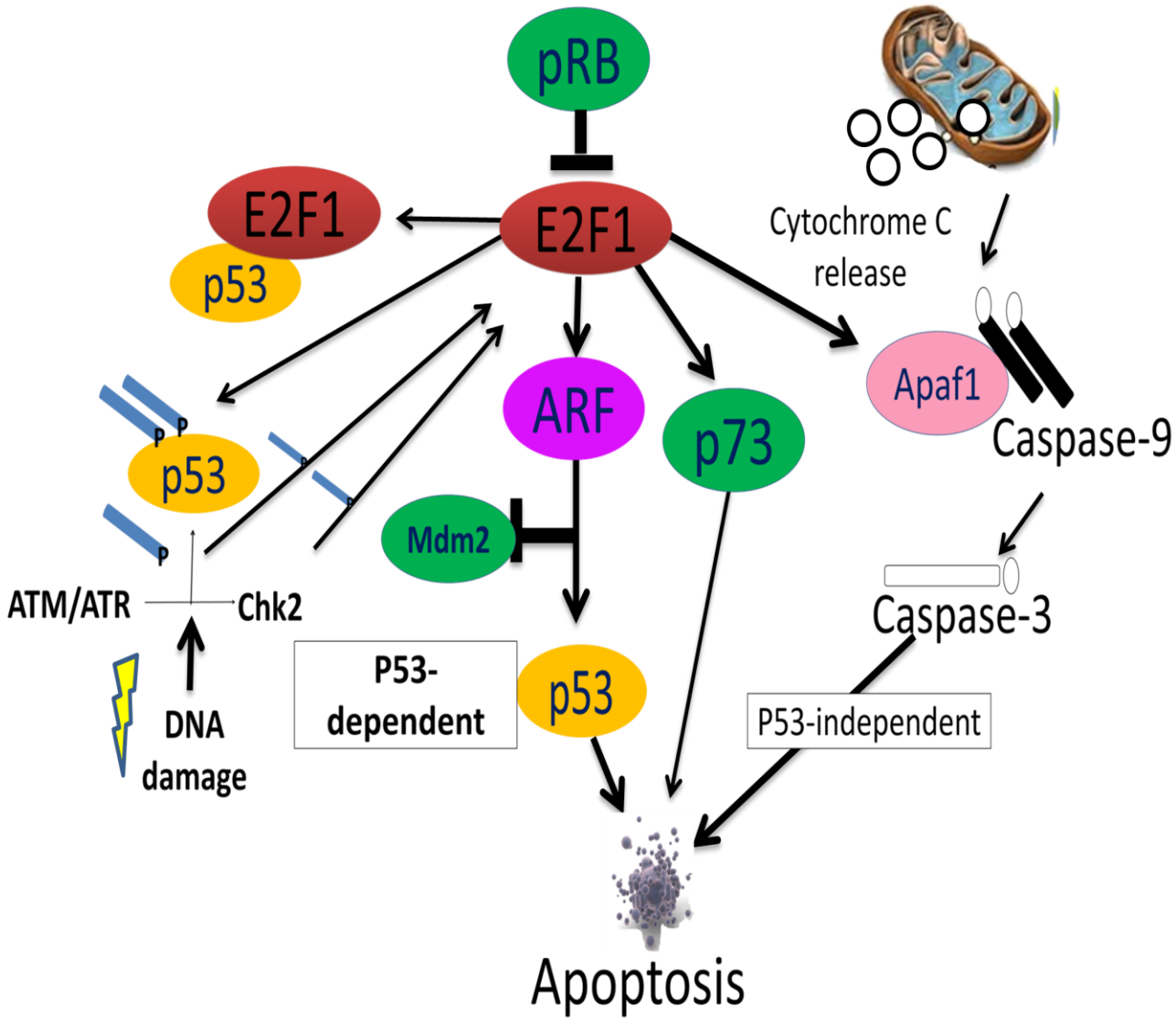


Fig 2.9 E2F1 mediated apoptosis [198].

Early studies indicated that expression of E2F1 was found to be increased in some cancer cell line [199, 200]. Subsequently, it observed that increased E2F1 expression was also linked with the increased levels of apoptosis and inversely associated with cell proliferation, mostly when sequential or semi- sequential sections were analyzed [201-203]. The expression levels of E2F1 are higher in lung metastasis [204, 205]. The stable higher levels of E2F1 in pancreatic cancer cell lines may be responsible for the increase in chemotherapy-induced apoptosis [206]. When pancreatic cancer cell lines were infected with E2F1-expressing adenoviral vector, the apoptosis levels increased in gemcitabine and etoposide or roscovitine treated cells [207, 208].

Higher levels of E2F-1 protein in hepatocellular carcinoma (HCC) are associated with increased apoptosis in tumor cell. This association support the pro-apoptotic role of E2F1 in HCC [209-211]. Other researcher also demonstrated that endogenous c-myc was overexpressed in the early stages of hepatocarcinogenesis, while p53 was upregulated in the tumors, suggesting that both E2F1-dependent proliferation and apoptosis are effective but at different stages of hepatocarcinogenesis [212]. Results clearly shows that overexpression of E2F1 increases widespread apoptosis in gastric carcinoma cells, probably by upregulation of PTEN expression *via* activation of caspases like caspase-3 and caspase-9 particularly when combined with CDK inhibitors, for example roscovitine [213-215]. In general, these conclusions propose that the loss of E2F1 activity promote tumor development by reducing cellular apoptosis. E2F1 mediated apoptosis is given in Fig 2.9.

MATERIAL AND METHOD

3.1.1 SNP information retrieval

Information regarding *E2F1* gene, especially SNP was obtained from databases like dbSNP [55], UniProt [56], Ensemble genome browser [54], and catalogue of somatic mutations in cancer (COSMIC) database [57]. We also checked the information related to functional annotation of each SNP in these databases, such as whether it is present in exons or introns, in the 5' untranslated region (UTR), 3' untranslated region, upstream or downstream of the *E2F1* gene. Total 1366 SNPs were reported in the *E2F1* gene from different databases. We retrieved the crystal structure of E2F1 (PDB ID: 2AZE) [216] and E2F4 (PDB ID: 1CF7) [217] from the Protein Data Bank (PDB) [218]. To analyze protein structural changes in protein by mutations using various computational tools.

3.2.1 Sequence-based prediction

There are 143 amino acid variants reported in E2F1 protein in different databases like NCBI, COSMIC, Ensembl, and UniProt. We checked the effect of single amino acid change at a time in FASTA sequence of E2F1 protein by using different computational tools such as SIFT [219], PhD-SNP [220], PROVEAN [221], Align Grantham Variation and Grantham Deviation (GVGD) [222, 223], fathmm [224], Polyphen-2.0 [225], SNAP [226], SNPs&GO [227], SuSpect [228], and I-Mutant 3.0 [229]. SIFT can be used in natural occurring nsSNPs as well as in somatic mis-sense mutations. SIFT prediction is based on multiple sequence alignment (MSA) information. The amino acid positions essential for protein function should be conserved in an alignment of the protein family, while insignificant positions should show variation in an alignment. SIFT assigns scores for all residues ranging from zero to one. According to the SIFT analysis, if a variant has score less than 0.05 then it is considered as deleterious mutation; while a mutation having >0.05 is considered as tolerated mutation [219].

PROVEAN (Protein Variation Effect Analyzer) predicts the effect of amino acid substitution on the biological function of a protein [221]. Polyphen-2.0 is a Bayesian classifier predicts the probable impact of amino acid substitutions on the structure and function of a human protein based on the sequence, phylogenetic and structural information characterizing the substitution

[225]. The SNAP (screening for non-acceptable polymorphism) method is based on neural networks (NN) and utilises an advanced machine-learning approach to predict the functional impact of amino acids substitution on proteins [226]. SNPs&GO is a method based on SVM that predicts disease-associated mutations from a protein sequence profile and functional information in gene ontology (GO) terms. In this method, a probability score >0.5 predicts that the mutation will have deleterious effect on protein function [227]. PhD-SNP is a single-sequence SVM based method that differentiates disease-related mutations on the basis of sequence and profile based information. This tool predicts whether an amino acid substitution is neutral or a disease-associated [220]. Fathmm combines sequence conservation in hidden Markov models (HMMs), which express the alignment of homologous sequences and conserved protein domains with "pathogenicity weights that represent the tolerance of the amino acid substitutions [224]. SuSpect uses sequence, structure and systems-biology approach to predict the phenotypic effects of mis-sense mutation [228].

3.2.1 Biophysical characterisation

Align-GVGD predicts whether an amino acid substitution is deleterious or neutral on the basis of evolutionary conservation and the chemical characteristic of amino acid residues. Align-GVGD assigns class of probability on the basis of this analysis. The C-score, assign seven different grades C0 to C65, which states that mutations that are least likely to be neutral are present under (class 65) and those that are most likely to be neutral present under (class 0) in provisions of the function of the protein [11,12].

3.2.2 Sequence-based protein stability prediction

The sequence level protein stability analysis was completed by analysing Gibbs free energy change. The more negative value of Gibbs free energy indicates higher stability while positive Gibbs free energy change indicates destabilizing mutation. In this study, we used I-Mutant 3.0 [32]. I-Mutant 3.0 classification was based on SVM and it trained with the inclusive dataset derived from ProTherm [230]. This method predicts protein stability changes caused by amino acid substitution and analyze the change in free energy between native and mutant sequences. I-Mutant 3.0 classification is of three types, neutral mutations (-0.5≤ Kcal/Mol), mutations that

produces a large decrease in Gibbs free energy (-0.5<Kcal/Mol), and mutations that produces a large increase (0.5>Kcal/Mol).

3.2.3 Analysis of evolutionary conservation profile

The structure of protein is maintained by amino acid sequence of the protein which depends on the degree of conservation of the amino acid residue. We checked the amino acid conservation profile by ConSurf [231], and by the analysis of MSA profile of E2F1 protein sequence in different vertebrate species which calculates the degree of conservation at each aligned position. ConSurf provides the conservation profile of protein and nucleic acid by determining the conserved positions and quantify the evolutionary conservation rate using an empirical Bayesian inference. ConSurf assign score from 1 to 9 according to the conservation profile of amino acid residue 1 indicates rapidly evolving sites, 5 indicates average evolving sites and 9 denotes highly conserved or slow-evolving sites.

3.2.4 Primary sequence analysis

We performed multiple sequence alignment using MUSCLE (Multiple Sequence Comparison by Log-Expectation), to verify the amino acid conservation profile of E2F1 protein sequence. MUSCLE is most commonly used to check conservation profile of protein sequence in different species. It is a web-based tool which can be used to align multiple sequences from different species [232]. The primary sequence information is important in cases where there is no structural information available. It gives information regarding possible functional mutation sites. We retrieved the protein sequence information of E2F1 protein sequence of different vertebrates from National Centre for Biotechnology Information (NCBI). The analysis of sequence was performed by using BioEdit program [233]. This program displays a graphical representation of amino acids, patterns in a set of aligned sequences.

3.3.1 Structural and functional analysis

To date, numerous computational methods are available for predicting the phenotypic effects of nsSNP the most of them are based on protein sequence and evolutionary conservation [49]. Structure-based computational methods give better insight on phenotypic effect of protein. In this study, all the nsSNP were not present on the available crystal structure of E2F1 protein.

Therefore, in the next phase of analysis, we used seven structure and functional analyses based protein tools (discussed in the next section).

3.3.2 Ranking scheme adapted by structural and functional analysis of nsSNPs

We adopted a ranking scheme to classify the nsSNPs as most deleterious by scores obtained through different computational tools. We used seven structure and functional analyses based computational tools, Atomic Non-Local Energy Assessment (ANOLEA) calculate the non-local energy profile (NL-profile) of a protein predicting changes in the stability of proteins caused by mutations [234]. *Site-Directed Mutator* (SDM), I-Mutant 3.0, ERIS, SNP effect 4.0, Polyphene-2, mutation assessor and PROVEAN, Site-Directed Mutator (SDM) calculates the difference in the stability scores for the folded and unfolded state for the native and mutant protein structures [235]. I-Mutant 3.0 can predict that to which extent a mutation in a protein sequence will or will not affect the stability of the folded protein. It classifies protein stability on the basis of free energy change, if DDG values for neutral mutation is -0.5=<DDG=<0.5 while changes <-0.5 are classified as large decrease and change >0.5 are classified as large increase [229]. Eris server calculate the protein stability change by amino acid substitution ($\Delta\Delta G$) by utilizing the recently developed medusa modelling suite [236]. SNP effect 4.0 is a database for phenotyping human SNPs. Which is mainly centred on the molecular characterization and annotation of disease and polymorphic variants in the human proteome [237]. Polyphen-2 Predicts the possible impact of amino acid substitutions on the stability and function of human proteins using structural and comparative evolutionary considerations [225]. Mutation-assessor prediction is based on evolutionary conservation of the affected amino acid such as mutations discovered in cancer or missense polymorphisms [238]. Provean is a tool which evaluates the impact of biological function of a protein by amino acid substitution [221].

We assigned a nsSNP as most deleterious if a minimum of 6 predictions agree that the SNP is damaging. Identified damaging mutations alongwith native proteins were run in Gromacs for Molecular Dynamics simulation analysis.

3.3.3 Molecular dynamics simulation

A molecular dynamic simulation study was performed for in depth insights about structural changes and was performed in GROMACS 5.1.2 software [239]. OPLS-AA/L all-atom force field [240],were used in native 2AZE and mutant Cys227Phe, Val295Asp and Cys298Tyr and AMBER99SB-ILDN force field were used in (1CF7) native and two another mutant proteins (R56W and Y59C) [241]. These protein complexes were solvated in a triclinic 1.5 nm of SPC water molecules [242]. A periodic boundary condition was applied in such a way that particle number, temperature and pressure remained constant during simulation period. The simulation framework was neutralized by the addition of sodium ions to the system; this can be achieved by adding Na^+ ion to both the wild- type and mutated E2F1 protein topology files. Protein and protein-DNA complex atoms were placed at an equal distance of 1.5nm from the triclinic box edges. The minimized simulation setup was then equilibrated for 100ps at 300K. Then equilibrated simulation framework was subjected to MD simulation for 50ns. During the simulation period, the temperature was kept constant at 300K.

3.4 Population-based case-control study:

3.4.1 Selection of SNPs

In this study, we included four nsSNP of *E2F1* gene rs3213172, rs3213173, rs3213176 and rs574956843 and one intronic variant rs2071054 for case-control association study. Two nsSNP rs3213172 and rs574956843 were highly damaging at the sequence as well as structural analysis. The rs3213173 and rs3213176 were predicted as disease risk SNP by National Institute of Environmental Health Sciences SNPs program (NIEHS SNPs Program) also. In this study, one intronic variant, rs2071054 of *E2F1* gene having higher minor allele frequency (MAF) in the different ethnic group is also included for association analysis.

3.4.2 Study Population

This case-control study comprised histopathologically proven cases of 163 lung, 142 head and neck cancer and 90 cervical cancer patients and 230 cancer-free control samples. Peripheral blood samples from patients, who were being treated at Regional Cancer Centre, Indira Gandhi Medical College (IGMC) Shimla were collected. Information regarding age, sex, smoking status, drinking status, occupational status, area inhabited, and eating habits were recorded from patients

file. All the participants signed an informed consent form and completed a questionnaire providing information on gender, race, ethnicity, education, religion, marital status, smoking habits (current smoker, ex-smoker), addiction to tobacco, occupational exposures, first-degree family history, history of benign disease etc. Information regarding stage of the disease was obtained from medical records of the patients. Patient are not including in study if they are suffering from other disease in spite of cancer. Patients and controls both are excluded from study if their ethnicity is not Himachali. The study was approved by Institutional Ethics Committee (IEC/project no-25-2015) of Jaypee University of Information Technology (H.P).

3.4.3 Sample collection and DNA isolation: 5ml of peripheral blood samples were collected in EDTA coated vial with the help of sterile and disposable syringes. The collected blood samples were transported to the laboratory in a mini cooler. Processing of blood samples was performed in Biosafety Cabinet Class II. DNA was isolated from the blood following method given by Miller at al [243].

3.5.1.1 Reagents Required for DNA isolation

- **Tris (hydroxymethyl) aminomethane-chloride *(Tris-Cl)*** To make (1 M; pH 8.0) Tris-Cl, 12.11 gm of Tris base was dissolved in 75 ml of milli-Q water and pH was set to 8.0 by adding drops of 1N HCl and the final volume made 100 ml with milli-Q.
- ***Tris-Cl* (1 M; pH 7.3):** As mentioned above, 12.11 gm of Tris base was dissolved in 75 ml of milli-q water and pH was set to 7.3 by adding drops of 1N HCl and final volume was made to 100 ml.
- **Ammonium chloride: (1M; NH_4Cl)** 5.35 gm of ammonium chloride was dissolved in final 100 ml of milli-Q water.
- **0.5M (Na_2EDTA; pH 8.0):** Dissolved 18.61 gm of Na_2EDTA in 50 ml of milli-Q water in a 250 ml flask and placed on a magnetic stirrer and added 10 M of NaOH dropwise until the pH reached at 8.0. To dissolve Na_2EDTA properly a gentle shaking was done and the final volume was set to 100 ml.
- **RBC Lysis buffer:** Add 10 ml Tris (1 M; pH 8.0), 2 ml 0.5 M EDTA and 125 ml of NH_4Cl (1 M) in 1000 ml flask and final volume adjusted is up to 1000 ml.
- **TE- buffer pH (8.0):** This buffer was prepared by mixing 10 ml Tris-Cl (1M) and 2 ml of 0.5M EDTA in 700 ml of milli-q water and final volume was adjusted up to 1000 ml.
- **TE- buffer pH (7.3):** This buffer was prepared by adding 10 ml Tris-Cl (1M; pH 7.3) and 2

ml of 0.5 M EDTA in 700 ml milli-q water and final volume adjusted up to 1000 ml.

- **10% Sodium dodecyl sulfate (SDS):** 10 gm of SDS salt was dissolved in 70 ml milli-q water and final volume up to 100 ml.
- **7.5M Ammonium acetate:** To prepare 7.5M Ammonium acetate, 28.9 gm of ammonium acetate salt was dissolved in 20 ml milli-q water and final volume adjusted to 1000 ml.
- **Chilled absolute ethanol:** Absolute ethanol kept at -20˚C or -80˚C before its use.
- **(70%) Ethanol:** 70 ml of absolute ethanol was mixed in 30 ml of sterile milli-Q water.

3.5.1.2 Genomic DNA Isolation Protocol:

- In a 2 ml conical micro-centrifuge tube, 300 µl of a blood sample was taken and 900 µl of RBC lysis buffer was added and the tube was kept on a rocker at room temperature (RT) on a rocker for appximately 15 minutes. The bright red color appears after lysis of RBCs.
- The sample was centrifuged at 13,000 rpm for 2 minutes and WBCs pellet was obtained.
- The supernatant was discarded and WBC pellet was resuspended in 300 µl TE buffer (pH 8.0).
- After vortexing, the above mixture 22µl of 10% SDS solution was added and the mixture was incubated at 56ºC for 30 minutes.
- Added 160 µl of 7.5M ammonium acetate and mixture was vigorously shaken for about 1 minute on a vortex machine. This mixture was centrifuged at 13,000 rpm, for 15 minutes at 25˚C.
- The supernatant containing DNA was transferred to a fresh 1.5 ml centrifuge tube, and added the double volume of chilled absolute ethanol. By gentle shaking DNA was precipitated from the mixture.
- The precipitated DNA was centrifuged at 13,000 rpm for 10 minutes to obtain the DNA pellet,
- This pallet was washed with 500 µl of 70 % ethanol, to remove contamination from DNA pallet.
- The pellet was then air-dried at RT for about 10-15 minutes.
- The obtained DNA pellet was dissolved in 100µl of TE buffer (pH 7.3) by incubating it at 65º C for 10 minutes.
- Dissolved DNA was finally stored at -20°C for further use.

3.6.1 Quantification of DNA: Quantification of DNA was determined by nanodrop. Multiskan GO 1.00.40 (µDrop Plate, Thermo Scientific) was used for the quantification of DNA in micro-

litre volumes.

3.6.1.1 Procedure:

1) Blank solution was taken in which DNA was dissolved (TE buffer pH-7.3).
2) Now place 1µl of the blank solution on the sample loading surface and close the lid. Set it as a blank.
3) Again open the lid wipe the sample loading surface and load 1µl of DNA sample and close the lid.
4) Measure the reading.
5) Note the reading of concentration (ng/µl) of DNA, absorbance (A260/280) and 260/230.
6) Again repeat steps 3-4 for another sample.
7) The good quality of DNA was determined by the analysis of 260/280 ratio. This ratio depicted the contamination of EDTA protein and a phenolic compound. If this ratio is between (1.8 and 2.0) that means isolated DNA have good quality. For a good quality sample, the value should be between 1.8 and 2.0.

3.7.1 Agarose gel electrophoresis (Sambrook and Russell, 2001)

The integrity of DNA was determined by agarose gel electrophoresis (0.8%).

3.7.1.1 Required Reagent

- **Agarose:** SeaKem® LE Agarose-Lonza **was** used for the agarose gel electrophoresis.
- **Ethidium bromide (EtBr) (10mg/ml):** 100 mg of EtBr was dissolved in 10 ml of milli-Q water and stored in dark bottle at room temperature.
- **Gel loading dye (10x):** 0.25 grams of xylene cyanol was dissolved in 50 ml of milli-Q water and added 50 ml of 50% glycerol and make up the *volume* to 100 ml with milli-Q water.
- **Gel Loading Dye (6x):** 0.25 grams of the bromophenol blue was dissolved in 50 ml of milli-Q water and then added 50 ml of 50% glycerol.
- **TAE buffer (50x):** TAE buffer was prepared by dissolving 242 grams of Tris base in 500 ml of milli-Q water and then added 100 ml of 0.5M EDTA. After that, 57.1 ml of glacial acetic acid was added, and total volume was made up to 1000 ml. Prepared TAE buffer was filtered and stored at room temperature.

3.7.1.2 Procedure

- 1X TAE buffer is prepared from 50X TAE buffer.
- Weighed 0.4g agarose and add it to 50 ml of 1X TAE buffer and heated in the microwave until a cleared solution is not formed.
- Added 3µl (0.5 ug/ml final concentration) of EtBr to the lukewarm gel solution and mixed.
- The solution is carefully poured towards opposite side of comb to avoid entrapment of any bubbles near the wells.
- The gel was allowed to solidify; comb was gently removed after the gel was solidified.
- Added 1µl loading dye in 5 µl DNA sample and dispend it to well after proper mixing.
- The voltage supplied is 10 x distance between the two electrodes of the unit and kept for 30 minutes at 100 volts.
- Observed the bands of DNA in Gel Documentation System to check the integrity of isolated DNA samples.

3.7.1.3 Photo documentation

- After completing gel run, the gel was removed from the tank and DNA was visualized by placing gel on UV-transilluminator (Alpha Innotech, New Delhi India).
- Bands were visualized under UV-B (302 nm) in Gel documentation system.

3.8.1 Genotyping of SNPs

The primers were designed for prioritized SNPs manually and using primer blast tool. The set of forward 5'-$A_{(15)}C_{(15)}$TCTGCAGGGTCTGCAATG**G**TA-3' and the reverse 5'-$A_{(15)}C_{(15)}$ TCCCAACCTCCTACCCACTCA-3' primers were used to genotype rs3213172 nsSNP. We created an artificial restriction site for *Rsa*I restriction enzyme to genotype this nsSNP by substituting C with **G** in forward primer.

The rs3213173 nsSNP was also genotyped by above mention genotyping method. The pair of primers: forward 5'-$A_{(15)}C_{(15)}$ATCTCACCTCCGAAGAGTCC**C**-3' and reverse 5'-GTC CTGACACGTCACGTAGGC-3' were used to genotype rs3213173 nsSNP. We substituted A with **C** in forward primer for creating restriction site of *Msp*I restriction enzyme. There was natural restriction site present for rs3213176 nsSNP. The strategy of restriction site creation by using mutagenic primer presented in Fig 3.1. The forward 5'-TGGACGAGGACCGCC TGT-3' and reverse 5'-CTCACCTTGTCTCTGCAGCC-3' primer sequences were used to genotype rs3213176 nsSNP.

The set of forward and reverse primers sequences (forward 5'-$A_{(15)}C_{(15)}$ CTGGATAAATTGCTTAACCTC**C**-3' and reverse 5'-AATGTTCAACACAACTCCAGG-3') used for genotyping of rs2071054 (C/T) intronic variant of *E2F1* gene. The bold and underline nucleotide **C** is substitution of T in forward primer. We added 5'-$A_{(15)}C_{(15)}$3' tail in all mutagenic primer so, that digested PCR fragments were easily visualized on gel.

PCR reaction was performed for each SNP in a total volume of 25µl; containing 40ng of genomic DNA, 12.5 µl green master mix of (Promega), and 0.16mM of each primer were used for amplification.

Standardization of DNA amplification conditions and optimization of annealing temperature for the set of primers were achieved through Gradient PCR. All the PCR amplifications were performed using reagents: green master mix (Promega), a total reaction volume of 25 µl containing nearly 40 ng genomic DNA, and 10 pmol of each working primer. The PCR amplification profile consisted of an initial denaturing step at 94°C for 3 minutes, followed by 32 cycles of amplification which included denaturing at 94°C for 40 sec, annealing at 52 - 59°C for 40 sec, extension at 72°C for 1 minute and a final extension at 72°C for 5 minutes. The PCR products were resolved on 1.2% agarose gel in (1X) TAE buffer at 100 volts for 1.0 hours.

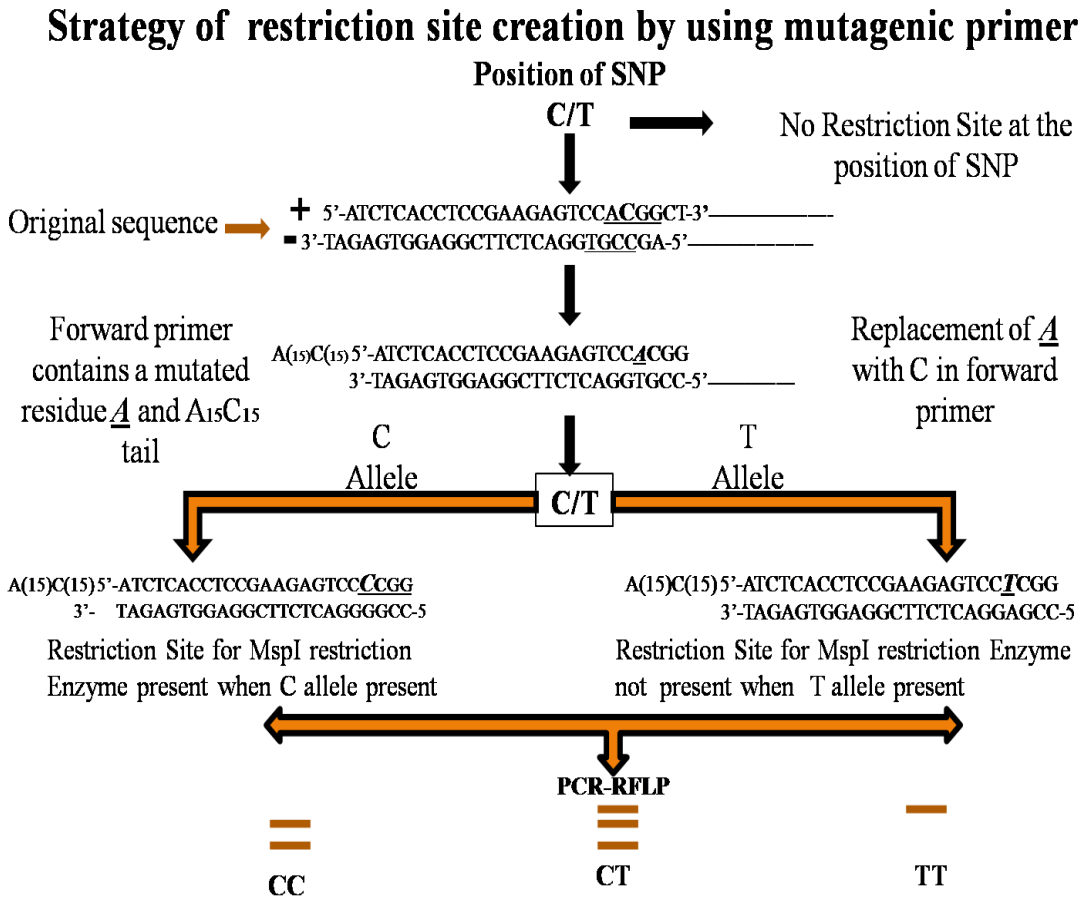


Fig 3.1 Representation of restriction site creation strategy for rs3213173 nsSNP

These amplified PCR products were digested overnight at 37˚C by *Msp*I (New England Biolabs) restriction enzyme used for rs3213173, rs3213176, and rs2071054 SNP and *Rsa*I (New England Biolabs) for rs3213172 SNP restriction enzyme) and *Acc*65I (New England Biolabs) restriction enzyme used to genotype rs574956843 SNP. The PCR cycle conditions primer detail and fragment size detail were shown in Table 3.1. Digested PCR products were analyzed by agarose gel electrophoresis (3 % agarose).

In the case of the rs3213172, the forward primer was specially designed to create the recognition site of the restriction enzyme *Rsa*I by replacing a G with a C which presented as bold color and underline in primer. The PCR reaction was carried out in a total volume of 12.5µl, containing 50ng of 1µl of genomic DNA, 0.25µl of each primer, 5.5 µl of redistilled water, and 5.5µl of GoTaq Green PCR Master Mix (Promega). The amplified DNA fragments were digested overnight at 37°C with the indicated restriction enzyme. DNA fragments were resolved on an agarose gel and stained with ethidium bromide. Amplified product length, optimal annealing temperature, and banding pattern of genotypes have been shown in Table 3.1

3.8 Statistical Analysis

The multiple logistic regression analyses were used to assess the relations of the *E2F1* genetic variant with the cancer risk. Odd ratio and 95% confidence interval were used to check the fold risk. Each of the statistical tests was two-sided and the P value of <0.05 was considered significant. Power of the study was analyzed by using the Online Sample Size Estimator (OSSE).

Table 3.1: Details of the primers, thermocycler conditions, and fragment size

rs ID	Primer sequence	Thermo cycler condition	Fragment size (bp)
rs3213172	**FP:** 5'-$A_{(15)}C_{(15)}$TCTGCAGGGTCTGCAATG**G**TA3' **RP:** 5' TGAGTGGGTAGGAGGTTGGGA 3'	94˚- 4 min, 94˚- 40 sec,59-40sec,72-1 min,72-5 min	CC = 151 and 51 CT = 202, 151 and 51 TT = 202
rs3213173	**FP:** 5'-$A_{(15)}C_{(15)}$ATCTCACCTCCGAAGAGTCC**C**-3' **RP:**5-GCCTACGTGACGTGTCAGGAC-3'	94˚-4 min,94˚-40 se 58˚-40sec,72˚-1 min 72˚-5 min	CC = 117 and 51 CT = 168, 117 and 51 TT = 168
rs3213176	**FP:** 5'-TGGACGAGGACCGCCTGT-3' **RP:** 5'-GGCTGCAGAGACAAGGTGAG-3'	94˚-4 min,94˚-40 se 59˚-40sec,72˚-1 min 72˚-5 min	GG = 191+71 GA = 262+191+71 AA = 262
rs574956843	FP: 5'TGAGTAGAGACTGGCTGGGAT 3' RP: 5'-$A_{(15)}C_{(15)}$AGGCCCGATCGATGTTTTCC**G**G3	94˚-4 min,94˚-40 se 57˚-40sec,72˚-1 min 72˚-5 min	CC = 215 CT = Not observed TT = Not observed
rs2071054	**FP:** 5'-$A_{(15)}C_{(15)}$CTGGATAAATTGCTTAACCTC**C**-3 **RP:**5'-AATGTTCAACACAACTCCAGG-3'	94˚-4 min,94˚-40 se 52˚-40sec,72˚-1 min 72˚-5 min	CC = 276+53 CT = 329+276+53 TT = 329

RESULTS

This study is focused on *E2F1* gene to find SNPs that play role in the cancer susceptibility that could be used as markers for identification of patients who are at higher risk of cancer; so that they could take preventive measure well in advance (at least onset of the disease could be delayed). The study is divided into two parts: (1) prediction of damaging nsSNPs by *in silico* methods (less number of SNP would have to be typed to find alleles which are associated with the disease), and (2) analysis of these predicted SNPs in the cancer patients.

4.1 *In silico* analysis for prediction of deleterious nsSNPs

4.1.1 Prediction of deleterious nsSNPs

This study is focused on nsSNPs of *E2F1* gene. An outline of the different methods used in the study is presented schematically in Fig 4.1. Distribution of SNPs on this gene was checked and the observed number of various variants in decreasing order was: intronic variants (IV) > downstream gene variants (DGV) > upstream gene variants (UGV) > missense variants (MV) > synonymous gene variants (SGV) > 3' untranslated region (3'UTR) > 5' UTR (Fig 4.2A). There are 143 nsSNPs reported in *E2F1* gene in various computational databases. The nucleotides substitutions of 143 missense variants of *E2F1* gene were checked, and C/T transition was one of the most frequent transitions in 143 missense variants. The nsSNPs in *E2F1* gene are responsible for transitions and transversion (Fig 4.2B).

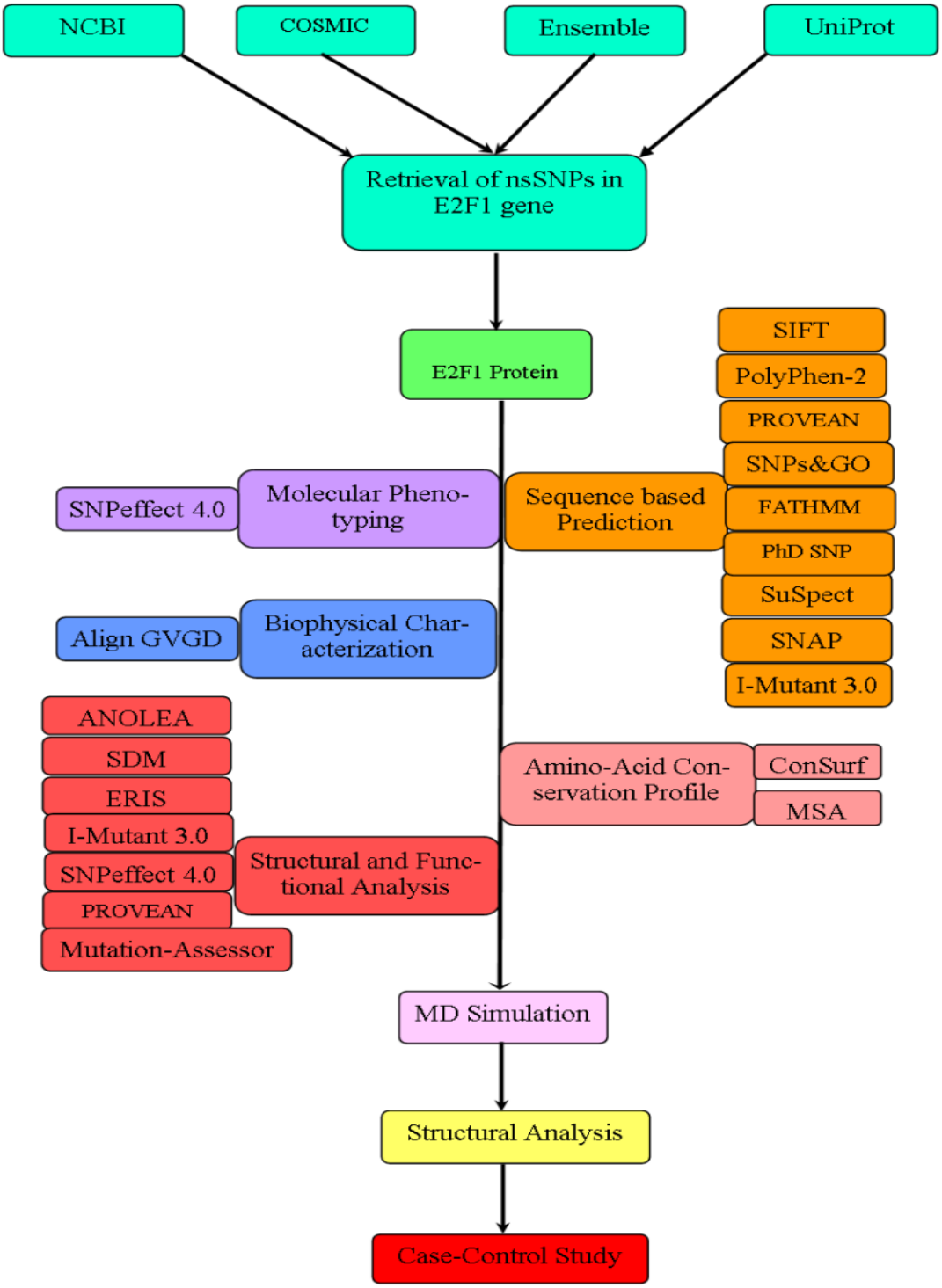


Fig 4.1 A systematic *in silico* approaches to classify damaging mutations in *E2F1* gene.

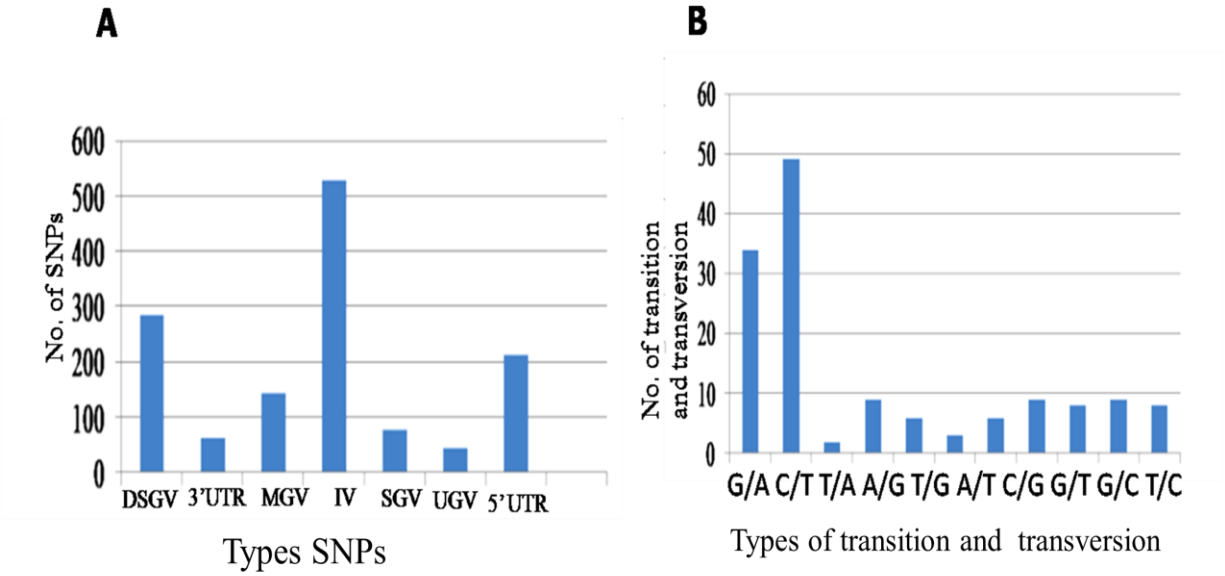


Fig 4.2 Distribution of SNPs along with various substitutions in E2F1. (A) Distribution of SNPs on E2F1 gene was determined by computational databases. The highest number of SNPs were present in IV followed by DSGV, 5` UTR, MGV, SV, 3` UTR and UGV. (**B)** The nsSNPs were further checked for various substitutions and are presented in this fig. The transitions and transversions to C/T was highest followed by G/A; C/A, C/G and C/G were present in same number. T/C and G/T transitions were also present in same in number, while T/G and T/A transitions were rare.

Deleterious and neutral SNPs predicted by computational tools are presented in Fig 4.3. About 30-50% nsSNPs were predicted as deleterious by computational tools. A total of 129 (90%) nsSNPs were predicted as damaging by at least one of the computational tools (Appendices 4.1). Moreover, most of these damaging mutations were located at the conserved region of this gene.

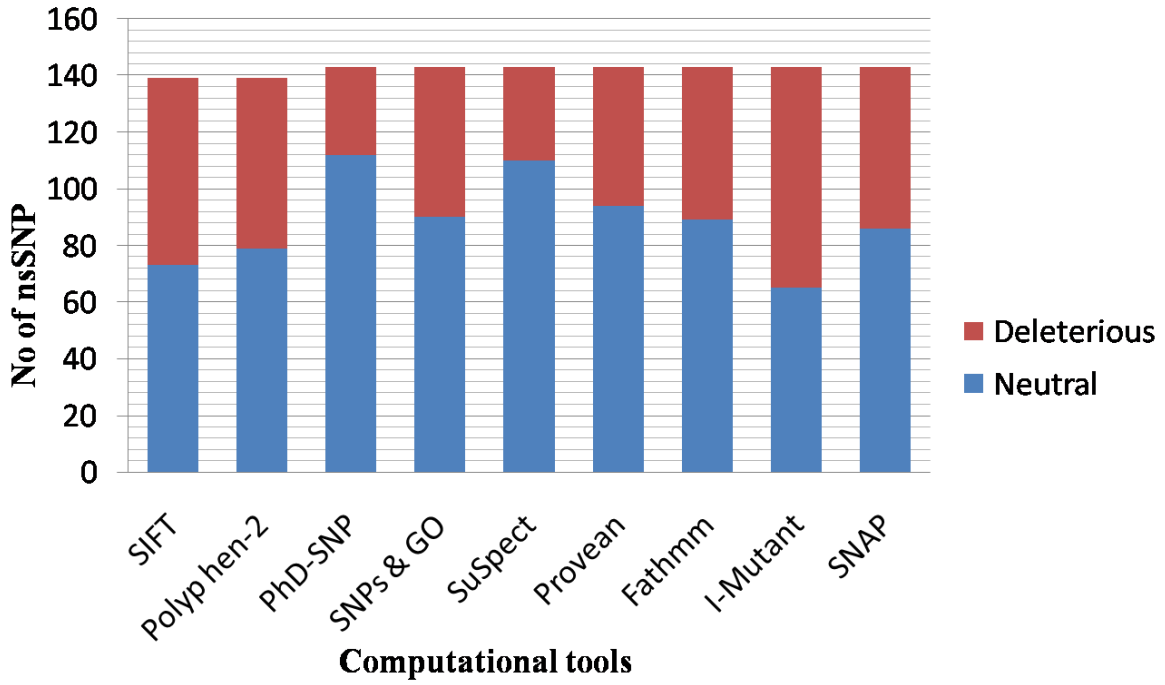


Fig 4.3 Deleterious and neutral predicted by different computational tools.

4.1.2 Biophysical characterisation

Align-GVGD provides a class probability which depends on chemical characteristics and evolutionary conservation of amino acid residues [222, 223]. This method compares the chemical and physical properties obtained by amino acids substitutions. The C-score provides a spectrum of seven classes (C0, C15, C25, C35, C45, C55 and C65). The mutation would be categorised as least likely to be neutral if amino acids substitutions belong to C65, while mutations that are the most likely to be neutral would belong to C0 in terms of protein's function. In total, 70 nsSNPs (49%) occurred in a conserved region of the E2F1 protein sequence had a C-score of 65. The remaining SNPs have classified as class 0 2 (1.39%), 15 15 (10.48%) and least likely to be neutral (Appendices 4.2).

4.1.3 Amino acids composition of native and mutated proteins

To evaluate the effect of deleterious amino acid variants in the E2F1 protein, we determined the amino acids composition of the protein in both its native and mutated forms using statistical analysis of protein sequences in ms-excel work sheet. Out of 437 AAs, the incidence of leucine was found to be highest, followed by Lys, Pro, Ala, Gly, and Asp. The mutant proteins represent a higher relative composition of Lys and Pro, followed by Arg and

Val. Ala→Thr, Arg→His, Pro→Leu, Val→Met, Arg →Gly, Ala Arg →Trp, Glu→Lys and Glu →Asp were the most common substitutions in E2F1 (Fig 4.4). The Ala to Thr was the most frequent substitution, followed by Arg to His. The Arg to His substitution was damaging in most of the *in silico* prediction analyses. In addition, the formation of disulfide bonds in the E2F1 protein was calculated using SCRATCH Protein Predictor [58]. Three of the Cys residues present at positions 12, 227 and 298 in E2F1 (total 5 Cys residues were present at positions 12, 59,248,227 and 298) were found to be involved in disulfide bond formation. Mutations at positions C227F and C298Y were predicted deleterious by SIFT, PolyPhen2, PhD-SNP, SNAP, SNPs & GO, Fathmm, Provean and I-mutant 3.

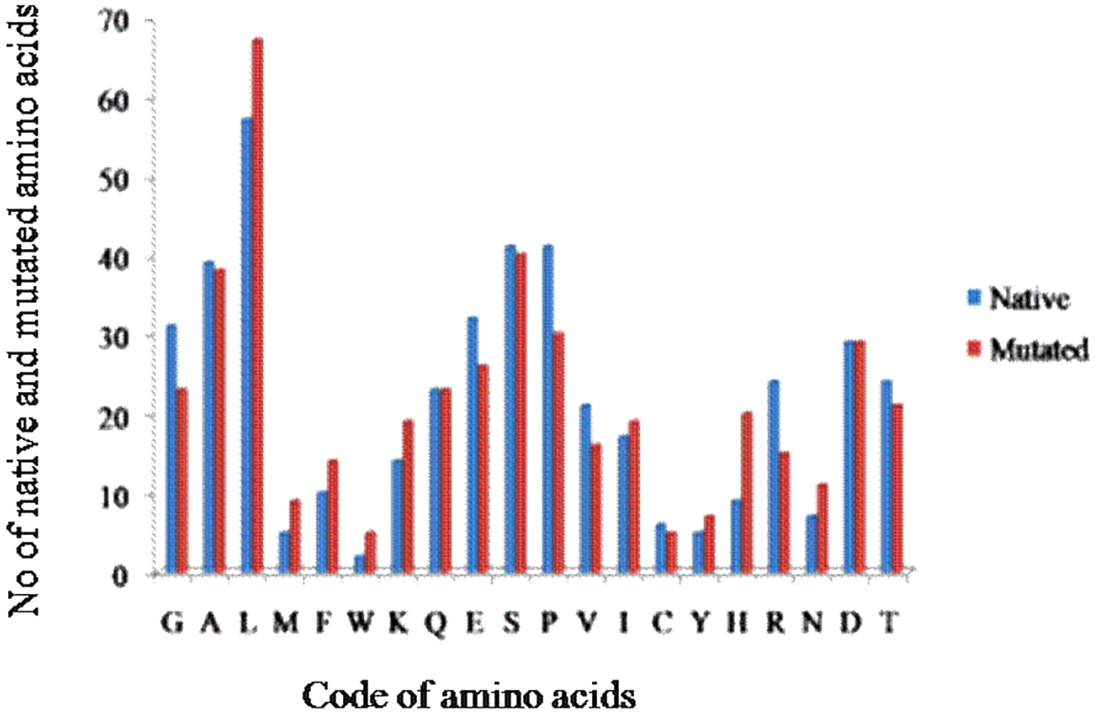


Fig 4.4 Native and mutated residues in E2F1 protein: blue colour indicates native amino acids and fire brick colour indicates mutated amino acids in E2F1 protein.

4.1.4 E2F1 protein sequence conservation analysis

E2F1 transcription factor has four domains: DNA-binding domain (127-192 residues) at N-terminal and in C-terminal it contain coiled-coil (CC) region (177-217 residues), transactivation domain (202-301residues). It also contains a marked-box domain (252-384 residues) [39]. Evolutionary conservation profile of E2F1 protein domain (Fig 4.5) was analyzed by using ConSurf. The result of ConSurf indicated that amino acid residues 88-308 were highly conserved during the evolution process [231]. We further analysed the conservation profile of amino acids residues by multiple sequence alignment tool and got result somewhat similar results like ConSurf.

In the analysis of multiple sequence alignment to other homologues of E2F1 protein gives a glimpse of protein's evolution [232]. The ConSurf result indicated that the amino acids of DNA binding domain, coiled-coil (CC) region and transactivation domain of E2F1 protein were highly conserved.

Table 4.1 E2F1 Homologous proteins sequences used in the MUSCLE analysis.

S.No	Different Species	Protein/Gene ID	Protein Length	NCBI Accession
1	Homo Sapienens	Uniprot/swissprot Q01094.1	437	NP-005216
2	Mus musculus (house mouse)	Uniprot/swissprot Q61501.1	430	NP-031917
3	Rattus norvegicus (Norway rat)	NCBI ref.Seq XP-00876058.1	375	XP-008760585
4	Cavia porcellus (domestic guinea pig)	NCBI ref.Seq XP-013005865.1	389	XP-013005865
5	Gallus gallus (chicken)	NCBI ref.Seq XP-015151824.1	406	XP-015151824
6	Xenopus laevis (African clawed frog)	NCBI ref.Seq NP-001090608	426	NP-00109608
7	Danio rerio (Zebra fish)	NCBI ref.Seq XP-005174590.1	431	XP-00174590

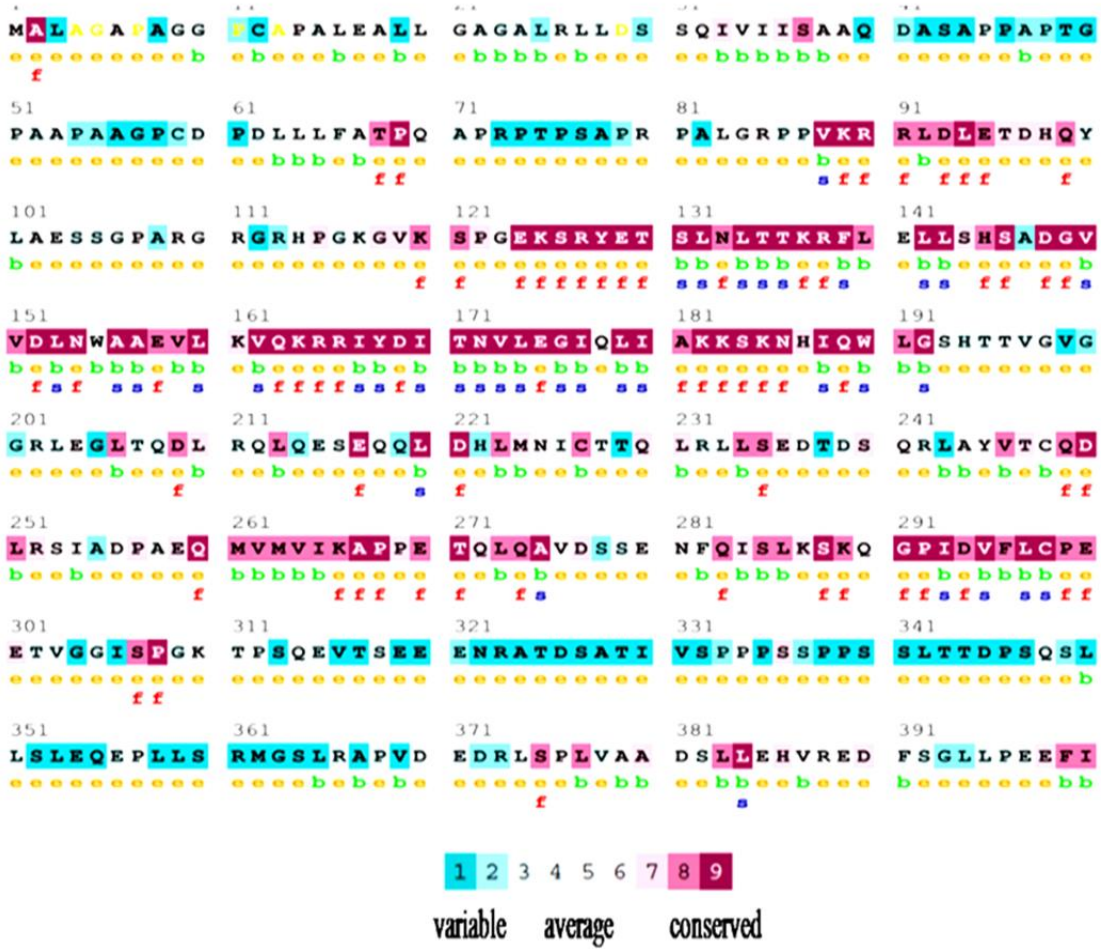


Fig 4.5 Amino acids conservation profile of E2F1 protein: Turquoise blue colour indicates rapidly evolving site; white colour indicates average evolving site and maroon colours indicates evolutionarily conserved site. **e** - An exposed residue, **b** - A buried residue, **f** - A predicted functional residue and **s** - A predicted structural residue

It is generally accepted that a substitution at any of the conserved residue position mostly causes deleterious effect on the protein. Thus, we investigated the conservation profile of E2F1 protein in ConSurf and MUSCLE. The conservation profile of E2F1 protein was analyzed in MUSCLE. For the validation homologous sequences of this protein were used from different vertebrate species (Table 4.1). In this analysis, we got nearly similar result of amino acids conservation like ConSurf (Fig.4.5).

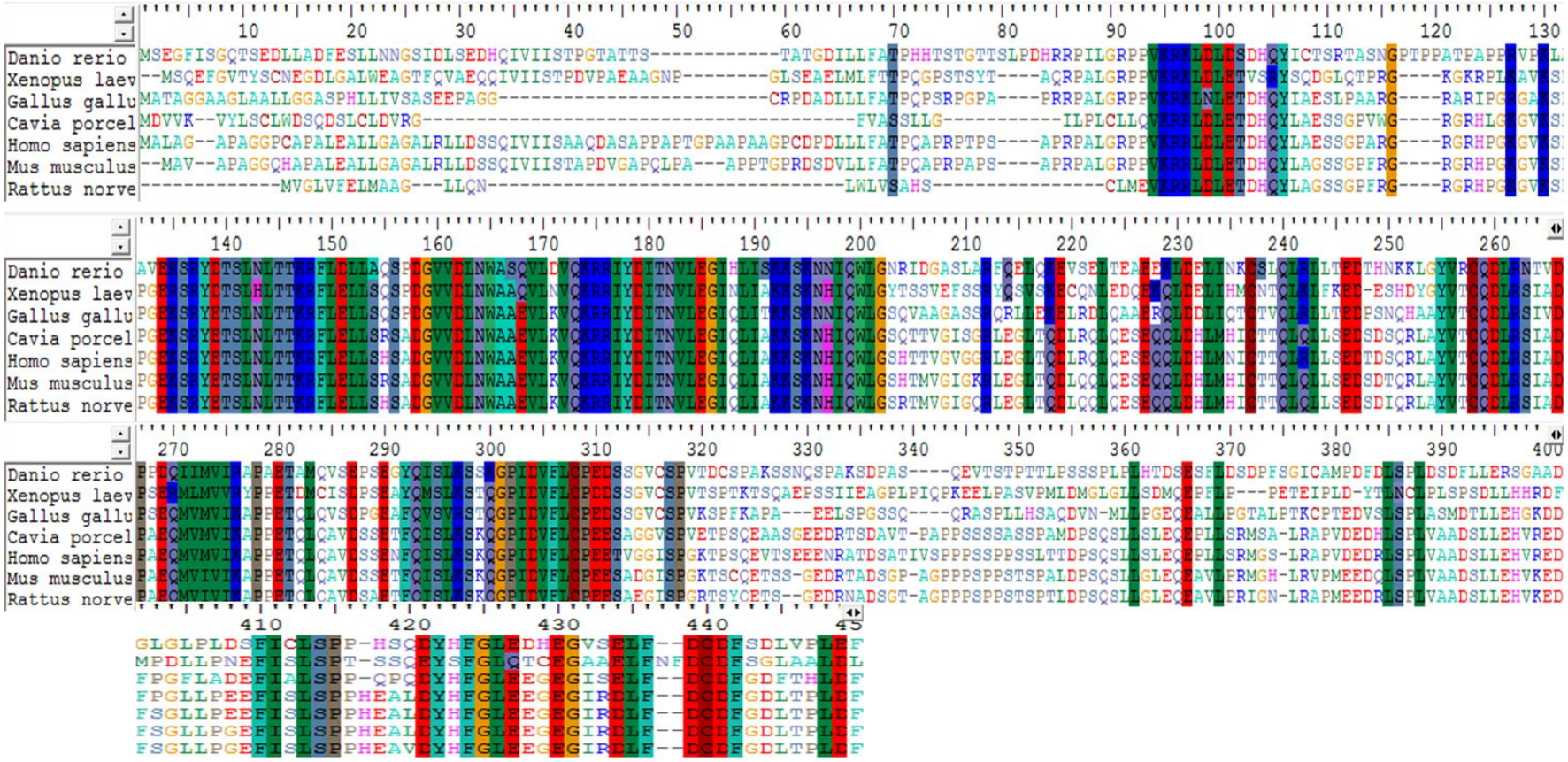


Fig 4.6 Multiple sequence alignment of seven selected E2F1 protein sequences from different species: The sequence alignment was performed with MUSCLE. The accession numbers of proteins used are as follows: *Homo sapiens (*NP-005216), *Mus musculus* (NP-031917), *Rattus norvegicus* (XP-008760585), *Cavia porcellus* (XP-013005865), *Gallus gallus* (XP-015151824), *Xenopus laevis* (NP-00109608) and Danio rerio (XP-00174590).

4.1.5 Molecular Phenotype of E2F1 protein

The SNPeffect4.0 tool was used to check the molecular phenotypes which identify polymorphic variants associated with the disease. The SNPeffect4.0 classification of amino acid variants is based on changes in aggregation, amyloidogenicity, chaperone binding sites and structural stability. These criteria facilitate the prediction of the mutants that would affect the protein structure [237]. The amino acid variants encoded by nsSNPs were mapped onto the available structure of E2F1 and E2F4. Total of 47 amino acid variants were located on available structure of E2F1 and E2F4, and 17 (36 %) of these were found to be associated with a significant structural changes (Appendices 4.3).

4.1.6 Ranking scheme adapted for structural and functional analysis of nsSNPs

The nsSNPs were classified as most deleterious by adopting a consensus ranking system on the basis of result obtained by various computational predictions tools, such as Provean [221], PolyPhen-2.0 [225], I-Mutant 3.0 [229], Atomic Non-Local Energy Assessment (ANOLEA) [234], Site Directed Mutator (SDM) [235], ERIS server [236], SNPeffect 4.0 [237], and Mutation-assessor [238]. The nsSNP was assigned as most deleterious if at least 6 out of 11 tools predicted it as a damaging SNP. Following this criteria six amino acid variants of E2F1, C227F, Arg252His, Val295Asp, Cys298Tyr, Arg56Trp, and Tyr59Cys were selected for further analysis (Appendices 4.4). These six mutant structures along with two native protein structures were further analyzed by molecular dynamics.

4.1.7 Molecular dynamics of mutant structures

To get insight about the structural changes, molecular dynamics simulation study was conducted. Top ranked 6 mutants (Cys227Phe, Arg252His, Val295Asp, Cys298Tyr, Arg56Trp and Tyr59Cys) along with 2 native (2AZE and 1CF7) structures were further analyzed by 50ns molecular dynamic simulation (MD) by GROMACS 5.1.2 software [239]. A significant change in the global parameters like, root mean square deviation (RMSD), root mean square fluctuation (RMSF), solvent accessible surface area (SASA) and radius of gyration were observed.

4.1.8 Impact of mutations to RMSD and RMSF

The global parameter RMSD has predicted the protein stability which was higher at different time points during simulation in mutants like, Arg252His and Tyr59Cys, while the RMSD of Val295Asp, Cys298Tyr and Arg56Trp showed relatively lesser values than their corresponding native structure (Fig. 4.6). Molecular dynamics simulation also provides important information about thermal fluctuation and structural movements. The RMSF provide information on residue level [244]. The thermal fluctuation of a native 2AZE lie between ~0.2 nm to ~1.9 nm, while in the mutant setup, Cys298Tyr showed lower fluctuation than its corresponding native structure and Cys227Phe, Arg252His and Val295Asp showed higher fluctuations than native structure. The thermal fluctuation of native 1CF7 lie between ~0.15 nm to ~0.33 nm, while its corresponding mutant Arg56Trp and Tyr59Cys showed significant fluctuations at the mutated as well as other residues positions (Fig. 4.7).

4.1.9 Impact of mutations on the solvent accessible surface area (SASA)

SASA describes the protein integrity and folding [245]. SASA of Cys298Tyr and Arg56Trp mutant was decreased during nearly entire simulation period, while SASA was increased in case of Arg252His and Tyr59Cys mutations. Furthermore, the SASA of Cys227Phe and Val295Asp was nearly similar like native E2F1 structure (Fig. 4.8).

These results indicate that Arg252His and Tyr59Cys increase unfolding in E2F1 and E2F4 proteins, while Cys298Tyr and Arg56W mutation increase folding in E2F1 and E2F4 protein.

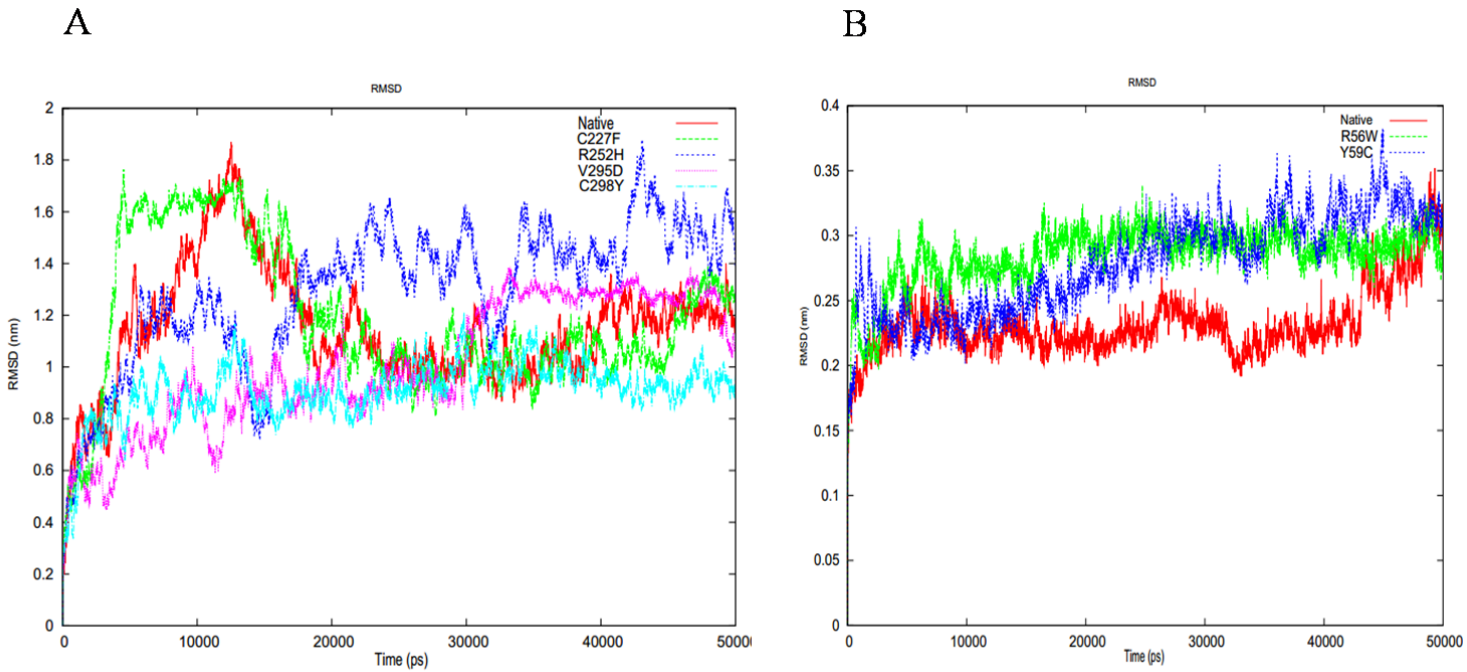


Fig 4.7 RMSD of the native and mutant structures of the E2F1 protein: The Y-axis represents RMSD (nm), and the X-axis represents time (ps). In **A,** the red, green, blue, pink, turquoise lines signify the native (2AZE) and mutant C227F, R252H, V295D and C298Y structures, respectively and in **B,** the red, green, and blue lines indicate native (1CF7) and mutant R56W and Y59C structures.

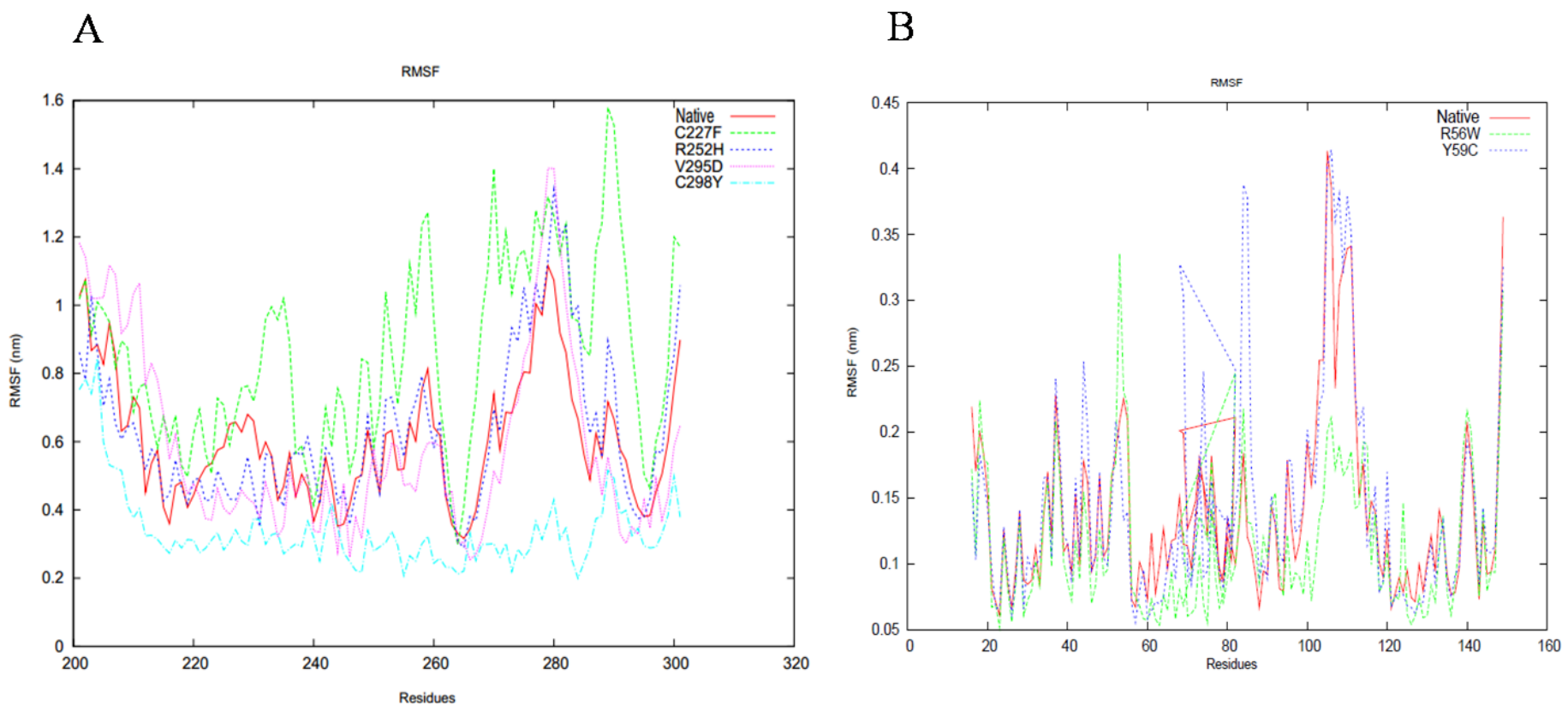


Fig 4.8 RMSF of the native and mutant structures of the E2F1 protein: The Y-axis represents RMSF (nm), and the X-axis represents residues. In **A,** the red, green, blue, pink, turquoise appearance shows the native (2AZE) and mutant (C227F, R252H, V295D and C298Y) structures, respectively. In **B,** the red, green, and blue lines indicate native (1CF7) and mutant (R56W and Y59C) structures, respectively.

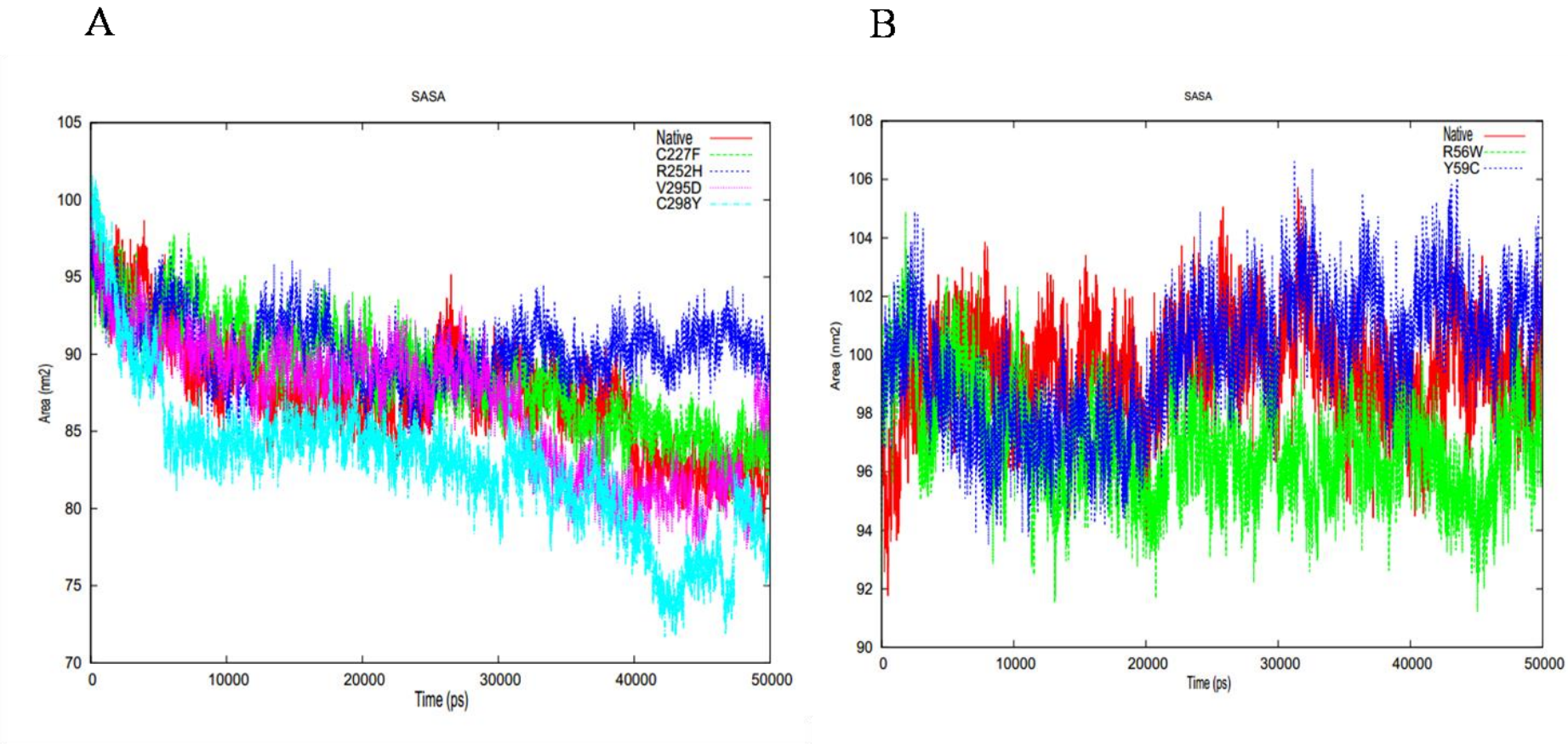


Fig 4.9 SASA of the native and mutant structures of the E2F1 protein: The Y-axis represents Area (nm^2), and the X-axis represents time (ps). In **A,** the red, green, blue, pink, turquoise lines specify the native (2AZE) and mutant (C227F, R252H, V295D and C298Y) structures. In **B,** the red, green, and blue colour indicates native (1CF7) and mutant (R56W and Y59C) structures, respectively.

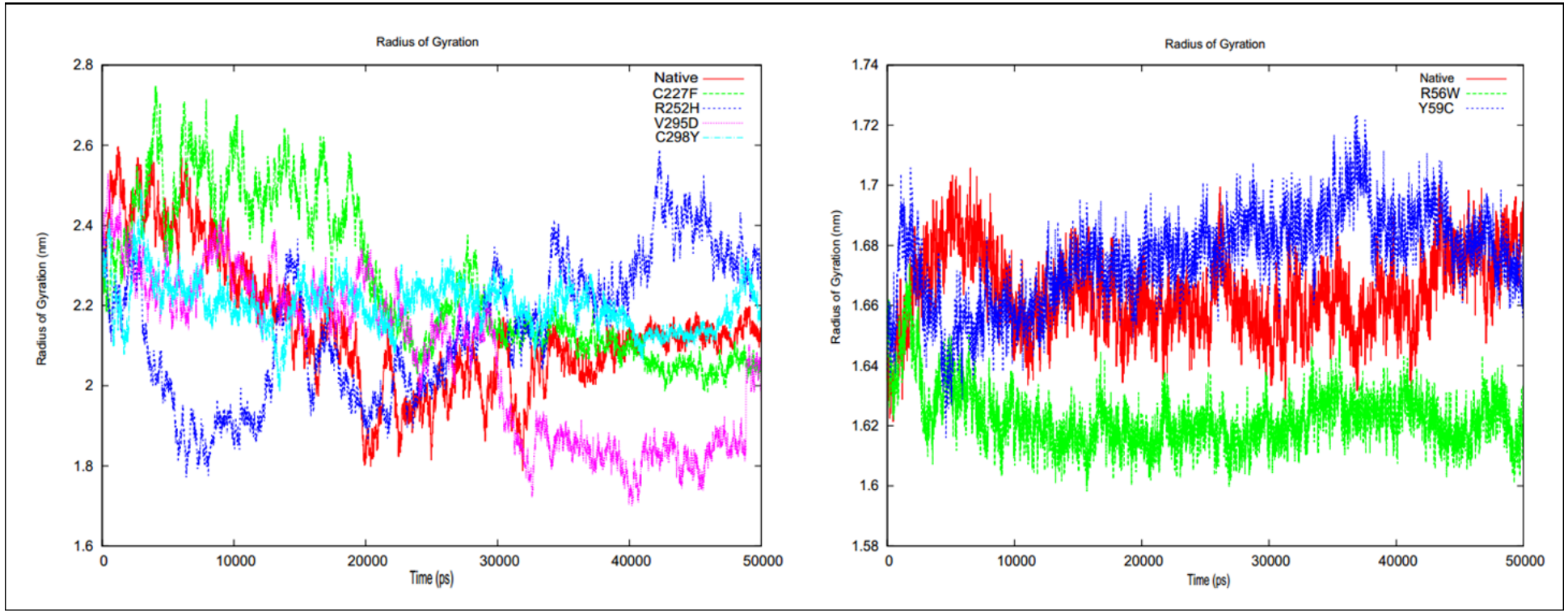


Fig 4.10 Radius of gyration of the native and mutant structures of the E2F1 protein: The Y-axis radius of gyration (nm^2) and the X-axis represents time (ps). In **A,** the red, green, blue, pink, turquoise lines specify the native (2AZE) and mutant (C227F, R252H, V295D and C298Y) structures respectively. In **B,** the red, green, and blue lines indicate native (1CF7) and mutant (R56W and Y59C) mutant structures.

4.1.10 Impact of mutations on the radius of gyration

The mass weighted root mean square distance of set of atom from their basic focal point refer as radius of gyration (Rg) [246] and reveal the compactness of a proteins. We found a significant difference in Rg of mutant structures (Arg252His and Tyr59Cys) (Fig. 4.10). The Rg of Cys227Phe and Cys298Tyr did not show a considerable change; however, Arg252His and Tyr59Cys show higher Rg than their native structures. These results indicate that high structural deviation of mutants (Arg252His and Tyr59Cys) in comparison to the native structure, while Val295Asp and Arg56Trp mutant structures showed lower Rg than their native (2AZE and 1CF7) structures.

4.1.11 Comparison of native and mutant structures before and after mutation

Both native structures 2AZE and 1CF7 show slight change in conformation when these were compared before and after MD run (Fig 4.11 C&D). Analysis of the native 2AZE and 1CF7 structures did not show significant changes in orientation. Proper orientations of protein are important for their normal functioning. This analysis shows that if there is no mutation present in the 2AZE and 1CF7 structure, MD simulation does not affect the confirmation of protein. However, mutant structure Cys227Phe, Arg252His, Val295Asp, and Cys298Tyr, Arg56Trp, and Tyr59Cys shows greater deviation in orientation from the native 1CF7 and 2AZE structure. These mutations change normal orientation of E2F1 and E2F4 protein that change might affect the proper functioning of these proteins. There were no significant structural changes observed in the native setup of 2AZE and 1CF7 in before and after MD run (Fig 4.11), while the mutant structures (C227F, Arg252His, Val295Asp, Cys298Tyr, Arg56Trp, and Tyr59Cys) showed alteration in confirmation in the rotamer (Fig 4.12).

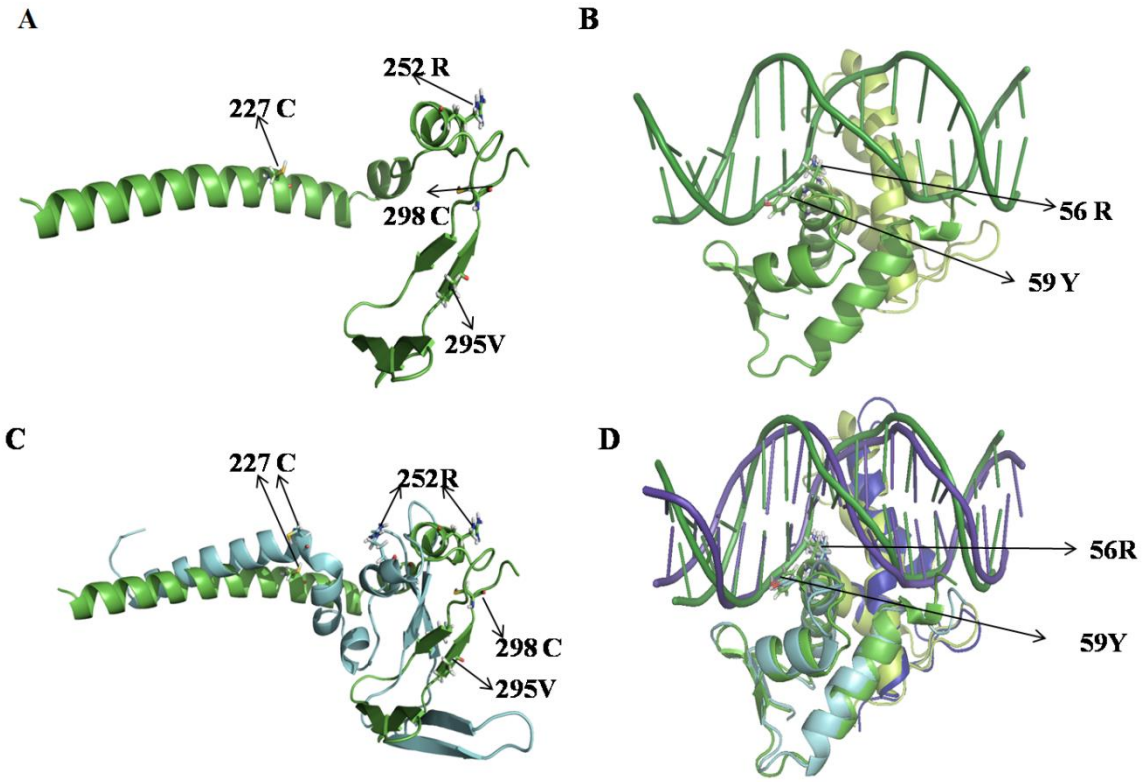


Fig 4.11 Orientation of native amino acids residue before and after MD run: A, represents MD run structure of 2AZE. **B,** represents before MD run structure of 1CF7. **C**, represents superimposed structure of native 2AZE after MD run. **D,** represents superimposed structure of native 1CF7 after MD run.

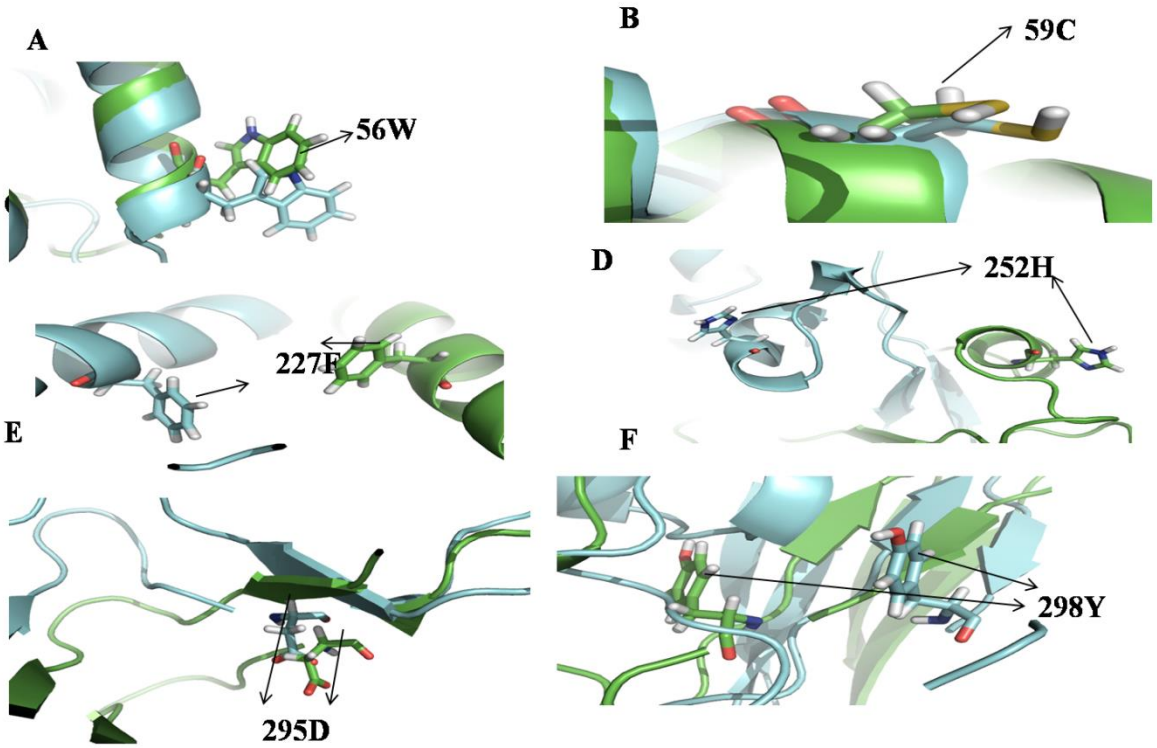


Fig 4.12 Superimposed structures of mutated residues 56W, 59C, 227F, 252H, 295D, and 298Y before and after the MD run. **A,** represents Superimposed structures of 56W before and after the MD run. **B,** represents Superimposed structures of 59C before and after the MD run. **C**, represents Superimposed structures of 227F before and after the MD run. **D,** represents Superimposed structures of 252H before and after the MD run. **E,** represents Superimposed structures of 295D before and after the MD run. **F,** represents Superimposed structures of 298Y before and after the MD run.

4.2 Case-control study to investigate SNPs rs3213172, rs3213173, rs3213176, rs57496843 and rs2071054 play role in predisposing individuals to cancer in Himachal Pradesh population

4.4.1 Demographic characteristic: Demographic characteristics of cases like smoking status, alcohol addict, age, gender, clinical stage of cancer are presented in Table 4.2. In the studied population approximately 93% of the cancer patients were smokers and around 67% of the cancer patients were alcohol addicts and more than 80% of the cancer patients were diagnosed in advance stage of cancer (III and IV). In the demographic variable clinicopathological parameters of cancer patients like, stage of cancer, size of tumor and lymph node metastasis are given in Table 4.2

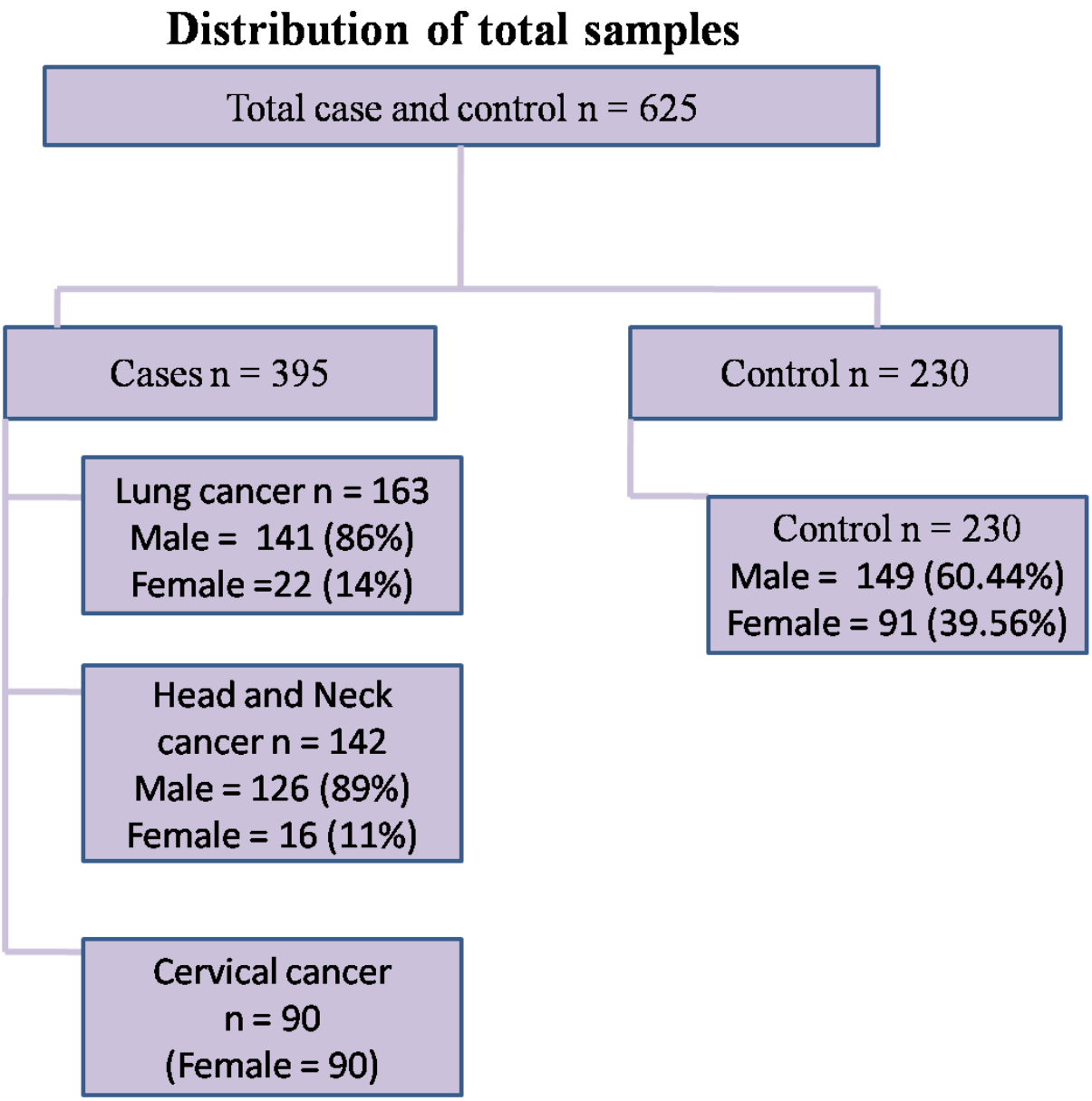


Fig. 4.13 Classification of total samples

Table 4.2: Demographic characteristic of cases and control group

Variables	**Patients (n/N[%]) (N=305)**	**Patients (n/N[%]) (N=90) (CaCx)**	**Controls (n/N[%]) (N=230)**	**P-value**
Gender				
Male	267 (87.54)		149 (60.44)	
Female	38 (12.45)	90	91 (39.56)	
Age	58±9	55±8	48±10	
Smoking Status				
Yes	284 (93.11)	8 (89)	125 (54)	**<0.01**
No	21 (6.88)	82 (91)	105 (46)	
Alcohol drinking				
Yes	205 (67.21%)	2 (2.2)	60 (31%)	**<0.01**
No	100 (32.78%)	88 (98)	131 (69%)	
NA			29	
Stage				
I+II	57 (19.33%)	36 (42.35)		
III+IV	238 (80.67%)	49 (57.64)		
NA	10	5		
Tumor T Status				
T1+T2	82 (27.80%)	NA		
T3+T4	213 (72.20%)	NA		
NA	10			
Lymph node Status				
N1+N2+N3	219 (74.23%)	NA		
N0	76 (25.77%)	NA		
NA	10			
Metastasis				
M0	273 (92.54%)	NA		
M1	22 (7.46%)	NA		
NA	10			

NA= Not available, LC = Lung Cancer, HNC = Head and Neck Cancer, CaCx = Cervical Cancer

4.2.2 Genomic DNA isolation and quantification

Genomic DNA from collected blood samples was isolated applying inorganic salting out method [52]. Genomic DNA quality and quantity was checked on 0.8 % agarose gel stained with EtBr and by spectrophotometric analysis, respectively. The concentration of isolated DNA in the samples ranged from 100-900ng/µl. The 260/280 ratio for all the DNA samples was in the range of 1.6-2.0. The representative image of the genomic DNA isolated from various samples is shown in the Fig 4.14.

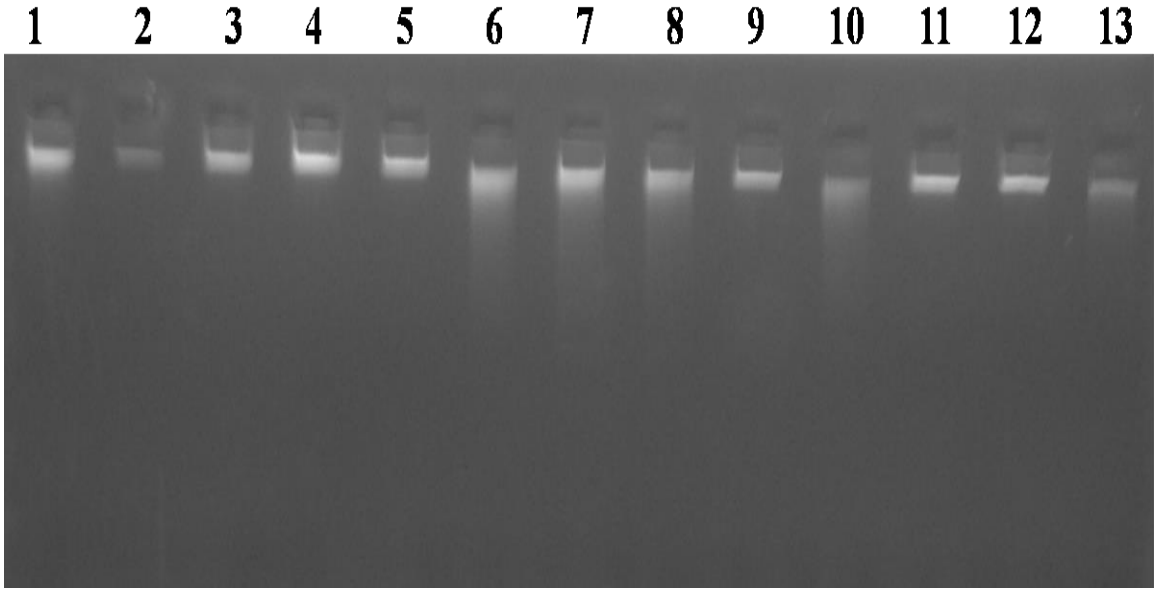


Fig 4.14 Check gel of genomic DNA: In this fig L1 represent the positive control genomic DNA and L2-L13 represent the gel check of isolated DNA sample.

4.2.3 Selection of SNPs for Genotyping:

We analyzed the frequencies of genotypes of identified deleterious nsSNPs (rs3213172 and rs574946483) by *in silico* approach, NIEHS predicted disease risk nsSNPs (rs3213173 and rs3213176) and an intronic variant rs2071054 of *E2F1* gene in cases (LC, HNC and CaCx patients) and control groups. *In silico* analysis indicated that Arg252His and Cys298Tyr amino acid variants of E2F1 were highly damaging in sequence as well as structural analysis and these two variants are germline mutations of *E2F1* gene which were further analysed in the association study.

4.2.4 Genotyping of rs3213172 nsSNP:

The rs3213172 nsSNP was genotyped by using mutagenic primers. The details of primer designing are described in section material and methods. The optimized PCR conditions and RFLP pattern of rs3213172 nsSNP are given in Table 4.3 and Fig 4.15. We observed the all three expected banding patterns; homozygous wild (CC = 151 and 51), heterozygous (CT = 202, 151 and 51) and homozygous mutant (TT = 202) in cases and controls groups.

Table 4.3: Standardized PCR condition for nsSNP rs3213172 (R252H).

rs ID	PCR components	Gradient range	Optimal condition
rs3213172 **(C/T)**	Annealing temperature	55–59 ℃	58 ℃
	Primers	-	0.16mM

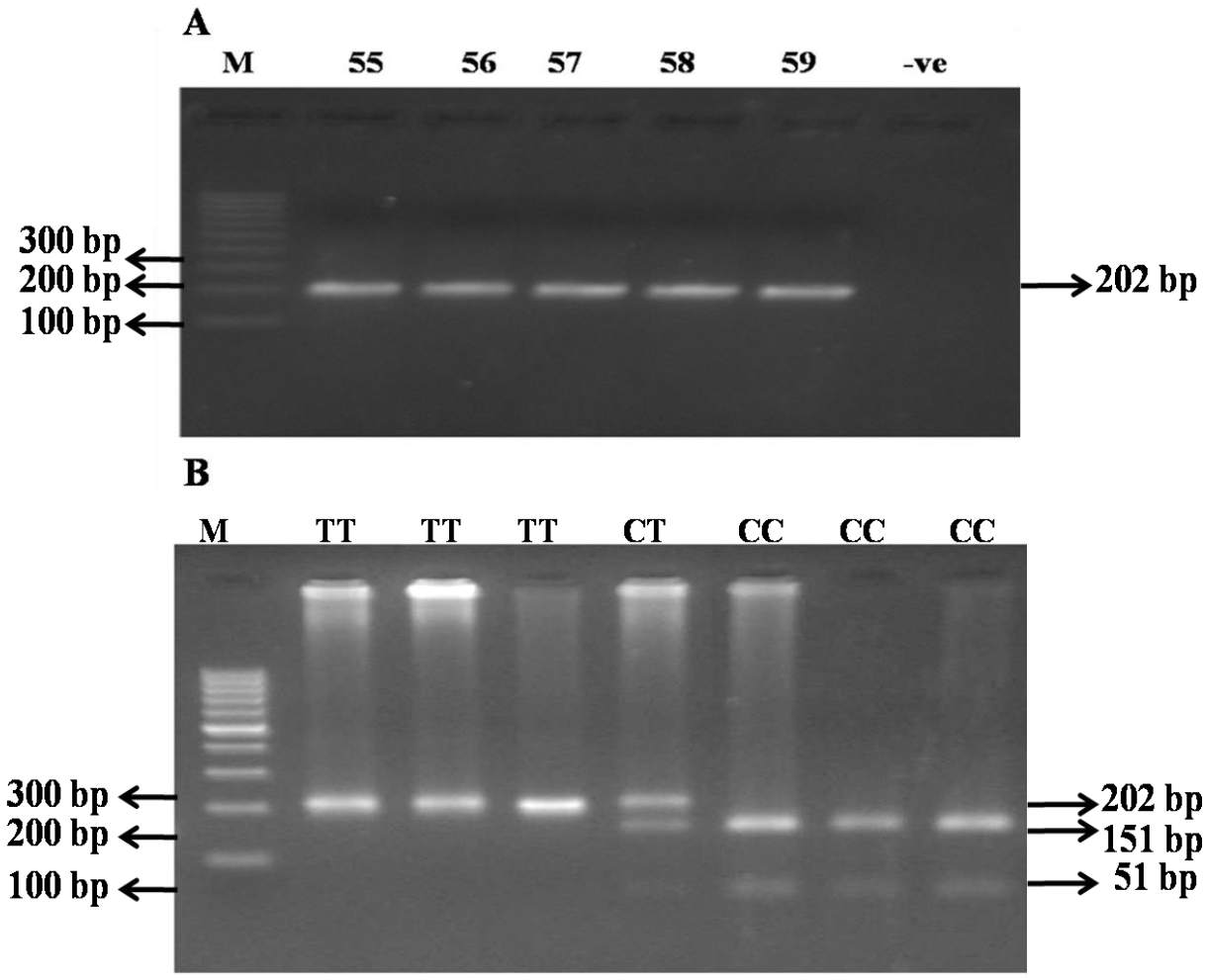


Fig 4.15: Representative gel images of optimization of PCR conditions and genotyping of rs3213172 applying optimized conditions. A, represents the gradient PCR gel image. 59˚C annealing temperature was used for amplification. **B,** represents the RFLP pattern. The 202 bp PCR fragment containing the *E2F1* rs3213172(C/T) polymorphism was digested using *Rsa*I restriction enzyme. The digested products of 202bp, 151 bp and 51 bp fragments indicated heterozygous (CT) genotype (L5). The presence of 151 bp and 51 bp indicate the homozygous wild genotype CC (L6, L7, L8 and L9) and presence of single band 202 bp indicate the homozygous mutant genotype TT (L2, L3 and L4). M =100 base pair ladder.

4.2.4 Genetic association analyses of *E2F1* genetic variants rs3213172 with cancer

4.2.4.1 Genetic association analysis of *alleles* with cancer risk

Power of the study for minor *allele* T is 87%, 81.9% and 30.5%, respectively in LC, HNC and cervical cancer patients. The frequencies of mutant *allele* (T) of rs3213172 (C/T) polymorphism was 50% in lung cancer patients, 49% in head and neck cancer and 34% in control. We observed significant association in *allelic* model for this nsSNP in LC (OR = 1.916, 95% CI = 1.434-2.561, P<0.01) and HNC patients (OR = 1.866, 95% CI = 1.380-2.524, <0.01). The frequency of T *allele* of rs3213172 (C/T) polymorphism was 30% and 40% in control and cervical cancer patients, respectively. The T *allele* was also a genetic risk factor for cervical cancer (OR = 1.575, 95% CI = 1.020-2.432, P = 0.04).

4.2.4.2 Genetic association analysis of genotypes with cancer risk

The rs3213172 (C/T) polymorphism was significantly associated with LC risk in homozygous model (OR = 4.868, 95% CI = 2.479-9.560, P<0.01), heterozygous model (OR = 1.859, 95% CI = 1.158-2.984, P<0.01) and dominant model (OR = 2.267, 95% CI = 1.158-2.984, P<0.01). There was also a significant association observed in HNC risk in all the analyzed genetic models i.e. in homozygous model (OR = 4.593, 95% CI = 2.295-9.192, <0.01), heterozygous model (OR = 1.735, 95% CI = 1.062-2.834, P = 0.02), and dominant model (OR = 2.123, 95% CI = 1.325-3.400, P <0.01) (Table 4.6). The rs3213172 (C/T) genetic variant was significantly associated with the cervical cancer risk in heterozygous (CC vs. CT; OR = 1.906, 95% CI = 1.015–3.578, P = 0.04) and dominant (CC vs. CT+TT; OR = 1.969, 95% CI = 1.076–3.601, P = 0.02) models.

In this study, we observed significant association of rs3213172 (C/T) polymorphism with the risk of lung, cervical and head and neck cancers in *allelic* as well as genotypic models (Table 4.4). In the *in silico* analysis, this SNP shows damaging effect in most of the sequence as well as structural analysis and even in molecular dynamic simulation analysis. To clarify the

influence of the rs3213172 (C/T) polymorphism with the clinicopathological status, such as TNM clinical staging, tumor size, lymph node metastasis. These parameters were further analyzed by conducting association analysis of this polymorphism with clinical parameters. No significant association was observed in clinical parameter and lung, cervical and head and neck cancers (Table 4.5a&4.5b).

Table 4.4: Genetic association analysis of the rs3213172 (C/T) polymorphism with lung and head and neck cancer risk.

Variable			**OR (95% CI)**	**P-value**
Lung Cancer				
rs3213172	Control (n/N[%]) (N= 230)	Cases (LC) (n/N [%]) (N=163)		
CC	90 (39.14)	36 (22)	Ref.	
CT	121(52.60)	90 (55.21)	**1.859 (1.158-2.984)**	**0.01**
TT	19 (8.26)	37 (22.69)	**4.868 (2.479-9.560)**	**<0.01**
CT+TT	140 (60.86)	127 (77.90)	**2.267 (1.158- 2.984)**	**<0.01**
C	301(65.44)	162 (49.70)	Ref.	
T	159 (34.56)	164 (50.30)	**1.916 (1.434-2.561)**	**<0.01**
Head and Neck Cancer				
rs3213172	Control (n/N[%]) (N=230)	Cases (HNC) (n/N [%]) (N=142)		
CC	90 (39.14)	33 (23.23)	Ref.	
CT	121(52.60)	77 (54.22)	**1.735 (1.062-2.834)**	**0.02**
TT	19 (8.26)	32 (22.53)	**4.593 (2.295-9.192)**	**<0.01**
CT+TT	140 (60.86)	109 (76.76)	**2.123 (1.325-3.400)**	**<0.01**
C	301(65.44)	143 (50.35)	Ref.	
T	159 (34.56)	141 (49.64)	**1.866 (1.380-2.524)**	**<0.01**
Cervical Cancer				
rs3213172	Controls (n/N[%]) (N=91)	Patients (n/N[%]) (N=90)		
CC	44 (48.35)	29 (32.22)	Ref.	
CT	39 (42.85)	49 (54.44)	**1.906 (1.015-3.578**	**0.04**
TT	8 (8.791)	12 (13.33)	2.275 (0.828-6.248)	0.11
CT+TT	47 (51.64)	61 (67.77)	**1.969 (1.076-3.601)**	**0.02**
C	127(69.78)	107 (59.44)	Ref.	
T	55 (30.22)	73 (40.56)	**1.575 (1.020-2.432)**	**0.04**

P-values <0.05 indicate the significance association

Table 4.5a: Genetic association analysis of the rs3213172 (C/T) polymorphism with clinical stage.

Variable			**OR (95% CI)**	**P-value**
Clinical Stage (LC)				
rs3213172	Stage I+II (n=37) (%)	Stage III+IV (n=124) (%)		
CC	6 (16.21)	30 (24.19)	Ref.	
CT	22 (59.45)	66 (53.22)	0.600 (0.220-1.631)	0.31
TT	9 (24.32)	28 (22.58)	0.622 (0.196-1.973)	0.42
CT+TT	31 (83.78)	94 (75.80)	0.606 (0.230-1.593)	0.31
C	34 (46)	126 (50.80)	Ref.	
T	40 (54)	122 (49.20)	0.823 (0.489-1.385)	0.46
Clinical Stage (HNC)				
rs3213172	Stage I+II (n=20) (%)	Stage III+IV (n=114) (%)		
CC	3 (15)	27 (23.68)	Ref.	
CT	12 (60)	61 (53.50)	0.564 (0.147-2.165)	0.40
TT	5 (25)	26 (22.80)	0.577 (0.125-2.666)	0.48
CT+TT	17 (85)	87 (76.30)	0.568 (0.154-2.088)	039
C	18 (45)	115 (50.44)	Ref.	
T	22 (55)	113 (49.56)	0.804 (0.409-1.578)	0.52
Clinical Stage (CaCx)				
rs3213172	Stage I+II (n =36) (%)	Stage III+IV (n =49) (%)		
CC	12 (33.33)	14 (28.57)	Ref.	
CT	18 (50.00)	29 (59.18)	1.381 (0.523-3.642)	0.51
TT	6 (16.67)	6 (12.24)	0.857 (0.218-3.370)	0.82
CT+TT	24 (66.67)	35 (71.42)	1.250 (0.493- 3.167)	0.63
C	42 (58.33)	57 (58.16)	Ref.	
T	30 (41.67)	41 (41.84)	1.007 (0.543-1.866)	0.98

P-values <0.05 indicate the significance association

Table 4.5b: Genetic association analysis of the rs3213172 (C/T) polymorphism with clinical variables

Variable			**OR (95% CI)**	**P-value**
Tumor Size (LC)				
rs3213172	T1+T2 (n=29) (%)	T3+T4 (n=132) (%)		
CC	3 (10.34)	32 (24.24)	Ref.	
CT	19 (65.51)	70 (53.03)	0.345 (0.095-1.251)	0.10
TT	7 (24.13)	30 (22.72)	0.401 (0.095-1.698)	0.21
CT+TT	26 (89.65)	100 (75.75)	0.360 (0.102-1.270)	0.11
C	25 (43.10)	134 (50.75)	Ref.	
T	33 (76.74)	130 (49.25)	0.735 (0.414-1.303)	0.29
Tumor Size (HNC)				
rs3213172	T1+T2 (n=52) (%)	T3+T4 (n=82) (%)		
CC	10 (19.23)	20 (24.39)	Ref.	
CT	27 (51.92)	46 (56.09)	0.851 (0.347-2.085)	0.72
TT	15 (28.84)	16 (19.51)	0.533 (0.414-1.303)	0.23
CT+TT	42 (80.76)	62 (75.60)	0.738 (0.314-1.734)	0.48
C	47 (45.19)	86 (52.43)	Ref.	
T	57 (54.80)	78 (47.57)	0.747 (0.456-1.224)	0.24
Lymph node Metastasis (LC)				
rs3213172	N0 (n=33) (%)	N1+N2+N3 (n=128) (%)		
CC	5 (15.15)	30 (23.43)	Ref	
CT	21 (63.63)	68 (53.12)	0.539 (0.185-1.566)	0.25
TT	7 (21.21)	30 (23.43)	0.714 (0.203-2.503)	0.59
CT+TT	28 (84.84)	98 (76.56)	0.583 (0.207-1.643)	0.30
C	31 (47)	128 (50)	Ref.	
T	35 (53)	128 (50)	0.885 (0.515-1.522)	0.66
Lymph node Metastasis (HNC)				
rs3213172	N0 (n=45) (%)	N1+N2+N3 (n=89) (%)		
CC	9 (20)	20 (22.47)	Ref.	
CT	27 (60)	47(52.80)	0.783 (0.312-1.961)	0.60
TT	9 (20)	22 (24.71)	1.100 (0.364-3.320)	0.86
CT+TT	36 (80)	67 (75.28)	0.837 (0.345-2.029)	0.69
C	45 (50)	87 (49)	Ref.	
T	45 (50)	91 (51)	1.046 (0.630-1.736)	0.86

P-values <0.05 indicate the significance association

4.2.5 Genotyping of nsSNP rs574956843:

This nsSNP was also genotyped by using mutagenic primer. The detail of primer designing is described in section material and methods (Chapter-3). The gradient temperature range and optimized PCR conditions shown in Table 4.6 were used to amplify DNA fragment of *E2F1* gene containing rs574956843(C/T) polymorphism. We observed the only homozygous wild (CC = 215 bp) genotype in all samples of cases and controls groups. The representative gel image was presented in Fig 4.16

Table 4.6: Standardized PCR condition for nsSNP rs574956843 (C298Y)

rs ID	PCR components	Gradient range	Optimal condition
rs574956843 **(C/T)**	Annealing temperature	57–61 ℃	57 ℃
	Primers	-	0.16mM

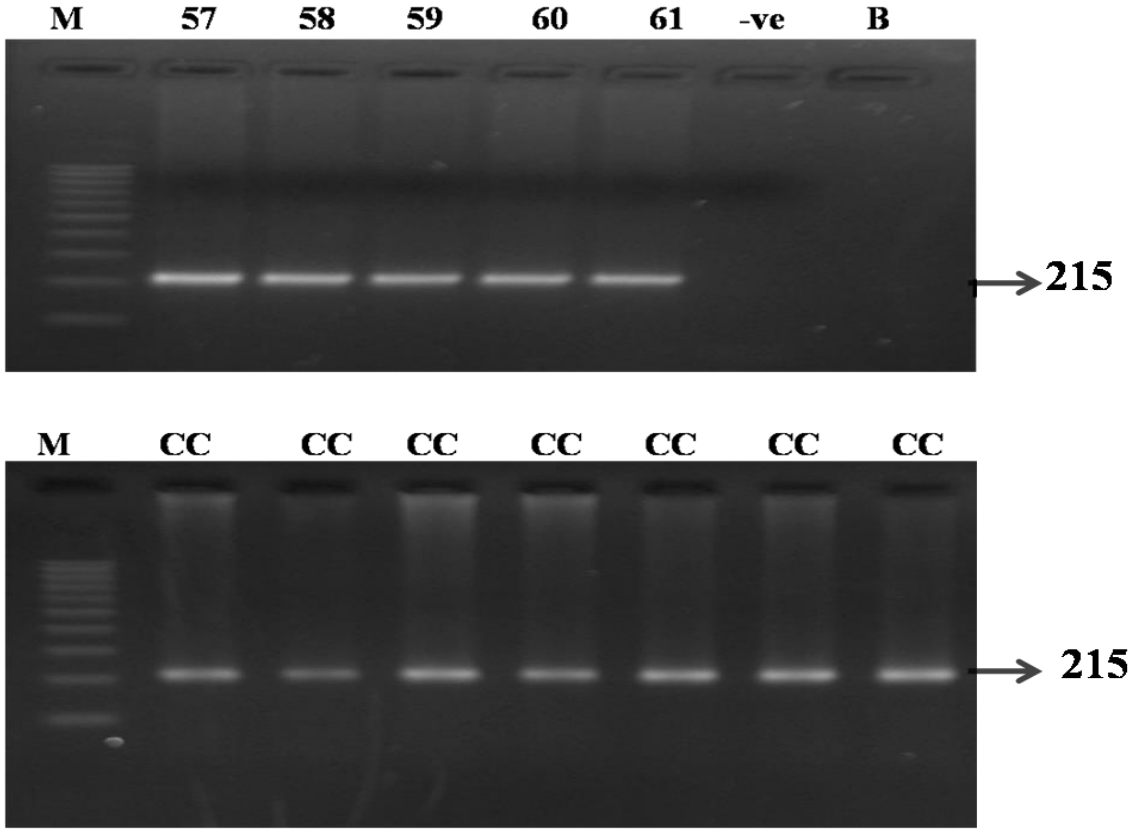


Fig 4.16: Representative gel images of optimization of PCR conditions and genotyping of rs574956843 applying optimized conditions. The **A,** shows the gradient PCR gel image, 57˚C annealing temperature used for further amplification. **B,** shows the RFLP pattern. The 215 bp PCR fragment containing the *E2F1* rs574956843 (C/T) polymorphism was digested using *Acc65*I restriction enzyme. The presence of single band (215 bp) indicates the homozygous wild genotype CC (L2-L7). M =100 base pair ladder.

4.2.6 Genetic association analyses of *E2F1* genetic variants rs574956843 with cancer

The rs574956843 (C/T) polymorphism of *E2F1* gene is highly damaging at sequence as well as structural levels in *in silico* analyses. But we can't analyze this nsSNP further in case-control study because in genotypic analysis only wild type genotype (CC) was observed in both cases and control group.

4.2.7 Genotyping of nsSNP rs3213173: This nsSNP was genotyped by using mutagenic primer. The detail of primers designing is described in section materials and methods (chapter-3). The optimized PCR conditions presented in Table 4.7 were used to amplify DNA fragment of cases and controls samples in order to genotype nsSNP rs3213173. We observed the all types of expected genotypes (Fig 4.17B).

Table 4.7: Standardized PCR condition for nsSNP rs3213173 (V276M).

rs ID	PCR components	Gradient range	Optimal condition
rs3213173 **(C/T)**	Annealing temperature	53–58 ℃	57 ℃
	Primers	-	0.16mM

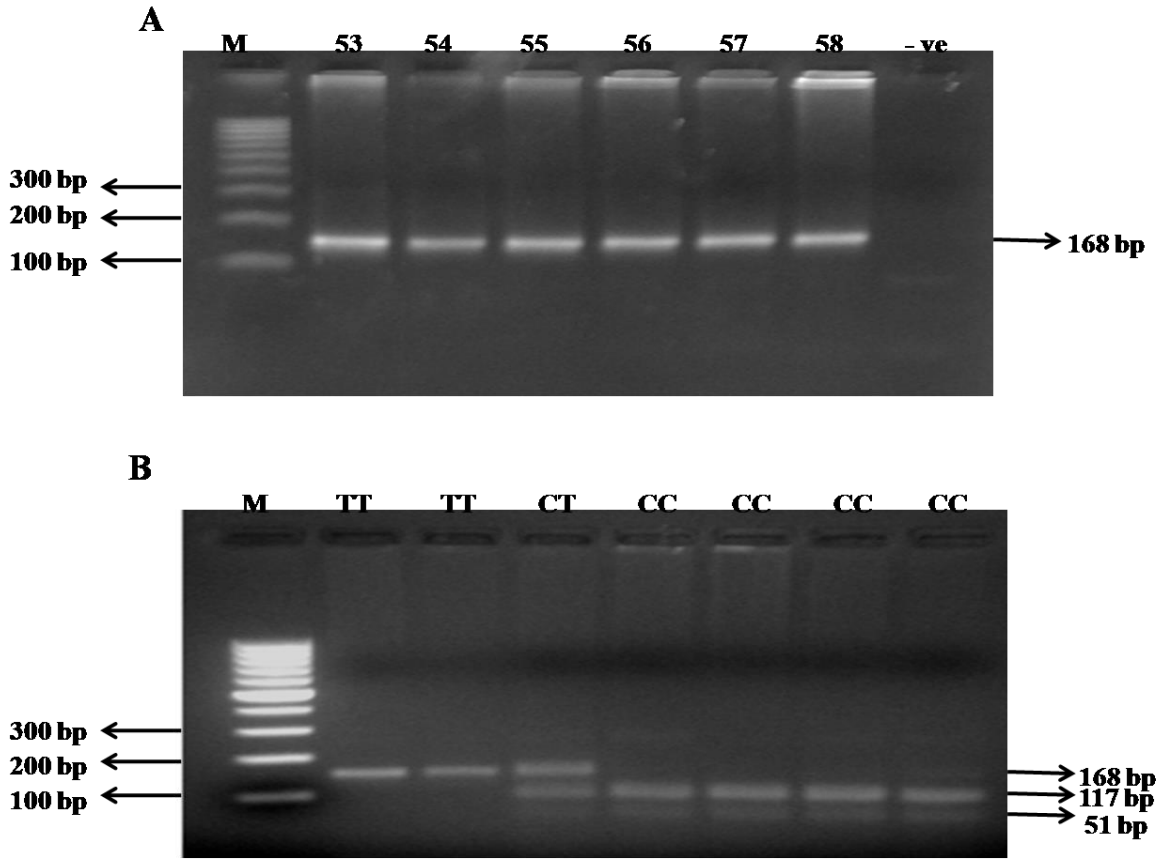


Fig 4.17 Representative gel images of optimization of PCR conditions and genotyping of rs3213173 applying optimized conditions. A, shows the gradient PCR gel image, 57˚C annealing temperature used for further amplification. **B,** shows the RFLP pattern. The 168 bp PCR fragment amplified for rs3213173 nsSNP genotyping and it was digested with *Msp*I restriction enzyme. The digested products of 168 bp, 117 bp and 51 bp fragments indicate CT heterozygous genotype (L4). The presence of 117 bp and 51 and bp indicate the homozygous genotype CC (L5, L6, L7 and L8) and presence of single band 168 bp indicate the homozygous mutant genotype TT (L2). M =100 base pair ladder.

4.2.8 Genetic association analyses of *E2F1* genetic variants rs3213173 with cancer

4.2.8.1 Genetic association analysis of alleles with cancer risk

We checked the power of study for rs3213173 (C/T) polymorphism and power for minor *allele* T was 86%, 2.6% and 34.2%, respectively, in LC, HNC and cervical cancer patients. The frequencies of the mutant *allele* (T) of rs3213173 (C/T) polymorphism were 40.79% and 26% in LC and HNC patients, respectively, and 26% in controls. The frequency of T *allele* was 38% in cervical cancer patients and 27% in controls. This genetic variant of *E2F1* gene was significantly associated with LC risk in *allelic* model (OR = 1.952, 95% CI = 1.441-

2.644; P<0.01). The rs3213173 (C/T) polymorphism was associated with cervical cancer risk in *allelic* model (OR = 1.641, 95% CI = 1.053-2.555; P = 0.02) (Table 4.8)

4.2.8.2 Genetic association analysis of genotypes with cancer risk

The statistical analysis of genetic models indicated that the rs3213173 (C/T) polymorphism was significantly associated with LC risk in homozygous model (CC vs. TT; OR = 2.166; 95% CI = 1.023-4.588; P = 0.04), heterozygous model (CC vs. CT; OR = 3.719; 95% CI = 2.377-5.818; P<0.01) and dominant model (CC vs. CT+TT; OR = 3.408; 95% CI = 2.215-5.246; P<0.01). The rs3213173 (C/T) polymorphism was not associated with HNC cancer risk in any of the analyzed genetic models (Table 4.8). This genetic variant was also associated with cervical cancer risk in homozygous model (CC vs. TT; OR = 2.710; 95% CI = 1.114-6.589; P = 0.02). However, this genetic variant was not associated with cervical cancer risk in heterozygous and dominant models (Table 4.9a&4.9b).

Table 4.8: Association analysis of the rs3213173 nsSNP with cancer risk

Variable			**OR (95% CI)**	**P-value**
Lung Cancer				
rs3213173	Control (n/N[%]) (N= 230)	Cases (LC) (n/N [%]) (N=163)		
CC	130 (56.5)	45 (27.60)	Ref.	
CT	80 (34.7)	103 (63.19)	**3.719 (2.377-5.818)**	**<0.01**
TT	20 (8.8)	15 (9.20)	**2.166 (1.023-4.588)**	**0.04**
CT+TT	100 (42.5)	118 (72.39)	**3.408 (2.215-5.246)**	**<0.01**
C	340 (73.91)	193 (59.21)	Ref.	
T	120 (26.09)	133 (40.79)	**1.952 (1.441-2.644)**	**<0.01**
Head and Neck Cancer				
rs3213173	Control (n/N[%]) (N=230)	Cases (HNC) (n/N [%]) (N=142)		
CC	130 (56.5)	81 (57)	Ref.	
CT	80 (34.7)	48 (33.80)	0.963 (0.612-1.514)	0.87
TT	20 (8.8)	13 (9.15)	1.043 (0.492-2.211)	0.91
CT+TT	100 (42.5)	61 (42.95)	0.979 (0.641-1.493)	0.92
C	340 (73.91)	210 (74)	Ref.	
T	120 (26.09)	74 (26)	0.998 (0.712-1.398)	0.99
Cervical Cancer				
rs3213173	Controls (n/N[%]) (N=91)	Patients (n/N[%]) (N=90)		
CC	50 (54.94)	41(45.55)	Ref.	
CT	32 (35.16)	29 (32.22)	1.105 (0.576-2.117)	0.76
TT	9 (9.890)	20 (22.22)	**2.710 (1.114- 6.589)**	**0.02**
CT+TT	41 (45.05)	49 (54.44)	1.457 (0.811-2.617)	0.20
C	132 (72.53)	111(61.67)	Ref.	
T	50 (27.47)	69 (38.33)	**1.641 (1.053-2.555)**	**0.02**

P-values <0.05 indicate the significance association

Table 4.9a: Genetic association analysis of the rs3213173 nsSNP with clinical variable

Variable			OR (95% CI)	P-value
Clinical Stage (LC)				
rs3213173	Stage I+II (n=37) (%)	Stage III+IV (n=124) (%)		
CC	17 (46)	42 (33.87)	Ref.	
CT	15 (40.54)	73 (58.87)	1.969 (0.892-4.346)	0.09
TT	5 (13.51)	9 (7.25)	0.718 (0.213-2.492)	0.61
CT+TT	20 (54.05)	82 (66.12)	1.659 (0.787-3.498)	0.18
C	49 (66.21)	157 (63)	Ref.	
T	25 (33.79)	91 (37)	1.136 (0.657-1.962)	0.64
Clinical Stage (HNC)				
rs3213173	Stage I+II (n=20) (%)	Stage III+IV (n=114) (%)		
CC	10 (50)	50 (43.85)	Ref.	
CT	7 (35)	47 (41.22)	1.342 (0.472-3.817)	0.58
TT	3 (15)	17 (14.91)	1.133 (0.278-4.608)	0.86
CT+TT	10 (50)	64 (56.14)	1.280 (0.494-3.314)	0.61
C	27 (67.5)	147 (64.47)	Ref.	
T	13 (32.5)	81 (35.52)	1.144 (0.559-2.339)	0.71
Clinical Stage (CaCx)				
rs3213173	Stage I+II (n =36) (%)	Stage III+IV (n =49) (%)		
CC	17 (47.22)	21 (42.85)	Ref.	
CT	11 (30.55)	16 (32.65)	1.177 (0.433- 3.197)	0.74
TT	8 (22.22)	12 (24.48)	1.214 (0.459 -3.205)	0.69
CT+TT	19 (52.77)	28 (57.14)	1.193 (0.502-2.833)	0.68
C	45 (62.5)	58 (59)	Ref.	
T	27 (37.5)	40 (41)	1.149 (0.615-2.146)	0.66
Tumor Size (LC)				
rs3213173	T1+T2 (n=29) (%)	T3+T4 (n=132) (%)		
CC	7 (24.13)	38 (28.78)	Ref.	
CT	20 (68.96)	82 (62.12)	0.755 (0.294-1.938)	0.55
TT	2 (6.89)	12 (9.09)	1.105 (0.201-6.051)	0.90
CT+TT	22 (75.86)	94 (71.21)	0.787 (0.310-1.995)	0.61
C	34 (58.62)	158 (59.84)	Ref.	
T	24 (41.38)	106 (40.15)	0.950 (0.533-1.693)	0.86
Tumor Size (HNC)				
rs3213173	T1+T2 (n=52) (%)	T3+T4 (n=82) (%)		
CC	27 (51.92)	47 (57.31)	Ref.	
CT	20 (38.46)	27 (33.33)	0.775 (0.367-1.637)	0.50
TT	5 (9.61)	8 (9.75)	0.919 (0.273-3.093)	0.89
CT+TT	25 (48.07)	35 (42.68)	0.804 (0.400-1.616)	0.54
C	74 (71.15)	121 (73.78)	Ref.	
T	30 (28.84)	43 (26.22)	0.876 (0.506-1.517)	0.63

Table 4.9b: Genetic association analysis of the rs3213173 nsSNP with clinical variable

Variable			OR (95% CI)	P-value
Lymph node Metastasis (LC)				
rs3213173	N0 (n=33) (%)	N1+N2+N3 (n=128) (%)		
CC	8 (24.25)	37 (28.90)	Ref	
CT	21 (63.63)	81 (63.28)	0.834 (0.338-2.056)	0.69
TT	4 (12.12)	10 (7.81)	0.540 (0.134-2.167)	0.38
CT+TT	25 (75.75)	91 (71.09)	0.787 (0.325-1.903)	0.59
C	37 (56)	155 (60.54)	Ref.	
T	29 (44)	101 (39.46)	0.831 (0.481-1.436)	0.50
Lymph node Metastasis (HNC)				
rs3213173	N0 (n=45) (%)	N1+N2+N3 (n=89) (%)		
CC	20 (44.44)	55 (61.79)	Ref.	
CT	20 (44.44)	26 (29.21)	0.472 (0.217-1.026)	0.05
TT	5 (11.11)	8 (8.98)	0.581 (0.170-1.988)	0.38
CT+TT	25 (55.55)	34 (38.20)	0.494 (0.239-1.023)	0.05
C	60 (66.66)	136 (76.40)	Ref.	
T	30 (33.34)	42 (23.59)	0.617 (0.353-1.079)	0.09

P-values <0.05 indicate the significance association

4.2.9 Genotyping of rs3213176 nsSNP: This nsSNP was genotyped by using PCR-RFLP. The optimized PCR conditions presented in Table 4.10 were used to amplify DNA fragment from cases and controls samples in order to genotype rs3213176 nsSNP. We observed the all types of expected genotypes (Fig 4.18B).

Table 4.10: The standardized PCR condition for nsSNP rs3213176 (G393S).

rsID	PCR components	Gradient range	Optimal condition
rs3213176 **(G/A)**	Annealing temperature	56–66 ℃	56 ℃
	Primers	-	0.16mM

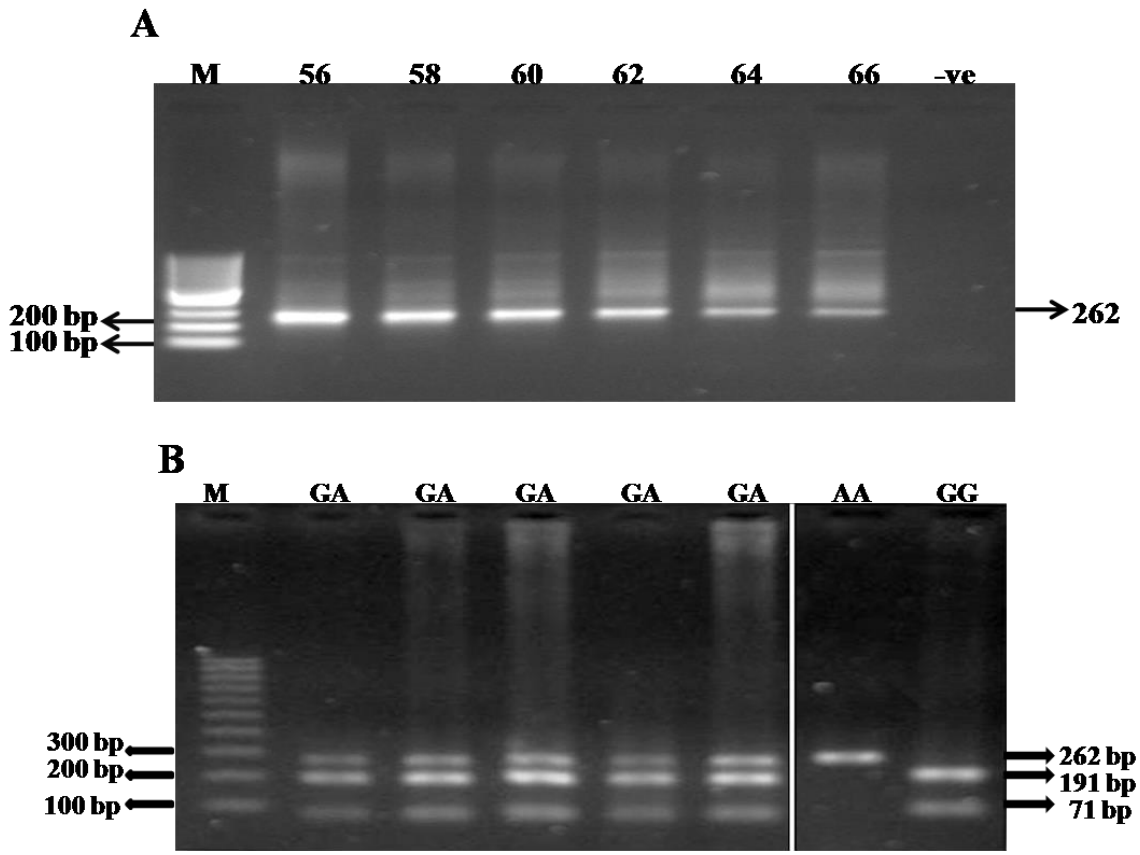


Fig 4.18 Representative gel images of optimization of PCR amplification conditions and genotyping of rs3213176 applying optimized conditions. **A,** represents the gel image of gradient PCR, 56˚C annealing temperature used for gene amplification. **B,** represents the RFLP pattern of rs3213176 nsSNP. The amplified 262 bp gene product was digested over night with *Msp*I restriction enzyme. The presence of 191 bp and 71 bp fragments indicated homozygous wild genotype GG (L8). The presence of 262 bp, 191 and 71 bp specify the heterozygous genotype GA (L2, L3, L4, L5 and L6) and presence of single band 262 bp indicated the homozygous mutant genotype AA (L7). M =100 base pair ladder.

4.2.10 Genetic association analyses of *E2F1* genetic variants rs3213176 (G/A) with cancer

4.2.10.1 Genetic association analysis of rs3213176 (G/A) alleles with cancer risk

Power of study for rs3213176 (G/A) polymorphism for minor *allele* A was 96%, 100% and 0.18%, respectively in LC, HNC and cervical cancer patients. The *allelic* frequencies of mutant *allele* (A) of rs3213176 (G/A) polymorphism was 48% in LC, 58% in HNC and 29.79% in healthy controls. The rs3213176 (G/A) polymorphism was associated with LC risk in *allelic* model (OR = 2.230; 95% CI = 1.661-2.995; P<0.01) and this polymorphism also increased the risk of HNC cancer in *allelic* model (OR = 3.316; 95% CI = 2.434-4.518; P<0.01). However, there was no significant association was observed in *allelic* model for cervical cancer risk (OR = 1.357; 95% CI = 0.864-2.130; P<0.01) (Table 4.10).

4.2.10.2 Genetic association analysis of rs3213176 (G/A) genotypes with cancer risk

The statistical analysis showed that rs3213176 (G/A) polymorphism of *E2F1* gene was associated with LC and HNC cancer risk. This nsSNP was associated with LC risk in homozygous model (GG vs. AA; OR = 4.789; 95% CI = 2.354-9.743; P<0.01), heterozygous model (GG vs. GA; OR = 3.712; 95%CI = 2.298-5.994; P<0.01) and dominant model (GG vs. GA+AA; OR = 3.885 95% CI = 2.440-6.186; P<0.01). This nsSNP increased the risk of HNC in homozygous model (GG vs. AA; OR = 12.32; 95% CI = 6.100-24.90; P<0.01), heterozygous model (GG vs. GA; OR = 3.805 95% CI = 2.201-6.578; P<0.01) and dominant model (GG vs. GA+AA; OR = 5.177; 95% CI = 3.068-8.734; P<0.01). There was no significant association observed in genotypic model for cervical cancer risk (Table 4.12a & 4.12b).

Table 4.11: Association analysis of the rs3213176 (G/A) with cancer risk

Variable			OR (95% CI)	P-value
Lung Cancer				
rs3213176	Control (n/N[%]) (N=230)	Cases (LC) (n/N [%]) (N=163)		
GG	112 (48.69)	32 (19.63)	Ref.	
GA	99 (43.00)	105 (64.41)	**3.712 (2.298-5.994)**	**<0.01**
AA	19 (8.26)	26 (15.95)	**4.789 (2.354-9.743)**	**<0.01**
GA+AA	118 (51.30)	131 (80.36)	**3.885 (2.440-6.186)**	**<0.01**
G	323 (70.21)	169 (51.84)	Ref.	
A	137 (29.79)	157 (48.15)	**2.230 (1.661-2.995)**	**<0.01**
Head and Neck Cancer				
rs3213176	Control (n/N[%]) (N=230)	Cases (HNC) (n/N [%]) (N=142)		
GG	112 (48.69)	22 (15.49)	Ref.	
GA	99 (43.00)	74 (52.11)	**3.805 (2.201-6.578)**	**<0.01**
AA	19 (8.26)	46 (32.39)	**12.32 (6.100-24.90)**	**<0.01**
GA+AA	118 (51.30)	120 (84.50)	**5.177 (3.068-8.734)**	**<0.01**
G	323 (70.21)	118 (41.54)	Ref.	
A	137 (29.79)	166 (58.45)	**3.316 (2.434-4.518)**	**<0.01**
Cervical Cancer				
rs3213176	Controls (n/N[%]) (N=91)	Cases (CaCx) (n/N[%]) (N=90)		
GG	49 (53.84)	41 (45.55)	Ref.	
GA	35 (38.46)	38 (42.22)	1.297 (0.698-2.409)	0.40
AA	7 (7.692)	11 (12.22)	1.878 (0.667- 5.284)	0.23
GA+AA	42 (46.15)	49 (54.44)	1.394 (0.777-2.502)	0.26
G	133 (73.08)	120 (66.67)	Ref.	
A	49 (26..92)	60 (33.33)	1.357 (0.864- 2.130)	0.18

P-values <0.05 indicate the significance association

4.2.10.2 Association of rs3213176 (G/A) genotypes and alleles with Clinical Parameters:

There was no association observed in genotypic as well as *allelic* models in clinical parameters such as clinical staging, tumor size and lymph node metastasis of LC, HNC and cervical cancer (Tables 4.12a & 4.12b). The mutant *allele* (A) showed protective association with advance stage of HNC (Table 4.12a)

Table 4.12a: Genetic association analysis of the rs3213176 SNP with clinical variables

Variable			**OR (95% CI)**	**P-value**
Clinical Stage (LC)				
rs3213176	Stage I+II (n=37) (%)	Stage III+IV (n=124) (%)		
GG	8 (21.62)	24 (19.35)	Ref.	
GA	24 (68.86)	79 (63.70)	1.097 (0.436-2.757)	0.84
AA	5 (13.51)	21 (16.93)	1.400 (0.396-4.943)	0.60
GA+AA	29 (78.37)	100 (80.63)	1.149 (0.467-2.828)	0.76
G	40 (54)	127 (51.20)	Ref.	
A	34 (46)	121 (49.80)	1.120 (0.666-1.886)	0.66
Clinical Stage (HNC)				
rs3213176	Stage I+II (n=20) (%)	Stage III+IV (n=114) (%)		
GG	1 (5)	19 (16.66)	Ref.	
GA	8 (40)	62 (54.38)	0.407 (0.047-3.472)	0.41
AA	11 (55)	33 (28.94)	0.157 (0.018-1.320)	0.08
GA+AA	19 (95)	95 (83.32)	0.263 (0.033-2.086)	0.20
G	10 (25)	100 (43.86)	Ref.	
A	30 (75)	128 (56.14)	0.426 (0.199-0.914)	**0.02**
Clinical Stage (CaCx)				
rs3213176	Controls (n/N[%]) (N=91)	Patients (n/N[%]) (N=90)		
GG	20 (55.55)	23 (46.93)	Ref.	
GA	13 (36.11)	18 (36.73)	1.204 (0.474-3.056)	0.69
AA	3 (8.333)	8 (16.32)	2.318 (0.540- 9.944)	0.25
GA+AA	16 (44.44)	26 (53.06)	1.413 (0.595-3.353)	0.43
G	53 (73.61)	64 (65.30)	Ref.	
A	19 (26.38)	34 (34.70)	1.007 (0.543-1.866)	0.98
Tumor Size (LC)				
rs3213176	T1+T2 (n=29) (%)	T3+T4 (n=132) (%)		
GG	6 (20.68)	26 (19.69)	Ref.	
GA	19 (65.51)	84 (63.63)	1.020 (0.368-2.823)	0.96
AA	4 (13.79)	22 (16.66)	1.269 (0.317-5.079)	0.73
GA+AA	23 (79.31)	106 (80.30)	1.063 (0.393- 2.878)	0.90
G	31 (53.44)	136 (51.51)	Ref.	
A	27 (46.56)	128 (48.48)	1.080 (0.611-1.910)	0.78

P-values <0.05 indicate the significance association

Table 4.12b: Genetic association analysis of the rs3213176 SNP with clinical variables

Variable			OR (95% CI)	P-value
Tumor Size (HNC)				
rs3213176	T1+T2 (n=52) (%)	T3+T4 (n=82) (%)		
GG	6 (11.54)	14 (17.03)	Ref.	
GA	23 (44.23)	47 (57.31)	0.875 (0.297-2.575)	0.80
AA	23 (44.23)	21 (25.60)	0.391 (0.127-1.204)	0.10
GA+AA	46 (88.46)	68 (82.91)	0.633 (0.226-1.769)	0.80
G	35 (33)	75 (45.74)	Ref.	
A	69 (67)	89 (54.26)	0.601 (0.361-1.002)	0.05
Lymph node Metastasis (LC)				
rs3213176	N0 (n=33) (%)	N1+N2+N3 (n=128) (%)		
GG	11 (33.33)	22 (17.18)	Ref.	
GA	18 (54.54)	84 (65.62)	2.333 (0.963-5.652)	0.06
AA	4 (12.12)	22 (17.18)	2.750 (0.758-9.969)	0.12
GA+AA	22 (66.66)	106 (82.81)	2.409 (1.022-5.676)	0.04
G	40 (60.60)	128 (50)	Ref.	
A	26 (39.39)	128 (50)	1.538 (0.886-2.669)	0.12
Lymph node Metastasis (HNC)				
rs3213176	N0 (n=45) (%)	N1+N2+N3 (n=89) (%)		
GG	5 (11.11)	15 (16.85)	Ref.	
GA	24 (53.33)	46 (51.68)	0.638 (0.207-1.970)	0.43
AA	16 (35.55)	28 (31.46)	0.583 (0.178-1.905)	0.37
GA+AA	40 (88.88)	74 (83.14)	0.616 (0.208-1.820)	0.38
G	34 (37.78)	76 (42.69)	Ref.	
A	56 (62.22)	102 (53.30)	0.814 (0.484-1.369)	0.43

P-values <0.05 indicate the significance association

4.2.11 Genotyping of rs2071054 (C/T)

This intronic variant was genotyped by using mutagenic primer. The detail of primer designing is presented in section Materials and Methods (chapter-3). The optimized PCR conditions presented in Table 4.13 were used to amplify DNA fragment from cases and controls samples in order to genotype rs2071054 SNP. We observed the all three banding pattern homozygous wild (CC = 276 and 53), heterozygous (CT = 329, 276 and 53) and homozygous mutant (TT = 329) in cases and controls groups. The representative gel image is presented in Fig 4.19.

Table 4.13: Standardized PCR condition for rs2071054 intronic variant

rs ID	PCR components	Gradient range	Optimal condition
rs2071054 **(C/T)**	Annealing temperature	52–57 ℃	52 ℃
	Primers	-	0.16mM

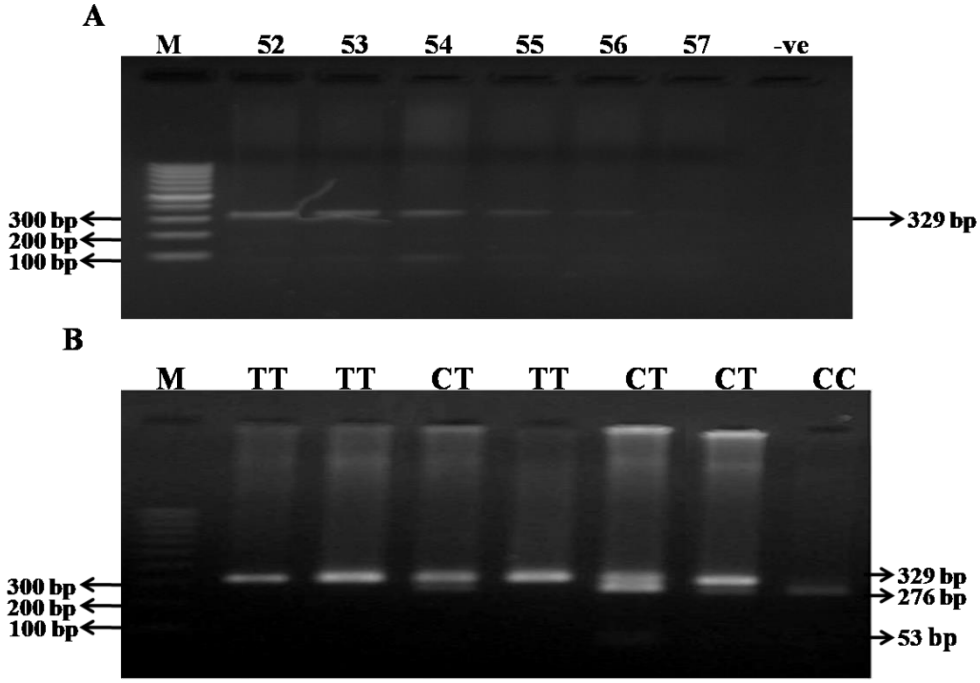


Fig 4.19 Representative gel images of optimization of PCR amplification conditions and genotyping of rs2071054 applying optimized conditions. A, shows the gradient PCR gel image, 57˚C annealing temperature used for further amplification. **B,** shows the RFLP pattern. The amplified PCR-product length is 329. PCR fragment containing the *E2F1* rs2071054 (C/T) polymorphism was digested over night by *Msp*I restriction enzyme. The digested fragment of 276 bp and 53 bp indicated (CC) homozygous wild genotype (L8). The presence of 329 bp, 276 bp and 51 bp fragment indicated the heterozygous genotype CT (L4, L6 and L7). Presence of single 329 bp band indicated the homozygous mutant allele TT (L2 and L3) M =100 base pair ladder.

4.2.12 Genetic association analysis of rs2071054 (C/T) polymorphism with cancer risk

4.2.12.1 Genetic association analysis of alleles with cancer risk

Power of the study for rs2071054 (C/T) polymorphism for minor *allele* T was 17.9%, 8.6% and 20.7%, respectively in LC, HNC and cervical cancer patients. The allele frequency of mutant *allele* T of rs2071054 (C/T) polymorphism was 65% in control group, 68.66% HNC patients (OR = 1.392, 95% CI = 1.020-1.900, P = 0.03). The distribution of mutant *allele* (T) was also different in LC patients (70%) and control group (65%). The differences in *allelic* distribution observed in cervical cancer patients and control group and this difference was also statistically significant (Table 4.14).

4.2.12.2 Genetic association analysis of rs2071054 (C/T) genotypes with cancer risk

The rs2071054 (C/T) polymorphism was not associated with risk of cancer (LC, HNC and CaCx). No significant association was observed in analyzed genotypic models i.e. homozygous (OR = 2.211, 95% CI 0.663-7.371, P = 0.99), heterozygous (OR = 1.584, 95% CI = 0.482-5.204, P = 0.44) and dominant (OR = 1.806, 95% CI = 0.556-5.864, P = 0.32) models. The logistic regression analysis showed that rs2071054 (C/T) polymorphism was not associated with HNC risk in any of the genetic models i.e. homozygous (OR = 2.106 95% CI = 0.569-7.786, P = 0.26), heterozygous (OR = 2.393, 95% CI = 0.629-9.095, P = 0.20) and dominant (OR = 2.106, 95% CI = 0.569-7.786, P = 0.26). This SNP was not associated with the cervical cancer risk in homozygous model (OR = 2.222, 95% CI = 0.347-14.19, P = 0.39) heterozygous model (OR = 1.180, 95% CI = 0.189-7.348, P = 0.85) and in dominant model (OR = 1.500, 95% CI = 0.244-9.197, P = 0.66) (Table 4.14).

4.2.12.3 Association of genotypes and *allele* with tumor stage

We didn't observe any association of rs2071054 (C/T) polymorphism with clinical parameters. This polymorphisms was not associated with clinical stage of cancer lung cancer in any genetic model i.e. homozygous model (OR = 0.733, 95% CI = 0.060-8.832, P = 0.80), heterozygous model (OR = 1.433, 95% CI = 0.121-16.96, P = 0.77) and dominant model (OR = 1.083, 95% CI = 0.094-12.41, P = 0.94). The rs2071054 (C/T) polymorphism was not

associated with clinical stage of HNC although frequencies of mutant genotype was higher in stages third and fourth of the cancer in patients (Table 4.15a). However, we observed a significant association in cervical cancer patients when comparing cancer stages with *allele* (OR = 2.072, 95% CI = 1.057-4.062, P = 0.03) (Table 4.15a).

Table 4.14: Genetic association analysis of *E2F1* intronic variant rs2071054 in cases and control group

Variable			**OR (95% CI)**	**P-value**
(Lung Cancer)				
rs2071054	Control (n/N[%]) (N=230)	Cases (LC) (n/N [%]) (N=163)		
CC	10 (4.34)	4 (2.45)	Ref.	
CT	142 (61.73)	90 (55.21)	1.584 (0.482-5.204)	0.44
TT	78 (33.91)	69 (42.33)	2.211 (0.663-7.371)	0.99
CT+TT	220 (95.65)	159 (97.54)	1.806 (0.556-5.864)	0.32
C	162 (35)	98 (30)	Ref.	
T	298 (65)	228 (70)	1.264 (0.932-1.714)	0.13
(Head and Neck Cancer)				
rs2071054	Control (n/N[%]) (N= 230)	Cases (HNC) (n/N [%]) (N=142)		
CC	10 (4.34)	3 (2.11)	Ref.	
CT	142 (61.73)	83 (58.45)	1.948 (0.521-7.281)	0.32
TT	78 (33.91)	56 (39.43)	2.393 (0.629-9.095)	0.20
CT+TT	220 (95.65)	139 (97.88)	2.106 (0.569-7.786)	0.26
C	162 (35)	89 (31.33)	Ref	
T	298 (65)	195 (68.66)	**1.392 (1.020-1.900)**	**0.03**
(Cervical Cancer)				
rs2071054	Control (n/N[%]) (N= 91)	(Cervical Cancer) (n/N [%]) (N= 90)		
CC	3 (3.29)	2 (2.22)	Ref.	
CT	61 (67.03)	48 (53.33)	1.180 (0.189-7.348)	0.85
TT	27 (29.67)	40 (44.44)	2.222 (0.347-14.19)	0.39
CT+TT	88 (96.70)	88 (97.77)	1.500 (0.244-9.197)	0.66
C	67 (36.81)	52 (28.88)	Ref.	
T	115 (63.19)	128 (71.12)	1.434(0.922-2.229)	0.10

P-values <0.05 indicate the significance association

Table 4.15a: Genetic association analysis of the rs2071054 SNP of *E2F1* gene with stage of the cancer

Variable			**OR (95% CI)**	**P-value**
Clinical Stage of (LC)				
rs2071054	Stage I+II (n=37) (%)	Stage III+IV (n=124) (%)		
CC	2 (5.40)	2 (1.61)	Ref.	
CT	18 (48.64)	70 (56.45)	3.888 (0.512-29.53)	0.18
TT	17 (45.94)	52 (41.93)	3.058 (0.399-23.40)	0.28
CT+TT	35 (94.58)	122 (98.38)	3.485 (0.473-25.64)	0.22
C	22 (29.73)	74 (30)	Ref.	
T	52 (70.27)	174 (70)	0.994 (0.563-1.755)	0.98
Clinical Stage of HNC				
rs2071054	Stage I+II (n=20) (%)	Stage III+IV (n=114) (%)		
CC	0	3 (2.63)	Ref.	
CT	15 (75)	64 (56.14)	0.594 (0.029-12.11)	0.73
TT	5 (25)	47 (41.22)	1.233 (0.056-27.16)	0.89
CT+TT	20 (100)	111 (97.36)	0.777 (0.038-15.61)	0.86
C	15 (37.5)	70 (30.70)	Ref.	
T	25 (62.5)	158 (69.30)	1.354 (0.673-2.725)	0.39
Clinical Stage of Cervical Cancer				
rs2071054	Stage I+II (n=36) (%)	Stage III+IV (n=49) (%)		
CC	1 (2.77)	1 (2.04)	Ref.	
CT	25 (69.44)	20 (40.81)	0.800 (0.047-13.60)	0.87
TT	10 (27.77)	28 (57.14)	2.800 (0.159-49.10)	0.48
CT+TT	35 (97.21)	48 (97.95)	1.371 (0.082-22.68)	0.82
C	27 (37.5)	22 (22.44)	Ref.	
T	45 (62.5)	76 (77.56)	**2.072 (1.057-4.062)**	**0.03**
Tumor Size of LC				
rs2071054	T1+T2 (n=29) (%)	T3+T4 (n=132) (%)		
CC	1 (3.44)	3 (2.27)	Ref.	
CT	16 (55.18)	72 (54.54)	1.500 (0.146-15.37)	0.73
TT	12 (41.38)	57 (43.18)	1.583 (0.151-16.55)	0.70
CT+TT	28 (96.56)	129 (97.72)	1.535 (0.154-15.31)	0.71
C	18 (31.03)	78 (29.54)	Ref.	
T	40 (68.96)	186 (70.45)	1.073 (0.579-1.986)	0.82
Tumor Size of HNC				
rs2071054	T1+T2 (n=52) (%)	T3+T4 (n=82) (%)		
CC	0	3 (3.65)	Ref.	

CT	33 (63.47)	46 (56.09)	0.198 (0.009-3.968)	0.28
TT	19 (36.53)	33 (40.24)	1.073 (0.579-1.986)	0.82
CT+TT	52 (100)	79 (96.34)	0.216 (0.010-4.274)	0.31
C	33 (71.73)	52 (31.70)	Ref.	
T	71 (31.73)	112 (68.30)	1.001 (0.590-1.697)	0.99

P-values <0.05 indicate the significance association

Table 4.15b: Genetic association analysis of the rs2071054 (C/T) of *E2F1* gene with lymph node metastasis

Lymph node Metastasis (LC)				
rs3213173	N0 (n=33) (%)	N1+N2+N3 (n=128) (%)		
CC	2 (6.06)	2 (1.56)	Ref.	
CT	16 (48.48)	72 (56.25)	4.500 (0.589-34.38)	0.14
TT	15 (45.45)	54 (42.19)	3.600 (0.467- 27.73)	0.21
CT+TT	31 (93.93)	126 (98.44)	4.064 (0.550-30.00)	0.16
C	20 (30.30)	76 (29.69)	Ref.	
T	46 (69.70)	180 (70.31)	1.029 (0.571-1.856)	0.92
Lymph node Metastasis (HNC)				
rs3213173	N0 (n=45) (%)	N1+N2+N3 (n=89) (%)		
CC	1 (2.22)	2 (2.24)	Ref.	
CT	27 (60)	53 (59.55)	0.981 (0.085-11.31)	0.98
TT	17 (37.78)	34 (38.20)	1.000 (0.084-11.82)	1.000
CT+TT	44 (97.78)	87 (97.75)	0.988 (0.087-11.20)	0.99
C	29 (32.22)	57 (32.03)	Ref.	
T	61(67.78)	121 (67.97)	1.009 (0.586-1.736)	0.97

P-values <0.05 indicate the significance association

4.2.13 Haplotype analysis for lung cancer

We used SHEsis software for haplotypes analysis. In the logistic regression analysis we considered most frequent haplotype CCG as a reference haplotype. Haplotypes analysis revealed that TTA (OR = 9.194, 95% CI = 4.798-17.61), CCA (OR = 6.810, 95% CI = 4.106-11.29), TCA (OR = 5.837, 95% CI =3.286-10.36), TTG (OR = 4.852, 95% CI = 2.753-8.553), TCG (OR = 3.729, 95% CI = 2.221-6.262), and CTG (OR = 7.415 95% CI = 4.162-13.20) were significantly associated with lung cancer risk (Table 4.16). Linkage Disequilibrium plots are given in appendices 4.5

4.2.14 Haplotype analysis for head and neck cancer

The haplotypes TTA (OR = 5.355, 95% CI = 2.869-9.994), CCA (OR = 2.868, 95% CI = 1.779-4.625), TCA (OR = 5.580, 95% CI = 3.364-9.255) are associated with head and neck cancer risk. The overall results of this analysis indicates that TTA, CCA, TCA, TTG, TCG and CTG haplotypes were associated with lung cancer risk, while TTA, CCA and TCA haplotypes were significantly associated with head and neck cancer risk (Table 4.17).

4.2.15 Haplotype analysis for cervical cancer

The frequencies of CTG, TCG and TTA were significantly higher in patients group than control (CTG: OR = 2.898, 95% CI = 1.469-5.717, P = 0.002, TCG: OR = 2.893, 95% CI = 1.469-5.717, P = 0.002; TTA: OR = 3.178 95% CI = 1.386-7.285; P = 0.006). The results indicated that CTG, TCG and TTA haplotypes may be risk factors for cervical cancer (Table 4.18).

Table 4.16: Haplotype analysis of rs3213172, rs3213173 and rs3213176 nsSNP in cases lung cancer and controls group

Haplotypes	**Cases (LC) (n/N[%])**	**Controls (n/N[%])**	**OR (95% CI)**	**P-value**
CCG	37 (11.49)	189 (41.44)	Ref.	
CTA	10 (3.10)	29 (6.35)	1.761 (0.791-3.922)	0.16
TTA	36 (11.18)	20 (4.38)	**9.194 (4.798-17.61)**	**<0.01**
CCA	68 (21.11)	51 (11.18)	**6.810 (4.106-11.29)**	**<0.01**
TCA	40 (12.42)	35 (7.67)	**5.837 (3.286-10.36)**	**<0.01**
TTG	40 (12.42)	38 (8.33)	**4.852 (2.753-8.553)**	**<0.01**
TCG	46 (14.28)	63 (13.81)	**3.729 (2.221-6.262)**	**<0.01**
CTG	45 (13.97)	31 (6.39)	**7.415 (4.162-13.20)**	**<0.01**

P-values <0.05 indicate the significance association

Table 4.17: Haplotype analysis of rs3213172, rs3213173 and rs3213176 nsSNP in cases (head and neck cancer) and controls group

Haplotype	Cases (HNC) (n/N[%])	Controls (n/N[%])	OR (95% CI)	P-value
CCG	60 (21.42)	189 (41.44)	Ref.	
CTA	12 (4.28)	29 (6.35)	1.303 (0.626-2.712)	0.47
TTA	34 (7.45)	20 (4.38)	**5.355 (2.869-9.994)**	**<0.01**
CCA	56 (20)	51 (11.18)	**2.868 (1.779-4.625)**	**<0.01**
TCA	62 (22.14)	35 (7.67)	**5.580 (3.364-9.255)**	**<0.01**
TTG	13 (4.64)	38 (8.33)	1.077 (0.538-2.156)	0.83
TCG	30 (10.71)	63 (13.81)	1.500 (0.889-2.530)	0.12
CTG	13 (4.64)	31 (6.39)	1.321 (0.649-2.686)	0.44

P-values <0.05 indicate the significance association

Table 4.18: Haplotype analysis of selected nsSNPs (rs3213172, rs3213173 and rs3213176) in cases (cervical cancer) and controls group

Haplotype	Cases (CaCx) (n/N[%])	Control (n/N[%])	OR (95% CI)	P-value
CCG	56 (31.28)	89 (49.44)	Ref.	
CTG	21(11.73)	11 (6.111)	**2.893 (1.469-5.717)**	**0.002**
CCA	14 (7.777)	13 (7.222)	1.711 (0.749-3.908)	0.20
CTA	16 (8.938)	13 (7.222)	1.956 (0.874-4.373)	0.10
TCG	31(17.31)	17 (9.444)	**2.898 (1.469-5.717)**	**0.002**
TCA	9 (5.027)	12 (6.666)	1.192 (0.471-3.011)	0.37
TTG	12 (6.703)	15 (8.333)	1.271 (0.554-2.914)	0.57
TTA	20 (11.17)	10 (5.555)	**3.178 (1.386-7.285)**	**0.006**

P-values <0.05 indicate the significance association.

DISCUSSION

The SNPs are one of the key hereditary variations and more than 10 million SNPs have been reported in the human genome [247]. The nsSNPs are the most frequent identifiable group of SNPs within coding region of the gene which are responsible for substitutions of amino acids in a protein. In a majority of the cases amino acid substitutions associated with thermodynamic stability, folding and aggregation of proteins. These amino acid substitutions are associated with human diseases in many cases, therefore, identification of deleterious nsSNPs is important [248, 249]. Deleterious nsSNPs could be identified using computational tools which are probably involved in disease susceptibility [250]. Various studies have documented that most of the cancers arise due to the mutations in cell-cycle-related genes because cancerous cell proliferation is controlled by this machinery [251]. It was reported that a number of common SNPs of *E2F1* are notably correlated with cancer risk [51, 252-254]. The transcription factor E2F1, is a well characterize cell-cycle regulatory gene among E2Fs family which are involved in cell-cycle regulation, cell growth and differentiation [9].

It also activates DNA repair pathway during DNA damage response, which is stabilized by ATM/ATR and Checkpoint kinase (Chk2/1) mediated phosphorylation [255]. It also induces apoptosis if damaged DNA is accumulated in the cell [255]. Marked box domain and its nearby region of E2F1 are responsible for its distinctive ability to strongly induce apoptosis [256]. According to a recent study, expression level of *E2F1* gene is a key determinant of the cell fate [257]. Higher level of *E2F1* induce apoptosis by increasing the expression of apoptotic genes (e.g. *APAF1*, *PUMA*, *HRK*, and *BIM*) [257]. Above discussed studies revealed the importance of *E2F1* gene in cellular homeostasis and SNPs also plays key role in disease susceptibility. However, there is no integrated computational and genetic association study undertaken regarding this gene. Therefore in this study, we prioritized SNPs for case-control study on the basis of computational as well as NIHES prediction.

5.1.1 Identification of deleterious nsSNPs and amino acids conservation profile

We used a systematic computational approach to identify deleterious nsSNPs in *E2F1* gene. There are 143 nsSNPs reported in different databases. These nsSNPs were analyzed at the sequence level and around 30-50% nsSNP were damaging in most of the computational tools (Fig 4.3). We checked the conservation profile of E2F1 protein by using ConSurf and multiple sequence alignment tools. The amino acid conservation profile of E2F1 is nearly

similar in both computational tools (Fig 4.5 and Fig 4.6). Align-GVGD [222, 223] provides a class probability which depends on amino acids conservation and its chemical nature. It provides seven distinct grade C0 to C65 which specify the mutations that are least likely to be neutral (class 65) to those that are the most likely to be neutral (class 0). There are 70 (52%) nsSNPs occurred in a conserved region of the E2F1 protein sequence having a C-score at least (C65) (Appendices 4.2).

5.1.2 Structural and functional analysis: The structural and functional analysis provides better insight about the functional consequences of nsSNP [49]. Therefore in this study, we prioritize nsSNPs based on the scores predicted by different computational methods (Appendices 4.3). By this approach, we identified six nsSNPs which were most destructive in the structural and functional analysis. These amino acid changes were further analyzed by conducting MD simulations analysis. The E2F1 functional protein has four domains DNA binding (128-192), coiled-coil (177-217) region, marked-box domain (252-367) and transactivation domain (368-437) [39, 40]. DNA binding domain is involved in the regulation of transcriptional activity of genes essential for G1 to S transition. This domain also increases the expression of apoptotic genes, such as P73 and APAF1 [258, 259]. The transactivation domain is involved in the regulation of DNA synthesis, cell-cycle progression, cellular proliferation and apoptosis [40, 260]. Marked box and its adjacent region on the E2F1 protein interacts with Jab1 and this region of E2F1 is associated with apoptosis induction [42].

Four missense mutations (Cys227Phe, Arg252His Val295Asp and Cys298Tyr) of 2AZE (E2F1 available structure) which present at coiled-coiled region and marked box domain and two missense mutations (Arg56Trp and Tyr59Cys) of DNA binding domain of 1CF7 (E2F1 available structure) were further analyzed by the analysis of protein global parameters (RMSD, RMSF, radius gyration, SASA). The MD simulation analysis result revealed that all mutant E2F1 (Cys227Phe, Arg252His, Val295Asp and Cys298Tyr) protein showed structural deviation in protein global parameters; from its corresponding native structures. According to the HOPE analysis mutated residues (Cys227Phe, Arg252His, Val295Asp and Cys298Tyr) located at the domain which is important for contact with other domain. These mutated residues might affect the contact with other domain [261]. The mutant structures (Arg56Trp and Tyr59Cys) of DNA binding domain also showed a significant structural deviation from its corresponding native structures. These two mutant structures of DNA binding domain (Arg56Trp and Tyr59Cys) were involved in hydrogen binding and these residues are also essential for interaction of E2F1 transcription factor to the DNA. These residues (RRIYD) are

critical for DNA binding and highly conserved among E2Fs family of transcription factor [262]. According to the HOPE analysis these amino acids are located within a stretch of residues annotated in UniProt as a special motif: DEF box and this stretch also interact with BIRC2/c-IAP1 protein [261]. The mutated residue is located in a domain that is important for the E2F1 activity and in contact with another domain [261]. Interaction between these domains could be disturbed by the mutation, which might affect the function of E2F1 protein. There are no significant structural changes observed in the native setup of 2AZE and 1CF7 in before and after MD run which were shown in (Fig 4.11) while, the mutant structures (C227F, Arg252His, Val295Asp, Cys298Tyr, Arg56Trp, and Tyr59Cys) shows alteration in confirmations in the rotamer (Fig 4.12).

Four mutations Cys227Phe, Val295Asp, Arg56Trp, and Tyr59Cys are somatic mutation of *E2F1* gene and these mutations are associated in different type of cancers and are reported in COSMIC database also [260].

These analyses indicated that these mutations might affect the E2F1 mediated cellular homeostasis. These changes might make individuals more susceptible to cancer. To check the genotypes frequencies of these nsSNPs, we also conducted a case-control association study in, Himachal Pradesh (a North Indian state) population, which accurately determines the impact of a genetic polymorphism with cancer risk. In this association study, we specifically focused on rs3213172 (C/T) and 574956843 (C/T) nsSNPs which were prioritized after computational approach and other three genetic variant rs3213173, rs3213176 (G/A) and rs2071054 (C/T) were also included for genetic association study. Detail of SNP prioritization presented in method section (chapter-2).

5.1.3 Genetic association analysis

In the current study, we checked the genotypes frequencies of all these genetic variants (rs574956843, rs3213172, rs3213173, rs3213176, and rs2071054). In the genetic analysis of rs574956843 (C/T) polymorphism only CC genotype was presented in both cases and control group, the results indicate that this SNP is might be monomorphic in Indian population. The genotypes frequencies data for different ethnic group are not reported in the 1000 genomes. Hence, this was excluded for further analysis. We observed all expected genotypes for other four SNPs. The frequency of mutant genotype TT was higher in cases than control group for rs3213172 (C/T) polymorphism (Table 4.4). The MD simulation analysis revealed that Arg252His substitution put maximum instability in E2F1 protein. The RMSD of

Arg252His is highest in than that of other mutants like Cys227Phe, Val295Asp, Cys298Tyr flexibility change in the form of RMSF is higher than that of other two mutant structure (Val295Asp, Cys298Tyr and native). The solvent accessible surface area of Arg252His is also maximum than that of all other mutant and native structures, the compactness of protein was also highest in Arg252His mutant structure. As mention earlier, the frequency of mutant *allele* T is significantly higher in cases (LC, HNC and CaCx) than control. This nsSNP increased risk of lung cancer to 4.8 fold in homozygous model 1.8 fold heterozygous model, in homozygous, (2.2 fold) in dominant model and (1.9 fold) in *allelic* model (Table 4.4). The rs3213172 (C/T) polymorphism increased risk of head and neck cancer to (4.5 fold) in homozygous, (1.7 fold) in heterozygous, and (2.1 fold) in dominant and (1.8 fold) in *allelic* model (Table 4.4).

This SNP also increased the risk of cervical cancer to (1.90 fold) in heterozygous, (1.96 fold) in dominant model and (1.57 fold) in *allelic* model (Table 4.4). This *allele* (T) is responsible for the conversion of Arginine to Histidine at the position of 252 in E2F1 protein. This conversion is highly damaging according to the *in silico* analysis. All these findings revealed that this *allele* might be genetic risk factor for cancer. The rs3213172 (C/T) polymorphism was not associated with clinical parameters (TNM clinical staging, lymph node metastasis and tumor size). The rs574956843 (C/T) polymorphism of *E2F1* gene also shows damaging effect on E2F1 protein but we did not observed change frequency distribution of *allele* in cases and control group. We observed only (C) *allele* in both cases and control group in Indian population. Some SNPs are polymorphic in some ethnic groups but some of them are monomorphic in other populations i.e. the rs6413419 SNP located on CYP2E1 which is monomorphic in Indian population. The minor *allele* frequency of rs80142782 is 0.05 in East Asians population but is monomorphic in both European and African populations [263, 264].

In this study, we found statistically significant differences between the cases and control groups for rs3213173 (C/T) and rs3213176 (G/A) genetic variant. The *alleles* and genotypes frequencies of the two nsSNPs exhibited statistically significant differences between the cases and the control groups (Table 4.8 and Table 4.10). The rs3213173 (C/T) polymorphism has also increased the risk of lung cancer to (2.1 fold) in homozygous model, (3.7 fold) in heterozygous (3.4 fold) in dominant model and (1.95 fold) in *allelic* model (Table 4.8). This nsSNP also increased the risk of cervical cancer to 2.7 fold in homozygous model and 1.6 fold in *allelic* model. The rs3213173 (C/T) polymorphism was not associated with head and

neck cancer risk (Table 4.8). The genetic variant rs3213176 (G/A) increased lung cancer risk to 4.7 fold, 3.7 fold 3.8 fold and 2.2 fold in homozygous, heterozygous, dominant and *allelic* models, respectively. This nsSNP increased the head and neck cancer risk more prominently than lung cancer. It increased risk of head and neck cancer risk by 12.3 fold, 3.8 fold, 5.1 fold and 3.3 fold in homozygous, heterozygous, dominant and *allelic* models, respectively. There is no significant association observed regarding cervical cancer risk for this nsSNP (Table 4.11).

A recent study on Chinese population has also shown the association of (rs35301225 (C/A) genetic variant of *E2F1* gene with colorectal cancer [264]. Another study also demonstrated a significant association of *E2F1* gene polymorphism rs3213180 (C/G) with HPV16 seropositivity [265]. In the functional validation of promoter region SNP (rs3213180 (C/G) of *E2F1* gene was associated with increased expression of this gene [266].

The rs3213172 (C/T), rs3213173 (C/T) and rs3213176 (G/A) of *E2F1* gene that are located in marked box domain and its adjacent region of E2F1 protein. This domain has been shown to interact with Jun activation-domain binding protein1 (Jab1) (cofactor factor) and is responsible for apoptosis induction [42]. These genetic variants of *E2F1* gene might alter the interaction of E2F1 protein with Jab1 leading to disturbed apoptosis and individuals carrying mutant genotypes might be more susceptible to various cancers. These nsSNPs of *E2F1* gene were also identified as disease risk SNPs by NIEHS SNPs program. We did not observe the association of these nsSNPs with clinical parameter (Table 4.5, 4.9 and 4.12). In this study, we also checked the association of *E2F1* intronic variant rs2071054 (C/T) with cancer risk. This genetic variant increase the risk of head and neck cancer to 1.3 fold in *allelic* model. The rs2071054 (C/T) polymorphism was also associated with advance stage of cervical cancer risk in *allelic* model (Table 4.15). According to the 1000 genomes the MAF of rs2071054 is higher in the different ethnic groups and even in Asian and Indian population. The MAF range of rs2071054 (C/T) is 0.26-1 in different ethnic groups [267]. It is 0.76 and 0.66, respectively, for Telugu and Gujarati Indian populations. In current study, we observed MAF for this genetic variant is 0.65 in Himachal Pradesh (Indian) population [39]. Thus, the existence of mutant *allele* in the different population is possible and this SNP could be used as a valuable marker after conducting a case-control study regarding this variant in various ethnic groups. Prior to this study, one genetic association study about rs2071054 (C/T) genetic variant showed significant association with a decreased risk of the nonsyndromic cleft lip (NSCL/P) in Polish population [267].

In the current study, we checked the association of *E2F1* nsSNPs with cervical cancer risk. The HPV viral protein E6 and E7 are one of main risk factors of cervical cancer [268, 269]. The HPV viral protein E7 binds with E2F1 protein with high affinity the E7-E2F1 interaction favour the transformation of infected cell to cancerous cell. The amino acids ranges (1-368, 191-368 and 284-437) of E2F1 interact with E7 viral protein. The analyzed nsSNPs of this gene are present in the high affinity binding region of E2F1 [270]. Genetic variants at this region of protein might increase the binding affinity of HPV-E7 protein and favour cervical cell carcinogenesis.

5.1.4 E2F1 haplotypes and cancer risk

Haplotype analysis is more informative than single SNP association analysis [271]. Therefore, in current study, we also analyzed the association of haplotypes of rs3213172, rs3213173 and rs3213176 nsSNPs with LC and HNC and cervical cancer risk. The haplotype analysis indicates that TTA, CCA, TCA, TTG, TCG and CTG haplotypes were associated with lung cancer risk (Table 4.16). The prevalence of TTA, CCA and TCA haplotypes were higher in HNC patients than controls group. These haplotypes are significantly associated with HNC risk (Table 4.17).

The TCG, CTG and TTA haplotypes of *E2F1* genetic variant present significantly higher in number in cervical cancer patients than female control group (Table 4.18).

The rs3213176 (G/A) polymorphism was not associated in genotypic and *allelic* models in cervical cancer. However, this SNP might be playing a role in cervical cancer susceptibility; TTA haplotype is associated significantly with risk of cervical cancer while haplotype TTG is not (Table 4.18). Mostowska et al (2014) found a protective association of rs3213180 and rs2071054 haplotypes of *E2F1* genetic variant in Polish population with Nonsyndromic Cleft Lip with or without Cleft Palate [267]. To the best of our knowledge, this is the first study which reports the association of E2F1 nonsynonymous genetic variants with cancer (LC, HNC and CaCx) risk [253, 254]. In this study, one of the intronic variants rs2071054 (C/T) was also investigated by conducting association study. The genetic association study revealed that rs2071054 (C/T) polymorphism was associated with HNC risk in *allelic* model but not associated with lung and cervical cancer risk. Although the frequency of mutant *allele* (T) is higher in cases lung and cervical cancer (Table 4.14) but this frequency distribution among cases and control groups are not significant. One of the earlier studies also analyzed the association of this genetic variant with nonsyndromic cleft lip with or without cleft palate [267].

Combined genetic association studies of SNP of promoter region of *E2F1* and *E2F2* genes showed association with the risk of HNC in American population [51]. Similarly, another genetic association study on *E2F1* gene on rs35301225 (C/A) genetic variants was associated with colorectal cancer [264]. The rs35301225 (C/A) polymorphism might increase the cell proliferation by increasing the percentage of S phase. In addition, this SNP was also associated with tumor differentiation, as well as metastasis in colorectal cancer patients [264]. Regulatory SNPs also play an important role in the expression pattern of the gene such as SNP of promoter region C to T substitution at position -897 of *E2F1* gene mostly linked with the reduction in expression of *E2F1* gene in HeLa cell line [52].

The rs3213180 SNP of *E2F1* gene is associated with OSCC susceptibility. The Ins/Del or Ins/Ins genotype of *E2F1* (rs3213180) and HPV seropositivity may jointly increase risk of OSCC in the pronounced among never-smokers or never-drinkers [265]. This study also revealed a significant association of *E2F1* genetic variants (rs3213172 (C/T) and rs3213173 (C/T) with cervical cancer risk [253]. In this study, individuals carrying heterozygous (CT) or mutant (TT) genotype of rs3213173 (C/T) polymorphism are associated with elevated risk of lung and cervical cancer and individual carrying heterozygous (GA) or mutant (AA) genotype of rs3213176 (G/A) polymorphism are susceptible to lung as well as HNC [254]. According to the results of association study, the rs3213172 (C/T), rs3213173 (C/T) and rs3213176 (G/A) polymorphisms could be used as a lung, cervical and HNC susceptibility marker but further study on more number of samples is needed. The rs2071054 (C/T) polymorphism was also associated with HNC and cervical cancer susceptibility. Thus, the present study provides valuable information regarding E2F1 genetic variants and cancer (LC, HNC and CaCx) risk.

CONCLUSION

In this study, we are basically focused on SNP prioritization for case-control association study and give a systematic approach for SNP selection. This approach can be relevant for any gene, to prioritization of SNPs and it might be helpful to save time and cost of SNPs validation through experimental analysis. In this study, we used this approach for *E2F1* gene which is critical for maintaining cellular homeostasis. The genome variations have been reported, in many studies, as an important marker to predict the individuals risk for cancer and could be used to understand the underlying mechanism of carcinogenesis.

We selected five SNPs (rs3213172, rs3213173, rs3213176, rs574956843, and rs2071054) of *E2F1* gene for case-control study. Additionally, we have developed a new artificial-restriction fragment length polymorphism based method for genotyping of rs3213172, rs3213173, rs574956843 and rs2071054 SNPs which would be helpful to study these SNP in other populations as well. This study is the first one to report the association of rs3213172 (C/T) polymorphism with LC, HNC and cervical cancer risk. The current study, also revealed a significant association between the rs3213173 (C/T) polymorphism with LC and CaCx risk and rs3213176 (G/A) polymorphism with (LC and HNC) cancer risk. The rs2071054 (C/T) polymorphism of *E2F1* gene was associated with HNC and advance stage of cervical cancer. Overall study, suggested that these polymorphisms of *E2F1* gene might serve as a valuable prognostic biomarker for genetic susceptibility to LC, HNC and CaCx these needs to be explored further in replication studies. Identified *E2F1* genetic variants could be used as prognostic marker after conducting follow-up studies for cancer treatment. Future studies regarding *E2F1* genetic variants in larger samples size in different ethnic groups could provide better insight about *E2F1*gene polymorphisms and cancer risk.

REFERENCES

[1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, *et al.*, "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012," *Int J Cancer,* vol. 136, pp. E359-86, Mar 1 2015.

[2] "Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015," *Lancet,* vol. 388, pp. 1659-1724, Oct 8 2016.

[3] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," *CA Cancer J Clin,* vol. 65, pp. 87-108, Mar 2015.

[4] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J Clin,* vol. 68, pp. 394-424, Nov 2018.

[5] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA Cancer J Clin,* vol. 61, pp. 69-90, Mar-Apr 2011.

[6] "The International HapMap Project," *Nature,* vol. 426, pp. 789-96, Dec 18 2003.

[7] M. D. Garrett, "Cell cycle regulation and cancer " *CURRENT SCIENCE,* vol. VOL. 81, NO. 5,, 10 SEPTEMBER 2001 2001.

[8] C. Whibley, P. D. Pharoah, and M. Hollstein, "p53 polymorphisms: cancer implications," *Nat Rev Cancer,* vol. 9, pp. 95-107, Feb 2009.

[9] N. Dyson, "The regulation of E2F by pRB-family proteins," *Genes Dev,* vol. 12, pp. 2245-62, Aug 01 1998.

[10] K. Helin, "Regulation of cell proliferation by the E2F transcription factors," *Curr Opin Genet Dev,* vol. 8, pp. 28-35, Feb 1998.

[11] J. DeGregori, "The genetics of the E2F family of transcription factors: shared functions and unique roles," *Biochim Biophys Acta,* vol. 1602, pp. 131-50, Jun 21 2002.

[12] M. V. Frolov and N. J. Dyson, "Molecular mechanisms of E2F-dependent activation and pRB-mediated repression," *J Cell Sci,* vol. 117, pp. 2173-81, May 1 2004.

[13] S. Polager, Y. Kalma, E. Berkovich, and D. Ginsberg, "E2Fs up-regulate expression of genes involved in DNA replication, DNA repair and mitosis," *Oncogene,* vol. 21, pp. 437-46, Jan 17 2002.

[14] N. B. La Thangue, "DP and E2F proteins: components of a heterodimeric transcription factor implicated in cell cycle control," *Curr Opin Cell Biol,* vol. 6, pp. 443-50, Jun 1994.

[15] L. Wu, C. Timmers, B. Maiti, H. I. Saavedra, L. Sang, G. T. Chong, *et al.*, "The E2F1-3 transcription factors are essential for cellular proliferation," *Nature,* vol. 414, pp. 457-62, Nov 22 2001.

[16] D. Ginsberg, "E2F1 pathways to apoptosis," *FEBS Lett,* vol. 529, pp. 122-5, Oct 2 2002.

[17] G. P. Dimri, K. Itahana, M. Acosta, and J. Campisi, "Regulation of a senescence checkpoint response by the E2F1 transcription factor and p14(ARF) tumor suppressor," *Mol Cell Biol,* vol. 20, pp. 273-85, Jan 2000.

[18] C. Stevens and N. B. La Thangue, "The emerging role of E2F-1 in the DNA damage response and checkpoint control," *DNA Repair (Amst),* vol. 3, pp. 1071-9, Aug-Sep 2004.

[19] T. F. Kowalik, J. DeGregori, J. K. Schwarz, and J. R. Nevins, "E2F1 overexpression in quiescent fibroblasts leads to induction of cellular DNA synthesis and apoptosis," *J Virol,* vol. 69, pp. 2491-500, Apr 1995.

[20] S. Ishida, E. Huang, H. Zuzan, R. Spang, G. Leone, M. West, *et al.*, "Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis," *Mol Cell Biol,* vol. 21, pp. 4684-99, Jul 2001.

[21] H. Muller, A. P. Bracken, R. Vernell, M. C. Moroni, F. Christians, E. Grassilli, *et al.*, "E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis," *Genes Dev,* vol. 15, pp. 267-85, Feb 01 2001.

[22] B. Ren, H. Cam, Y. Takahashi, T. Volkert, J. Terragni, R. A. Young, *et al.*, "E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints," *Genes Dev,* vol. 16, pp. 245-56, Jan 15 2002.

[23] A. S. Weinmann, P. S. Yan, M. J. Oberley, T. H. Huang, and P. J. Farnham, "Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis," *Genes Dev,* vol. 16, pp. 235-44, Jan 15 2002.

[24] T. L. Sladek, "E2F transcription factor action, regulation and possible role in human cancer," *Cell Prolif,* vol. 30, pp. 97-105, Mar-Apr 1997.

[25] F. S. Donaires, P. R. Godoy, G. S. Leandro, D. Puthier, and E. T. Sakamoto-Hojo, "E2F transcription factors associated with up-regulated genes in glioblastoma," *Cancer Biomark,* vol. 18, pp. 199-208, 2017.

[26] E. Ozono, H. Komori, R. Iwanaga, T. Tanaka, T. Sakae, H. Kitamura, *et al.*, "Tumor suppressor TAp73 gene specifically responds to deregulated E2F activity in human normal fibroblasts," *Genes Cells*, vol. 17, pp. 660-72, Aug 2012.

[27] M. Y. Lee, C. S. Moreno, and H. I. Saavedra, "E2F activators signal and maintain centrosome amplification in breast cancer cells," *Mol Cell Biol*, vol. 34, pp. 2581-99, Jul 2014.

[28] S. J. Field, F. Y. Tsai, F. Kuo, A. M. Zubiaga, W. G. Kaelin, Jr., D. M. Livingston, *et al.*, "E2F-1 functions in mice to promote apoptosis and suppress proliferation," *Cell*, vol. 85, pp. 549-61, May 17 1996.

[29] L. Yamasaki, T. Jacks, R. Bronson, E. Goillot, E. Harlow, and N. J. Dyson, "Tumor induction and tissue atrophy in mice lacking E2F-1," *Cell*, vol. 85, pp. 537-48, May 17 1996.

[30] S. Polager and D. Ginsberg, "p53 and E2f: partners in life and death," *Nat Rev Cancer*, vol. 9, pp. 738-48, Oct 2009.

[31] J. Floter, I. Kaymak, and A. Schulze, "Regulation of Metabolic Activity by p53," *Metabolites*, vol. 7, May 20 2017.

[32] Z. E. Stine, Z. E. Walton, B. J. Altman, A. L. Hsieh, and C. V. Dang, "MYC, Metabolism, and Cancer," *Cancer Discov*, vol. 5, pp. 1024-39, Oct 2015.

[33] N. N. Pavlova and C. B. Thompson, "The Emerging Hallmarks of Cancer Metabolism," *Cell Metab*, vol. 23, pp. 27-47, Jan 12 2016.

[34] S.-M. Fendt, "Is There a Therapeutic Window for Metabolism-Based Cancer Therapies?," *Frontiers in Endocrinology*, vol. 8, 2017-July-04 2017.

[35] S. Bates, A. C. Phillips, P. A. Clark, F. Stott, G. Peters, R. L. Ludwig, *et al.*, "p14ARF links the tumour suppressors RB and p53," *Nature*, vol. 395, pp. 124-5, Sep 10 1998.

[36] J. L. Russell, J. T. Powers, R. J. Rounbehler, P. M. Rogers, C. J. Conti, and D. G. Johnson, "ARF differentially modulates apoptosis induced by E2F1 and Myc," *Mol Cell Biol*, vol. 22, pp. 1360-8, Mar 2002.

[37] T. Hershko and D. Ginsberg, "Up-regulation of Bcl-2 homology 3 (BH3)-only proteins by E2F1 mediates apoptosis," *J Biol Chem*, vol. 279, pp. 8627-34, Mar 5 2004.

[38] P. J. Real, C. Sanz, O. Gutierrez, C. Pipaon, A. M. Zubiaga, and J. L. Fernandez-Luna, "Transcriptional activation of the proapoptotic bik gene by E2F proteins in cancer cells," *FEBS Lett*, vol. 580, pp. 5905-9, Oct 30 2006.

[39] J. Magae, C. L. Wu, S. Illenye, E. Harlow, and N. H. Heintz, "Nuclear localization of DP and E2F transcription factors by heterodimeric partners and retinoblastoma protein family members," *J Cell Sci,* vol. 109 (Pt 7), pp. 1717-26, Jul 1996.

[40] A. Marchler-Bauer, Y. Bo, L. Han, J. He, C. J. Lanczycki, S. Lu, *et al.*, "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures," *Nucleic Acids Res,* vol. 45, pp. D200-d203, Jan 4 2017.

[41] V. Gupta, B. C. Sarin, H. Changotra, and P. K. Sehajpal, "Association of G-308A TNF-alpha polymorphism with bronchial asthma in a North Indian population," *J Asthma,* vol. 42, pp. 839-41, Dec 2005.

[42] T. C. Hallstrom and J. R. Nevins, "Jab1 is a specificity factor for E2F1-induced apoptosis," *Genes Dev,* vol. 20, pp. 613-23, Mar 1 2006.

[43] M. W. Nachman, "Single nucleotide polymorphisms and recombination rate in humans," *Trends Genet,* vol. 17, pp. 481-5, Sep 2001.

[44] H. J. Hsu, Y. H. Yang, T. Y. Shieh, C. H. Chen, Y. H. Kao, C. F. Yang, *et al.*, "TGF-beta1 and IL-10 single nucleotide polymorphisms as risk factors for oral cancer in Taiwanese," *Kaohsiung J Med Sci,* vol. 31, pp. 123-9, Mar 2015.

[45] F. Li, J. Wang, and M. Chen, "Single nucleotide polymorphisms in DNA repair genes and the risk of laryngeal cancer: A meta-analysis," *Biomed Pharmacother,* vol. 78, pp. 92-100, Mar 2016.

[46] F. Ma, H. Zhang, Y. Zhai, W. Huang, C. Zhao, S. Ou, *et al.*, "Functional polymorphism -31C/G in the promoter of BIRC5 gene and risk of nasopharyngeal carcinoma among chinese," *PLoS One,* vol. 6, p. e16748, Feb 3 2011.

[47] N. G. Bediaga, X. Marichalar-Mendia, N. Rey-Barja, A. Setien-Olarra, J. A. Gonzalez-Garcia, M. M. de Pancorbo, *et al.*, "Polymorphisms in alcohol and tobacco metabolism genes in head and neck cancer in the Basque Country," *J Oral Pathol Med,* vol. 44, pp. 769-75, Nov 2015.

[48] E. Lai, "Application of SNP technologies in medicine: lessons learned and future challenges," *Genome Res,* vol. 11, pp. 927-9, Jun 2001.

[49] D. C. George, C. Chakraborty, S. A. Haneef, N. Nagasundaram, L. Chen, and H. Zhu, "Evolution- and structure-based computational strategy reveals the impact of deleterious missense mutations on MODY 2 (maturity-onset diabetes of the young, type 2)," *Theranostics,* vol. 4, pp. 366-85, 2014.

[50] C. M. Yates and M. J. Sternberg, "The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions," *J Mol Biol,* vol. 425, pp. 3949-63, Nov 1 2013.

[51] M. Lu, Z. Liu, H. Yu, L. E. Wang, G. Li, E. M. Sturgis, *et al.*, "Combined effects of E2F1 and E2F2 polymorphisms on risk and early onset of squamous cell carcinoma of the head and neck," *Mol Carcinog,* vol. 51 Suppl 1, pp. E132-41, Oct 2012.

[52] H. Belanger, P. Beaulieu, C. Moreau, D. Labuda, T. J. Hudson, and D. Sinnett, "Functional promoter SNPs in cell cycle checkpoint genes," *Hum Mol Genet,* vol. 14, pp. 2641-8, Sep 15 2005.

[53] T. A. Lehman, B. G. Haffty, C. J. Carbone, L. R. Bishop, A. A. Gumbs, S. Krishnan, *et al.*, "Elevated frequency and functional activity of a specific germ-line p53 intron mutation in familial breast cancer," *Cancer Res,* vol. 60, pp. 1062-9, Feb 15 2000.

[54] P. Flicek, I. Ahmed, M. R. Amode, D. Barrell, K. Beal, S. Brent, *et al.*, "Ensembl 2013," *Nucleic Acids Res,* vol. 41, pp. D48-55, Jan 2013.

[55] S. T. Sherry, M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, *et al.*, "dbSNP: the NCBI database of genetic variation," *Nucleic Acids Res,* vol. 29, pp. 308-11, Jan 01 2001.

[56] A. Bairoch and R. Apweiler, "The SWISS-PROT protein sequence data bank and its new supplement TREMBL," *Nucleic Acids Res,* vol. 24, pp. 21-5, Jan 01 1996.

[57] S. A. Forbes, D. Beare, P. Gunasekaran, K. Leung, N. Bindal, H. Boutselakis, *et al.*, "COSMIC: exploring the world's knowledge of somatic mutations in human cancer," *Nucleic Acids Res,* vol. 43, pp. D805-11, Jan 2015.

[58] P. Anand, A. B. Kunnumakkara, C. Sundaram, K. B. Harikumar, S. T. Tharakan, O. S. Lai, *et al.*, "Cancer is a preventable disease that requires major lifestyle changes," *Pharm Res,* vol. 25, pp. 2097-116, Sep 2008.

[59] M. Plummer, C. de Martel, J. Vignat, J. Ferlay, F. Bray, and S. Franceschi, "Global burden of cancers attributable to infections in 2012: a synthetic analysis," *Lancet Glob Health,* vol. 4, pp. e609-16, Sep 2016.

[60] F. Clavel-Chapelon and M. Gerber, "Reproductive factors and breast cancer risk. Do they differ according to age at diagnosis?," *Breast Cancer Res Treat,* vol. 72, pp. 107-15, Mar 2002.

[61] H. R. Harris, M. S. Rice, A. L. Shafrir, E. M. Poole, M. Gupta, J. L. Hecht, *et al.*, "Lifestyle and Reproductive Factors and Ovarian Cancer Risk by p53 and MAPK Expression," *Cancer Epidemiol Biomarkers Prev,* vol. 27, pp. 96-102, Jan 2018.

[62] A. T. Ali, "Reproductive factors and the risk of endometrial cancer," *Int J Gynecol Cancer,* vol. 24, pp. 384-93, Mar 2014.

[63] S. McGuire, "World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015," *Adv Nutr,* vol. 7, pp. 418-9, Mar 2016.

[64] N. M. Lindor, M. L. McMaster, C. J. Lindor, and M. H. Greene, "Concise handbook of familial cancer susceptibility syndromes - second edition," *J Natl Cancer Inst Monogr,* pp. 1-93, 2008.

[65] D. M. Parkin, P. Pisani, A. D. Lopez, and E. Masuyer, "At least one in seven cases of cancer is caused by smoking. Global estimates for 1985," *Int J Cancer,* vol. 59, pp. 494-504, Nov 15 1994.

[66] P. National Center for Chronic Disease, S. Health Promotion Office on, and Health, "Reports of the Surgeon General," in *The Health Consequences of Smoking-50 Years of Progress: A Report of the Surgeon General*, ed Atlanta (GA): Centers for Disease Control and Prevention (US), 2014.

[67] D. Hoffmann and I. Hoffmann, "The changing cigarette, 1950-1995," *J Toxicol Environ Health,* vol. 50, pp. 307-64, Mar 1997.

[68] C. J. Smith, T. A. Perfetti, M. A. Rumple, A. Rodgman, and D. J. Doolittle, ""IARC group 2A Carcinogens" reported in cigarette mainstream smoke," *Food Chem Toxicol,* vol. 38, pp. 371-83, Apr 2000.

[69] C. J. Smith, T. A. Perfetti, M. A. Mullens, A. Rodgman, and D. J. Doolittle, ""IARC group 2B Carcinogens" reported in cigarette mainstream smoke," *Food Chem Toxicol,* vol. 38, pp. 825-48, Sep 2000.

[70] A. Rivenson, D. Hoffmann, B. Prokopczyk, S. Amin, and S. S. Hecht, "Induction of lung and exocrine pancreas tumors in F344 rats by tobacco-specific and Areca-derived N-nitrosamines," *Cancer Res,* vol. 48, pp. 6912-7, Dec 1 1988.

[71] E. L. Wynder and D. Hoffmann, "Smoking and lung cancer: scientific challenges and opportunities," *Cancer Res,* vol. 54, pp. 5284-95, Oct 15 1994.

[72] J. Xue, S. Yang, and S. Seng, "Mechanisms of Cancer Induction by Tobacco-Specific NNK and NNN," *Cancers (Basel),* vol. 6, pp. 1138-56, May 14 2014.

[73] G. Akopyan and B. Bonavida, "Understanding tobacco smoke carcinogen NNK and lung tumorigenesis," *Int J Oncol,* vol. 29, pp. 745-52, Oct 2006.

[74] J. A. Shapiro, E. J. Jacobs, and M. J. Thun, "Cigar smoking in men and risk of death from tobacco-related cancers," *J Natl Cancer Inst,* vol. 92, pp. 333-7, Feb 16 2000.

[75] C. Iribarren, I. S. Tekawa, S. Sidney, and G. D. Friedman, "Effect of cigar smoking on the risk of cardiovascular disease, chronic obstructive pulmonary disease, and cancer in men," *N Engl J Med,* vol. 340, pp. 1773-80, Jun 10 1999.

[76] N. J. Wald and H. C. Watt, "Prospective study of effect of switching from cigarettes to pipes or cigars on mortality from three smoking related diseases," *Bmj,* vol. 314, pp. 1860-3, Jun 28 1997.

[77] S. Sun, J. H. Schiller, and A. F. Gazdar, "Lung cancer in never smokers--a different disease," *Nat Rev Cancer,* vol. 7, pp. 778-90, Oct 2007.

[78] H. A. Wakelee, E. T. Chang, S. L. Gomez, T. H. Keegan, D. Feskanich, C. A. Clarke, *et al.,* "Lung cancer incidence in never smokers," *J Clin Oncol,* vol. 25, pp. 472-8, Feb 10 2007.

[79] J. A. Zell, S. H. Ou, A. Ziogas, and H. Anton-Culver, "Epidemiology of bronchioloalveolar carcinoma: improvement in survival after release of the 1999 WHO classification of lung tumors," *J Clin Oncol,* vol. 23, pp. 8396-405, Nov 20 2005.

[80] R. Peto, S. Darby, H. Deo, P. Silcocks, E. Whitley, and R. Doll, "Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies," *Bmj,* vol. 321, pp. 323-9, Aug 5 2000.

[81] P. C. Gupta, P. R. Murti, R. B. Bhonsle, F. S. Mehta, and J. J. Pindborg, "Effect of cessation of tobacco use on the incidence of oral mucosal lesions in a 10-yr follow-up study of 12,212 users," *Oral Dis,* vol. 1, pp. 54-8, Mar 1995.

[82] B. Liu, Q. Xue, Y. Tang, J. Cao, F. P. Guengerich, and H. Zhang, "Mechanisms of mutagenesis: DNA replication in the presence of DNA damage," *Mutation research. Reviews in mutation research,* vol. 768, pp. 53-67, Apr-Jun 2016.

[83] K. A. Warnakulasuriya, N. W. Johnson, K. M. Linklater, and J. Bell, "Cancer of mouth, pharynx and nasopharynx in Asian and Chinese immigrants resident in Thames regions," *Oral Oncol,* vol. 35, pp. 471-5, Sep 1999.

[84] F. S. Mehta, P. C. Gupta, D. K. Daftary, J. J. Pindborg, and S. K. Choksi, "An epidemiologic study of oral cancer and precancerous conditions among 101,761 villagers in Maharashtra, India," *Int J Cancer,* vol. 10, pp. 134-41, Jul 15 1972.

[85] T. Axell, "Occurrence of leukoplakia and some other oral white lesions among 20,333 adult Swedish people," *Community Dent Oral Epidemiol,* vol. 15, pp. 46-51, Feb 1987.

[86] S. S. Hecht, "Tobacco carcinogens, their biomarkers and tobacco-induced cancer," *Nat Rev Cancer,* vol. 3, pp. 733-44, Oct 2003.

[87] N. A. Jafarey, Z. Mahmood, and S. H. Zaidi, "Habits and dietary pattern of cases of carcinoma of the oral cavity and oropharynx," *J Pak Med Assoc,* vol. 27, pp. 340-3, Jun 1977.

[88] A. Z. Keller, "Residence, age, race and related factors in the survival and associations with salivary tumors," *Am J Epidemiol,* vol. 90, pp. 269-77, Oct 1969.

[89] T. V. Evstifeeva and D. G. Zaridze, "Nass use, cigarette smoking, alcohol consumption and risk of oral and oesophageal precancer," *Eur J Cancer B Oral Oncol,* vol. 28b, pp. 29-35, Jul 1992.

[90] P. R. Murti, R. B. Bhonsle, J. J. Pindborg, D. K. Daftary, P. C. Gupta, and F. S. Mehta, "Malignant transformation rate in oral submucous fibrosis over a 17-year period," *Community Dent Oral Epidemiol,* vol. 13, pp. 340-1, Dec 1985.

[91] J. H. Jeng, M. C. Chang, and L. J. Hahn, "Role of areca nut in betel quid-associated chemical carcinogenesis: current awareness and future perspectives," *Oral Oncol,* vol. 37, pp. 477-92, Sep 2001.

[92] E. Bernzweig, J. B. Payne, R. A. Reinhardt, J. K. Dyer, and K. D. Patil, "Nicotine and smokeless tobacco effects on gingival and peripheral blood mononuclear cells," *J Clin Periodontol,* vol. 25, pp. 246-52, Mar 1998.

[93] M. A. Jaber, S. R. Porter, M. S. Gilthorpe, R. Bedi, and C. Scully, "Risk factors for oral epithelial dysplasia--the role of smoking and alcohol," *Oral Oncol,* vol. 35, pp. 151-6, Mar 1999.

[94] M. Soler, C. Bosetti, S. Franceschi, E. Negri, P. Zambon, R. Talamini, *et al.*, "Fiber intake and the risk of oral, pharyngeal and esophageal cancer," *Int J Cancer,* vol. 91, pp. 283-7, Feb 1 2001.

[95] E. De Stefani, A. Ronco, M. Mendilaharsu, and H. Deneo-Pellegrini, "Diet and risk of cancer of the upper aerodigestive tract--II. Nutrients," *Oral Oncol,* vol. 35, pp. 22-6, Jan 1999.

[96] E. Negri, S. Franceschi, C. Bosetti, F. Levi, E. Conti, M. Parpinel, *et al.*, "Selected micronutrients and oral and pharyngeal cancer," *Int J Cancer,* vol. 86, pp. 122-7, Apr 1 2000.

[97] P. D. James and I. O. Ellis, "Malignant epithelial tumours associated with autoimmune sialadenitis," *J Clin Pathol,* vol. 39, pp. 497-502, May 1986.

[98] R. Sankaranarayanan, B. Mathew, C. Varghese, P. R. Sudhakaran, V. Menon, A. Jayadeep, *et al.*, "Chemoprevention of oral leukoplakia with vitamin A and beta carotene: an assessment," *Oral Oncol,* vol. 33, pp. 231-6, Jul 1997.

[99] S. E. Benner, R. J. Winn, S. M. Lippman, J. Poland, K. S. Hansen, M. A. Luna, *et al.*, "Regression of oral leukoplakia with alpha-tocopherol: a community clinical oncology program chemoprevention study," *J Natl Cancer Inst,* vol. 85, pp. 44-7, Jan 6 1993.

[100] I. Cruz, A. J. Van Den Brule, A. A. Brink, P. J. Snijders, J. M. Walboomers, I. Van Der Waal, *et al.*, "No direct role for Epstein-Barr virus in oral carcinogenesis: a study at the DNA, RNA and protein levels," *Int J Cancer,* vol. 86, pp. 356-61, May 1 2000.

[101] G. M. Swanson and S. H. Belle, "Cancer morbidity among woodworkers in the U.S. automotive industry," *J Occup Med,* vol. 24, pp. 315-9, Apr 1982.

[102] L. Denny, "Cervical cancer: prevention and treatment," *Discov Med,* vol. 14, pp. 125-31, Aug 2012.

[103] T. Hoory, A. Monie, P. Gravitt, and T. C. Wu, "Molecular epidemiology of human papillomavirus," *J Formos Med Assoc,* vol. 107, pp. 198-217, Mar 2008.

[104] C. de Martel, J. Ferlay, S. Franceschi, J. Vignat, F. Bray, D. Forman, *et al.*, "Global burden of cancers attributable to infections in 2008: a review and synthetic analysis," *Lancet Oncol,* vol. 13, pp. 607-15, Jun 2012.

[105] N. Munoz, F. X. Bosch, S. de Sanjose, R. Herrero, X. Castellsague, K. V. Shah, *et al.*, "Epidemiologic classification of human papillomavirus types associated with cervical cancer," *N Engl J Med,* vol. 348, pp. 518-27, Feb 6 2003.

[106] V. Cogliano, R. Baan, K. Straif, Y. Grosse, B. Secretan, and F. El Ghissassi, "Carcinogenicity of human papillomaviruses," *Lancet Oncol,* vol. 6, p. 204, Apr 2005.

[107] M. Schiffman, G. Clifford, and F. M. Buonaguro, "Classification of weakly carcinogenic human papillomavirus types: addressing the limits of epidemiology at the borderline," *Infect Agent Cancer,* vol. 4, p. 8, Jun 1 2009.

[108] H. zur Hausen, "Papillomaviruses and cancer: from basic studies to clinical application," *Nat Rev Cancer,* vol. 2, pp. 342-50, May 2002.

[109] P. Guan, R. Howell-Jones, N. Li, L. Bruni, S. de Sanjose, S. Franceschi, *et al.*, "Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer," *Int J Cancer,* vol. 131, pp. 2349-59, Nov 15 2012.

[110] A. K. Chaturvedi, "Beyond cervical cancer: burden of other HPV-related cancers among men and women," *J Adolesc Health,* vol. 46, pp. S20-6, Apr 2010.

[111] A. P. Sowjanya, M. Jain, U. R. Poli, S. Padma, M. Das, K. V. Shah, *et al.*, "Prevalence and distribution of high-risk human papilloma virus (HPV) types in invasive squamous cell carcinoma of the cervix and in normal women in Andhra Pradesh, India," *BMC Infect Dis,* vol. 5, p. 116, Dec 22 2005.

[112] S. S. Kulkarni, S. S. Kulkarni, P. P. Vastrad, B. B. Kulkarni, A. R. Markande, G. S. Kadakol, *et al.*, "Prevalence and distribution of high risk human papillomavirus (HPV) Types 16 and 18 in Carcinoma of cervix, saliva of patients with oral squamous cell carcinoma and in the general population in Karnataka, India," *Asian Pac J Cancer Prev,* vol. 12, pp. 645-8, 2011.

[113] T. Gheit, S. Vaccarella, M. Schmitt, M. Pawlita, S. Franceschi, R. Sankaranarayanan, *et al.*, "Prevalence of human papillomavirus types in cervical and oral cancers in central India," *Vaccine,* vol. 27, pp. 636-9, Jan 29 2009.

[114] P. Basu, S. Roychowdhury, U. D. Bafna, S. Chaudhury, S. Kothari, R. Sekhon, *et al.*, "Human papillomavirus genotype distribution in cervical cancer in India: results from a multi-center study," *Asian Pac J Cancer Prev,* vol. 10, pp. 27-34, Jan-Mar 2009.

[115] N. Bhatla, N. Lal, Y. P. Bao, T. Ng, and Y. L. Qiao, "A meta-analysis of human papillomavirus type-distribution in women from South Asia: implications for vaccination," *Vaccine,* vol. 26, pp. 2811-7, Jun 2 2008.

[116] P. Datta, N. Bhatla, R. M. Pandey, L. Dar, A. R. Patro, S. Vasisht, *et al.*, "Type-specific incidence and persistence of HPV infection among young women: a prospective study in North India," *Asian Pac J Cancer Prev,* vol. 13, pp. 1019-24, 2012.

[117] A. B. Moscicki, "Impact of HPV infection in adolescent populations," *J Adolesc Health,* vol. 37, pp. S3-9, Dec 2005.

[118] S. Dutta, R. Begum, D. Mazumder Indra, S. S. Mandal, R. Mondal, J. Biswas, *et al.*, "Prevalence of human papillomavirus in women without cervical cancer: a population-based study in Eastern India," *Int J Gynecol Pathol,* vol. 31, pp. 178-183, Mar 2012.

[119] M. H. Einstein, J. T. Schiller, R. P. Viscidi, H. D. Strickler, P. Coursaget, T. Tan, *et al.*, "Clinician's guide to human papillomavirus immunology: knowns and unknowns," *Lancet Infect Dis,* vol. 9, pp. 347-56, Jun 2009.

[120] M. Lehoux, C. M. D'Abramo, and J. Archambault, "Molecular mechanisms of human papillomavirus-induced carcinogenesis," *Public Health Genomics,* vol. 12, pp. 268-80, 2009.

[121] A. J. Klingelhutz and A. Roman, "Cellular transformation by human papillomaviruses: lessons learned by comparing high- and low-risk viruses," *Virology,* vol. 424, pp. 77-98, Mar 15 2012.

[122] D. M. Parkin, F. Bray, J. Ferlay, and P. Pisani, "Global cancer statistics, 2002," *CA Cancer J Clin,* vol. 55, pp. 74-108, Mar-Apr 2005.

[123] J. T. Schiller, X. Castellsague, L. L. Villa, and A. Hildesheim, "An update of prophylactic human papillomavirus L1 virus-like particle vaccine clinical trial results," *Vaccine,* vol. 26 Suppl 10, pp. K53-61, Aug 19 2008.

[124] S. Garte, "The role of ethnicity in cancer susceptibility gene polymorphisms: the example of CYP1A1," *Carcinogenesis,* vol. 19, pp. 1329-32, Aug 1998.

[125] D. M. Parkin, "International variation," *Oncogene,* vol. 23, pp. 6329-40, Aug 23 2004.

[126] P. Kheirandish and F. Chinegwundoh, "Ethnic differences in prostate cancer," *Br J Cancer,* vol. 105, pp. 481-5, Aug 9 2011.

[127] V. L. Shavers and M. L. Brown, "Racial and ethnic disparities in the receipt of cancer treatment," *J Natl Cancer Inst,* vol. 94, pp. 334-57, Mar 6 2002.

[128] D. Ford, D. F. Easton, M. Stratton, S. Narod, D. Goldgar, P. Devilee, *et al.*, "Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium," *Am J Hum Genet,* vol. 62, pp. 676-89, Mar 1998.

[129] G. Mitchell, A. C. Antoniou, R. Warren, S. Peock, J. Brown, R. Davies, *et al.*, "Mammographic density and breast cancer risk in BRCA1 and BRCA2 mutation carriers," *Cancer Res,* vol. 66, pp. 1866-72, Feb 1 2006.

[130] G. Rennert, S. Bisland-Naggan, O. Barnett-Griness, N. Bar-Joseph, S. Zhang, H. S. Rennert, *et al.*, "Clinical outcomes of breast cancer in carriers of BRCA1 and BRCA2 mutations," *N Engl J Med,* vol. 357, pp. 115-23, Jul 12 2007.

[131] P. J. O'Donovan and D. M. Livingston, "BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair," *Carcinogenesis,* vol. 31, pp. 961-7, Jun 2010.

[132] M. R. Spitz, W. K. Hong, C. I. Amos, X. Wu, M. B. Schabath, Q. Dong, *et al.*, "A risk model for prediction of lung cancer," *J Natl Cancer Inst,* vol. 99, pp. 715-26, May 2 2007.

[133] M. R. Spitz, C. J. Etzel, Q. Dong, C. I. Amos, Q. Wei, X. Wu, *et al.*, "An expanded risk prediction model for lung cancer," *Cancer Prev Res (Phila),* vol. 1, pp. 250-4, Sep 2008.

[134] A. Matakidou, T. Eisen, and R. S. Houlston, "Systematic review of the relationship between family history and lung cancer risk," *Br J Cancer,* vol. 93, pp. 825-33, Oct 3 2005.

[135] J. E. Bailey-Wilson, C. I. Amos, S. M. Pinney, G. M. Petersen, M. de Andrade, J. S. Wiest, *et al.*, "A major lung cancer susceptibility locus maps to chromosome 6q23-25," *Am J Hum Genet,* vol. 75, pp. 460-74, Sep 2004.

[136] L. Le Marchand, C. Guo, S. Benhamou, C. Bouchardy, I. Cascorbi, M. L. Clapper, *et al.*, "Pooled analysis of the CYP1A1 exon 7 polymorphism and lung cancer (United States)," *Cancer Causes Control,* vol. 14, pp. 339-46, May 2003.

[137] S. Benhamou, W. J. Lee, A. K. Alexandrie, P. Boffetta, C. Bouchardy, D. Butkiewicz, *et al.*, "Meta- and pooled analyses of the effects of glutathione S-transferase M1 polymorphisms and smoking on lung cancer risk," *Carcinogenesis,* vol. 23, pp. 1343-50, Aug 2002.

[138] Z. Ye, H. Song, J. P. Higgins, P. Pharoah, and J. Danesh, "Five glutathione s-transferase gene variants in 23,452 cases of lung cancer and 30,397 controls: meta-analysis of 130 studies," *PLoS Med,* vol. 3, p. e91, Apr 2006.

[139] Y. Gao, F. Gao, T. T. Hu, G. Li, and Y. X. Sui, "Combined effects of glutathione S-transferase M1 and T1 polymorphisms on risk of lung cancer: Evidence from a meta-analysis," *Oncotarget,* vol. 8, pp. 28135-28143, Apr 25 2017.

[140] C. I. Amos, X. Wu, P. Broderick, I. P. Gorlov, J. Gu, T. Eisen, *et al.*, "Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1," *Nat Genet,* vol. 40, pp. 616-22, May 2008.

[141] W. Zhou, G. Liu, S. W. Thurston, L. L. Xu, D. P. Miller, J. C. Wain, *et al.*, "Genetic polymorphisms in N-acetyltransferase-2 and microsomal epoxide hydrolase, cumulative cigarette smoking, and lung cancer," *Cancer Epidemiol Biomarkers Prev,* vol. 11, pp. 15-21, Jan 2002.

[142] M. R. Spitz, Q. Wei, Q. Dong, C. I. Amos, and X. Wu, "Genetic susceptibility to lung cancer: the role of DNA damage and repair," *Cancer Epidemiol Biomarkers Prev,* vol. 12, pp. 689-98, Aug 2003.

[143] X. Wu, G. L. Delclos, J. F. Annegers, M. L. Bondy, S. E. Honn, B. Henry, *et al.*, "A case-control study of wood dust exposure, mutagen sensitivity, and lung cancer risk," *Cancer Epidemiol Biomarkers Prev,* vol. 4, pp. 583-8, Sep 1995.

[144] C. S. Dela Cruz, L. T. Tanoue, and R. A. Matthay, "Lung Cancer: Epidemiology, Etiology, and Prevention," *Clinics in chest medicine,* vol. 32, p. 10.1016/j.ccm.2011.09.001, 2011.

[145] M. P. Copper, A. Jovanovic, J. J. Nauta, B. J. Braakhuis, N. de Vries, I. van der Waal, *et al.*, "Role of genetic factors in the etiology of squamous cell carcinoma of the head and neck," *Arch Otolaryngol Head Neck Surg,* vol. 121, pp. 157-60, Feb 1995.

[146] G. Hernandez, L. Arriba, C. Jimenez, J. V. Bagan, B. Rivera, M. Lucas, *et al.*, "Rapid progression from oral leukoplakia to carcinoma in an immunosuppressed liver transplant recipient," *Oral Oncol,* vol. 39, pp. 87-90, Jan 2003.

[147] D. P. Bartel, "MicroRNAs: genomics, biogenesis, mechanism, and function," *Cell,* vol. 116, pp. 281-97, Jan 23 2004.

[148] E. Berezikov, V. Guryev, J. van de Belt, E. Wienholds, R. H. Plasterk, and E. Cuppen, "Phylogenetic shadowing and computational identification of human microRNA genes," *Cell,* vol. 120, pp. 21-4, Jan 14 2005.

[149] X. Zhang, R. He, F. Ren, R. Tang, and G. Chen, "Association of miR-146a rs2910164 polymorphism with squamous cell carcinoma risk: a meta-analysis," *J buon,* vol. 20, pp. 829-41, May-Jun 2015.

[150] S. Zhang, F. Hu, H. Liang, Y. Liu, J. Yang, and W. Zhou, "Association between a miRNA-146a polymorphism and susceptibility to head and neck squamous cell carcinoma in Chinese patients: A meta-analysis of 8 case-control studies," vol. 12, p. e0186609, 2017.

[151] D. W. Crabb, H. J. Edenberg, W. F. Bosron, and T. K. Li, "Genotypes for aldehyde dehydrogenase deficiency and alcohol sensitivity. The inactive ALDH2(2) allele is dominant," *Journal of Clinical Investigation,* vol. 83, pp. 314-316, 1989.

[152] K. Liu, G. Song, X. Zhu, X. Yang, Y. Shen, W. Wang, *et al.*, "Association between ALDH2 Glu487Lys polymorphism and the risk of esophageal cancer," *Medicine (Baltimore),* vol. 96, p. e6111, Apr 2017.

[153] J. Massague, "TGFbeta signalling in context," *Nat Rev Mol Cell Biol,* vol. 13, pp. 616-30, Oct 2012.

[154] Y. Y. Gu, H. Wang, and S. Wang, "TGF-beta1 C-509T and T869C polymorphisms and cancer risk: a meta analysis," *Int J Clin Exp Med,* vol. 8, pp. 17932-40, 2015.

[155] X. Zhuo, J. Song, J. Liao, W. Zhou, H. Ye, Q. Li, *et al.*, "Does CYP2E1 RsaI/PstI polymorphism confer head and neck carcinoma susceptibility?: A meta-analysis based on 43 studies," *Medicine (Baltimore),* vol. 95, p. e5156, Oct 2016.

[156] R. Wattleworth, "Human papillomavirus infection and the links to penile and cervical cancer," *J Am Osteopath Assoc,* vol. 111, pp. S3-10, Mar 2011.

[157] A. Gadducci, C. Barsotti, S. Cosio, L. Domenici, and A. Riccardo Genazzani, "Smoking habit, immune suppression, oral contraceptive use, and hormone replacement therapy use and cervical carcinogenesis: a review of the literature," *Gynecol Endocrinol,* vol. 27, pp. 597-604, Aug 2011.

[158] Q. Li and G. Lozano, "Molecular pathways: targeting Mdm2 and Mdm4 in cancer therapy," *Clin Cancer Res,* vol. 19, pp. 34-41, Jan 1 2013.

[159] G. L. Bond, W. Hu, E. E. Bond, H. Robins, S. G. Lutzker, N. C. Arva, *et al.*, "A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans," *Cell,* vol. 119, pp. 591-602, Nov 24 2004.

[160] J. J. Manfredi, "The Mdm2-p53 relationship evolves: Mdm2 swings both ways as an oncogene and a tumor suppressor," *Genes Dev,* vol. 24, pp. 1580-9, Aug 1 2010.

[161] X. Zhuo, J. Ren, D. Li, Y. Wu, and Q. Zhou, "MDM2 SNP309 variation increases cervical cancer risk among Asians," *Tumour Biol,* vol. 35, pp. 5331-7, Jun 2014.

[162] A. Thomas, T. Giesler, and E. White, "p53 mediates bcl-2 phosphorylation and apoptosis via activation of the Cdc42/JNK1 pathway," *Oncogene,* vol. 19, pp. 5259-69, Nov 2 2000.

[163] K. S. Srivenugopal and F. Ali-Osman, "The DNA repair protein, O(6)-methylguanine-DNA methyltransferase is a proteolytic target for the E6 human papillomavirus oncoprotein," *Oncogene,* vol. 21, pp. 5940-5, Aug 29 2002.

[164] A. Storey, M. Thomas, A. Kalita, C. Harwood, D. Gardiol, F. Mantovani, *et al.*, "Role of a p53 polymorphism in the development of human papillomavirus-associated cancer," *Nature,* vol. 393, pp. 229-34, May 21 1998.

[165] S. Habbous, V. Pang, L. Eng, W. Xu, G. Kurtz, F. F. Liu, *et al.*, "p53 Arg72Pro polymorphism, HPV status and initiation, progression, and development of cervical cancer: a systematic review and meta-analysis," *Clin Cancer Res,* vol. 18, pp. 6407-15, Dec 1 2012.

[166] F. Balkwill, "Tumour necrosis factor and cancer," *Nat Rev Cancer,* vol. 9, pp. 361-71, May 2009.

[167] H. Walczak, "TNF and ubiquitin at the crossroads of gene activation, cell death, inflammation, and cancer," *Immunol Rev,* vol. 244, pp. 9-28, Nov 2011.

[168] S. Fargion, L. Valenti, P. Dongiovanni, A. Scaccabarozzi, A. L. Fracanzani, E. Taioli, *et al.*, "Tumor necrosis factor alpha promoter polymorphisms influence the phenotypic expression of hereditary hemochromatosis," *Blood,* vol. 97, pp. 3707-12, Jun 15 2001.

[169] L. Valenti, A. L. Fracanzani, P. Dongiovanni, G. Santorelli, A. Branchi, E. Taioli, *et al.*, "Tumor necrosis factor alpha promoter polymorphisms and insulin resistance in nonalcoholic fatty liver disease," *Gastroenterology,* vol. 122, pp. 274-80, Feb 2002.

[170] L. J. Abraham and K. M. Kroeger, "Impact of the -308 TNF promoter polymorphism on the transcriptional regulation of the TNF gene: relevance to disease," *J Leukoc Biol,* vol. 66, pp. 562-6, Oct 1999.

[171] J. P. Bayley, T. H. Ottenhoff, and C. L. Verweij, "Is there a future for TNF promoter polymorphisms?," *Genes Immun,* vol. 5, pp. 315-29, Aug 2004.

[172] M. Li, Y. Han, T.-T. Wu, Y. Feng, and H.-B. Wang, "Tumor Necrosis Factor Alpha rs1800629 Polymorphism and Risk of Cervical Lesions: A Meta-Analysis," *PLoS ONE,* vol. 8, p. e69201, 08/27

[173] T. R. Smith, M. S. Miller, K. Lohman, E. M. Lange, L. D. Case, H. W. Mohrenweiser, *et al.*, "Polymorphisms of XRCC1 and XRCC3 genes and susceptibility to breast cancer," *Cancer Lett,* vol. 190, pp. 183-90, Feb 20 2003.

[174] Z. Hu, H. Ma, F. Chen, Q. Wei, and H. Shen, "XRCC1 polymorphisms and cancer risk: a meta-analysis of 38 case-control studies," *Cancer Epidemiol Biomarkers Prev,* vol. 14, pp. 1810-8, Jul 2005.

[175] J. Yin, U. Vogel, Y. Ma, R. Qi, Z. Sun, and H. Wang, "The DNA repair gene XRCC1 and genetic susceptibility of lung cancer in a northeastern Chinese population," *Lung Cancer,* vol. 56, pp. 153-60, May 2007.

[176] Y. Li, F. Liu, S.-Q. Tan, Y. Wang, and S.-W. Li, "X-Ray Repair Cross-Complementing Group 1 (XRCC1) Genetic Polymorphisms and Cervical Cancer Risk: A HuGE Systematic Review and Meta-Analysis," *PLoS ONE,* vol. 7, p. e44441, 09/12

[177] L. Y. Qin, X. Chen, P. Li, Z. Yang, and W. N. Mo, "Association between the XRCC3 Thr241Met polymorphism and cervical cancer risk: a meta-analysis," *Asian Pac J Cancer Prev,* vol. 14, pp. 6703-7, Jan 2014.

[178] A. Rao, M. Manikandan, G. Arunkumar, S. Revathidevi, and V. Vinothkumar, "Prevalence of p53 codon 72,p73 G4C14-A4T14 and MDM2 T309G polymorphisms and its association with the risk of oral cancer in South Indians.," *Gene Reports,* 17 March 2017 2017.

[179] F. Murray-Zmijewski, E. A. Slee, and X. Lu, "A complex barcode underlies the heterogeneous response of p53 to stress," *Nat Rev Mol Cell Biol,* vol. 9, pp. 702-12, Sep 2008.

[180] J. D. Weber, L. J. Taylor, M. F. Roussel, C. J. Sherr, and D. Bar-Sagi, "Nucleolar Arf sequesters Mdm2 and activates p53," *Nat Cell Biol,* vol. 1, pp. 20-6, May 1999.

[181] D. K. Dimova and N. J. Dyson, "The E2F transcriptional network: old acquaintances with new faces," *Oncogene,* vol. 24, pp. 2810-26, Apr 18 2005.

[182] S. Polager and D. Ginsberg, "E2F - at the crossroads of life and death," *Trends Cell Biol,* vol. 18, pp. 528-35, Nov 2008.

[183] J. Stanelle and B. M. Putzer, "E2F1-induced apoptosis: turning killers into therapeutics," *Trends Mol Med,* vol. 12, pp. 177-85, Apr 2006.

[184] S. van den Heuvel and N. J. Dyson, "Conserved functions of the pRB and E2F families," *Nat Rev Mol Cell Biol,* vol. 9, pp. 713-24, Sep 2008.

[185] H. Z. Chen, S. Y. Tsai, and G. Leone, "Emerging roles of E2Fs in cancer: an exit from cell cycle control," *Nat Rev Cancer,* vol. 9, pp. 785-97, Nov 2009.

[186] J. M. Trimarchi and J. A. Lees, "Sibling rivalry in the E2F family," *Nat Rev Mol Cell Biol,* vol. 3, pp. 11-20, Jan 2002.

[187] A. K. Biswas and D. G. Johnson, "Transcriptional and nontranscriptional functions of E2F1 in response to DNA damage," *Cancer Res,* vol. 72, pp. 13-7, Jan 1 2012.

[188] H. Chen, X. Gu, Y. Liu, J. Wang, S. E. Wirt, R. Bottino, *et al.*, "PDGF signalling controls age-dependent proliferation in pancreatic beta-cells," *Nature,* vol. 478, pp. 349-55, Oct 12 2011.

[189] A. K. Biswas, D. L. Mitchell, and D. G. Johnson, "E2F1 responds to ultraviolet radiation by directly stimulating DNA repair and suppressing carcinogenesis," *Cancer Res,* vol. 74, pp. 3369-77, Jun 15 2014.

[190] L. Ingram, S. Munro, A. S. Coutts, and N. B. La Thangue, "E2F-1 regulation by an unusual DNA damage-responsive DP partner subunit," *Cell Death Differ,* vol. 18, pp. 122-32, Jan 2011.

[191] J. Li, C. Ran, E. Li, F. Gordon, G. Comstock, H. Siddiqui, *et al.*, "Synergistic Function of E2F7 and E2F8 is Essential for Cell Survival and Embryonic Development," *Developmental cell*, vol. 14, pp. 62-75, 2008.

[192] N. A. Laurie, S. L. Donovan, C. S. Shih, J. Zhang, N. Mills, C. Fuller, *et al.*, "Inactivation of the p53 pathway in retinoblastoma," *Nature*, vol. 444, pp. 61-6, Nov 2 2006.

[193] A. M. Pierce, I. B. Gimenez-Conti, R. Schneider-Broussard, L. A. Martinez, C. J. Conti, and D. G. Johnson, "Increased E2F1 activity induces skin tumors in mice heterozygous and nullizygous for p53," *Proc Natl Acad Sci U S A*, vol. 95, pp. 8858-63, Jul 21 1998.

[194] A. M. Pierce, R. Schneider-Broussard, I. B. Gimenez-Conti, J. L. Russell, C. J. Conti, and D. G. Johnson, "E2F1 has both oncogenic and tumor-suppressive properties in a transgenic model," *Mol Cell Biol*, vol. 19, pp. 6408-14, Sep 1999.

[195] L. Yamasaki, R. Bronson, B. O. Williams, N. J. Dyson, E. Harlow, and T. Jacks, "Loss of E2F-1 reduces tumorigenesis and extends the lifespan of Rb1(+/-)mice," *Nat Genet*, vol. 18, pp. 360-4, Apr 1998.

[196] S. Munro, U. Oppermann, and N. B. La Thangue, "Pleiotropic effect of somatic mutations in the E2F subunit DP-1 gene in human cancer," *Oncogene*, vol. 33, pp. 3594-603, Jul 3 2014.

[197] T. Itoshima, T. Fujiwara, T. Waku, J. Shao, M. Kataoka, W. G. Yarbrough, *et al.*, "Induction of apoptosis in human esophageal cancer cells by sequential transfer of the wild-type p53 and E2F-1 genes: involvement of p53 accumulation via ARF-mediated MDM2 down-regulation," *Clin Cancer Res*, vol. 6, pp. 2851-9, Jul 2000.

[198] B. M. Putzer, "E2F1 death pathways as targets for cancer therapy," *J Cell Mol Med*, vol. 11, pp. 239-51, Mar-Apr 2007.

[199] T. Suzuki, W. Yasui, H. Yokozaki, K. Naka, T. Ishikawa, and E. Tahara, "Expression of the E2F family in human gastrointestinal carcinomas," *Int J Cancer*, vol. 81, pp. 535-8, May 17 1999.

[200] W. Yasui, J. Fujimoto, T. Suzuki, S. Ono, K. Naka, H. Yokozaki, *et al.*, "Expression of cell-cycle-regulating transcription factor E2F-1 in colorectal carcinomas," *Pathobiology*, vol. 67, pp. 174-9, Jul-Aug 1999.

[201] P. Zacharatos, A. Kotsinas, K. Evangelou, P. Karakaidos, L. V. Vassiliou, N. Rezaei, *et al.*, "Distinct expression patterns of the transcription factor E2F-1 in relation to

tumour growth parameters in common human carcinomas," *J Pathol,* vol. 203, pp. 744-53, Jul 2004.

[202] J. Bramis, P. Zacharatos, I. Papaconstantinou, A. Kotsinas, F. Sigala, D. P. Korkolis, *et al.*, "E2F-1 transcription factor immunoexpression is inversely associated with tumor growth in colon adenocarcinomas," *Anticancer Res,* vol. 24, pp. 3041-7, Sep-Oct 2004.

[203] A. Xanthoulis, A. Kotsinas, D. Tiniakos, A. Fiska, A. A. Tentes, A. Kyroudi, *et al.*, "The relationship between E2F family members and tumor growth in colorectal adenocarcinomas: A comparative immunohistochemical study of 100 cases," *Appl Immunohistochem Mol Morphol,* vol. 22, pp. 471-7, Jul 2014.

[204] D. Banerjee, R. Gorlick, A. Liefshitz, K. Danenberg, P. C. Danenberg, P. V. Danenberg, *et al.*, "Levels of E2F-1 expression are higher in lung metastasis of colon cancer as compared with hepatic metastasis and correlate with levels of thymidylate synthase," *Cancer Res,* vol. 60, pp. 2365-7, May 1 2000.

[205] M. Kasahara, Y. Takahashi, T. Nagata, S. Asai, T. Eguchi, Y. Ishii, *et al.*, "Thymidylate synthase expression correlates closely with E2F1 expression in colon cancer," *Clin Cancer Res,* vol. 6, pp. 2707-11, Jul 2000.

[206] T. Plath, M. Peters, K. Detjen, M. Welzel, Z. von Marschall, C. Radke, *et al.*, "Overexpression of pRB in human pancreatic carcinoma cells: function in chemotherapy-induced apoptosis," *J Natl Cancer Inst,* vol. 94, pp. 129-42, Jan 16 2002.

[207] M. J. Elliott, M. R. Farmer, C. Atienza, Jr., A. Stilwell, Y. B. Dong, H. L. Yang, *et al.*, "E2F-1 gene therapy induces apoptosis and increases chemosensitivity in human pancreatic carcinoma cells," *Tumour Biol,* vol. 23, pp. 76-86, Mar-Apr 2002.

[208] F. Rodicker, T. Stiewe, S. Zimmermann, and B. M. Putzer, "Therapeutic efficacy of E2F1 in pancreatic cancer correlates with TP73 induction," *Cancer Res,* vol. 61, pp. 7052-5, Oct 1 2001.

[209] Y. L. Chen, Y. H. Uen, C. F. Li, K. C. Horng, L. R. Chen, W. R. Wu, *et al.*, "The E2F transcription factor 1 transactives stathmin 1 in hepatocellular carcinoma," *Ann Surg Oncol,* vol. 20, pp. 4041-54, Nov 2013.

[210] C. Wang, Y. Xiao, Z. Hu, Y. Chen, N. Liu, and G. Hu, "PEG10 directly regulated by E2Fs might have a role in the development of hepatocellular carcinoma," *FEBS Lett,* vol. 582, pp. 2793-8, Aug 6 2008.

[211] R. M. Pascale, M. M. Simile, M. R. De Miglio, M. R. Muroni, D. F. Calvisi, G. Asara, *et al.*, "Cell cycle deregulation in liver lesions of rats with and without genetic predisposition to hepatocarcinogenesis," *Hepatology,* vol. 35, pp. 1341-50, Jun 2002.

[212] E. A. Conner, E. R. Lemmer, M. Omori, P. J. Wirth, V. M. Factor, and S. S. Thorgeirsson, "Dual functions of E2F-1 in a transgenic mouse model of liver carcinogenesis," *Oncogene,* vol. 19, pp. 5054-62, Oct 19 2000.

[213] C. Atienza, Jr., M. J. Elliott, Y. B. Dong, H. L. Yang, A. Stilwell, T. J. Liu, *et al.*, "Adenovirus-mediated E2F-1 gene transfer induces an apoptotic response in human gastric carcinoma cells that is enhanced by cyclin dependent kinase inhibitors," *Int J Mol Med,* vol. 6, pp. 55-63, Jul 2000.

[214] Q. Xiao, L. Li, Y. Xie, N. Tan, C. Wang, J. Xu, *et al.*, "Transcription factor E2F-1 is upregulated in human gastric cancer tissues and its overexpression suppresses gastric tumor cell proliferation," *Cell Oncol,* vol. 29, pp. 335-49, 2007.

[215] Y. Xie, C. Wang, L. Li, Y. Ma, Y. Yin, and Q. Xiao, "Overexpression of E2F-1 inhibits progression of gastric cancer in vitro," *Cell Biol Int,* vol. 33, pp. 640-9, Jun 2009.

[216] S. M. Rubin, A. L. Gall, N. Zheng, and N. P. Pavletich, "Structure of the Rb C-terminal domain bound to E2F1-DP1: a mechanism for phosphorylation-induced E2F release," *Cell,* vol. 123, pp. 1093-106, Dec 16 2005.

[217] N. Zheng, E. Fraenkel, C. O. Pabo, and N. P. Pavletich, "Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP," *Genes Dev,* vol. 13, pp. 666-74, Mar 15 1999.

[218] A. Kouranov, L. Xie, J. de la Cruz, L. Chen, J. Westbrook, P. E. Bourne, *et al.*, "The RCSB PDB information portal for structural genomics," *Nucleic Acids Res,* vol. 34, pp. D302-5, Jan 01 2006.

[219] P. Kumar, S. Henikoff, and P. C. Ng, "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm," *Nat Protoc,* vol. 4, pp. 1073-81, 2009.

[220] E. Capriotti, R. Calabrese, and R. Casadio, "Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information," *Bioinformatics,* vol. 22, pp. 2729-34, Nov 15 2006.

[221] Y. Choi, G. E. Sims, S. Murphy, J. R. Miller, and A. P. Chan, "Predicting the functional effect of amino acid substitutions and indels," *PLoS One,* vol. 7, p. e46688, 2012.

[222] S. V. Tavtigian, A. M. Deffenbaugh, L. Yin, T. Judkins, T. Scholl, P. B. Samollow, *et al.*, "Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral," *J Med Genet,* vol. 43, pp. 295-305, Apr 2006.

[223] E. Mathe, M. Olivier, S. Kato, C. Ishioka, P. Hainaut, and S. V. Tavtigian, "Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods," *Nucleic Acids Res,* vol. 34, pp. 1317-25, 2006.

[224] H. A. Shihab, J. Gough, D. N. Cooper, P. D. Stenson, G. L. Barker, K. J. Edwards, *et al.*, "Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models," *Hum Mutat,* vol. 34, pp. 57-65, Jan 2013.

[225] I. Adzhubei, D. M. Jordan, and S. R. Sunyaev, "Predicting functional effect of human missense mutations using PolyPhen-2," *Curr Protoc Hum Genet,* vol. Chapter 7, p. Unit7.20, Jan 2013.

[226] Y. Bromberg and B. Rost, "SNAP: predict effect of non-synonymous polymorphisms on function," *Nucleic Acids Res,* vol. 35, pp. 3823-35, 2007.

[227] R. Calabrese, E. Capriotti, P. Fariselli, P. L. Martelli, and R. Casadio, "Functional annotations improve the predictive score of human disease-related mutations in proteins," *Hum Mutat,* vol. 30, pp. 1237-44, Aug 2009.

[228] C. M. Yates, I. Filippis, L. A. Kelley, and M. J. Sternberg, "SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features," *J Mol Biol,* vol. 426, pp. 2692-701, Jul 15 2014.

[229] E. Capriotti, P. Fariselli, I. Rossi, and R. Casadio, "A three-state prediction of single point mutations on protein stability changes," *BMC Bioinformatics,* vol. 9 Suppl 2, p. S6, Mar 26 2008.

[230] K. A. Bava, M. M. Gromiha, H. Uedaira, K. Kitajima, and A. Sarai, "ProTherm, version 4.0: thermodynamic database for proteins and mutants," *Nucleic Acids Res,* vol. 32, pp. D120-1, Jan 01 2004.

[231] M. Landau, I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko, *et al.*, "ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures," *Nucleic Acids Res,* vol. 33, pp. W299-302, Jul 01 2005.

[232] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Res,* vol. 32, pp. 1792-7, 2004.

[233] A. A. Inc., "Biological sequence alignment editor Version 7.2.0; An Abbott Inc., California, USA."

[234] F. Melo and E. Feytmans, "Assessing protein structures with a non-local atomic interaction energy," *J Mol Biol,* vol. 277, pp. 1141-52, Apr 17 1998.

[235] C. M. Topham, N. Srinivasan, and T. L. Blundell, "Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables," *Protein Eng,* vol. 10, pp. 7-21, Jan 1997.

[236] S. Yin, F. Ding, and N. V. Dokholyan, "Eris: an automated estimator of protein stability," *Nat Methods,* vol. 4, pp. 466-7, Jun 2007.

[237] G. De Baets, J. Van Durme, J. Reumers, S. Maurer-Stroh, P. Vanhee, J. Dopazo, *et al.,* "SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants," *Nucleic Acids Res,* vol. 40, pp. D935-9, Jan 2012.

[238] B. Reva, Y. Antipin, and C. Sander, "Predicting the functional impact of protein mutations: application to cancer genomics," *Nucleic Acids Res,* vol. 39, p. e118, Sep 01 2011.

[239] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation," *J Chem Theory Comput,* vol. 4, pp. 435-47, Mar 2008.

[240] G. A. K. a. R. A. F. *, "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides†," *J. Phys. Chem. B, 2001, 105 (28), pp 6474–6487, Julian Tirado-Rives and William L. Jorgensen,* April 19, 2001 2001.

[241] L.-L. e. al, " Molecular Dynamics in practice with GROMACS -Prace Training Portal ., Proteins 78, 1950-58, 2010 " 2010.

[242] J. C. William L. Jorgensen, and Jeffry D. Madura, "Comparison of simple potential functions for simulating liquid water," *The Journal of Chemical Physics,* vol. Volume 79, , 1983.

[243] S. A. Miller, D. D. Dykes, and H. F. Polesky, "A simple salting out procedure for extracting DNA from human nucleated cells," *Nucleic Acids Res,* vol. 16, p. 1215, Feb 11 1988.

[244] Q. Cui and I. Bahar, *Normal mode analysis: theory and applications to biological and chemical systems*: CRC press, 2005.

[245] W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, *Computer simulation of biomolecular systems: theoretical and experimental applications* vol. 3: Springer Science & Business Media, 2013.

[246] M. J. Cheatham TE, Fox T, Darden TA, Kollman A "Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins.," *J Am Chem Soc* vol. 14, 1995.

[247] B. S. Shastry, "SNP alleles in human disease and evolution," *J Hum Genet,* vol. 47, pp. 561-6, 2002.

[248] S. Sunyaev, V. Ramensky, I. Koch, W. Lathe, 3rd, A. S. Kondrashov, and P. Bork, "Prediction of deleterious human alleles," *Hum Mol Genet,* vol. 10, pp. 591-7, Mar 15 2001.

[249] P. Yue and J. Moult, "Identification and analysis of deleterious human SNPs," *J Mol Biol,* vol. 356, pp. 1263-74, Mar 10 2006.

[250] M. W. Dimmic, S. Sunyaev, and C. D. Bustamante, "Inferring SNP function using evolutionary, structural, and computational methods," *Pac Symp Biocomput,* pp. 382-4, 2005.

[251] M. D. Garrett, "Cell cycle control and cancer," *CURRENT SCIENCE,* vol. VOL. 81, 10 SEPTEMBER 2001.

[252] J. M. Cunningham, R. A. Vierkant, T. A. Sellers, C. Phelan, D. N. Rider, M. Liebow, *et al.*, "Cell cycle genes and ovarian cancer susceptibility: a tagSNP analysis," *Br J Cancer,* vol. 101, pp. 1461-8, Oct 20 2009.

[253] S. Singh, M. Gupta, R. K. Seam, and H. Changotra, "E2F1 genetic variants and risk of cervical cancer in Indian women," *Int J Biol Markers,* p. 1724600818768459, Apr 1 2018.

[254] S. Singh, M. Gupta, A. Sharma, R. K. Seam, and H. Changotra, "The Nonsynonymous Polymorphisms Val276Met and Gly393Ser of E2F1 Gene are Strongly Associated with Lung, and Head and Neck Cancers," *Genet Test Mol Biomarkers,* vol. 22, pp. 498-502, Aug 2018.

[255] W. C. Lin, F. T. Lin, and J. R. Nevins, "Selective induction of E2F1 in response to DNA damage, mediated by ATM-dependent phosphorylation," *Genes Dev,* vol. 15, pp. 1833-44, Jul 15 2001.

[256] T. C. Hallstrom and J. R. Nevins, "Specificity in the activation and control of transcription factor E2F-dependent apoptosis," *Proc Natl Acad Sci U S A*, vol. 100, pp. 10848-53, Sep 16 2003.

[257] I. Shats, M. Deng, A. Davidovich, C. Zhang, J. S. Kwon, D. Manandhar, *et al.*, "Expression level is a key determinant of E2F1-mediated cell fate," *Cell Death Differ*, vol. 24, pp. 626-637, Apr 2017.

[258] T. Stiewe and B. M. Putzer, "Role of the p53-homologue p73 in E2F1-induced apoptosis," *Nat Genet*, vol. 26, pp. 464-9, Dec 2000.

[259] M. C. Moroni, E. S. Hickman, E. Lazzerini Denchi, G. Caprara, E. Colli, F. Cecconi, *et al.*, "Apaf-1 is a transcriptional target for E2F and p53," *Nat Cell Biol*, vol. 3, pp. 552-8, Jun 2001.

[260] S. A. Forbes, D. Beare, H. Boutselakis, S. Bamford, N. Bindal, J. Tate, *et al.*, "COSMIC: somatic cancer genetics at high-resolution," *Nucleic Acids Res*, vol. 45, pp. D777-d783, Jan 4 2017.

[261] H. Venselaar, T. A. Te Beek, R. K. Kuipers, M. L. Hekkelman, and G. Vriend, "Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces," *BMC Bioinformatics*, vol. 11, p. 548, Nov 8 2010.

[262] W. Yu, W. Chan-On, M. Teo, C. K. Ong, I. Cutcutache, G. E. Allen, *et al.*, "First somatic mutation of E2F1 in a critical DNA binding residue discovered in well-differentiated papillary mesothelioma of the peritoneum," *Genome Biol*, vol. 12, p. R96, Sep 28 2011.

[263] S. Lakkakula, R. Maram, A. K. Munirajan, R. M. Pathapati, S. B. Visveswara, and B. V. Lakkakula, "Functional PstI/RsaI polymorphisms in the CYP2E1 gene among south Indian populations," *Asian Pac J Cancer Prev*, vol. 14, pp. 179-82, 2013.

[264] H. Jiang, F. Ge, B. Hu, L. Wu, H. Yang, and H. Wang, "rs35301225 polymorphism in miR-34a promotes development of human colon cancer by deregulation of 3'UTR in E2F1 in Chinese population," *Cancer Cell Int*, vol. 17, p. 39, 2017.

[265] Y. Li, E. M. Sturgis, Y. Yuan, M. Lu, X. Cao, Q. Wei, *et al.*, "Effect of human papillomavirus seropositivity and E2F2 promoter variants on risk of squamous cell carcinomas of oropharynx and oral cavity," *Carcinogenesis*, Sep 08 2016.

[266] Y. Zhao, L. Tang, W. Nie, Z. Wang, and X. Guan, "Functional variants at the miRNA binding sites of the E2F1 gene and its mRNA expression," *Oncol Lett*, vol. 5, pp. 398-402, Jan 2013.

[267] A. Mostowska, K. K. Hozyasz, P. Wojcicki, D. Galas-Filipowicz, A. Lasota, I. Dunin-Wilczynska, *et al.*, "Genetic variants in BRIP1 (BACH1) contribute to risk of nonsyndromic cleft lip with or without cleft palate," *Birth Defects Res A Clin Mol Teratol*, vol. 100, pp. 670-8, Sep 2014.

[268] K. Munger and P. M. Howley, "Human papillomavirus immortalization and transformation functions," *Virus Res*, vol. 89, pp. 213-28, Nov 2002.

[269] V. M. Williams, M. Filippova, V. Filippov, K. J. Payne, and P. Duerksen-Hughes, "Human papillomavirus type 16 E6* induces oxidative stress and DNA damage," *J Virol*, vol. 88, pp. 6751-61, Jun 2014.

[270] S. G. Hwang, D. Lee, J. Kim, T. Seo, and J. Choe, "Human papillomavirus type 16 E7 binds to E2F1 and activates E2F1-driven transcription in a retinoblastoma protein-independent manner," *J Biol Chem*, vol. 277, pp. 2923-30, Jan 25 2002.

[271] Y. C. Yang, T. Y. Chang, T. C. Chen, W. S. Lin, S. C. Chang, and Y. J. Lee, "ITPR3 gene haplotype is associated with cervical squamous cell carcinoma risk in Taiwanese women," *Oncotarget*, vol. 8, pp. 10085-10090, Feb 7 2017.

APPENDICES

Appendices 4.1: Biophysical characterisation of E2F1 amino acid variant by align GVGD

S.No	IDs	Variants	Aligned GVGD		
			GV	GD	Prediction
1	rs574523664	C12S	0.00	111.67	Class C65
2	rs13041707	A55S	0.00	99.13	Class C65
3	rs13041706	A56S	0.00	99.13	Class C65
4	rs757413708	K89R	0.00	26.00	Class C25
5	rs560305222	D93E	0.00	44.60	Class C35
6	rs758760049	D97H	0.00	81.24	Class C65
7	COSM3545490	H98Y	0.00	83.33	Class C65
8	rs753190783	Y100F	0.00	21.61	Class C15
9	rs145741678	A102T	0.00	58.02	Class C55
10	rs577941795	E103D	0.00	44.60	Class C35
11	rs767288307	P107R	0.00	102.71	Class C65
12	rs149803612	R109W	0.00	101.29	Class C65
13	rs199740633	R109Q	0.00	42.81	Class C35
14	COSM3545488	G110S	0.00	55.27	Class C55
15	rs768373441	R113H	0.00	28.82	Class C25
16	rs774045899	R113C	0.00	179.53	Class C65
17	COSM1713365	S121F	0.00	154.81	Class C65
18	rs200324089	P122Q	0.00	75.14	Class C65
19	rs200324089	P122L	0.00	97.78	Class C65
20	rs748471303	E124G	0.00	97.85	Class C65
21	COSM244178	R127H	0.00	28.82	Class C25

22	rs779392416	T130S	0.00	57.75	Class C55
23	COSM	S131E	0.00	79.79	Class C65
24	rs755388533	T135I	0.00	89.28	Class C65
25	TCGA	R138C	0.00	179.53	Class C65
26	COSM1025973	S144R	0.00	109.21	Class C65
27	rs751015434	A147V	0.00	64.43	Class C55
28	COSM724002	D148N	0.00	23.01	Class C15
29	rs752291306	V151A	0.00	64.43	Class C55
30	rs758127423	V151I	0.00	29.61	Class C25
31	rs759306871	E158K	0.00	56.87	Class C55
32	rs371120550	V159M	0.00	21.52	Class C15
33	COSM577076	R165W	0.00	101.29	Class C65
34	COSM443676	R165Q	0.00	42.81	Class C35
35	COSM1025972	R166H	0.00	28.82	Class C25
36	COSM443675	Y168C	0.00	193.72	Class C65
37	rs377425346	I180V	0.00	29.61	Class C25
38	COSM1166737	I188L	0.00	4.86	Class C0
39	TCGA	Q189E	0.00	29.27	Class C25
40	rs776055929	G192A	0.00	60.00	Class C55
41	rs770049550	H194Y	0.00	83.33	Class C65
42	rs771638059	V199I	0.00	29.61	Class C25
43	rs754564586	G200D	0.00	93.77	Class C65
44	rs35385772	G200S	0.00	55.27	Class C55
45	rs753472103	G201R	0.00	125.13	Class C65
46	rs377637237	R202W	0.00	101.29	Class C65
47	rs373902244	R202Q	0.00	42.81	Class C35

48	rs546049774	L203H	0.00	98.69	Class C65
49	rs750422437	G205E	0.00	97.85	Class C65
50	rs767306081	R211Q	0.00	42.81	Class C35
51	rs200658903	Q212L	0.00	112.44	Class C65
52	rs764256438	S216N	0.00	46.24	Class C45
53	rs746424387	E217K	0.00	56.87	Class C55
54	rs199745489	M224I	0.00	10.12	Class C0
55	rs759773652	N225D	0.00	23.01	Class C15
56	COSM724004	C227F	0.00	204.39	Class C65
57	rs771273717	T229M	0.00	81.04	Class C65
58	rs771273717	T229K	0.00	77.74	Class C65
59	rs201920937	Q230H	0.00	24.08	Class C15
60	rs531113927	R232C	0.00	179.53	Class C65
61	IntOGen	R232L	0.00	101.88	Class C65
62	COSM310784	R232H	0.00	28.82	Class C25
63	rs745608299	R242H	0.00	28.82	Class C25
64	COSM1025970	V246A	0.00	64.43	Class C55
65	rs376645897	Q249E	0.00	29.27	Class C25
66	COSM122751	R252C	0.00	179.53	Class C65
67	rs3213172	R252H	0.00	28.82	Class C25
68	rs373383167	I254T	0.00	89.28	Class C65
69	rs758585438	D256H	0.00	81.24	Class C65
70	rs765506787	V262I	0.00	29.61	Class C25
71	rs755210188	K266R	0.00	26.00	Class C25
72	rs541956472	Q272H	0.00	24.08	Class C15
73	rs3213173	V276M	0.00	21.52	Class C15

74	rs183250565	S279L	0.00	144.08	Class C65
75	rs767969405	S279P	0.00	73.35	Class C65
76	COSM1025969	E280Q	0.00	29.27	Class C25
77	COSM3545486	D294N	0.00	23.01	Class C15
78	rs752075116	D294E	0.00	44.60	Class C35
79	COSM1025967	V295D	0.00	152.01	Class C65
80	rs574956843	C298Y	0.00	193.72	Class C65
81	COSM1025966	P299H	0.00	76.28	Class C65
82	rs770643353	V303I	0.00	29.61	Class C25
83	COSM577077	I306V	0.00	29.61	Class C25
84	rs773197012	P308S	0.00	73.35	Class C65
85	rs535573433	G309E	0.00	97.85	Class C65
86	rs3213174	T311N	0.00	64.77	Class C55
87	COSM225412	P312L	0.00	97.78	Class C65
88	rs768640915	S313P	0.00	73.35	Class C65
89	COSM1713364	S318F	0.00	154.81	Class C65
90	COSM1025965	E321D	0.00	44.60	Class C35
91	rs144481343	R323K	0.00	26.00	Class C25
92	rs553413048	A324T	0.00	58.02	Class C55
93	COSM310785	I330T	0.00	89.28	Class C65
94	COSM230404	P334L	0.00	97.78	Class C65
95	rs750641087	P334Q	0.00	75.14	Class C65
96	rs149272498	P335A	0.00	26.87	Class C25
97	rs757620294	P338S	0.00	73.35	Class C65
98	rs764626563	P339L	0.00	97.78	Class C65
99	rs752029737	P339A	0.00	26.87	Class C25

100	COSM1713363	S341F	0.00	154.81	Class C65
101	rs564825532	T343I	0.00	89.28	Class C65
102	rs765976516	T343P	0.00	37.56	Class C35
103	rs771807428	T344R	0.00	70.97	Class C65
104	rs761616906	D345G	0.00	93.77	Class C65
105	COSM3545484	P346S	0.00	73.35	Class C65
106	COSM3545485	P346L	0.00	97.78	Class C65
107	COSM350181	L350V	0.00	31.78	Class C25
108	rs745908531	P357L	0.00	97.78	Class C65
109	rs778271263	R361Q	0.00	42.81	Class C35
110	rs747187347	R361W	0.00	101.29	Class C65
111	COSM1307298	G363R	0.00	125.13	Class C65
112	rs755713339	R366Q	0.00	42.81	Class C35
113	rs779509637	R366W	0.00	101.29	Class C65
114	rs367723468	V369M	0.00	21.52	Class C15
115	rs373587139	E371K	0.00	56.87	Class C55
116	rs373587189	R373L	0.00	101.88	Class C65
117	rs371155232	V378M	0.00	21.52	Class C15
118	rs371155232	V378L	0.00	31.78	Class C25
119	rs770915873	A380V	0.00	64.43	Class C55
120	rs773280034	S382L	0.00	144.08	Class C65
121	rs748480309	E385D	0.00	44.60	Class C35
122	rs769226220	R388L	0.00	101.88	Class C65
123	rs779566653	R388W	0.00	101.29	Class C65
124	COSM3911138	E389K	0.00	56.87	Class C55
125	rs760692854	S392Y	0.00	143.11	Class C65

126	rs3213176	G393S	0.00	55.27	Class C55
127	rs751247203	E398K	0.00	56.87	Class C55
128	rs553240434	H406Y	0.00	83.33	Class C65
129	rs373228968	H406Q	0.00	24.08	Class C15
130	rs373228968	H406D	0.00	81.24	Class C65
131	rs373228968	H406N	0.00	68.35	Class C65
132	rs760618502	D410N	0.00	23.01	Class C15
133	rs535111025	F413L	0.00	21.82	Class C15
134	rs772350151	L415F	0.00	21.82	Class C15
135	COSM2757808	E416K	0.00	56.87	Class C55
136	rs774912473	G418S	0.00	55.27	Class C55
137	COSM3963493	E419D	0.00	44.60	Class C35
138	rs749840514	G420D	0.00	93.77	Class C65
139	rs777420006	D426H	0.00	81.24	Class C65
140	rs777420006	D426N	0.00	23.01	Class C15
141	rs151195835	T433I	0.00	89.28	Class C65
142	rs267605891	P434L	0.00	97.78	Class C65
143	COSM1532962	F437L	0.00	21.82	Class C15

*GD Grantham deviation (GD) ,*GV Grantham variation

Appendices 4.2: Summary of the scores obtained from the different computational tools for E2F1 nsSNPs

S.No	IDs	Variants	SIFT	Polyphen-2	PhD-SNP	SNAP	SNPs&GO	PROVEAN	I-MUTANT	SuSpect	Fathmm
1	rs574523664	C12S	1	0	N	N	**D**	0.198	**-1.73**	37	1.66
2	rs13041707	A55S	0.43	0.024	N	N	N	0.149	-0.33	12	1.47
3	rs13041706	A56S	0.82	0.017	N	N	N	0.096	-0.37	14	1.49
4	rs757413708	K89R	0	0.999	N	N	N	-2.509	0.07	34	2.60
5	rs560305222	D93E	0.25	0.21	N	N	N	1.431	-0.14	17	1.53
6	rs758760049	D97H	0.01	0.993	N	N	N	**-2.813**	**-0.98**	4	3.12
7	COSM3545490	H98Y	**0**	**0.968**	N	**E**	**D**	-0.968	0.72	21	1.42
8	rs753190783	Y100F	0.06	**0.976**	N	**E**	N	-2.187	-0.81	23	3.00
9	rs145741678	A102T	0.13	0.564	N	N	N	-0.547	**-0.98**	20	1.48
10	rs577941795	E103D	0.28	0.031	N	**E**	N	0.579	-0.14	12	1.45
11	rs767288307	P107R	0.03	0.434	N	N	N	-1.462	-0.48	37	3.00
12	rs149803612	R109W	**0.01**	0.066	**D**	**E**	**D**	**-3.588**	-032	26	1.37
13	rs199740633	R109Q	**0.03**	**0.842**	N	**N**	**N**	-1.349	-0.69	18	1.46
14	COSM3545488	G110S	**0.02**	**0.573**	**D**	N	N	-2.141	-0.02	41	1.43
15	rs768373441	R113H	0.01	0.95	N	N	N	-1.998	-0.07	33	2.95

16	rs774045899	R113C	0	0.95	N	N	N	-3.284	-0.60	28	2.90
17	COSM1713365	S121F	**0**	**0.977**	**D**	**E**	N	**-5.179**	0.16	**62**	1.16
18	rs200324089	P122Q	**0**	**0.979**	N	**E**	**D**	**-6.730**	**-1.65**	**59**	1.19
19	rs200324089	P122L	**0.01**	0.141	**D**	**E**	N	**-3.284**	-0.51	**62**	2.75
20	rs748471303	E124G	**0**	**0.999**	N	**E**	N	**-6.182**	-2.14	34	2.85
21	COSM244178	R127H	**0**	**1**	**D**	**E**	**D**	**-4.712**	**-0.66**	**87**	**-2.34**
22	rs779392416	T130S	**0**	**0.998**	N	N	N	**-3.769**	-0.36	**60**	**-1.99**
23	COSM	S131E	NA	NA	N	**E**	**D**	**-3.012**	-0.15	**92**	**-2.68**
24	rs755388533	T135I	0.29	**0.999**	**D**	N	N	0.577	-0.09	**60**	**-1.90**
25	TCGA	R138C	NA	NA	**D**	**E**	**D**	**-5.665**	**-0.70**	**95**	**-2.14**
26	COSM1025973	S144R	**0**	**0.955**	**D**	N	**D**	-2.424	**-0.94**	37	**-2.03**
27	rs751015434	A147V	**0.01**	0.686	N	N	N	-1.465	0.44	34	-2.04
28	COSM724002	D148N	**0**	**0.966**	N	E	**D**	**-7.672**	**-1.48**	37	**-2.08**
29	rs752291306	V151A	**0**	**0.994**	**D**	**E**	N	0.207	-2.54	**91**	**-2.05**
30	rs758127423	V151I	**0.01**	**0.992**	N	N	N	-2.620	-1.06	**62**	**-2.29**
31	rs759306871	E158K	0	0.49	**D**	**E**	N	3.468	-0.85	**63**	**-1.92**
32	rs371120550	V159M	**0**	**0.981**	N	N	**D**	**-7.672**	**-1.25**	**63**	**-2.04**

33	COSM577076	R165W	**0**	**1**	N	**E**	**D**	**-4.795**	**-0.40**	**93**	**-4.30**
34	COSM443676	R165Q	**0**	**1**	N	**E**	**D**	**-8.631**	**-0.56**	**78**	**-4.28**
35	COSM1025972	R166H	**0**	**1**	N	**E**	**D**	**-0.959**	**-1.98**	**97**	**-4.29**
36	COSM443675	Y168C	**0**	**1**	**D**	**E**	**D**	**-1.918**	**0.78**	**93**	**-4.21**
37	rs377425346	I180V	**0**	**0.987**	N	N	N	**-4.454**	**-0.56**	**88**	**-2.33**
38	COSM1166737	I188L	**0.01**	**0.649**	**D**	**E**	**D**	**-3.588**	**0.01**	**84**	**-2.26**
39	TCGA	Q189E	NA	**0.78**	**N**	**E**	**D**	**-5.520**	**0.99**	**71**	**-2.02**
40	rs776055929	G192A	**0.02**	**0.999**	**D**	**E**	**D**	**-4.111**	**-0.71**	**63**	**-2.87**
41	rs770049550	H194Y	0.07	0.222	N	**E**	N	-0.751	-0.24	**60**	3.10
42	rs771638059	V199I	0.64	0.001	N	N	N	-0.154	-0.78	6	3.16
43	rs754564586	G200D	0.6	0.066	N	**E**	N	0.426	-0.24	22	3.10
44	rs35385772	G200S	0.96	0.002	N	**E**	N	0.875	-0.36	6	1.57
45	rs753472103	G201R	0.05	0.032	N	NA	N	-1.082	**-0.99**	12	3.13
46	rs377637237	R202W	**0**	**0.874**	N	**E**	**D**	**-4.454**	0.36	38	**-2.04**
47	rs373902244	R202Q	0.6	0.008	N	N	**D**	0.226	**-1.03**	17	**-1.62**
48	rs546049774	L203H	0.58	0.005	N	N	**D**	0.232	0.01	14	**-1.96**
49	rs750422437	G205E	0.82	0.01	N	N	N	-1.015	**1.00**	12	**-1.64**

50	rs767306081	R211Q	0.81	0.001	N	N	N	-0.007	**0.85**	6	**-1.78**
51	rs200658903	Q212L	**0**	0.046	N	N	**D**	**-3.811**	**-0.09**	32	**-1.82**
52	rs764256438	S216N	0.3	0.002	N	NA	N	-1.163	**0.77**	36	**-1.80**
53	rs746424387	E217K	**0**	**1**	N	**E**	N	**-3.618**	-0.25	**84**	**-4.32**
54	rs199745489	M224I	1	0	N	N	N	**-7.616**	**0.69**	6	**-1.29**
55	rs759773652	N225D	0.05	0.021	**D**	N	**D**	-0.815	-0.43	21	**-1.69**
56	COSM724004	C227F	**0**	**0.952**	**D**	**E**	**D**	**-71616**	-0.12	**67**	**-2.09**
57	rs771273717	T229M	0.02	0.579	N	N	N	-1.077	-0.28	22	**-1.81**
58	rs771273717	T229K	0.02	0.126	N	N	N	-2.172	**-1.06**	16	**-1.75**
59	rs201920937	Q230H	**0.01**	**0.909**	N	**E**	**D**	**-2.934**	**0.12**	**63**	**-2.21**
60	rs531113927	R232C	**0**	**0.802**	**D**	**E**	**D**	**-5.225**	**-0.58**	27	**-2.09**
61	IntOGen	R232L	NA	**0.500**	**D**	**E**	N	**-3.588**	**0.99**	33	**-2.06**
62	COSM310784	R232H	**0.01**	**0.012**	N	**E**	N	**-2.934**	**-0.68**	27	**-2.07**
63	rs745608299	R242H	**0.12**	**0.002**	**N**	**E**	N	-2.446	**-1.14**	16	**-2.66**
64	COSM1025970	V246A	**0.01**	**0.619**	**D**	**E**	**D**	**-4.666**	**-2.19**	**61**	**-2.89**
65	rs376645897	Q249E	**0.02**	**0.003**	N	N	**D**	**-2.573**	0.53	18	**-1.56**
66	COSM122751	R252C	**0**	**1**	**D**	N	**D**	**-7.516**	0.52	27	0.62

67	rs3213172	R252H	**0.01**	**0.952**	N	N	**D**	**-4.666**	**-1.50**	32	0.64
68	rs373383167	I254T	**0.09**	**0.145**	**D**	**E**	**D**	**-4.019**	**-4.22**	31	**-2.49**
69	rs758585438	D256H	0.01	0.957	N	**E**	N	-2.506	-0.66	43	**-1.73**
70	rs765506787	V262I	0.04	0.152	N	N	N	-0.799	0.05	45	**-2.33**
71	rs755210188	K266R	0.33	0.057	N	N	N	-2.047	-0.42	30	**-2.40**
72	rs541956472	Q272H	**0.01**	**0.894**	N	N	**D**	-2.133	**-1.34**	23	**-1.95**
73	rs3213173	V276M	0.1	0.013	N	**E**	**D**	-0.200	-0.24	20	**-1.80**
74	rs183250565	S279L	0.12	0.108	N	N	**D**	-1.759	0.11	18	**-1.77**
75	rs767969405	S279P	0.07	0.167	N	N	N	-1.858	-1.10	15	**-1.83**
76	COSM1025969	E280Q	**0.02**	0.036	**D**	**E**	**D**	-2.308	-0.38	27	**-2.51**
77	COSM3545486	D294N	**0**	**0.661**	N	**E**	**D**	**-2.692**	-0.42	46	**-2.54**
78	rs752075116	D294E	0.68	0.157	N	N	N	-2.446	0.62	34	-2.27
79	COSM1025967	V295D	**0**	**0.93**	**D**	**E**	**D**	**-6.359**	**-1.73**	**87**	**-3.59**
80	rs574956843	C298Y	**0**	**0.947**	**D**	**E**	**D**	**-9.888**	-0.03	**84**	**-2.12**
81	COSM1025966	P299H	**0**	**0.871**	N	N	D	-7.831	-1.44	29	-1.83
82	rs770643353	V303I	0.07	0.03	N	N	N	-0.165	-1.11	15	- 1.80
83	COSM577077	I306V	0.13	0.002	N	N	N	0.103	-0.47	17	2.10

84	rs773197012	P308S	0.01	0.191	N	**E**	N	**-2.883**	-0.76	27	1.22
85	rs535573433	G309E	1	0.002	N	N	**D**	0.066	**0.76**	19	1.08
86	rs3213174	T311N	**0.03**	0.098	N	N	N	0.346	0.02	17	2.03
87	COSM225412	P312L	0.06	0.009	N	N	N	**-2.744**	**0.65**	28	2.26
88	rs768640915	S313P	0.31	0.013	N	N	N	-1.359	-0.67	15	1.03
89	COSM1713364	S318F	**0.01**	0.418	N	**E**	**D**	-2.023	**0.92**	26	1.77
90	COSM1025965	E321D	0.56	0.012	N	N	N	0.208	-0.38	8	2.13
91	rs144481343	R323K	0.93	0.003	N	N	N	0.008	-0.36	16	2.10
92	rs553413048	A324T	0.42	0	N	N	N	0.443	-0.36	7	2.15
93	COSM310785	I330T	0.61	0	N	N	N	0.824	**-2.57**	6	2.09
94	COSM230404	P334L	0.34	0.009	N	N	N	1.266	**-0.90**	26	2.20
95	rs750641087	P334Q	0.26	0.029	N	N	N	-1.154	**-2.17**	18	1.14
96	rs149272498	P335A	0.18	0.045	N	N	N	-0.679	**-1.98**	23	2.10
97	rs757620294	P338S	0.28	0.006	N	N	N	0.030	**-1.74**	18	1.33
98	rs764626563	P339L	0.27	0.006	N	N	N	0.504	-0.52	18	1.17
99	rs752029737	P339A	0.52	0.002	N	N	N	-0.661	**-1.42**	15	1.15
100	COSM1713363	S341F	0.06	**0.635**	N	**E**	N	-1.937	-0.55	24	2.25

101	rs564825532	T343I	0.22	0.031	N	N	N	-1.365	0.19	20	2.11
102	rs765976516	T343P	0.16	0.33	N	N	N	-0.606	-0.21	16	1.13
103	rs771807428	T344R	0.4	0.028	N	N	N	-1.186	0.16	18	1.13
104	rs761616906	D345G	0.2	0.002	N	N	N	-1.834	-1.45	19	1.02
105	COSM3545484	P346S	0.15	0.027	N	N	N	-1.759	**-1.40**	24	2.17
106	COSM3545485	P346L	0.15	**0.597**	N	N	N	**-3.197**	**-0.64**	33	2.04
107	COSM350181	L350V	0.35	0.043	N	N	N	-0.288	-0.90	14	2.08
108	rs745908531	P357L	0.05	0.356	N	N	N	-2.290	0.09	31	1.12
109	rs778271263	R361Q	0.58	0.277	N	N	N	-1.376	-1.08	19	1.32
110	rs747187347	R361W	**0.02**	0.889	N	**E**	N	-0.082	-0.45	19	1.02
111	COSM1307298	G363R	**0.03**	0.439	N	N	**D**	-1.060	**-0.62**	20	2.06
112	rs755713339	R366Q	0.62	0.018	N	N	N	-0.053	**-1.22**	14	1.00
113	rs779509637	R366W	**0.01**	0.889	N	**E**	N	-1.826	-0.58	22	0.94
114	rs367723468	V369M	0.53	0.004	N	N	**D**	-0.038	**-2.87**	6	2.11
115	rs373587139	E371K	0.29	**0.588**	N	N	**D**	-0.935	**-0.57**	19	2.02
116	rs373587189	R373L	0.37	0.111	N	N	D	-0.819	-0.64	21	0.99
117	rs371155232	V378M	0.05	**0.669**	N	N	N	-0.532	**-2.49**	19	2.04

118	rs371155232	V378L	0.17	0.009	N	N	N	-0.781	**-1.99**	19	1.07
119	rs770915873	A380V	0.79	0.047	N	N	N	-0.654	**1.19**	19	1.09
120	rs773280034	S382L	0.47	0.111	N	N	N	-1.731	-0.36	25	0.95
121	rs748480309	E385D	0.19	0.009	N	N	N	-1.019	-0.45	17	1.04
122	rs769226220	R388L	0.15	0.235	N	N	N	-1.731	**-1.04**	27	1.01
123	rs779566653	R388W	0	0.834	N	**E**	N	-2.077	**-1.04**	38	0.98
124	COSM3911138	E389K	**0.02**	0.023	N	**E**	**D**	-1.857	-0.48	17	2.09
125	rs760692854	S392Y	0.15	0.187	N	N	N	-2.065	-0.08	22	0.88
126	rs3213176	G393S	1	0.01	N	**E**	N	0.489	-0.44	13	2.29
127	rs751247203	E398K	0.24	0.827	N	**E**	N	-1.059	-0.46	22	1.00
128	rs553240434	H406Y	1	0.33	N	N	**D**	0.254	**1.99**	24	2.05
129	rs373228968	H406Q	0.33	0.008	N	N	**D**	-0.311	-0.32	26	2.12
130	rs373228968	H406D	0.22	0.075	N	**E**	N	-1.379	-0.19	22	1.21
131	rs373228968	H406N	0.34	0.007	N	**E**	N	-0.834	**-1.69**	21	1.09
132	rs760618502	D410N	0	**0.95**	**D**	**E**	**D**	**-3.982**	0.09	**68**	1.08
133	rs535111025	F413L	0.26	**0.935**	**D**	**E**	N	**-2.305**	**-1.70**	**55**	1.18
134	rs772350151	L415F	0	**0.999**	**D**	**E**	**D**	**-3.332**	**-0.79**	**74**	1.03

135	COSM2757808	E416K	**0.03**	**0.917**	N	N	**D**	-1.835	-0.31	39	1.88
136	rs774912473	G418S	0.16	0.034	N	N	N	**-5.060**	-1.21	22	1.01
137	COSM3963493	E419D	**0**	0.333	**D**	**E**	**D**	**-2.496**	**-0.86**	**69**	2.01
138	rs749840514	G420D	**0**	**1**	**D**	**E**	**D**	**-5.952**	**-2.01**	**81**	1.07
139	rs777420006	D426H	**0**	**0.998**	**D**	**E**	**D**	**-5.105**	**-0.76**	**68**	0.98
140	rs777420006	D426N	**0.01**	**0.84**	**D**	**E**	**D**	**-3.696**	**-1.04**	**63**	1.02
141	rs151195835	T433I	**0**	**0.806**	**D**	N	**D**	-0.374	**-0.64**	30	2.05
142	rs267605891	P434L	**0**	**0.919**	N	N	**D**	-0.852	**-0.70**	36	2.03
143	COSM1532962	F437L	0.16	0.115	N	N	**D**	-0.882	**-2.87**	10	1.46

† nsSNPs ID s highlighted in bold were found to be deleterious by the computational prediction methods; N-Neutral; D-Disease; NA-Not available

Appendices 4.3: Molecular phenotyping of nsSNPs of E2F1 protein by SNPeffect 4.0

SNP ID	**Variant**	**SNP Effect**			
		TANGO	WALTZ	LIMBO	FoldX
COSM244178	R17H	No effect	No effect	No effect	No effect
rs755388533	T27I	No effect	No effect	No effect	No effect
COSM1025973	S34R	No effect	No effect	No effect	No effect
rs751015434	A36V	No effect	No effect	No effect	No effect
COSM724002	D38N	No effect	No effect	No effect	**Enhances the protein stability**
rs752291306	V40I	No effect	No effect	No effect	No effect
rs758127423	V40A	No effect	No effect	No effect	No effect
COSM443676	R56W	No effect	No effect	No effect	No effect
COSM577076	R56Q	No effect	No effect	No effect	No effect
COSM1025972	R57H	No effect	No effect	No effect	No effect
COSM443675	R59C	No effect	No effect	No effect	**Reduces the protein stability.**
rs377425346	I71V	No effect	No effect	No effect	**Reduces the protein stability.**
COSM1166737	I79L	No effect	No effect	No effect	slightly enhances the protein stability
rs377637237	R202W	No effect	No effect	No effect	slightly enhances the protein stability
rs373902244	R202Q	No effect	No effect	No effect	No effect
rs546049774	L203H	No effect	No effect	No effect	slightly enhances protein stability
rs767306081	R211Q	No effect	No effect	No effect	No effect
rs200658903	Q212L	No effect	No effect	**decrease**	No effect
rs764256438	S216N	No effect	No effect	No effect	No effect
rs746424387	E217K	No effect	No effect	No effect	No effect
rs199745489	M224I	No effect	No effect	No effect	No effect
COSM724004	C227F	No effect	No effect	No effect	**Enhances the protein stability**

rs771273717	T229M	No effect	No effect	No effect	No effect
rs771273717	T229K	No effect	No effect	No effect	No effect
rs201920937	Q230H	No effect	No effect	No effect	slightly reduce the protein stability
COSM4097723	R232C	No effect	No effect	**decrease**	**Reduces the protein stability**
COSM310784	R232H	No effect	No effect	**decrease**	slightly reduce the protein stability
COSM1025970	R242H	No effect	No effect	No effect	No effect
COSM1025970	V246A	No effect	No effect	No effect	No effect
rs376645897	Q249E	No effect	No effect	No effect	No effect
COSM122751	R252C	No effect	No effect	No effect	**Reduces the protein stability**
rs3213172	R252H	No effect	No effect	No effect	**Reduces the protein stability**
rs373383167	R254T	No effect	No effect	No effect	slightly reduce the protein stability
rs758585438	D256H	No effect	No effect	No effect	No effect
rs765506787	V262I	No effect	No effect	No effect	No effect
rs755210188	K266R	No effect	No effect	No effect	No effect
rs541956472	Q272H	No effect	No effect	No effect	**Reduces the protein stability.**
rs3213173	V276M	No effect	No effect	No effect	No effect
rs183250565	S279L	No effect	No effect	No effect	No effect
rs767969405	S279P	No effect	No effect	No effect	No effect
COSM1025969	E280Q	No effect	No effect	No effect	**Reduces the protein stability**
rs752075116	D294E	No effect	No effect	No effect	No effect
COSM3545486	D294N	**Increase**	No effect	No effect	No effect
COSM1025967	V295D	No effect	No effect	No effect	**Reduces the protein stability**
rs574956843	C298Y	No effect	No effect	No effect	No affect
COSM1025966	P299H	No effect	No effect	No effect	No effect

Appendices 4.4: structural and functional analysis for deleterious mutation identification on E2F1 protein

S.No	Mutation	Native	ANOLEA	SDM	ERIS	I-MUTANT 3.0	PROVEAN	SNPeffect 4.0				Polyphen 2.0	Mution Assessor	Consensus
								TANGO	WALTZ	LIMBO	FOLD X kcal/mol			
1	R17H (1CF7)	-0.28	-0.28	0.42	0.91	**-1.07/Decrease**	**-4.712/Deleterious**	No effect	No effect	No effect	0.06	**1**	**3.785/high**	4\11
2	T27I (1CF7)	-0.28	-0.19	0.32	-1.1	-0.41/Decrease	0.577/Neutral	No effect	No effect	No effect	-0.84	**0.999**	1.19/low	1\11
3	A36V (1CF7)	-0.28	-0.3	-0.07	-1.65	-0.39/Decrease	-1.465/Neutral	No effect	No effect	No effect	-0.22	0.686	0.895/medium	0\11
4	D38N (1CF7)	-0.25	-0.25	0.07	-0.52	-0.31/Increase	**-4.462/Deleterious**	No effect	No effect	No effect	-1.71	**0.966**	1.56/low	2\11
5	V40I (1CF7)	-0.28	-0.15	0.22	0.38	-0.21/Decrease	0.207/Neutral	No effect	No effect	No effect	-0.22	**0.992**	0.87/low	1\11
6	V40A (1CF7)	-0.28	-0.12	**-1.99**	**3.31**	-0.85/Decrease	-2.620/Deletrious	No effect	No effect	No effect	0.4	**0.994**	**2.915/medium**	4\11
7	R56W (1CF7)	-0.25	**-0.06**	**1.29**	**6.91**	-0.72/Decrease	**-7.672/Deleterious**	No effect	No effect	**Decrease**	0.06	**1**	**4.265/high**	**7\11**
8	R56Q (1CF7)	-0.25	**-0.2**	-0.07	-0.34	**-1.24/Decrease**	**-3.836/Deleterious**	No effect	No effect	No effect	0.06	**1**	**3.92/high**	4\11
9	R57H (1CF7)	-0.25	-0.12	-0.24	-0.75	**-1.26/Decrease**	**-4.795/Deleterious**	No effect	No effect	No effect	0.06	**1**	**4.27/high**	4\11
10	Y59C (1CF7)	-0.25	**-0.44**	**2.38**	**1.34**	-0.6/Decrease	**-8.631/Deleterious**	No effect	No effect	No effect	0.06	**1**	**3.92/high**	**6\11**
11	I71V (1CF7)	-0.25	-0.21	-0.83	0.14	**-1.14/Decrease**	-0.959/Neutrral	No effect	No effect	No effect	0.06	**0.987**	1.545/low	1\11
12	I79L (1CF7)	-0.25	-0.27	-0.59	**4.19**	-0.96/Decrease	-1.918/Neutral	No effect	No effect	**Increase**	0.06	0.649	**2.905/medium**	3\11
13	R202W	-2.29	**-1.86**	**1.19**	**5.82**	-0.27/Decrease	**-4.454/Deleterious**	No	No effect	No effect	-0.78	**0.874**	2.165/medi	5\11

	(2aze)							effect					um	
14	R202Q (2aze)	-2.29	-2.04	0.08	0.26	-0.60/Decrease	-0.226/Neutral	No effect	No effect	No effect	-0.48	0.008	1.125/low	0\11
15	L203H (2aze)	-2.29	-2.34	-0.7	1.8	**-2.01/Decrease**	0.232/Neutral	No effect	No effect	No effect	**2.33**	0.005	1.82/low	3\11
16	G205E (2aze)	-2.29	-2.43	**3.6**	-3.41	-0.31/Decrease	1.015/Neutral	No effect	No effect	No effect	-0.92	0.01	1.565/medium	1\11
17	R211Q (2aze)	-2.29	-2.3	-0.07	**1.84**	-0.30/Decrease	0.007/Neutral	No effect	No effect	No effect	-0.06	0.001	0.035/medium	1\11
18	Q212L (2aze)	-2.29	-2.19	-0.11	-3.08	0.45/Increase	**-3.811/Deleterious**	No effect	No effect	No effect	-0.52	0.046	1.12/low	1\11
19	S216N (2aze)	-2.29	-2.31	-0.06	-0.17	-0.37/Decrease	-1.163/Neutral	No effect	No effect	No effect	-0.25	0.001	1.555/medium	0\11
20	E217K (2aze)	-2.29	-2.31	-0.72	0.55	**-1.67/Decrease**	-3.618/Deleterious	No effect	No effect	No effect	-0.43	0.002	**3.26/medium**	2\11
21	M224I (2aze)	-2.29	-2.29	-0.84	-0.01	0.08/Decrease	1.887/Neutral	No effect	**Increase**	No effect	0.19	0	-1.365/neutral	1\11
22	C227F (2aze)	-2.29	-2.2	**-2.4**	**1.93**	0.13/Decrease	**-7.616/Deleterious**	No effect	No effect	No effect	**11.52**	**0.952**	**2.36/medium**	**6\11**
23	T229K (2aze)	-2.29	-2.34	0.32	**3.44**	**-1.26/Decrease**	-2.172/Neutral	No effect	No effect	No effect	0.5	0.126	2.25/low	2\11
24	T229M (2aze)	-2.29	-2.28	**2.1**	-2.03	-0.82/Decrease	-1.077/Neutral	No effect	No effect	No effect	-0.62	0.579	**2.595/medium**	2\11
25	Q230H (2aze)	-2.29	-2.28	-0.91	-2.09	**-1.68/Decrease**	-2.614/Deleterious	No effect	No effect	**Decrease**	**2.78**	**0.909**	**2.345/medium**	5\11
26	R232C (2aze)	-2.29	-2.25	0.16	0.98	-0.53/Decrease	**-5.225/Deleterious**	No effect	No effect	No effect	**1.26**	0.802	2.085/medium	2\11
27	R232H (2aze)	-2.29	-2.29	-0.49	**2.73**	-0.68/Decrease	**-2.934/Deleterious**	No effect	No effect	**Decrease**	0.62	0.012	1.88/low	3\11
28	R242H (2aze)	-2.29	-2.22	-0.61	**2.75**	**-1.26/Decrease**	-2.446/Neutral	No effect	No effect	No effect	0.37	0.002	**2.62/medium**	3\11
29	V246A	-2.29	-2.24	-1.62	**1.3**	**-2.10/Decrease**	**-3.588/Deleterious**	No	No effect	No effect	**2.98**	0.619	**2.93/mediu**	5\11

	(2aze)							effect					**m**	
30	Q249E (2aze)	-2.29	-2.31	-0.84	-1.53	0.02/Increase	-2.573/Deleterious	No effect	No effect	No effect	0.06	0.003	0.87/low	0\11
31	R252C (2aze)	-2.29	**-1.61**	0.67	**2.22**	-0.58/Decrease	**-7.516/Deleterious**	No effect	No effect	No effect	**1.44**	1	1.245/low	4\11
32	R252H (2aze)	-2.29	**-1.73**	0.45	**2.92**	**-1.02/Decrease**	**-4.666/Deleterious**	No effect	No effect	No effect	**3.93**	**0.952**	1.59/low	**6\11**
33	I254T (2aze)	-2.29	-2.31	0.17	0.57	**-2.20/Decrease**	**-4.019/Deleterious**	No effect	No effect	No effect	**1.85**	0.145	**2.24/mediu m**	4\11
34	D256H (2aze)	-2.29	-2.14	1.03	0.63	0.12/Decrease	-2.506/Deleterious	No effect	No effect	No effect	0.05	**0.957**	**2.56/mediu m**	3\11
35	V262I (2aze)	-2.29	-2.31	0.01	0.85	-0.97/Decrease	0.799/Neutral	No effect	No effect	No effect	-0.33	0.152	2.11/low	0\11
36	K266R (2aze)	-2.29	-2.95	0.15	-2.66	-0.12/Decrease	2.047/Neutral	No effect	No effect	No effect	0.08	0.057	1.67/low	0\11
37	Q272H (2aze)	-2.29	-2.22	-0.52	0.75	-0.87/Decrease	-2.133/Neutral	No effect	No effect	No effect	0.90	**0.894**	1.39/low	1\11
38	V276M (2aze)	-2.29	-2.29	0.38	-1.58	**-1.66/Decrease**	-0.200/Neutral	No effect	No effect	No effect	0.48	0.013	1.04/low	0\11
39	S279P (2aze)	-2.29	-2.36	0.03	-1.04	-0.92/Decrease	-2.047/Neutral	No effect	No effect	No effect	-0.83	0.167	2.135/low	0\11
40	S279L (2aze)	-2.29	-2.22	0.45	-1.65	0.09/Increase	-1.759/Neutral	No effect	No effect	No effect	-0.12	0.108	1.545/low	0\11
41	E280Q(2a ze)	-2.29	-2.18	-0.02	0.05	-0.79/Decrease	-2.308/Neutral	No effect	No effect	No effect	0.56	0.036	**2.35/mediu m**	1\11
42	D294E (2aze)	-2.29	-2.28	**2.15**	-0.91	0.06/Decrease	-1.858/Neutral	No effect	No effect	No effect	0.03	0.157	0.81/low	1\11
43	D294N (2aze)	-2.29	-2.29	0.64	-0.83	-0.47/Decrease	-2.692/Deleterious	**Increase**	No effect	No effect	0.04	0.066 1	1.7/low	1\11
44	V295D (2aze)	-2.29	-2.29	**-2.99**	**4.75**	**-2.14/Decrease**	**-6.359/Deleterious**	No effect	No effect	No effect	**3.62**	**0.93**	**2.755/medi um**	**7\11**
45	C298Y	-2.29	**1.14**	**-4.19**	**>10**	-0.02/Decrease	**-9.888/Deleterious**	No effect	**Increase**	**Decrease**	**2.41**	**0.947**	2.44/mediu m	**8\11**

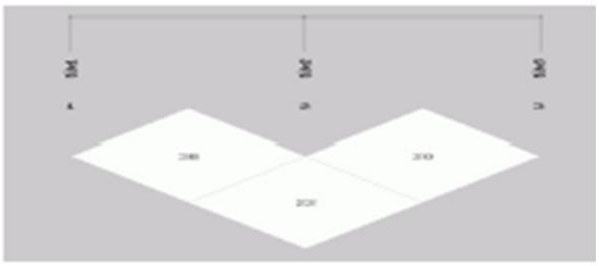
	(2aze)													
46	P299H (2aze)	-2.29	-2.25	**1.74**	-1.67	**-1.25/Decrease**	**-7.831/Deleterious**	No effect	No effect	No effect	0.29	**0.871**	1.965/medium	3\11
47	R17H (1CF7)	-0.28	-0.28	0.42	0.91	**-1.07/Decrease**	**-4.712/Deleterious**	No effect	No effect	No effect	0.06	**1**	**3.785/high**	4\11

Appendices 4.5: Linkage disequilibrium plot of cases and controls groups

230 controls & 163 cases observed

Linkage Disequilibrium tests

D': Site2 Site3
Site1 0.388 0.227
Site2 - 0.209

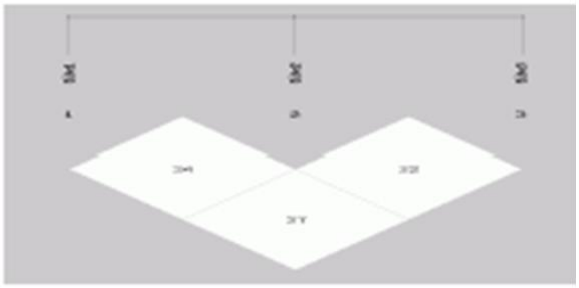


r2: Site2 Site3
Site1 0.123 0.030
Site2 - 0.021

230 controls & 142 cases observed

Linkage Disequilibrium tests

D': Site2 Site3
Site1 0.341 0.376
Site2 - 0.221

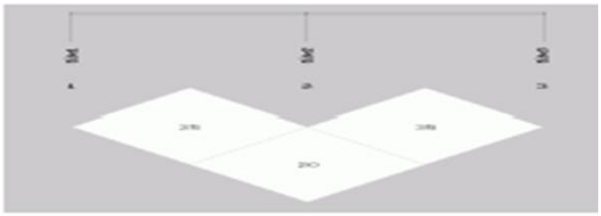


r2: Site2 Site3
Site1 0.066 0.065
Site2 - 0.013

91 controls & 90 case

Linkage Disequilibrium tests

D': Site2 Site3
Site1 0.259 0.202
Site2 - 0.352



r2: Site2 Site3
Site1 0.057 0.032
Site2 - 0.113

List of publications

- **Singh S**, Gupta M, Seam RK and Changotra H. E2F1 genetic variants and risk of cervical cancer in Indian women. *International Journal of biological Markers*. 2018; DOI: 10.1177/1724600818768459 **(I.F 1.456; SCOPUS)**
- **Singh S**, Gupta M, Sharma A, Seam RK and Changotra H. Non-synonymous polymorphisms Val276Met and Gly393Ser of *E2F1* gene are strongly associated with lung and head and neck cancers '' *Genetic Testing and Molecular Biomarkers* 2018; doi: 10.1089/gtmb.2018.0066. **(I.F 1.236; SCOPUS)**
- **Singh S**, Yennamalli RM, Gupta M, Seam RK and Changotra H. Identification of Genetic marker in *E2F1* Gene by a Computational approach predictive in Lung and Oral Cancer (due for submission)
- **Singh S**, S Kour, Gupta M, Seam RK and Changotra H. An intronic variant (rs2071054) of *E2F1*gene predisposes individuals to head and neck and cervical cancer (due for submission)

List of conferences

- **Singh S**, Seam RK, Gupta MK, Changotra H (2017). Association of G393S Variant of E2F1 gene with Lung and HNC cancer risk in North Indian Population. *Proceedings of the Punjab Science Congress* [20th: IET Bhaddal, Ropar, Punjab: 7-9 February 2017]
- **Singh S,** Seam RK, Gupta MK, Yennamalli RM, Changotra H (2017). Prediction of Genetic marker in E2F1 Gene by a Computational Approach. *Proceedings of the Annual Meeting of the Indian Society of Human Genetics & International Symposium on Trends in Human Genetic Research & Management* [42nd: Indian Institute of Sciences, Bangalore: March, 2017]
- **Singh S**, Seam RK, Gupta M, Changotra H (2018). Association of rs3213173 (C/T) polymorphism of E2F1 gene with lung cancer risk in Himachal Pradesh population. *Proceedings of the International Conference on Futuristic Trends in Network and Communication Technologies (FTNCT-2018)* [1st. : Jaypee University of Information Technology : 9-10 February 2018]