Computational and Molecular Analysis of TP53, PTEN and AR Genes in Prostate Adenocarcinoma

By

Mohammad Haroon Khan

DOCTOR OF PHILOSOPHY IN BIOINFORMATICS

DEPARTMENT OF BIOINFORMATICS & BIOSCIENCES
CAPITAL UNIVERSITY OF SCIENCE & TECHNOLOGY
ISLAMABAD, PAKISTAN
DECEMBER 2015

Computational and Molecular Analysis of TP53, PTEN and AR Genes in Prostate Adenocarcinoma

By

Mohammad Haroon Khan

A research thesis submitted to the Department of Bioinformatics & Biosciences,
Capital University of Science and Technology, Islamabad, Pakistan
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN BIOINFORMSTICS

DEPARTMENT OF BIOINFORMATICS & BIOSCIENCES
CAPITAL UNIVERSITY OF SCIENCE & TECHNOLOGY
ISLAMABAD, PAKISTAN
DECEMBER 2015

Computational and Molecular Analysis of TP53, PTEN and AR Genes in Prostate Adenocarcinoma

By

Mohammad Haroon Khan
(PI091001)

Prof. Dr. Jean-Baptiste Cazier
Chair Bioinformatics/Director, Centre for Computational Biology
University of Birmingham, UK

Dr. Christiana Kartsonaki
Senior Statistician
Nuffield Department of Population Health,
University of Oxford, UK

DEPARTMENT OF BIOINFORMATICS & BIOSCIENCES
CAPITAL UNIVERSITY OF SCIENCE & TECHNOLOGY
ISLAMABAD, PAKISTAN

DECEMBER 2015

DEDICATION

Dedicated to my family especially to my father because my Ph.D was his dream and he was always more concerned than me for the successful completion of my Ph.D.

Mohammad Haroon Khan

PUBLICATIONS

1.1 Impact factor publications:

1. **Khan MH**, Khalil A, Rashid H. 2015. Evaluation of the p53 Arg72Pro polymorphism and its association with cancer risk: a HuGE review and meta-analysis. Genet Res. 97: e7.
2. **Khan MH**, Rashid H, Mansoor Q, Hameed A, Ismail M. 2014. Association of rs1042522 Polymorphism with Increased Risk of Prostate Adenocarcinoma in the Pakistani Population and its HuGE Review. Asian Pac J Cancer P. 15(9): 3973-3980.

1.2 Non-impact factor publications:

1. **Khan MH**, Fazal S, Bano R**.** 2013. p53; A miracle if restored. Afr J Biotechnol. 12(24): 3745-3751.
2. **Khan MH,** Mansoor Q, Rashid H, Ismail M. 2012. The black face of TP53 in prostate cancer. Pak J Physiol. 2; 8(Suppl 1): 29.
3. **Khan MH,** Rashid H, Mir A. 2011. Phylogenetic analysis of human Tp53 gene using computational approach. Afr J Biotechnol. 10: 344-349.
4. **KhanMH,** Rashid H. 2011. Molecular Biology and computational analysis of P53 genetic alterations associated with prostate cancers in Pakistani population.10th Shaukat Khanum Memorial Cancer Symposium, 18-20th November, 2011. Lahore Pakistan: 79.

ACKNOWLEDGEMENT

I bow in deep reverence to “Almighty ALLAH”, the most merciful, most beneficent, the most adorable, deity and paramount to whom I am indebted. Allah who is omnipresent, omnipotent and omniscient and I bow my head before, “Almighty Allah” for paying my thanks.

This thesis represents not only my work at the keyboard, it is a milestone in more than half a decade of work at Capital University of Science and Technology (CUST), Islamabad and specifically within the Department of Bioinformatics and Biosciences. My experience at CUST has been nothing short of amazing. I have been given unique opportunities and taken advantage of them. It is my pleasure to take this opportunity to thank all those who have helped me directly or indirectly in bringing up this thesis. The work in hand is not the achievement of my sole efforts but is also the result of many experiences I have encountered at CUST from dozens of remarkable individuals who I also wish to acknowledge.

Words can’t express my sense of gratitude to my esteemed supervisor Professor Dr. Hamid Rashid, the Department of Bioinformatics and Biosciences, CUST for his constant guidance and inspiration. I am also grateful to Dr. Nayyer Massod Dar, HoD Computer Sciences, and my co-supervisor for always being so polite and kind to me.

I owe a sincere debt of gratitude to Dr. Muhammad Ismail, Director Institute of Biomedical and Genetic Engineering (IB&GE), Islamabad (External collaborator). He has been supportive since the days I began working in his lab. I remember he used to say something like "you’re always the first one in working on that project, you can do it!" to encourage me. Ever since, Dr. Ismail has supported me not only by providing research support, but also emotionally through the rough road to finish this thesis. I also wish to put my immense gratitude to Mr. Qaiser Mansoor (SSO, IB&GE), my associate advisor for his unconditional help and support during my stay in IB&GE. He taught me advanced molecular biology techniques and refined my basics. I am also thankful to all the other staff of IB&GE specially Mrs. Amara Javed and Mrs. Noureen Qais who helped me a lot to make the work possible.

I would like to acknowledge the scientists and staff members specially Mr. Haider and Mr. Zahoor of NORI hospital Islamabad for allowing me to the hospital facility and extending their kind help whenever required during the process of sample collection.

I place on record my special sense of gratitude to Capital University of Science and Technology in general and Prof. Dr. Muhammad Mansoor Ahmed (VC CUST) in specific for providing me with faculty development scholarship to pursue this degree. I am also thankful to Dr. Aamer Nadeem (HoD Bioinformatics and Biosciences), Dr. Shaukat Iqbal Malik (Associate Professor, Bioinformatics and Biosciences), Mr. Shahid Hussain, Mr. Muhammad Tahir Khan, Mr. Muhammad Aadil-ur-Rehman and others for their moral support and wholehearted cooperation.

Words cannot express my deepest thanks and gratitude to my parents who stood behind me in all the sweet and sour moments of life and were always there to pat my shoulder affectionately. Their love, patience and sacrifice are vested in every page of

this study. Whenever I felt tired and sense of despair gripped me, I always found them standing by my side, providing me words of immense consolation. My father was specially awaiting the day when I would be finished. Most of all, I take the opportunity to thank my younger brother, my brother Jani, Mr. Mohammad Ishaq Khan (Advocate), My sisters Ms. Robina Aasia and Ms. Laila for their grace and everlasting encouragement and supports. Their enthusiasms and heart-warming welcome will always remain fresh in my memory. My acknowledgment would be incomplete if I don't offer special attention to my younger sister Ms. Safia Motamin (Late), she was a treasure of care, love and prayers for our whole family. May her sour rest in peace and we will never forget her. I would also like to pay tribute to my grandmother who is always a source of inspiration for me.

A very special chapter is dedicated to my life partner, Mrs. Raisa Bano. I want to thank her for always standing by me everywhere. She waited up many late nights for my arrival after working late on my dissertation. It is hard to imagine having made it to this point without her. Last but not the least, I wish to thank my angel daughters Ms. Haya Haroon and Ms. Safia Haroon, a source of unending joy and love, who suffered a lot and were ignored during this period.

At last I wish to thank many other people whose names are not mentioned here but this does not mean that I have forgotten their help and support.

Mohammad Haroon Khan

DECLARATION

It is declared that this is an original piece of my own work, except where otherwise acknowledged in text and references. This work has not been submitted in any form for another degree or diploma at any university or other institution for tertiary education and shall not be submitted by me in future for obtaining any degree from this or any other University or Institution.

Mohammad Haroon Khan

Reg. No. PI091001

December 2015

ABSTRACT

Prostate adenocarcinoma (PCa) is one among the most common global disorders affecting men and is a leading cause of cancer related mortality. In the present study, TP53, PTEN and AR genes in 680 histologically confirmed prostate cancer patients and 500 controls from Pakistan were investigated to unravel their role in prostate cancer. Exon specific primers were designed for all the exons of the three genes along with flanking intronic and UTR regions. Sequence changes were initially screened by PCR-SSCP and then confirmed through automated sequencing. Our data showed that none of the samples had a complete deletion of any of the three genes. A total of twenty six novel exonic mutations and eight novel intronic variants were detected in the target genes along with the four previously reported exonic alterations. The intronic variants were observed both in cases and controls and thus were further investigated for their possible association with the disease through un-conditional logistic regression under different genetic models. The g.7675016T>A variant of TP53 showed significant association with increased PCa risk under allelic contrast (OR=1.84, 95%CI=1.29-2.63, p-value=0.001), heterozygous model (OR=2.25, 95%CI=1.37-3.69, p-value=0.001) and dominant model (OR=2.13, 95%CI=1.38-3.28, p-value=0.001) while its g.7674991T>A variant showed positive association in all genetic models except TT vs TC and recessive model. In our studied population, we observed non-significant association for g.87891382G>A variant of PTEN in all models except allelic contrast (OR=1.87, 95%CI=1.36-2.58, p-value=0.000), while the other three variants of PTEN showed positive correlation with increased PCa risk. Similarly, both the intronic variants of AR gene were also found associated with increased PCa risk. The AA genotypes of g.67637091T>A variant and CC of g.67724021T>C variant are more common in cases. Homozygous individuals for A allele has 3.68 times higher risk of prostate cancer.

Majority of the mutations are novel and thus establishing their pathogenicity is of prime diagnostic importance. A comprehensive structural and functional annotation was therefore reported in the present study for the observed mutations through the application of bioinformatics approaches. It was noticed that 5/8 of TP53 mutations (P152A, T170M, E171R, H179Q and V203G), 7/9 of PTEN mutations (A86P, E91K, H93Q, Q97H, E99X, H272F and E288F), while 11/13 of the AR gene mutations

(I870fs, I870_splice, D880Y, E884X, I900V, S909C, K906N, L908P, V912G, I915F and Y916S) were predicted to endorse changes in their respective protein structure and thus are functionally damaging.

Phenotypic data was collected in integration with the genotypic data from all the study participants and was statistically evaluated for their correlation and impact on individual's survival. Age, smoking CVDs, BMI, dairy products consumption, physical activity, family history and hypertension was observed responsible for increasing prostate cancer risk under uni-variate and multi-variate models. The diabetic patients were observed at lower risk of having PCa while others risk factors were found non-associated. Overall survival analysis showed a median survival time of 21 months for the PCa patients (95%CI= 18-26). Intra-group differences in patient's survival after adjusting for confounder age were confirmed through a proportional hazard model. It was observed that age, BMI, smoking, prostectomy and physical activities are associated with survival probability while the rest of risk factors have no effect on survival. It was further noticed that diabetic patients has better survival length as compared to the non-diabetic counterparts. Our results elucidate the significance of combining molecular and *in silico* approaches to fully distinguish pathogenic mutations from benign which will have profound effects on the patient management.

We also have developed an integrated platform, pakprostate.com during the course of study to provide users the facility to search literature, genes, drugs, to have access to download verified dataset, can upload data, can apply basic stats and visualize data just on click of a button. The repository pakprostate is freely available online at the URL www.pakprostate.com.

Case-control samples were collected from all the four provinces of Pakistan including Capital territory, Gilgit-baltistan and Kashmir. It is thus hoped that, this study is a good representation of Pakistani population.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AKT1	Gene encoding RAC-alpha serine/threonine-protein kinase
APS	Ammonium persulphate
AR	Androgen Receptor
ARE	Androgen Response Elements
BC	Betweenness centrality
BMI	Body Mass Index
CC	Closeness centrality
CHEK2	Checkpoint kinase 2
CpG	Regions of DNA where a cytosine nucleotide is followed by a guanine nucleotide in the linear sequence of bases along its 5'→3' direction; 5'—C—phosphate—G—3' , that is, cytosine and guanine separated by only one phosphate
CUST	Capital University of Science and Technology
CV	Cardiovascular
CVD	Cardiovascular Diseases
DEG	Differentially Expressed Genes
DHQ	District Head Quarter Hospital
DJBP	DJ-1-Binding Protein,
E006AA	highly tumorigenic African American prostate cancer cell line
EDTA	Ethylenediaminetetraacetic aci
EFCAB6	EF-hand calcium-binding domain-containing protein 6
EGFR	Epidermal growth factor receptor
ER	Estrogen Receptor
ERG	Erythroblast transformation-specific (ETS) Related gene
FocA1	Forkhead box protein A1
GO	Gene Ontology
GPCR	G protein–coupled receptors
HGMD	Database of Human Gene Mutation Data
IB&GE	Institute of Biomedical & Genetic Engineering
IRNUM	Institute of Radiotheraphy and Nuclear Medicine
LCI	Lower 95% Confidence Interval
LOR	Log Odd Ratio

MDM2	Mouse Double Minute 2 Homolog
NCOA1	Nuclear receptor coactivator 1
NORI	Nuclear Medicine Oncology & Radiotherapy Institute
NR0B2	Nuclear receptor subfamily 0, group B, member 2
NR-RTK	Nuclear receptors (NR)-Receptor tyrosine kinase (RTK)
OR	Odd Ratio
PARK7	Parkinson disease protein 7
PCa	Prostate Cancer/Adenocarcinoma
PCR	Polymerase Chain Reaction
PELP1	Proline-, glutamic acid- and leucine-rich protein 1
PI3K	Phosphoinositide-3-kinase
PPI	Protein-protein interaction
PQBP1	Polyglutamine-binding protein 1
PRMT2	Protein Arginine Methyltransferase 2
PROVEAN	Protein Variation Effect Analyzer
PSA	Prostate-specific Antigen
PTEN	Phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase
RANBP9	RAN Binding Protein 9
RBAK	RB-Associated KRAB Zinc Finger
RFLP	Restriction Fragment Length Polymorphism
RNF14	Ring finger protein 14
RNF6	Ring Finger Protein (C3H2C3 Type) 6)
RR	Risk Ratio
RREB1	Ras Responsive Element Binding Protein 1
SE	Standard Error
SIFT	Sorting Intolerant from Tolerant
SLC30A9	Solute Carrier Family 30 (Zinc Transporter), Member 9
SNP	Single Nucleotide Polymorphism
SPDEF	SAM Pointed Domain Containing ETS Transcription Factor
SRA1	Steroid Receptor RNA Activator 1
SSCP	Single-strand conformation polymorphism
STE	Sodium Chloride-Tris-EDTA
TBE	Tris/Borate/EDTA

TGFB1I1	Transforming Growth Factor Beta 1 Induced Transcript 1
TMPRSS2	Transmembrane protease, serine 2
TP53	Tumor Protein p53
TRIM24	Tripartite Motif Containing 24
UBE2C	Ubiquitin-conjugating enzyme E2 C
UCI	Upper 95% Confidence Interval
USP26	Ubiquitin Specific Peptidase 26
Var	Variance
VC	Vice Chancellor
ZMIZ2	Zinc Finger Miz-Domain Containing 2
ZNF318	Zinc finger protein 318
95%CI	95% Confidence Interval

Chapter 1

INTRODUCTION

Cancer is a highly-proliferating disorder, with the involvement of cellular transformation, dysregulation of cell proliferation, apoptosis, invasion, angiogenesis, and metastasis (Ries et al., 2003, ElShamy and Duhe, 2013). The mortality rate differs in economically developed and the developing countries, leading cause in the former and second leading mortality cause in the latter. The cancer burden is rapidly increasing in the developing countries as a result of population aging and growth, along with the adoption of cancer oriented lifestyles, which includes indulging in smoking and westernized diets. Remaining physically inactive for long periods of time also contributes to cancer development (Jemal et al., 2011). Cancer research is in an exciting phase of its evolution (Khan et al., 2013) and it is now clear that the root cause of all types of cancers are genetic mutations (Stratton et al., 2009).

The most common form of cancer in adult males is Prostate adenocarcinoma (Hsing et al., 2000), the number is still significantly increasing and thus also increasing the level of morbidity and mortality among men around the world (Jemal et al., 2005). Globally, it ranks third in cancer incidence and sixth in relevant mortality and this figure is increasing rapidly as the population continues to grow over the age of 50 (Van Haute et al., 2010). In Pakistan, prostate adenocarcinoma is extremely common like the rest of the world, it ranks third most common cancer diagnosed in males with a ratio of approximately 7% of all malignancies (Ahmad et al., 2009, Bashir et al., 2014). Therefore, there is a great need to improve the standard of diagnosis, prevention and timely treatment of the prostate cancer.

1.3 Overview

The prostate is an exocrine gland, conserved in all male mammals. It is not essential for the survival, but plays an important role in sexual reproduction and thus its excision abolishes fertilization through the natural ways (Jemal et al., 2011). It is also a fibromuscular stroma organ, surrounds the urethra just below the bladder. Together

with the seminal vesicles, it produces the gross of seminal fluid, or ejaculate, released during the emission phase of copulation. Its secreted fluid contains citric acid, lipids, proteolytic enzymes including PSA (Gsur et al., 2002), acid phosphatase etc., which are necessary for the motility and viability of sperm (Pennefather et al., 2000, Kavoussi and Wein, 2007). Development of the prostate gland depends on androgen and its early events are regulated through the AR (androgen receptors) in mesenchymal cells (Prins and Putz, 2008). Other steroid receptors like ER and retinoid receptors also contribute to the differentiation and morphogenesis of the prostate (Prins et al., 2002).

Prostatic intraepithelial neoplasia (PIN), invasive cancer and androgen dependent or independent metastases are some of the defined states of Prostate cancer development (Scher and Heller, 2000). Human tumors are heterogeneous and evolve through dynamic genetic mutations and selections but unfortunately their underlying genetic mechanisms remains poorly understood. PCa is also dynamic and shows a wide range of clinical behaviors, starting from slow-growing tumors of no clinical significance to aggressively metastasizing, eventually leading to lethal disease. Human PCa consists of a mature luminal phenotype and is characterized by the expression of androgen receptor and PSA production (Roudier et al., 2003). Accumulation of somatic mutations in prostate progenitor cells is considered a major factor while determining the aggressiveness of PCa (Lapointe et al., 2007).

PTEN (phosphatase and tensin homolog) gene, a tumor suppressor, is also one of the most common inactivated, lost or mutated genes in human cancer including prostate cancer (Steck et al., 1997). Aberrations in PTEN accounts for approximately 70% of the prostate cancers (Sircar et al., 2009). It functions as a major regulator of multiple signaling pathways controlling cell proliferation, survival and size (Keniry and Parsons, 2008). Constitutive abnormal PI3K activation of the signaling pathway is an established consequence of PTEN inactivation which is responsible for the uncontrolled cell growth, proliferation, and survival (Stambolic et al., 1998).

Aggressive forms of human PCa is characterized by loss of PTEN and mutation of TP53 (Schlomm et al., 2008). TP53 is a tumor suppressor and encodes a transcription factor which is activated in response to different stresses on cellular level and enforces multiple anti-proliferative functions (Vogelstein et al., 2000). It is one of the

major cell cycle regulators preventing abnormal cell proliferation and guarantees genome maintenance following cellular stress (Bhattacharyya et al., 2000). Recent genomic profiling research has shown some significant number of copy number losses of TP53 which could be homozygous or heterozygous. Such losses were shown to be present in 21% of PCa cases with primary prostate cancers (Taylor et al., 2010). Somatic and germline TP53 alterations are frequent in most human cancers (Hainaut and Hollstein, 1999). However majority of the TP53 alterations are found out to be missense substitutions when compared to other tumor.

1.4 Statement of Problem

Modern cancer research mainly focuses on the identification of underlying biological mechanisms for the acquisition of motility and invasiveness in cancerous cells. It is generally believed that cancers usually involve alterations in different genes and multiple associated SNPs have also been reported. Genetic screening of different cancers is therefore essential to better understand the scope and magnitude of the mutations and the complexities created by them. The task can be challenging because presence of mutation are not always problematic and thus we also have to focus on the frequency and functional impact of the mutations. Furthermore, some mutations also have indirect effect, so effective measurement should include the interrelationships between different mutations and it is thus hypothesized to provide a foundation for better therapeutics against cancers.

1.5 Purpose of the research

Cancer is an overwhelming global problem for millions of people and their families who are forced to face it each year. Prostate cancer is male specific which starts its growth in the prostate gland situated in the male reproductive system. This is an exocrine gland located directly under the urinary bladder in front of the rectum. It also produces a fluid which is part of the semen, protecting and nourishing the sperm. As the urethra goes through the prostate, it is also involved in urine control. PCa is the second most frequently diagnosed cancer and the sixth leading cause of cancer death in males worldwide.

The aims of this dynamic study are,

1. Screening of different sub-ethnic groups of Pakistani population for TP53, PTEN and AR gene mutations
2. Screening of Pakistani population for prostate associated and population unique polymorphism in prostate adenocarcinoma
3. To computationally analyze the effect of observed genetic alterations on TP53, PTEN and AR hub proteins on the individual protein structure and function and also on the protein-protein interactions in the relevant pathways
4. To investigate the potential association of various socio-demographic and clinical factors with prostate cancer in Pakistani population

This research on genes linked to prostate cancer will help to better understand how prostate cancer develops, it will also help to provide answers to the genetic alterations leading to PCa.

1.6 Applications of the research

In majority of human cancers, mutations arise in a single or few cells circumvented by unaffected neighbors. Expansion of mutated cells can sanction the accumulation of supplemental mutations. This research will help to better understand how prostate cancer develops and also will provide answers to the genetic changes that lead to prostate cancer. Genetic and genetic-epidemiological research will be performed to identify and assess genetic factors involved in prostate cancer and will also provide a strong foundation to make it possible to design better and more effective therapeutic strategies.

Chapter 2

LITERATURE REVIEW

2.1 Prostate Cancer

Prostate cancer is the second most frequently diagnosed malignancy in men around the world (Jemal et al., 2011). Carter and his colleagues for the first time classify prostate cancer into three categories on the basis of on family history i.e. hereditary prostate cancer, familial and non-familial or sporadic prostate cancer (Carter et al., 1993). It is well established that, both the incidence and mortality rates of prostate cancer vary considerably among different countries and ethnicities. The highest rate been observed in Australia, New Zealand, Western Europe and Northern America (Xu et al., 2013).

Although, incidence and mortality rate of PCa is comparatively lower in Asia, but unfortunately rate of prostate cancer in most native Asian populations have gradually increased in recent decades (Ito, 2014). A number of risk factors may contribute to the incidence rates of prostate cancer, including genetic factors, aging, Western diet, etc. (Xu et al., 2013).

2.2 Genetic alterations

2.2.1 TP53

The p53 codon 72 polymorphism, either independently or in integration with p21 C98A and C70T polymorphisms, alters prostate cancer risk in Slovak population. p53 and p21 genes were gontyped through PCR-RFLP in 300 prostate cancer patients and 446 healthy controls and it was established that none of the three SNPs were correlated individually with the prostate cancer risk. On the contrary, interaction between the SNPs of both the genes collectively decreased the risk of having prostate cancer with the odd ratio of 0.49(0.27-0.86) for cases having the p53 codon 72 Arg/ Pro+Pro/Pro and p21 C98A CA genotypes. Genotypes of p53 and p21 showed

non-significant differences in Gleason score and PSA levels (P>0.05) (Sivonova et al., 2015).

The association of p53 P72R polymorphism with prostate cancer was studied in a meta-analysis to evaluate the risk by using meta-regression, Galbraith plots, subgroup and sensitivity analysis. The study was based upon 17 case-control studies including 2,371 cases and 2,854 controls. Their results showed no association between p53 codon 72 polymorphism and PCa risk in all genetic models. When limiting the analysis to the studies following the Hardy-Weinberg equilibrium, significant association was observed in the pooled analysis in Caucasian population in co-dominant model (OR = 1.57, 95% CI = 1.08-2.28, P = 0.017) and recessive model (OR = 1.60, 95% CI = 1.12-2.27, P = 0.009). In subgroup analysis stratified by PCa stages and Gleason grades, a slight but significant association was found when advanced PCa was compared with localized PCa only in recessive model (OR = 1.51, 95% CI = 1.02-2.23, P = 0.039). This meta-analysis suggested that the Proline/Proline genotypes were associated with the increased risk of PCa among the Caucasians (Lu et al., 2014).

The transformed phenotypes resulting from the deletion of tumor suppressors PTEN and TP53 in prostate epithelium were characterized which described the histological and metastatic characteristics of primary tumors. The study showed that tumor heterogeneity resulted from the transformation of multi-potent progenitors (Martin et al., 2011).

Ecke and his co-workers measured the prognostic value of TP53 mutations and PSA in PCa patients by using ninety tissue samples of patients having radical prostatectomy. TP53 mutations were screened through temperature gradient gel electrophoresis (TGGE) via exon specific manner. In 35.6% patients, TP53 mutations were detected and it was observed that mutations in exon 7 and 8 mainly contribute to tumor progression in PCa (Ecke et al., 2010).

Zheng along with his co-workers screened samples of 84 primary PCa and detected nine somatic and two germline alterations of TP53 in the same group. They observed one novel germline mutation (c.A408T/p.Gln136His) and two novel somatic mutations (c.T1022G/p.Phe341Cys and c.108-109ins22/p.His37fsX13). It was further

observed that, TP53 and CHEK2 mutations were mutually substituted in these tumors in at least 25% of the cases (Zheng et al., 2006).

TP53 mutations can occur in about 1/3 of the early stage prostate cancer cases. Expression of HPV E6 or over expression of MDM2 contributes to the loss of TP53 function. High levels of TP53 mutation were observed in normal prostate tissues of prostate cancer patients. Prostatic intraepithelial neoplasia additionally accusing p53 mutations or loss as a primary event in tumorigenesis (Downing et al., 2003).

The pattern of TP53 mutations in primary PCa and distant metastases was studied using PCR-SSCP and sequencing methods to analyze TP53 exons 5-8 in 17 patients. Transitions of GC pair to AT were the most common (64%). 6/11 of the transitions were occurred at the CpG dinucleotides in five hot-spot codons (175, 245, 248, 273, and 282). These results suggest that specific mutations of TP53 play role in prostate cancer progression (Navone et al., 1999).

Another study characterized a group of 86 patients treated with radical prostatectomy to determine the potential role of TP53 inactivation in PCa. They explored the patterns of p53, mdm2 and p21/WAF1 expression by immune-histochemistry and found an association between p53-positive phenotypes and decreased time to PSA relapse ($P < 0.01$) in addition to p21-positive phenotype and high Ki67 proliferative index. They hypothesized that PCa pathway progression involves p53 inactivation due to mdm2 overexpression and summarized that mutations affecting the p53 pathway are common events in prostate cancer (Osman et al., 1999).

The differences in TP53 mutation spectra of PCa in between populations of Japan and Western countries were explored by a group of researchers. According to their study, PCa is more common neoplasm in USA and its frequency is comparatively low in Japan. Its frequency is increasing in Japan with the adaptation of western life style. They examined a series of 90 lesions through PCR-SSCP and found that six of the total was well, 34 were moderately and 50 were poorly differentiated adenocarcinomas. Median Gleason score was recorded as 7.9. 12% of the cases had mutations in exons 2-11. Ten substitutions in addition with an insertion were encountered. The greater proportion of transversions in the Japanese population

suggests that different factors are responsible for prostate adenocarcinoma in various countries (Watanabe et al., 1997).

TP53 mutations in human prostate adenocarcinoma were characterized to analyze alterations in exons 4-11 of TP53 mRNA in 44 prostate tissue samples and 4 metastatic lesions from PCa patients who had surgical resection. TP53 expression was analyzed by PCR, followed by confirmation through Southern blotting. Two overlapping regions i.e. exons 4-6 and 6-11 were analyzed through non-isotopic PCR-SSCP system. All the specimens with detected alteration through SSCP were sequenced bi-directionally to confirm the results. 6% of the specimen displayed nearly undetectable expression and 35% expressed mutated mRNA encoding amino acid substitutions within exons 4-11or deletions within the TP53 transcripts. The A>G or T>C transitions were the most frequent (Chi et al., 1994).

2.2.2 PTEN

The association between PTEN genomic deletion and a higher Gleason score or a higher possibility of capsular penetration was elucidated in a systematic. The overall meta-analysis demonstrated that PTEN genomic deletion was associated with a higher Gleason score (OR=0.319; 95%CI=0.153-0.666; P = 0.000) and a higher possibility of capsular penetration (OR=0.393; 95%CI=0.185-0.837; P = 0.015). None of the studies materially altered the original results and no evidence of publication bias was found (Wang and Dai, 2015).

Williams and his group analyzed the recurrent copy number alterations in PCa. Their study comprised 11 publications including 662 cases. They identified five genomic subgroups with majority of the samples having abnormal profiles with gains of 8q. The most common loss was 8p of NKX3.1. The distribution in other genomic subgroups was characterized by losses at 2q, 3p, 5q, 6q, 13q, 16q, 17p, 18q, and PTEN (10q), and acquisition of 21q deletions associated with the TMPRSS2-ERG fusion rearrangement. It was observed that PTEN deletion was adequate to enforce higher levels of copy number alterations. Furthermore, the overall ratios of the genomic alterations were significantly higher when PTEN was deleted (Williams et al., 2014). The inactivation of PTEN gene is heavily implicated in PCa while

up-regulation of the chemokine receptor 4 (CXCR4) is linked with the progression and metastasis of PCa. Loss of PTEN tumor suppressor function is mainly due to the genetic and epigenetic variations, as well as active site oxidation by ROS (Chetram et al., 2011).

The potential prognostic role of the combined variations in PTEN and EGFR, their prevalence, somatic mutations, amplification and expression was investigated in a series of prostate cancer trials. Eight EGFR and three PTEN mutations were detected in 98 and 92 prostate adenocarcinomas respectively. The mutual prevalence of both the genes was observed 11%. Mutations in EGFR–PTEN signaling pathway represent 1/3 of the prostate adenocarcinomas (de Muga et al., 2010).

The frequency of PTEN gene mutations in PCa has not been well studied in Asia. Fifty one Iranian patients with PCa were analyzed to scrutinize the role of PTEN aberrations in tumor progression. Formalin-fixed PCa specimens were used with the age of the patients ranging from 57 to 82 years. The PCR-SSCP analysis revealed band shifts in 6 tumors samples, two in exon 1, one in exon 2 and three in exon 5 (Pourmand et al., 2009).

Thirty two Chinese patients who were not diagnosed through the PSA test were analyzed to clarify the role of *PTEN* in primary PCa. Six cases from the American patients who died of prostate cancer were also analyzed. It is unknown whether the rate of *PTEN* mutations is different in PCa from Asian compared to Western men. Five of the 32 primary PCa from Chinese men and two of six metastases from American men displayed mutations in 10 codons of *PTEN*. It was concluded that mutations of *PTEN* occurs more often in primary PCa of Chinese population whose tumors are high grade (Dong et al., 2001).

2.2.3 AR gene

AR mutations are infrequent in organ confined tumors in Caucasian Americans, but occur at a higher rate in advanced stage. Somatic missense mutations were identified at a higher rate in African Americans (17/200) than in Caucasian Americans (2/100). In Caucasian Americans, the majority of these mutations (41.1%) were from Gleason 7 tumors. Germline AR mutations were also approximately 4 times higher in

Caucasian Americans. The expression of A-allele of E213 (G/A) polymorphism was 5.85 times higher in control samples of African Americans than in Caucasian Americans. Silencing of Ser597Gly somatic AR mutations in E006AA cell line revealed that analogous AR mutation can be associated simultaneously with both "gain-of-function" and a "loss-of-function" phenotype (Koochekpour et al., 2014).

Sun and Lee aimed to evaluate the effects of (CAG)n repeat polymorphisms on the prevalence of PCa in a meta-analysis included 47 studies comprising 13,346 cases and 15,172 controls. The study also includes 31 reports on Caucasians, ten on Asians, one on Hispanics and four on combined ethnic groups. The carriers of a shorter CAG repeat sequence had an increased risk of prostate cancer (OR=1.21, 95%CI=1.10-1.34) regardless of the length of CAG repeat. The risk of carrying a shorter CAG repeat was 1.10-1.83 times higher than carrying longer repeat sequence in Caucasians and Asians. Carriers of less than 22 repeats were observed having higher risk (OR= 1.16, 95%CI=1.04-1.29) than those having more than CAG repeat polymorphisms. This study proposes that shorter CAG repeat polymorphism increases prostate cancer risk, mostly observed in Asian and Caucasians (Sun and Lee, 2013). Study groups with more than 20, 22 and 23 CAG repeats were compared with carriers having less than 20, 22 or 23 repeats in another study. It was observed that carriers of 20, 22 or 23 repeats had 21% (P = 0.076), 5% (P = 0.508) and 5% (P = 0.681) reduced risk of PCa. After categorizing studies on the basis of geographic distribution, carriers of 20 repeats had 11% decreased risk in the USA population, 53% in Europe, and 20% in Asia (P > 0.05). Stratification by study designs showed no substantial changes in risk estimation (Gu et al., 2012).

The possibility of AR gene mutations in association with familial PCa in Afro-Americans was examined in sixty cases from thirty high risk Afro-American and Caucasian families through exon-specific PCR, bi-directional sequencing and RFLP genotyping. A germline substitution (A1675T/T559S) was identified in the DNA-binding domain in three cases. The said mutation contributes to the disease by modifying the DNA-binding affinity and/or its response to androgens. More research is desired to explain the incidence and role of A1675T allele in the early onset and familial prostate cancer in Afro-Americans (Hu et al., 2010). AR selectively up-regulates the genes governing M-phase of cell cycle in androgen-independent

cells. Epigenetic scripts at the UBE2C enhancer, explicitly histone H3K4 methylation and FoxA1 transcription factor binding are present in androgen-independent cells and direct AR-enhancer binding and UBE2C activation. It can thus be concluded that, the role of AR in androgen-independent cancer cells is not to guide the androgen-dependent gene expression without androgen, but to execute a different package resulting in androgen-independent growth (Lapointe et al., 2007).

Three novel splice variants of AR gene were identified which were lacking the AR3, AR4 and AR5 ligand-binding domains in hormone insensitive prostate cancer cells. Immuno-histochemistry analysis of 429 prostate samples in tissue microarrays showed that AR3 was significantly up-regulated during progression of PCa and its expression is associated with the tumor recurrence risk after radical prostatectomy. AR3 may also play a distinct but vital role in ablation-independent progression through the regulation of a unique set of genes, including AKT1(Guo et al., 2009).

Waltering and his colleagues showed through microarray-based transcript profiling and subsequent hierarchical clustering through unsupervised learning that, LNCaP-ARhi cells grouped with VCaP cells. These cells contain endogenous AR gene amplification and overexpression, indicating its crucial role in overall gene expression and regulation in prostate cancer. Their results confirms that, overexpression of AR gene sensitize castration-resistant PCa cells to the low levels of androgens (Waltering et al., 2009).

2.2.4 Computational Analysis

Hub proteins are considered key players in preserving the function and constancy of the protein-protein interaction (PPI) network. It is therefore essential to study PPIs from a structural perspective to unravel valuable information. PPI complexes of hub proteins of human NR-RTK network have already been predicted through comparative modeling and docking studies. The key role of Estrogen receptor-1 in signal transduction between human nuclear receptors and Receptor Tyrosine kinases have also been explored through structural analysis. The predicted method requires human interpolation and verdict, but can detect the interactions that could occur simultaneously (Choura and Rebai, 2011).

The evolutionary study of TP53 and MDM2 genes, to predict their role in tumorigenesis using *in silico* methods showed higher sequence similarity across the mammals. It indicates that these species probably share cancer causing mechanisms. They observed five clusters each in their individual unrooted trees. However, in contrast to MDM2, TP53 was found in a large number of species. It is therefore evident that, these genes play an important role in the process of tumorigenesis (Jayaraman et al., 2011).

High-throughput protein-protein interactions stored in public databases were studied in the context of metabolic networks. Topological distance based classification of reaction pairs revealed higher rate of enzyme-enzyme interactions (EEI) for directly neighbored reactions. Randomized networks were examined to determine the potential functional effects of these interactions. It was proposed that a functional significance of EEI could be established for those exhibiting low connectivity (Huthmacher et al., 2007).

A phylogenetic analysis of the TP53 coding sequences from 32 vertebrate species showed that directional selection can affect the codons and thus influence the efficiency of TP53 product binding to the DNA as a transcription factor. It showed that, the frequency of generative mutations statistically differs from the frequency of loss of function mutations in conserved regions (Pintus et al., 2006).

2.2.5 Risk Factors

Dairy products and calcium consumption in association with prostate cancer risk based on thirty-two cohort studies revealed its estimated Risk Ratios by using a random effects model, for total milk [1.03 (1.00-1.07), low-fat milk [1.06 (1.01-1.11) per 200 g/d], cheese [1.09 (1.02-1.18) per 50 g/d], and dietary calcium [1.05 (1.02-1.09). They found statistically significant association with increased PCa risk (Aune et al., 2015).

Davis and his group analyzed the prevalence of cardiac risk factors in men with localized prostate cancer undergoing androgen deprivation therapy in British Columbia, Canada. Data on CV risk factors and disease were collected and Framingham risk scores were calculated. The median age of the study cohort was 73 years. According to their study established CVD was present in 25% of patients.

Among patients without established CV disease, calculated Framingham risk was higher in 65%, intermediate in 33%, and low in 1%. Baseline hypertension was present in 58% of patients, dyslipidemia in 51%, and diabetes or impaired glucose tolerance in 24%. Hypertension was more prevalent in the study cohort than in an age- and sex-matched population sample (OR= 1.74, P = 0.006); diabetes had a similar prevalence (OR = 0.93, P = 0.8)(Davis et al., 2015).

Coffee consumption and prostate cancer risk was studied in a meta-analysis of prospective cohort studies including 8973 cases and 206,096 participants by using random effects model to compute the pooled risk. A pooled relative risk of PCa as OR=0.88 (0.82-0.95) was observed for regular coffee drinkers and thus it was concluded that coffee consumption is inversely associated with PCa risk (Cao et al., 2013).

The association of tobacco use and PCa mortality and incidence was investigated in a meta-analysis consisted of 51 articles having 11823 deaths, 50349 incident cases, and 4,082,606 cohort participants. Current smoking was observed positively associated with an enhanced risk of PCa death (OR=1.24; 95%CI=1.18-1.31). It was thus concluded that a modest but statistically significant association exists in between smoking and fatal PCa (Islami et al., 2014). Huncharek and co-workers assessed the correlation between smoking and PCa in a meta-analysis comprised of 24 cohort studies including 21579 cases. Relative risks and 95% CI were individually calculated for incidence and mortality. Current smokers were found non-significantly associated with increased risk of PCa incident (RR=1.04; 95%CI=0.87-1.24) in pooled analysis but were found significantly associated when data was stratified by amount of smoking. In comparison, former smokers were having an increased risk (RR= 1.09; 95%CI=1.02-1.16). Smokers with higher smoking ratios had a comparatively elevated risk of 24-30% of death from PCa than non-smokers (Huncharek et al., 2010).

The 8q24 (rs4242382) SNP was genotyped in Chinese men (335 PCa cases and 347 age-matched controls) by using polymerase chain reaction-high-resolution melting analysis. The association between minor allele and PCa risk along with clinical coverts were analyzed. In addition, they also performed a meta-analysis to define the association between PCa risk and the risk allele by using genotyping data of 1793

cases and 1864 controls from their own and previously published studies in American and European populations. The prevalence of risk allele was comparatively higher in cases (OR=0.222, P=7.3). The results thus suggesting that the polymorphism (rs4242382-A) was associated with disease risk in Chinese men in dominant genetic model (OR=2.03; 95%CI=1.42-2.91). The risk allele was also found associated with clinical covariates including age at diagnosis ≥65 years, PSA >10ng/ml, Gleason score <8, tumor stage and aggressive PCa. The study confirmed that there is a direct link in between rs4242382-A polymorphism and PCa risk (OR=1.62; 95%CI=1.39-1.88) across African American, Caucasian and Asian populations (Zhao et al., 2013).

Participants of the Health Professionals Follow-Up Study were scrutinized for any possible association between lethal prostate cancer and dairy product ingestion. The group comprised of 3918 men with localized PCa from 1986-2006 and followed to 2008. Total milk and dairy intake was observed not associated with a increased risk of lethal PCa. It was further stated that men with higher whole milk intake were at greater risk of progression (HR=2.15; 95%CI=1.28-3.60; $P_{(trend)} < 0.01$) (Pettersson et al., 2012). Another meta-analysis of Milk consumption as a risk factor for PCa in the United States including published studies between 1984 and 2003 showed their combined results as OR=1.68(1.34-2.12). Only little variation was observed after results stratification. It was concluded that there exists a positive association between milk consumption and PCa (Qin et al., 2004).

In King County, Washington, a population based case-control study of black and white men having confirmed PCa, ages ranged from 40-64 years (1993-1996) was conducted. The study included 753 cases from SEER cancer registry and 703 age-matched controls. Risk factors included in the study were medical history, sexual behavior and other potential PCa risk factors. They reported no association between sexual orientation and PCa. The risk directly increases with the lifetime number of female sexual partners ($p< 0.001$) but not with male partners ($p=0.62$). The hazard increases with decreasing age at first intercourse, but the trend disappeared after adjusting for the number of female partners. Prior infection with gonorrhea was positively related to PCa risk (OR=1.50; 95%CI=1.0-2.2), but other sexually transmitted diseases were found not associated. Furthermore, there was no correlation

between lifetime frequency of sexual intercourse and risk of PCa (Rosenblatt et al., 2001).

Cerhan and his co-workers analyzed data from 1050 men aged between 65-101 years with no known cancer in the past 10 years in 1982. There were 71 incidents of PCa in 1993. They recorded that cigarette smoking (RR=2.9) for current smokers taking ≥20 cigarettes/day, Higher BMI (RR= 1.7) and higher level of physical inactivity (RR= 1.9; $P_{(trend)}$=0.05) are self-determining predictors of PCa. Percent change in BMI from age 50 to baseline was positively linked with enhanced risk (P= 0.01) and the association seemed stronger in heavier men (Cerhan et al., 1997).

Chapter 3

METHODOLOGIES

3.1 Construction and analysis of interaction network related to PCa

Prostate adenocarcinoma is one of the most common type of cancers and a leading cause of cancer related deaths in men (Leongamornlert et al., 2012, Tafrihi et al., 2013, Khan et al., 2013). Due to the complex origins and causes of PCa, pinpointing the root cause of disease is difficult, but PCa often results due to diverse risk factors including age, culture, environmental factors and genetic mutations (Zhang et al., 2012). Epidemiologically, PCa can be classified into rare hereditary and much more common sporadic forms.

Expression profiling using microarray has swiftly emerged as a potent tool to reveal multiple gene expression signatures associated with different disorders, including cancers (LaTulippe et al., 2002, Singh et al., 2002, Glinsky et al., 2004). Global gene expression analysis and the use of microarray databases have allowed the simultaneous identification of hundreds and thousands of genes, which is helpful for identification and screening of genes that show changes in expression in PCa (Martin et al., 2011). However, the determination of a unique set of regulatory relations based on observed changes in gene expression is difficult due to multi-level complexities. Network analysis can be progressively applied to genome-wide expression profiles to explore regulatory associations amongst genes and to comprehend the root causes of multifarious diseases (Chen et al., 2008). This study involves microarray data analysis of multiple datasets integrated with text mining of online literature repositories to identify vital proteins and regulatory pathways that are associated with PCa. The molecular linkages among these pathways were then further evaluated by topological investigation of protein-protein interaction networks of the translated protein products of causal or susceptible genes involved in PCa development and progression.

3.1.1 Prostate cancer expression profiles:

Expression profiles GSE27914, GSE35373, GSE35324 were downloaded from the online public functional genomics data repository Gene Expression Omnibus of the National Center for Biotechnology Information (NCBI GEO, http://www.ncbi.nlm.nih.gov/geo). These expression profiles were based on the Affymetrix GDS4159, GDS4158, GDS4123 platform data (Affymetrix Human Genome U133A), and contained15, 16 and 5 samples respectively.

3.1.2 Expression profile pre-processing:

The original .CEL (Roudier et al., 2003) files and platform probe annotation information files were downloaded for the next step. The raw data was transformed into expression values using the affy package in R environment, followed by standardization of absent parts of the data. Expression data was then normalized using the normalization method of the affy package.

3.1.3 DEG Analysis:

Significance of differentially expressed genes (DEGs) between the disease samples and controls was tested using t-test through MeV 4.9 (Molecular Environment Visualizer), which is freely available software that is mainly used for the investigation of microarray data, and provides different analysis functionalities. Variance filters were applied to the datasets followed by K-Means clustering to cluster the up- and down-regulated genes from all three datasets.

3.1.4 Intersection of DEGs from the processed expression profiles:

Intersections of differentially expressed genes from the three microarray datasets were determined through BioVenn (www.cmbi.ru.nl/cdd/biovenn/) to strengthen the odds, and only those genes were processed in subsequent steps. The DAVID (The Database for Annotation, Visualization and Integrated Discovery)(http://david.abcc.ncifcrf.gov) Gene Functional Classification Tool was used for functional annotation of the DEGs dataset. This tool uses a clustering algorithm to sort a gene list or related biological terms into classes of associated genes or terms, called biological modules. This tool is a potent method for grouping functionally associated genes and terms into a

convenient number of biological modules for the effective elucidation of gene datasets in a network perspective.

3.1.5 Mining of candidate PCa-associated genes from the literature:

Candidate PCa-associated genes were extracted from online literature repositories through PolySearch (http://wishart.biology.ualberta.ca/polysearch/), which is an online text mining tool specifically designed for the extraction and analysis of text-derived relationships among human diseases, organs, tissues, genes, proteins, mutations, SNPs, metabolites, drugs and pathways from multiple databases including PubMed, DrugBank, SwissProt, HGMD and Entrez SNP (Cheng et al., 2005). We used the PolySearch mining system to extract PCa-associated genes by using query of 'Disease-Gene/Protein Association' and the query keyword 'prostate cancer'. PolySearch returned 482 publications after an initial search which were manually checked to confirm accuracy. A list of 47 candidate PCa-associated genes was compiled (Table 3.1)

3.1.6 Extraction of most important candidate genes from the expression profiles and text-derived gene list:

DEGs extracted from the expression profiles were compared with the text-derived candidate PCa-associated genes list to re-confirm the significance of identified genes. The candidate genes that were common to both lists were termed as "most important genes".

3.1.7 Scanning PPIs through STRING:

In the cell, proteins may have multiple functional associations, many of which serve regulatory purposes that in turn can be part of networks. The size and complexity of these networks provides an exceptional opportunity to gain deeper insights into genomes (Pattin and Moore, 2009, Szklarczyk et al., 2011). The PCa-associated candidate genes were transformed into seed proteins and their interactions were obtained from the STRING database (http://string-db.org/), which encompasses approximately 2.5million proteins from a diverse range of organisms (Jensen et al., 2009).

3.1.8 Network reconstruction from PPI data:

An extended network of seed proteins that includes both neighboring proteins and their interactions was constructed using Pajek software (Batagelj and Mrvar, 1998). Pajek is a Windows based program designed to allow analysis and visualization of large networks and is free for non-commercial users.

3.1.9 Network reconstruction from DEG expression profiles:

A comprehensive network of DEGs identified from the microarray data was constructed using the GeneMANIA (http://www.genemania.org) online resource, an online tool that, among other applications, is used for generating hypothesis about gene functions and analyzing gene lists. GeneMANIA basically extends the query list with functionally linked genes identified from online accessible data and also provides weights that signify predicted values(Warde-Farley et al., 2010).

3.1.10 Topological analysis of PPI networks:

Nodes in the network were evaluated on the basis of different parameters related to network theory, including K (degree of connectivity), BC (betweenness centrality) and CC (closeness centrality). K is the basic attribute of a node in a network and quantitatively reflects adjoining links or interactions connecting a node to its neighbors. BC represents the number of shortest paths passing through each node and is a measure of the frequency of a node occurring on the shortest paths. A high BC node has significant influence over the network and thus plays an important role as a global property for detecting bottlenecks. CC is the inverse of the average length of the shortest paths to or from all other nodes in the graph and reflects the topological center of the network. Average degree and diameter (D) were used as global topological measurements to characterize the network. <k> is the representation by means of all degree values of network nodes, while D is the longest path among all the shortest paths.

3.1.11 Identification of large BC nodes to create a backbone network:

The nodes with higher BC are of prime importance and thus these proteins, along with their interactions, will be used to compile the backbone network. These nodes and

their interactions were identified and used for the said purpose. The threshold for a high BC was set at 5% (Goni et al., 2008, Kim and Kim, 2009).

3.1.12 Comparison of networks constructed through GeneMANIA and Pajek:

As a final step, both networks were compared in a network perspective using Cytoscape (Shannon et al., 2003) to validate the networks and their authenticity. Cytoscape is an open source platform for visualization of molecular interactions either in networks or biological pathways, and also allows integration of these interactions with annotations and expression profiles as well as other parameters.

Table 3.1: List of prostate cancer associated candidate genes derived through text mining

S. No.	Symbol	Description
1.	BRAF	v-raf murine sarcoma viral oncogene homolog B
2.	MDM2	MDM2 oncogene, E3 ubiquitin protein ligase
3.	AR	androgen receptor
4.	PTEN	phosphatase and tensin homologue
5.	TPX2	TPX2, microtubule-associated
6.	BRCA2	breast cancer 2, early onset
7.	CTGF	connective tissue growth factor
8.	EGF	epidermal growth factor
9.	ATM	ataxia telangiectasia mutated
10.	CTNNB1	catenin (cadherin-associated protein), beta 1, 88kDa
11.	TRRAP	transformation/transcription domain-associated protein
12.	CAV1	caveolin 1, caveolae protein, 22kDa
13.	KDM6A	lysine (K)-specific demethylase 6A
14.	CDKN2A	cyclin-dependent kinase inhibitor 2A
15.	BRCA1	breast cancer 1, early onset
16.	HRAS	Harvey rat sarcoma viral oncogene homolog
17.	GSTP1	glutathione S-transferase pi 1
18.	RNASEL	ribonuclease L (2',5'-oligoisoadenylate synthetase-dependent)
19.	CDK1	cyclin-dependent kinase 1
20.	EGFR	epidermal growth factor receptor
21.	MYST4	K(lysine) acetyltransferase 6B
22.	AKAP9	A kinase (PRKA) anchor protein 9
23.	CCNA2	cyclin A2
24.	ERBB2	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2
25.	TP53	tumor protein p53
26.	BCL2	B-cell CLL/lymphoma 2
27.	KRAS	Kirsten rat sarcoma viral oncogene homolog
28.	MED12	mediator complex subunit 12
29.	MLL	lysine (K)-specific methyltransferase 2A

30.	LEPR	leptin receptor
31.	SMAD4	SMAD family member 4
32.	IGF1	insulin-like growth factor 1 (somatomedin C)
33.	EGR1	early growth response 1
34.	PMS2	PMS2 postmeiotic segregation increased 2
35.	MSH2	mutS homolog 2
36.	APC	adenomatous polyposis coli
37.	ELAC2	ElaC ribonuclease Z2
38.	SRF	serum response factor
39.	RAD51	RAD51 recombinase
40.	RB1	retinoblastoma 1
41.	MLL2	myeloid/lymphoid or mixed-lineage leukemia 2
42.	CSAD	cysteine sulfinic acid decarboxylase
43.	AURKA	aurora kinase A
44.	CCL2	chemokine (C-C motif) ligand 2
45.	MSR1	macrophage scavenger receptor 1
46.	FOXA1	forkhead box A1
47.	MLL3	myeloid/lymphoid or mixed-lineage leukemia 3

3.2 Blood samples

This population based case-control study was carried out at Capital University of Science and Technology, Islamabad, Pakistan in collaboration with Institute of Biomedical and Genetic Engineering (IB&GE), Islamabad, Pakistan. The study was approved by the Departmental Scientific Committee of the University. Whole blood samples (5ml) were collected from 680 prostate cancer patients along with 500 controls from different ethnic groups of Pakistan. Blood samples were collected with relevant information and informed signed consent (Annexure-2) from histologically confirmed prostate cancer patients along with age matched controls by using disposable syringes under strict aseptic conditions and were transferred to ACD (acid citrate dextrose) vacutainer tubes. Each tube was labeled properly by using the ID given to each individual. The tubes were kept at 4°C, until the genomic DNA extraction was started. Control samples were collected for comparison and polymorphism analysis. Samples were collected from NORI (Nuclear Medicine Oncology and Radiology Institute) Islamabad, DHQ (District Head Quarter Hospital) Rawalpindi, IRNUM (The Institute of Radiotherapy and Nuclear Medicine) Peshawar and visiting different villages in Pakistan. The vital status of study participants was determined through the hospital records and making personal inquiries to determine whether the participant were alive or dead and to record the date of all deaths. All the wet lab part of this study was conducted at the labs of Institute of Biomedical and Genetic Engineering (IBGE), Islamabad, Pakistan. IBGE is a dedicated facility devoted to basic biomedical and translational research to promote the understanding and prevention of human diseases. It is working in the field of medical genetics for more than 25 years.

3.3 DNA extraction

Genomic DNA was extracted from all the samples by using the standard phenol-chloroform method (Organic method) (Sambrook and Russell, 2001), followed by ethanol precipitation. The protocol was slightly modified as per requirement. Three days were given to each batch (20 samples) of blood samples to get reliable results and large amount of pure DNA.

3.3.1 Day 1:

1. All the samples were arranged in order according to their ID's and were transferred to labeled (50 ml) falcon tubes.
2. Quantity of each sample was carefully recorded. Cell lysis buffer (KHCO3, NH4Cl and 0.5M EDTA) was added to each sample for the lysis of cells and removal of hemoglobin in a quantity equal to three times of the blood samples i.e. in 5ml blood, 15 ml of cell lysis buffer was added.
3. All the tubes were kept on ice for 30 minutes after addition of cell lysis buffer to stop the cellular metabolic activities followed by centrifugation (Eppendorf Refrigerated Centrifuge 5130) at 1200 rpm for 10 minutes at 4°C.
4. The supernatant was discarded and pallets were re-suspended followed by washing of the pallet with 15ml of cell lysis buffer in the same manner as mentioned above to remove the haemoglobin and complete lysis of the remaining cells.
5. The supernatant was discarded and the pellet was re-suspended.
6. 4.75ml of Sodium Tris EDTA (STE) was added to each sample.
7. Then 250µl of 10% Sodium dodecyl sulfate (SDS) was added while vertexing the samples.
8. To the above mixture 10 µL of proteinase K enzyme (20mg/ml) (Fermentas, Lithuania) was added and samples were incubated at 55℃ for overnight in shaking water bath (Orbit Shaker Bath, Lab-Line, USA).

3.3.2 Day 2:

1. The samples were extracted with 5ml (equal quantity) of equilibrated phenol (pH=8).
2. Agitated for 10 minutes and kept on ice or freezer for 10 minutes.
3. Then the samples were centrifuged (Eppendorf Refrigerated Centrifuge 5130) at 3200 rpm for 30 minutes at 4°C.
4. The aqueous phase was transferred to labeled (15ml) falcon tubes with cut tips method.

5. 5ml of chilled chloroform:isoamyl alcohol (24:1) was added to each sample and agitated for 10 minutes and kept on ice for 10 minutes.
6. Samples were centrifuged at 3200rpm for 30 minutes at 4℃ in refrigerated centrifuge (Eppendorf Refrigerated Centrifuge 5130).
7. The supernatant was separated with cut tip into the separate labeled 15ml centrifuge tubes.
8. To the isolated supernatant, 500µL of 10M ammonium acetate and 5ml of chilled Isopropanol was added and agitated until DNA precipitates as visible white threads.
9. The samples were then placed overnight at -20℃ (or for 15 minutes at -70℃).

3.3.3 Day 3:

1. The stored samples were centrifuged at 3200 rpm for 1 hour at 4°C.
2. The supernatant was discarded and the DNA pellet was re-suspended by tapping the 15ml centrifuge tube.
3. 5ml of chilled 70% ethanol was added to the washed samples and again was centrifuged at 3200rpm for 40 minutes at 4°C.
4. The supernatant was again discarded and DNA was dried until the last drop disappeared.
5. After drying the pallets, TE buffer (10mM Tris; 1mM EDTA) was added to each sample according to the size of the pellet.
6. The complete DNA dissolving in TE buffer was achieved by incubating the sample tube in shaking water bath at 55℃ (Orbit Shaker Bath, Lab-Line, USA)DNA to expand freely.
7. Then whole solution was transferred to (1.5ml) tubes and was labeled as stock solution.
8. 20µl of each DNA was added to 80µl of ddH_2O to make 20% DNA dilution as working solution in a labeled (1.5ml) tubes.

3.4 Assessment of Quality and Quantity of DNA

Each sample was diluted 50 folds by adding 6 µl DNA sample to 294 µl distilled water and quantified on UV Spectrophotometer (U-3210, Hitachi, Japan) at 260nm and 280nm wavelength.

Optical density (OD) ratio for each sample was calculated as:

OD=Absorbance at 260nm / Absorbance at 280nm

The ratio from above formula should lie between 1.7-1.9 for good quality DNA. The concentration of DNA samples was calculated as:

DNA concentration (µg/ml) =Absorbance at 260nm × dilution factor ×correction factor

3.4.1 Working Solution of DNA:

40ng/µl working solution of DNA was prepared from the stock DNA solution by using the following formula:

$$C1V1=C2V2$$

3.5 Primer designing

3.5.1 Retrieval of sequences

Nucleotide sequences of TP53 (Ensembl Id: ENST00000269305), PTEN (Ensembl Id: ENST00000371953) and AR (Ensembl Id: ENST00000374690) genes along with introns, 5 and 3' flanking regions were retrieved from Ensembl (http://www.ensembl.org/index.html), to be used for primer designing, selection of restriction enzymes etc. Ensembl is a collaborative project of European Bioinformatics Institute (EBI) and the Wellcome Trust Sanger Institute, launched in 1999 in response to the imminent completion of the Human Genome Project. The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Exon-specific primers covering intron/exon boundaries were designed for all the three genes by using the online resource, Primer3 version 0.4.0 (http://frodo.wi.mit.edu).

Primer3 is an open source, widely used program and can also design hybridization probes and sequencing primers. It has many different input parameters that researcher can control for making good primers. The primers used are shown in Table 4.5, Table 4.6 and Table 4.7.

3.6 PCR

Before carrying out a PCR for the samples (Annexure-3), the conditions were carefully optimized for each primer. All the samples (each 1.5µl) in diluted form were added to the labeled PCR tubes. All the contents of PCR were kept on ice tray to avoid degradation of enzymes and other contents while preparing master mix. PCR recipe was prepared by using the contents given below in Table 3.2.

Table 3.2: PCR Recipe

Reagents	**Stock Conc.**	**Required Conc.**	**Final Conc. for 1 reaction**	**Final Conc. for 27 reactions**
dH2O			4.6 µl	124.2 µl
PCR Buffer	10X	1X	1 µl	27 µl
MgCl2	25mM	1.5mM	1 µl	27 µl
dNTP's	0.8 µ/µl	0.2mM	1 µl	27 µl
Taq Polymerase	2 µ/µl	0.8 µ/µl	o.3 µl	8.1 µl
forward Primer	2O µM	600nM	0.3 µl	8.1 µl
Reverse Primer	20 µM	600nM	0.3 µl	8.1 µl

8.5 µl of PCR Master Mix was added to each sample of DNA (1.5µl) been added to PCR tubes so total volume of DNA and Master mix was reached to 10µl.The caps were tightly closed on strips and were kept in PCR machine for 30 cycles at following temperatures;

3.6.1 PCR Conditions:

- 95°C for 5 minutes.
- 95°C for 1-2 minutes for 30 cycles.
- 54-57°C for 2 minutes.
- 72°C for 1 minutes and final extension at 72°C for 10 minutes.
- After completion of polymerase chain reaction the strips were taken out of PCR and were stored at 4°C till next process.

3.6.2 Gel Electrophoresis

Gel loading dye (5µl) was added to each PCR product and was dissolved with tip sucking method. Then 10µl of each DNA sample was loaded into gel wells by using separate tips for each sample. All the samples were transferred and electric current (120 volts) was applied for a mean time depending upon the size of the product sequence. TAE (Tris/Acetate/EDTA) and TBE (Tris/Borate/EDTA) were used in electrophoresis as they are the most commonly used buffers for nucleic acids electrophoresis. After completion of electrophoresis, the DNA molecules were stained in the gel with ethidium bromide to make them visible. Ethidium bromide is an intercalating agent and makes the DNA fluorescent under UV light and thus each and containing ~20 ng DNA became distinctly visible. The images were saved for further reading and processing.

3.7 Single Strand Conformation Polymorphism (SSCP)

The mobility of a particle in a gel depends both on its size and shape. In non-denaturing conditions, ssDNA has a folded structure due to intra-molecular interactions. A mutated sequence can thus be detected as a change of mobility caused by the altered folded structure in polyacrylamide gel electrophoresis in SSCP analysis. The amplified PCR product was denatured and resolved by polyacrylamide gel

electrophoresis, and mutations were detected as altered mobility of separated single strands. The overall procedure is rapid and simple.

In our study, we used the allele-Specific enzymatic cleavage, one of the earliest and most widely used genotyping methods to analyze the previously reported SNPs like P72R of TP53 in our samples. The required endonucleases were selected by using online tools.

3.8 SNP analysis

Single nucleotide polymorphisms (SNPs) represent the most frequent and abundant form of human genetic variations and their patterns are more likely to influence many phenotypes. Large scale SNP based association studies are therefore expected to help in the identification of genes affecting complex diseases. Different methods are available for SNP genotyping. In our study, we used the allele specific enzymatic cleavage (PCR-RFLP), for already reported SNPs while direct sequencing for the rest of experiments. The required endonucleases were selected by using WATCUT (http://watcut.uwaterloo.ca/watcut/watcut/template.php), an online tool for SNP-RFLP analysis.

Allelic and genotypic frequencies were calculated, ORs with 95% CIs were calculated to assess the strength of association between the observed polymorphism and PCa susceptibility by using the unconditional logistic regression under multiple models in cases-controls groups. $P<0.05$ was considered statistically significant and different genetic model including Homozygous, Heterozygous, Dominant and Recessive were used to evaluate the risk and allelic associations.

3.9 Purification of PCR products for sequencing:

The PCR product was transferred to new labeled tubes, 2.5 µl EDTA (125mM) was added to it, then 30µl absolute ethanol was added and vertexing was done. The solution was kept at -4°C for 15 minutes, and then its spinning was done at 13000 rpm for 20 minutes. The supernatant was discarded and pallet was vigorously vertexed after adding 100µl (70%) chilled ethanol. Then its spinning was done for 10 minutes at 13000 rpm, again supernatant was discarded and pallet was dried in oven at 55°C. Finally 10µl Hi-Di formamide was added. The pallet was then denatured by PCR at

95°C for 4 minutes. Orange dye (250 µl) was added to sample and sequencer machine was switched on the samples were kept in tray of sequencer and software was monitored for sequencing the samples.

3.10 DNA sequencing

Band shifts detected by SSCP were used for sequencing analysis. DNA was sequenced with both forward and reverse primers. Sequencing was performed according to the standard protocol provided with the Big Dye TerminatorKit® v3.1 (ABI 3130 Genetic analyzer). The products were then analyzed through SeqScape 2.5 software (Applied Biosystems) and other freely available softwares like FinchTV etc. The sequence data obtained was then analyzed through different Bioinformatics techniques for their *in silico* characterization and annotations.

3.11 Computational Analysis

3.11.1 Structure prediction and functional impact analysis of genetic alterations

Proteins are always exposed to alterations which can pose diverse effects either by changing their interactions, residues or folding pattern. Structure predictions and comparisons will provide a detailed insight into the molecular, tertiary level and folding pattern variations in proteins. We considered the prostate cancer related mutations and associated SNPs of TP53, PTEN and AR genes in Pakistani population and were annotated for their ability and intensity to affect structures and functions of the respective proteins. It was thus hoped that it will help to enhance our understanding towards the prospective molecular causes, and guide to develop new treatment and management strategies against genetic disorders. Systematic computations on the basis of sequences and structures were applied to study the effect of genetic variations as they affect the protein structures and other physiochemical properties, thus damaging their interactions.

Biochemical assays, though are used globally to identify deleterious alterations, but are time consuming and laborious. Accurate and precise computational analysis of the possible functional outcomes of genetic variants can significantly decrease the time

complexity by ranking them probably to be deleterious or vice-versa. In our study we used automated methods i.e. SIFT (Ng and Henikoff, 2003), PolyPhen2 (Adzhubei et al., 2010), MutationAssessor (Reva et al., 2011) and PROVEAN (Choi et al., 2012) capable of discriminating between driver and passenger mutations.

SIFT (Sorting Intolerant from Tolerant) uses sequence homology for the prediction of functional impacts due to an amino acid substitution. It presumes that functionally significant residues may be conserved in a protein family, and thus substitutions at conserved positions are most probably anticipated as deleterious. Substitutions with scores less than 0.05 are predicted as deleterious (http://www.blocks.fhcrc.org/sift/SIFT.html) (Thomas et al., 2006). PolyPhen2 annotate substitutions by using the integration of sequence, phylogenetic and structural information with its position in the protein (http://coot.embl.de/PolyPhen/). MutationAssessor annotates the functional impact of protein residue substitutions by assessing its evolutionary conservation. It has already been validated on OMIM data sets (http://mutationassessor.org/). PROVEAN (Protein Variation Effect Analyzer) has the ability to predict the functional impact of amino acid substitutions and indel and is handy to classify functionally important nonsynonymous or indel variants (Choi et al., 2012).

3.11.2 Phylogenetic analysis

With the increasing number of whole genome sequences availability, mapping genes to their corresponding counterpart becomes even more important. Orthology is usually used to find these analogous genes, conservativeness of gene and to allow transfer of annotation from the known to the unknown gene (Schreiber et al., 2013).

We used PhylomeDB (http://phylomedb.org/) and TreeFam (http://www.treefam.org/) to confirm the conservativeness of TP53, PTEN and AR genes across the model organisms. PhylomeDB is a public database for complete catalogs of gene phylogenies and allows users to explore the phlogenetic history of genes through visualizing the evolutionary trees along with multiple sequence alignments. It also provides genome-wide paralogy predictions based on the phylogenetic trees. TreeFam is a database of phylogenetic trees inferred from animal genomes, providing orthology/parology predimeactions along with the evolutionary history of genes.

3.11.3 The protein-protein interaction (PPI) network analysis

PPIs are the major mechanism controlling biological systems and are therefore attractive for exploring functions and therapeutics. With the remarkable growth of human PPI data, these methods are in use to understand the molecular mechanics of diseases and related sub-networks. In PPIs, proteins are symbolized as nodes and their relations as edges. Usually proteins have only a small number of interactions in the PPI network but some proteins called hubs share a large number of connections. Hub proteins are vital to the normal function and stability of the interaction network.

It was reported that that TP53, PTEN and AR are hub proteins and play key roles in the protein interaction networks. TP53 is a principal hub of cell proliferation and apoptotic pathways, respond to potential oncogenic stresses. Similarly PTEN is the main hub of PI3K-mediated signaling pathway, while AR is also a hub protein connecting androgen signaling to FlnA and integrin beta 1 in NIH3T3 fibroblasts. Mutations in hub protein are much more crucial than any other protein. Research has proved that hub protein deleterious mutations are more likely to be lethal (Jeong et al., 2001). It is therefore important to structurally and functionally analyze the effect of different genetic alterations on TP53, PTEN and AR hub proteins in the network.

3.12 Analysis of socio-demographic and clinical factors

Blood samples were collected from the prostate cancer patients and controls with relevant information on an informed signed consent on a specifically designed proforma. The proforma was designed to gather the required information about age, weight, BP, ethnicity, family history, smoking, health conditions, PSA level, tumor size, any other complication if any.

Cancer is a complex multistep disorder, resulting from a combination of factors (Mahan and Escott-Stump, 2004). The associated etiological factors of prostate cancer including aging, immune function disorders, environmental and lifestyle factors are poorly investigated as compared to other common cancers (Agalliu et al., 2009). To address this issue, we have collected detailed information about all the possible risk factors including demographic and lifestyle characteristics. An advantage of our study over the previous studies is that it was specifically designed to investigate the association of each factor independently and in combination with other as well.

Literature survey was systematically carried out to have a deeper insight towards the contribution of different risk factors including family history (Powell, 2011), dietary factors (Mandair et al., 2014). The data was statistically analyzed for the association of risk factors through logistic regression, cox-proportional hazard and survival analysis through Kaplen-mier curve. All statistical analysis was performed using R version 3.1.1 (R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna (2013)).

Chapter 4

RESULTS AND DISCUSSION

4.1 Construction and analysis of interaction network related to PCa

4.1.1 Prostate cancer expression profile datasets:

Since the expression profiles of prostate cancer downloaded from NCBI GEO were not in a compatible format for the analysis software, they were first converted into a compatible (.txt) format through the affy library (http://www.bioconductor.org) of the R suite. R is an integrated suite of software facilities for statistical computing, data handling, management, computation and graphical display (http://www.r-project.org/). The datasets were processed for normalization with the same suite and among the normalized samples of the dataset there were 54675 genes.

The data was then loaded to MeV in the .txt format for filtration to explore the DEGs above the threshold. After variance filtration with a parameter of 50, the number of filtered genes for the next step of analysis was 27337. A t-test was performed to evaluate the significance of the genes after filtration followed by K-Means and Hierarchical Clustering treatment of filtered data to yield different clusters of DEGs (up- or down-regulated expression). K-Means clustering yielded 10 clusters for the three datasets (Table 4.1). After analyzing intersection of the DEGs in all three datasets and functional annotation of chip IDs using the DAVID tool, a total of 96 genes were obtained.

4.1.1 Network Re-construction:

The network constructed with Pajek consisted of 47 nodes with 200 connections. The backbone network has only 14 nodes with 41 connections (Figure 4.1). Parameters including number of nodes (N), average degree (<k>) and diameter (D) were explored for the network with the largest degree in the network being 62, and an average

degree of 19.31. This network was comprised of a small number of highly inter linked genes/proteins, while most of the others had only a small number of connections.

The nodes with either higher degree or BC value were designated as key nodes, which represented 14 out of the 47. Twenty-one nodes have a large degree (Table 4.2) and14 have high BC values (Table 4.3), while only 7 have both higher BC and larger degree (Table 4.4). In order to determine the role of these nodes in the network, they were coded with different colors and varying sizes (Figure 4.1). TP53 has the largest degree and its product is a bottleneck protein with highest BC and CC values. It indicates that TP53 is located at the center of the network. The backbone network consisted of 14 high BC nodes (EGR1, IGF1, AR, CTNNB1, EGFR, HRAS, CDK1, BRCA1, BRCA2, SRF, RNASEL and KDM6A) connected through 41 links (Figure 4.2). P53was found to have 9 neighbors (EGR1, IGF1, AR, CTNNB1, EGFR, HRAS, CDK1, BRCA1 and BRCA2) that compose its respective pathway.

Different interaction networks such as physical interaction, genetic interaction and pathway interaction were also constructed using GeneMANIA and the DEGs from expression profiles to validate the results.TP53 was again found to be an important player in PCa, thus recapitulating the results derived from the backbone network that was constructed from text-derived data using Pajek software (Batagelj and Mrvar, 1998). In all the networks reconstructed here, the central gene TP53 was found to be linked with the highest number of genes, either directly or indirectly, because of its significance in the network and multi-functional character. The networks reconstructed both through Pajek and GeneMANIA were compared in a network perspective with Cytoscape software to reconfirm and strengthen the hypothesis. This analysis showed that14 genes involved in the backbone construction were common to both networks, thus confirming their importance and significance and validating our network results.

Table 4.1: Gene clusters for the three gene expression datasets GSE27914, GSE35373, GSE35324.

Dataset 1			Dataset 2			Dataset 3		
Clusters	**No. of Genes**	**Genes %**	**Clusters**	**No. of Genes**	**Genes %**	**Clusters**	**No. of Genes**	**Genes %**
Cluster 1	1581	6	**Cluster 1**	4136	15	**Cluster 1**	1732	6
Cluster 2	2716	10	**Cluster 2**	2430	9	**Cluster 2**	2315	8
Cluster 3	2628	10	**Cluster 3**	1454	5	**Cluster 3**	3997	15
Cluster 4	2149	8	**Cluster 4**	3514	13	**Cluster 4**	3518	13
Cluster 5	1841	7	**Cluster 5**	3493	13	**Cluster 5**	3200	12
Cluster 6	1743	6	**Cluster 6**	3179	12	**Cluster 6**	1427	5
Cluster 7	4935	18	**Cluster 7**	3095	11	**Cluster 7**	2904	11
Cluster 8	2508	9	**Cluster 8**	2443	9	**Cluster 8**	1279	10
Cluster 9	6174	23	**Cluster 9**	1637	6	**Cluster 9**	2435	9
Cluster 10	1062	4	**Cluster 10**	1956	7	**Cluster 10**	3017	11

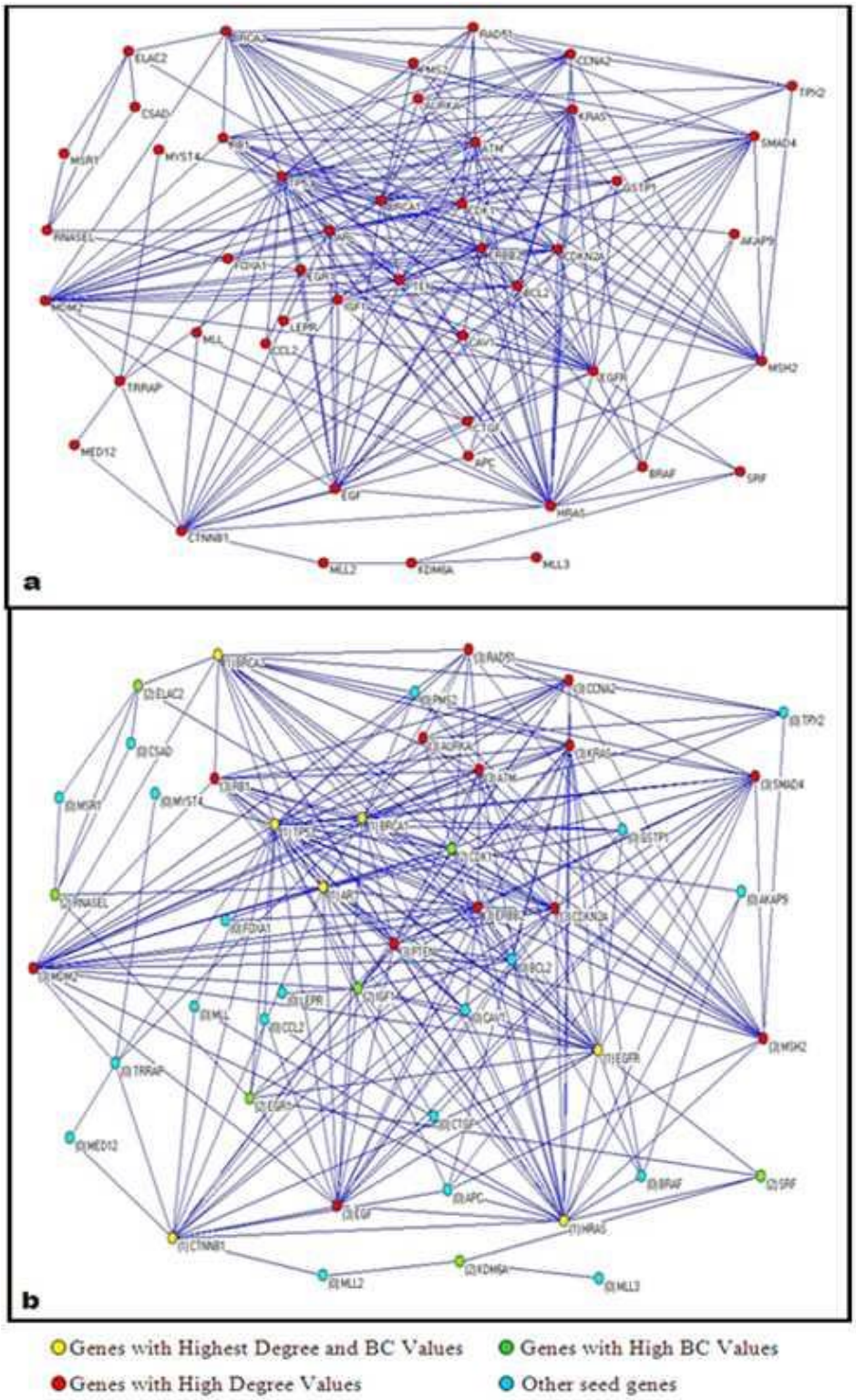

Figure 4.1: Gene networks constructed through Cytoscape

a. Overview of Complete Network. The network includes the seed genes. The nodes with label are seed genes converted from the candidate genes listed in Table 3.1.

b. The topology of the network.

Table 4.2: The list of high BC nodes and their CC Values

SN	SYMBOL	BC	CC Value
1	TP53	0.163816	0.69697
2	HRAS	0.099444	0.666667
3	CTNNB1	0.095746	0.567901
4	BRCA1	0.089278	0.630137
5	AR	0.088829	0.597403
6	EGFR	0.076464	0.638889
7	SRF	0.067248	0.455446
8	IGF1	0.054749	0.554217
9	RNASEL	0.050525	0.422018
10	KDM6A	0.044593	0.330935
11	ELAC2	0.038173	0.418182
12	CDK1	0.036381	0.554217
13	EGR1	0.033376	0.505495
14	BRCA2	0.033221	0.582278

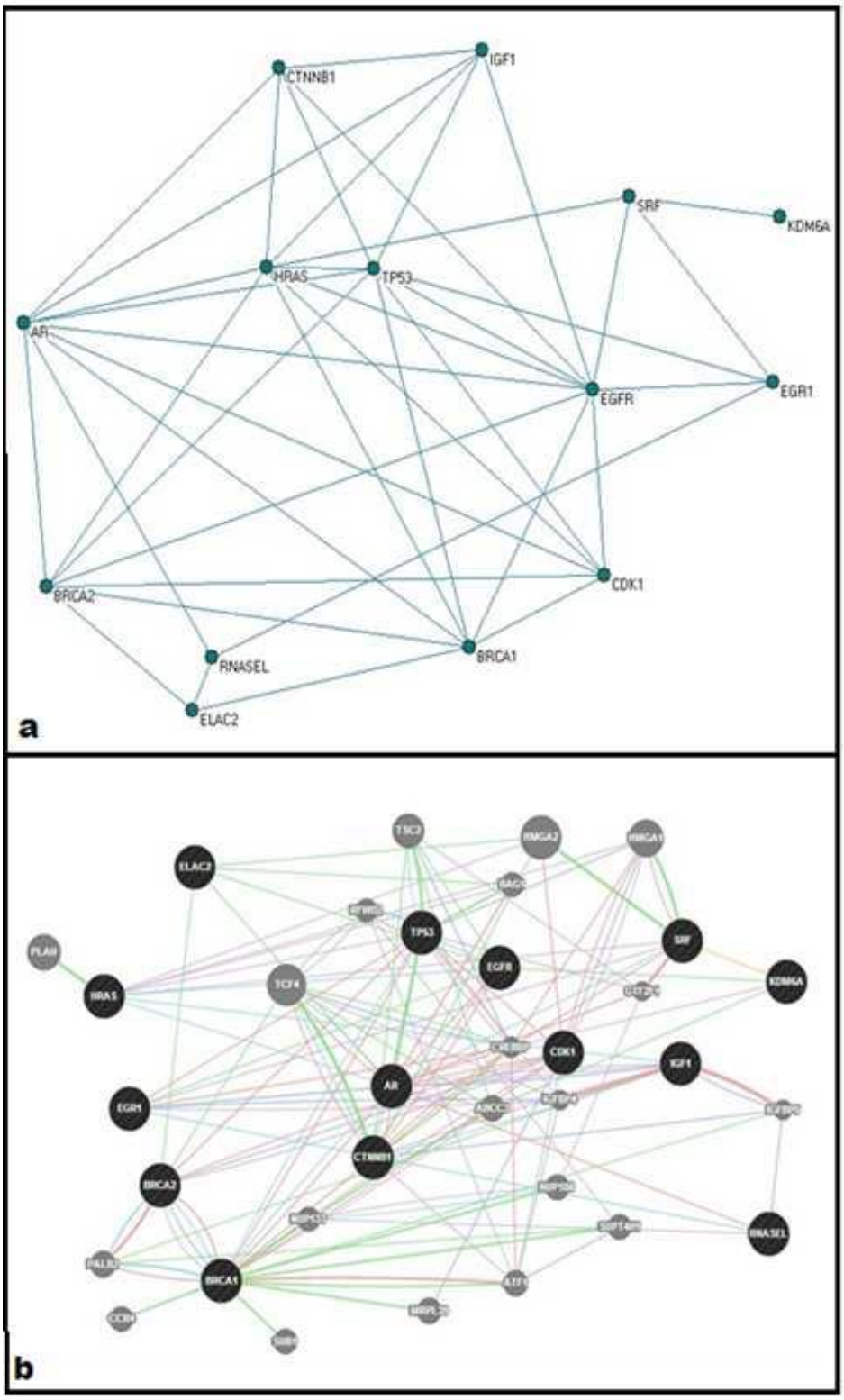

Figure 4.2: Representation of network topology.

a. The topology of the backbone network. The backbone network consists from 14 nodes with high BC value.
b. Complete network of the highest BC values genes from the GeneMANIA

Table 4.3: The list of large degree nodes and their CC Values

S. No	Symbol	Degree	CC Value	S. No	SYMBOL	Degree	CC Value
1	TP53	62	0.69697	12	AR	30	0.597403
2	BRCA1	50	0.630137	13	PTEN	30	0.605263
3	EGFR	46	0.638889	14	KRAS	30	0.589744
4	HRAS	45	0.666667	15	ATM	29	0.528736
5	CDK1	42	0.554217	16	RAD51	28	0.484211
6	MSH2	35	0.534884	17	EGF	26	0.582278
7	BRCA2	34	0.582278	18	CDKN2A	24	0.589744
8	ERBB2	34	0.597403	19	SMAD4	23	0.560976
9	MDM2	33	0.582278	20	RB1	23	0.528736
10	CTNNB1	31	0.567901	21	AURKA	23	0.464646
11	CCNA2	31	0.528736				

Table 4.4: The list of large degree nodes and their description.

S. No	**Symbol**	**Function Description**
1.	TP53	encodes a tumor suppressor protein
2.	BRCA1	encodes a nuclear phospho-protein
3.	EGFR	encoding a transmembrane glyco-protein
4.	HRAS	functioning in different signaling pathways
5.	CDK1	encodes a member of the Ser/Thr kinase family
6.	MSH2	Component of mismatch repair machinery
7.	BRCA2	It repairs double-strand breaks and/or function in homologous recombination
8.	ERBB2	part of several cell surface receptor complexes, ligand binding receptor
9.	MDM2	mediates ubiquitination of p53/TP53
10.	CTNNB1	component of the canonical Wnt signaling pathway
11.	CCNA2	Play role in cell cycle control
12.	AR	affect cellular proliferation and differentiation in target
13.	PTEN	tumor suppressor gene
14.	KRAS	code protein that bind GDP/GTP and possess intrinsic GTPase activity
15.	ATM	activates checkpoint signaling upon double strand breaks
16.	RAD51	participates in a common DNA damage response pathway
17.	EGF	stimulates the growth of various epidermal and epithelial tissues
18.	CDKN2A	induce cell cycle arrest in G1 and G2 phases
19.	SMAD4	Co-activator and mediator of signal transduction by TGF-beta
20.	RB1	tumor suppressor gene

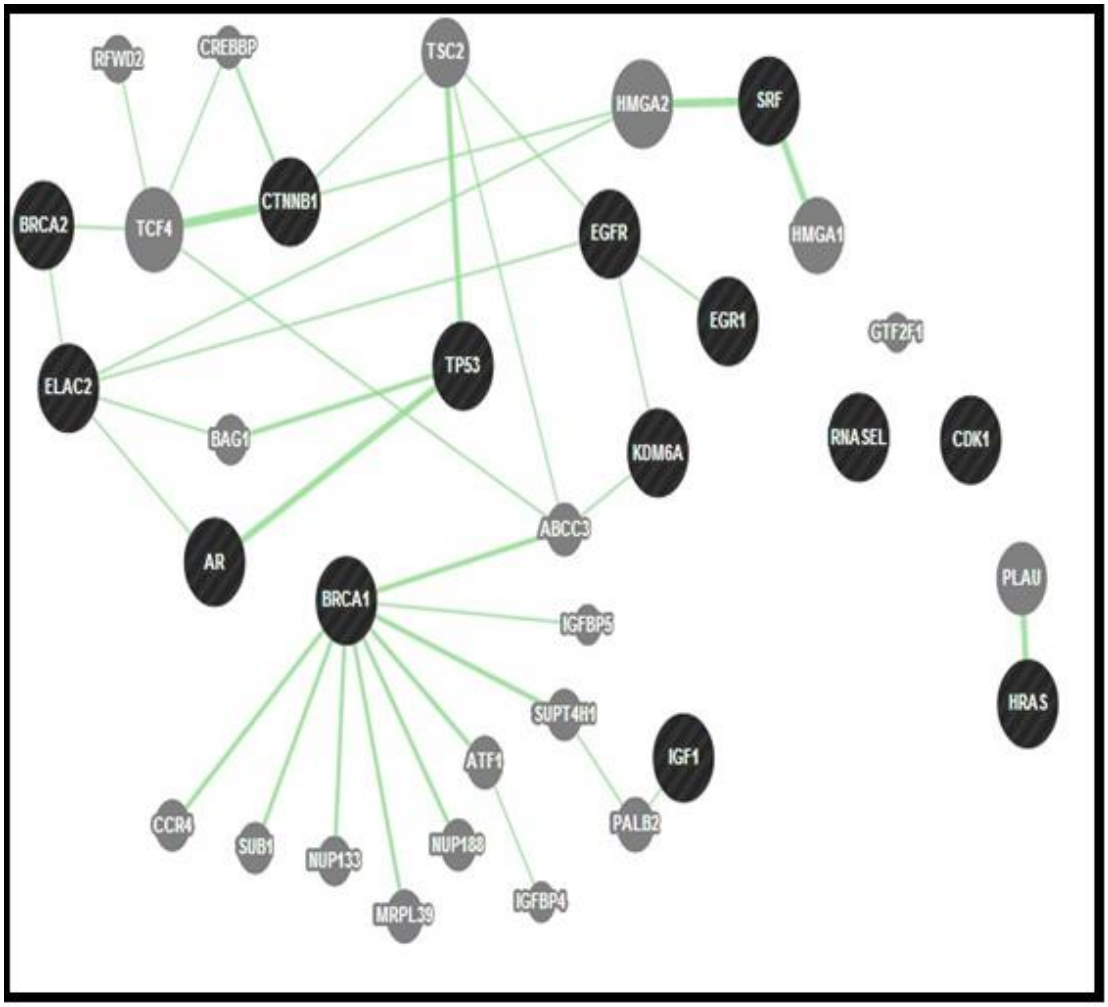

Figure 4.3: Network of Genetic Interactions of the highest BC values genes of the GeneMANIA

Figure 4.4: Nodes after merging network constructed on expression data and text mining through Cytoscape

Although many articles have been published concerning the molecular mechanisms of PCa, and in turn a number of causal genes have been identified as participating in this cancer, PCa pathogenesis nonetheless remains vague. We proposed here that the products of the reported genes may determine respective pathways through their interactions. This study was thus designed to analyze the contribution of candidate genes to prostate cancer and also to discover other key players that cooperate with these genes through a network topological analysis. BC and CC were used as major parameters in this study to evaluate the nodes in the interaction network, which is similar to fundamental measures in network theory previously used by others (Goni et al., 2008, Nguyen et al., 2011, Hashimoto et al., 2009, Li et al., 2012).

A total of 47 genes were shortlisted from the initial list as main causative or susceptible genes for PCa. While the study was performed in a systematic way, there may be some missing values and genes from either the literature search or expression profiles and thus other responsible genes that participate in PCa may still remain to be discovered. Moreover, this study cannot discount the possibility that false positives and false negatives appeared in the results. However, biological networks are robust with respect to arbitrary adjustment of nodes, but susceptible to hub deletion (Lima-Mendez and van Helden, 2009). There were 21 proteins with higher degree and 14 with higher BC values in the network, while only 7 nodes have both larger degree and higher BC (Table 4.2, Table 4.3, Table 4.4 and Figure 4.3). Genes/proteins in the network were systematically classified into four categories i) hub–bottlenecks (those with large degree and high BC values); ii) non-hub–bottlenecks (those with small degree and high BC); iii) non-hub–non-bottlenecks (small degree and low BC); and iv) hub–non-bottlenecks (large degree but low BC). In a similar previous study, genes/proteins were classified into two subtypes as highly connected genes/proteins: i) hub–bottlenecks, which are likely to be date-hubs, wherein the protein products bind differently with different partners at different times or locations; and ii) hub–non-bottlenecks, or “party-hubs” that have the most simultaneous interactions with their partners. Further confirmation of the space-time consequences of these nodes will likely facilitate the discovery of new drug targets and biomarkers in PCa (Han et al., 2004).

TP53 has the largest degree and highest BC value and thus ranks highest in both the large degree nodes and high BC nodes list. Meanwhile, androgen receptor (AR) ranks 5th in the BC nodes list and 12th in large degree node list, while phosphatase and tensin homolog (PTEN) ranks 13th in the large degree node list. TP53 encodes a tumor suppressor protein with domains that govern transcriptional activation, DNA binding, and oligomerization. As such TP53 is a vital actor in cancer pathways and also is a frequent object of diverse genetic alterations in many human cancers.TP53is a well-known transcription factor that is critical for diverse cellular activities, including transcriptional regulation of target gene expression, induction of cell cycle arrest, DNA repair, apoptosis and senescence (Khan et al., 2013).In cancers, p53 ubiquitously loses function as a result of direct genetic aberrations or malfunctions in upstream or downstream signaling pathways (Toledo and Wahl, 2006, Bourdon, 2007, Vousden and Lane, 2007). AR encodes a protein with three key functional domains, i.e. the N-terminal, DNA-binding and androgen binding domains (Callewaert et al., 2003). It expresses itself in bone marrow, mammary gland, prostate, testicular and muscle tissues. It exists as a hetero dimer with Hsp90 and HMGB proteins. AR binds to nuclear response elements in its activated state for transcriptional regulation, but also can affect the target genes by interacting with transcription factors like as AP-1, NF-kappaB and STAT. Insulin-like growth factor 1 and some other genes responsible for the advancement of primary and secondary male sexual characters are its main target genes. It is activated by steroid-hormone and has a key role in the growth and development of the prostate gland. It also has a function in prostate cancer development as demonstrated by its differential expression in majority of primary prostate cancers (Lonergan and Tindall, 2011). PTEN is a ubiquitously expressed tumor suppressor having dual specificity. It antagonizes the PI3K signaling pathway through its lipid phosphatase activity, and negatively regulates the MAPK pathway through its protein phosphatase activity (Pezzolesi et al., 2007). PTEN also antagonizes signal transduction downstream of the PI-3 kinase by dephosphorylating phosphatidylinositol-phosphate (PtdInsP) and modulating cell cycle progression and cell survival, while negatively regulating cell interactions with the extracellular matrix. PTEN has been found to be mutated in multiple advanced cancers, including prostate carcinoma.

To an extent, we can carefully hypothesize that PCa originates from the malfunctioning of general gene networks or changes in local gene function that occur with precise timing and localization. TP53, with the largest CC value, locates at the core of the main and backbone network derived from high BC genes, which emphasizes its noteworthy role in PCa. In this study, the reconstructed backbone network centering on TP53 is a pathway to regulate PCa development (Figure 4). The proteins with direct links to TP53 include EGR1, IGF1, AR, CTNNB1, EGFR, HRAS, CDK1, BRCA1 and BRCA2. This study also showed that p53 works as a significant checkpoint in the cell cycle and carries out the chief function of tumor suppression by detecting problems in cell division. The backbone network presents an unambiguous and visual synopsis by showing all vital genes and related regulatory pathways for PCa and their respective interactions. For additional confirmation of the role of major genes in the backbone network, a sub-network was constructed using GeneMANIA, which showed that some genes other than the candidate genes had links with TP53 and others in the network. It is also to be noted that TP53 has the highest BC value and the top 9 BC nodes in this sub-network overlapped with the 14 large BC nodes of the main network. Our finding suggested that prostate cancer was coordinated by an integrated interaction network centered on TP53.

4.2 Blood samples and DNA extraction

Case control Samples (680 cases and 500 controls) were collected from NORI (Nuclear Medicine Oncology and Radiology Institute) Islamabad, DHQ (District Head Quarter Hospital) Rawalpindi and visiting different villages in Pakistan. Genomic DNA was extracted from all the samples by using the standard phenol-chloroform method. Each sample was quantified by using UV Spectrophotometer. 40ng/µl working solution of DNA was prepared from the stock DNA solution.

4.3 Primer designing for point (missense) mutations

4.3.1 Retrieval of sequences

Nucleotide sequences of TP53 (ENST00000269305), PTEN (ENST00000371953) and AR (ENST00000374690) genes along with introns and 5’ and 3’ flanking regions were retrieved from Ensembl (http://www.ensembl.org/index.html), and were used for primer designing and selection of restriction enzymes. Exon-specific primers covering

intron/exon boundaries were designed for all the three genes to amplify the exons by using the online resource, Primer3 version 0.4.0 (http://frodo.wi.mit.edu). The primers used are shown in Table 4.5, Table 4.6 and Table 4.7.

4.4 DNA sequencing

PCR was performed on all the exons of TP53, PTEN and AR genes with our own designed primers (Table 4.5, Table 4.6 and Table 4.7) from both the cases and controls. The samples of TP53 were sequenced at the Institute of Biomedical and Genetic Engineering (IBGE), Islamabad, Pakistan on ABI 3130 Genetic analyzer system while PTEN and AR genes were sequenced through Macrogen Inc. Korea. After sequencing analysis we found a number of polymorphisms, both reported and novel mutations in all the three genes on different exons (Table 4.8,

Table 4.15,
Table 4.18, Table 4.20, Table 4.23 and Table 4.25), discussed in detail in the proceeding sections one by one.

Table 4.5: TP53 gene primers

Primer ID	Sequence	Length	Product Size
TP53_2-3_F	TCTCATGCTGGATCCCCACT	20	344 bp
TP53_2-3_R	AGTCAGAGGACCAGGTCCTC	20	
TP53_4_F	TGCTCTTTTCACCCATCTAC	20	353 bp
TP53_4_R	ATACGGCCAGGCATTGAAGT	20	
TP53_5-6_F	TGTTCACTTGTGCCCTGACT	20	467 bp
TP53_5-6_R	TTAACCCCTCCTCCCAGAGA	20	
TP53_7_F	CTTGCCACAGGTCTCCCCAA	20	237 bp
TP53_7_R	AGGGGTCAGCGGCAAGCAGA	20	
TP53_7A_F	AGGCGCACTGGCCTCATCTT	20	177 bp
TP53_7AR	TGTGCAGGGTGGCAAGTGGC	20	
TP53_8-9_F	TTGGGAGTAGATGGAGCCT	19	445 bp
TP53_8-9_R	AGTGTTAGACTGGAAACTTT	20	
TP53_10_F	CAATTGTAACTTGAACCATC	20	260 bp
TP53_10_R	GGATGAGAATGGAATCCTAT	20	
TP53_11_F	AGACCCTCTCACTCATGTGA	20	245 bp
TP53_11_R	TGACGCACACCTATTGCAAG	20	

Table 4.6: PTEN gene primers

Primer ID	Sequence	Length	Product Size
PTEN_Ex1_F	TTCCATCCTGCAGAAGAAGC	20	236
PTEN_Ex1_R	TTTTCGCATCCGTCTACTCC	20	
PTEN_Ex2_F	CTCCAGCTATAGTGGGGAAAA	21	376
PTEN_Ex2_R	CTGTATCCCCCTGAAGTCCA	20	
PTEN_Ex3_F	CCATAGAAGGGGTATTTGTTGG	22	364
PTEN_Ex3_R	CAATGCTCTTGGACTTCTTGA	21	
PTEN_Ex4_F	TGGGGGTGATAACAGTATCTACTTA	25	399
PTEN_Ex4_R	TGCAATACTTTTTCCTAAAACACAA	25	
PTEN_Ex5_F	GGAATCCAGTGTTTCTTTTAAATACC	26	473
PTEN_Ex5_R	TGTCAATTACACCTCAATAAAACTGA	26	
PTEN_Ex6_F	CCTTTGAATAAATGGGTTGTTATG	24	499
PTEN_Ex6_R	CCTGCATAAATTTCAAATGTGG	22	
PTEN_Ex7_F	TCCATATTTCGTGTATATTGCTGA	24	398
PTEN_Ex7_R	AGCAAAACACCTGCAGATCTAA	22	
PTEN_Ex8_F	TCATGTGAATGAAAATGCAACA	22	475
PTEN_Ex8_R	ACAAGTCAACAACCCCCACA	20	
PTEN_Ex9_F	TGTTCATCTGCAAAATGGAATAA	23	387
PTEN_Ex9_R	CACAATGTCCTATTGCCATTAAA	23	

Table 4.7: AR gene primers

Primer ID	Sequence	Length	Product Size
AR_Ex1a_F	GCCTGTTGAACTCTTCTGAGC	21	398
AR_Ex1a_R	GTAGCCTGTGGGGCCTCTA	19	
AR_Ex1b_F	GCAGCAGCAGCAAGAGACTA	20	464
AR_Ex1b_R	TCGAAGTGCCCCCTAAGTAA	20	
AR_Ex1c_F	GCAGGAAGCAGTATCCGAAG	20	444
AR_Ex1c_R	GACAGGGTAGACGGCAGTTC	20	
AR_Ex1d_F	AAAGGGCTAGAAGGCGAGAG	20	380
AR_Ex1d_R	CCTTCTTCGGCTGTGAAGAG	20	
AR_Ex1e_F	GGTTCTGGGTCACCCTCAG	19	435
AR_Ex1e_R	GCAGGTAGGAGCCGCTAGAT	20	
AR_Ex2_F	GACCTGAGACTTCACTTGCCTA	22	300
AR_Ex2_R	GCCCTGAAAGGTTAGTGTCTC	21	
AR_Ex3_F	TGTTCTAGAAATACCCGAAGAAAGA	25	297
AR_Ex3_R	CCTTGGAAGCATCAAAGAAGA	21	
AR_Ex4_F	CTGTGACCAGGGAGAATGGT	20	497
AR_Ex4_R	GGCAGAAAAGCACCAGACAT	20	
AR_Ex5_F	CCCAACAGGGACTCAGACTT	20	300
AR_Ex5_R	GTCACCCCATCACCATCAC	19	

AR_Ex6_F	GGGATGGCAATCAGAGACAT	20	285
AR_Ex6_R	TTAATGGCAAAAGTGGTCCTC	21	
AR_Ex7_F	CCCAAGCACACAGACTTCAA	20	383
AR_Ex7_R	ACTCAAAGCCAGAGGGGAAT	20	
AR_Ex8_F	TTGGGGAAGAGGCTAGCAG	19	300
AR_Ex8_R	TGCAGAGTTATAACAGGCAGAA	22	

4.5 TP53

The discovery the TP53 gene, located on chromosome 17p13.1, in 1979 by three independent groups studying tumorigenesis is one of the greatest discoveries in the combat against cancer (Sarkar et al., 2002). It is universally known as “guardian of the genome” and regarded as a hope for novel therapeutics by many scientists.

It functions as a transcription factor and responds to a diverse set of stresses by inducing cell-cycle arrest, apoptosis, senescence, DNA repair and/or changes in metabolism. p53 remains inactive in normal conditions through the action of MDM2, which promote its degradation through ubiquitination and/or inhibiting its transcriptional activity. Its activities are modulated through several post-translational modifications including phosphorylation and acetylation. p53 loss its activities in human cancer either through its sequence mutations or by losing upstream or downstream cell signaling (Vousden and Lane, 2007).

Reisman and his co-workers characterized TP53 and identified two promoters in it. Promoter-1 is located 100-250bp upstream of the non-coding first exon while the second one which is comparatively stronger, is located in the first intron (Reisman et al., 1988). It has 11 exons, two transcriptional start sites in exon 1 and an alternative splicing site in intron 2 and between exons 9 and 10. It also contains an internal transcription initiation site in intron-4 (Bourdon et al., 2005).

We confirmed the conservativeness of p53 across different species phylogenetically through PhylomeDb4 and TreeFam (Figure 4.5) which confirmed that the gene/protein is conserved across the species.

p53 acts as a tumor suppressor, located in the nucleus throughout the body. It directly binds to DNA regulating cell division and avoiding uncontrolled growth thus preventing tumorigenesis (Ismaeel, 2013). It is p53 who decides whether the damaged DNA will either be repaired or cell with damage DNA will suffer apoptosis (Michael and Oren, 2003).

In this study, all the TP53 exons along with flanking regions from case-control samples of Pakistani population were sequence and analyzed. Eight exonic and two intronic variants were observed in the sequence analysis. Out of the 8-exonic variants

only one i.e. P72R (g.7676154) in the exon 4 is a well reported polymorphism of TP53 in different studies with reference to different disease in different ethnic group while the rest of the seven are novel mutations (Table 4.8). All the 8 mutations are in the DNA binding domain (Figure 4.6). Similarly both the intronic variants are also novel of our study and are most probably SNPs as were recorded in high frequency in both cases and controls (

Table 4.15).

Proteins are complex molecules, vital to the structure and function of every organism including viruses. They usually exist as compact and folded structures rather than as linear polypeptides. Genetic changes can affect protein structurally at different levels and thus we predicted the secondary structures of wild and mutant p53 proteins. All the predicted structures were then compared in 2D to highlight changes imposed by the genetic alterations. Function of a protein is defined by their overall three-dimensional conformation (Waheed et al., 2012). Structure of p53 has already been determined through crystallography and thus is present in PDB. The crystal structure was obtained from the PDB (Figure 4.6) with accession ID of 1TUP and was used as a template for predicting the mutant structures through homology modeling.

Structural knowledge of a protein provides information about its interactions (Aydin et al., 2011), which ultimately define its biological role (Cheng et al., 2005). It can be anticipated that, if not all, some of the genetic changes can get translated into higher order conformations of the respective proteins and thus tertiary structures were predicted for all the sequences. Structures are more conserved than sequence (Capriotti and Marti-Renom, 2010), thus mutant structures were aligned against the wild to measure the level of similarity among them (Figure 4.7)

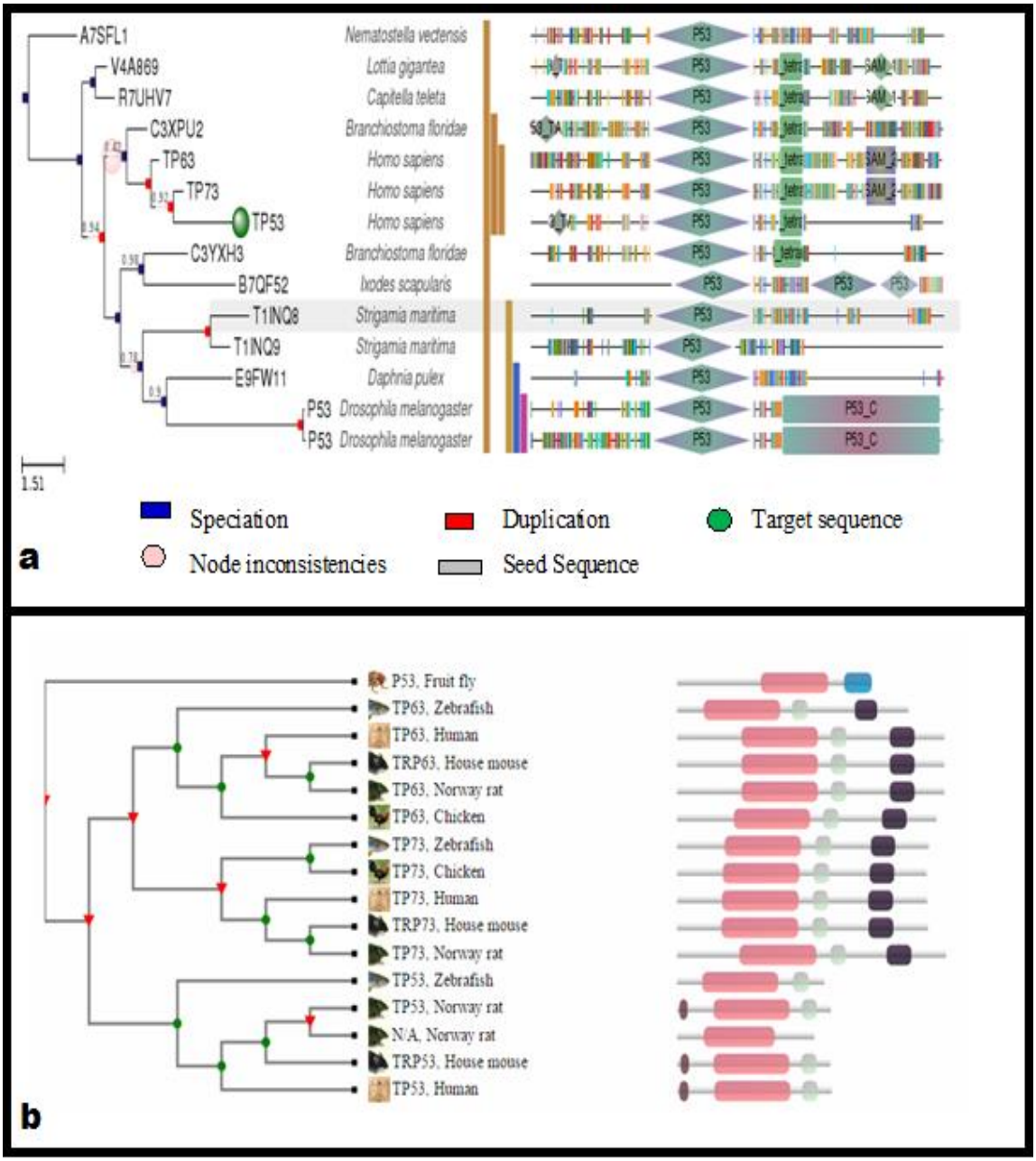

Figure 4.5: Evolutionary history of TP53 along with pfam domains.

a. TP53 along with pfam domains through PhylomDB4

b. TP53 gene tree through TreeFam

Table 4.8: TP53 Exonic variants

S. No	Exon	Ref Allele	VarAllele	Chr StartPosition	ChrEnd Position	Residue change	MutType	Reported By
1.	4	C	G	7676154	7676154	P72R	Missense	Reported
2.	5	-	G	7675075	7675075	H179Q	Frame Shift	Novel
3.	5	-	A	7675101	7675101	E171R	Frame Shift	Novel
4.	5	C	T	7675103	7675103	T170M	Missense	Novel
5.	5	C	G	7675153	7675153	P153P	Silent	Novel
6.	5	C	G	7675158	7675158	P152A	Missense	Novel
7.	6	T	G	7674923	7674923	V203G	Missense	Novel
8.	6	T	C	7674958	7674958	P191P	Silent	Novel

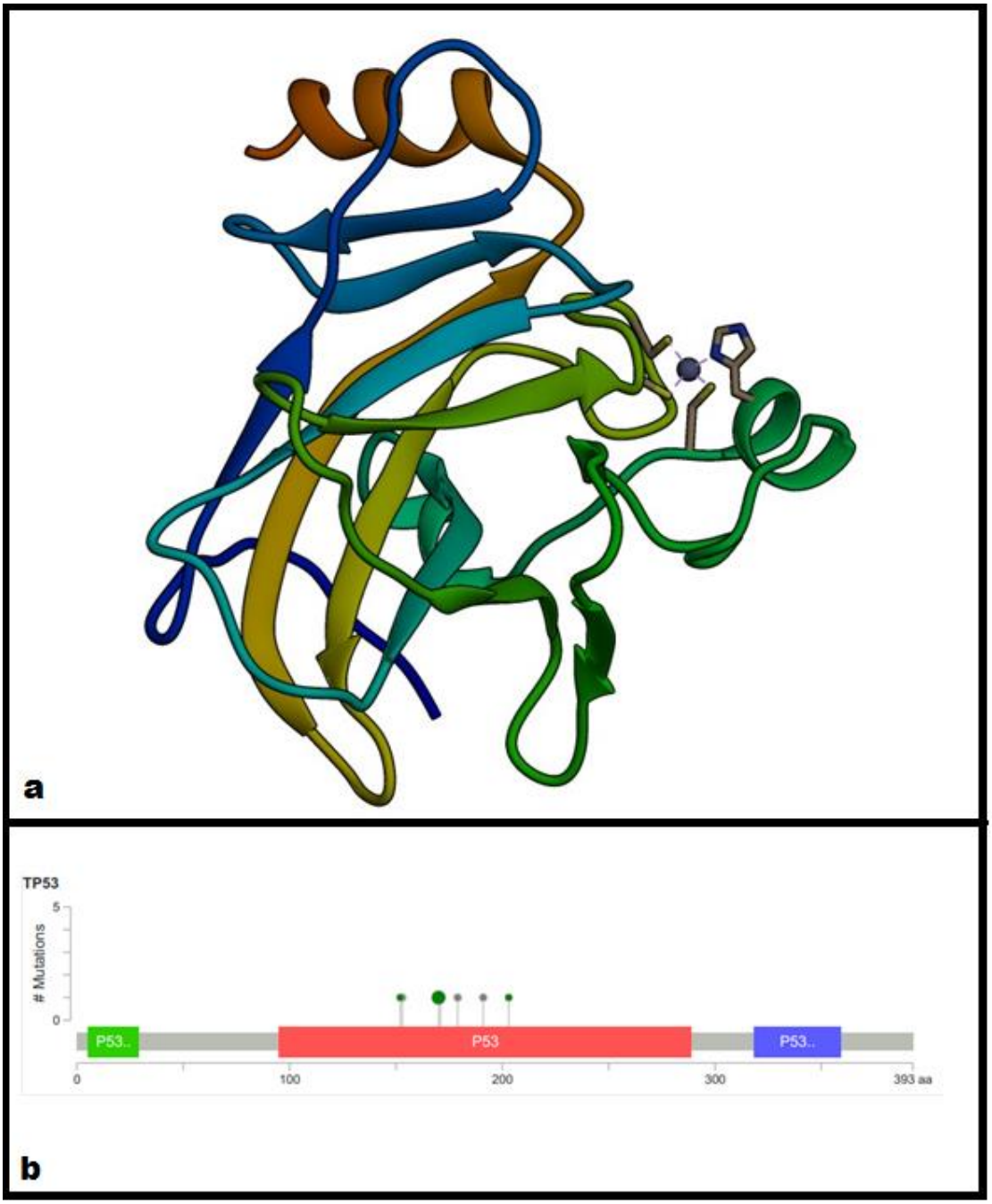

Figure 4.6: Structure of p53 protein.

a. Crystal structure of p53 (PDB ID: 1TUP) bound with Zinc ion

b. Lollipop plot of TP53 mutations of exonic regions in the sampled population, all the seven mutations are in the DNA binding domain.

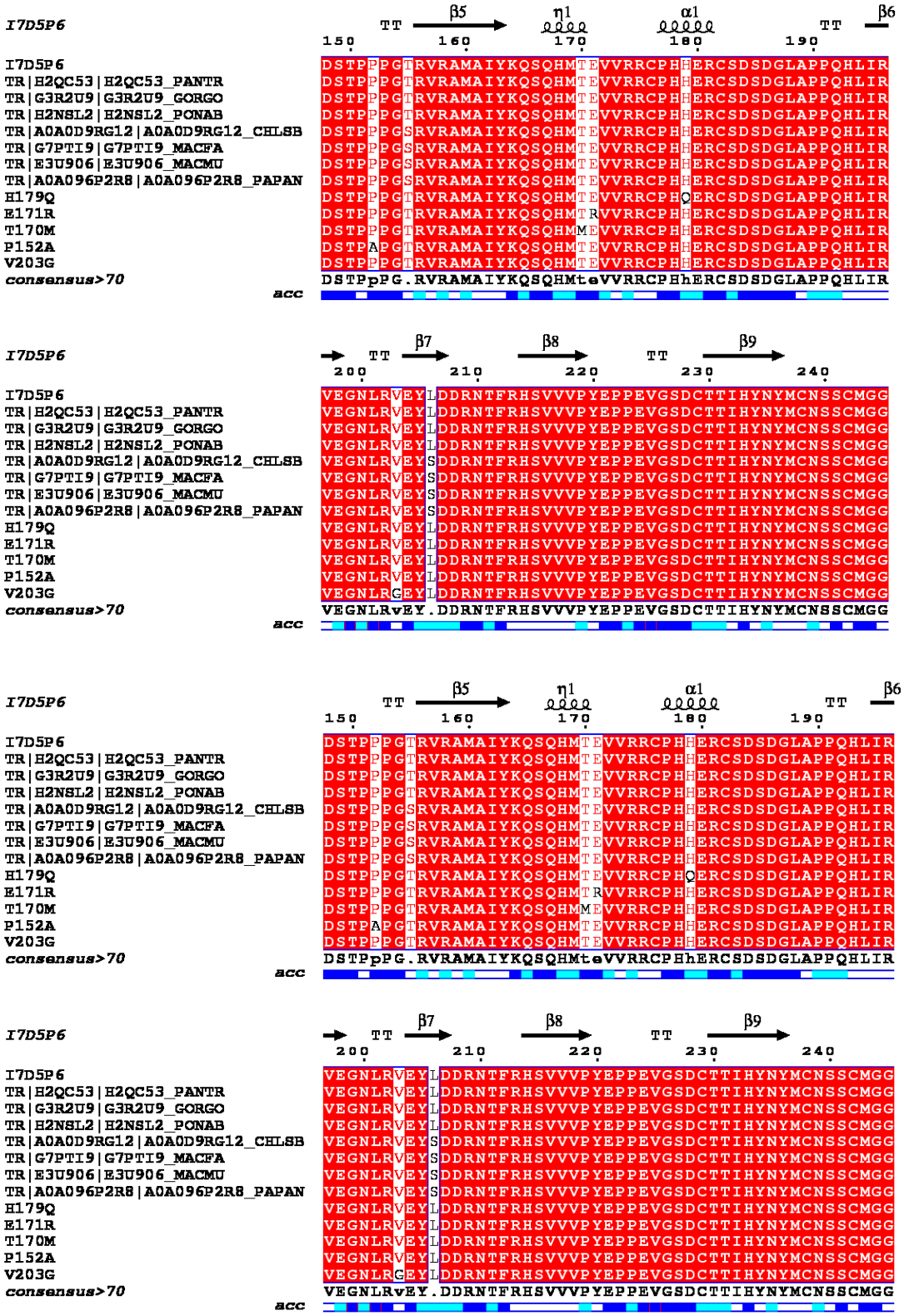

Figure 4.7: Partial snap of secondary structure analysis of wild and mutant p53 proteins along with some other species (mutations in the exonic regions only)

Table 4.9: Status of TP53 Exonic variants confirmed through HGMD and Ensembl along with mutants functional impact analysis through different servers

S. No	Residue change	Reported By	SIFT	PROVEAN	Mutation Assessor	PolyPhen2
1.	P72R	Reported	Tolerated	neutral	neutral	Benign
2.	H179Q	Novel	Damaging	Deleterious	Medium	Probably Damaging
3.	E171R	Novel	Damaging	Deleterious	Medium	Probably Damaging
4.	T170M	Novel	Damaging	Deleterious	Medium	Benign
5.	P152A	Novel	Damaging	Deleterious	Medium	Probably Damaging
6.	V203G	Novel	Damaging	neutral	neutral	Benign
7.	P153P	Novel	Tolerated	neutral	neutral	Benign
8.	P191P	Novel	Tolerated	neutral	neutral	Benign

4.5.2 C>G (g.7676154) P72R

P72R is a most common polymorphism which alters CCC>CGC at codon 72 in exon-4 of p53. It alters the p53 functionality and has recorded differentially distributed worldwide. It results in substitution of proline to arginine residue in the proline rich region, vital for p53-mediated apoptosis (Ricks-Santi et al., 2012). P72R polymorphism has been well studied by large number of researchers in different ethnic groups (Chosdol et al., 2002) but unfortunately, no attempt has been made for its association with prostate cancer in Pakistani population to the best of our knowledge. Critical review from different ethnic groups motivated us to explore its status and association in Pakistani population.

A total of 212 samples (120 cases and 92 controls) were genotyped for detecting P72R polymorphism through PCR-RFLP by using 5'TGCTCTTTTCACCCATCTAC3' (Forward) and 5'ACTTCAATGCCTGGCCGTAT3' (Reverse) primers. The detection was carried to distinguish C allele from G. The purified PCR product was digested BstU1 (Fermentas, Vilnius, Lithuania) enzyme at 57°C for 16hrs and the digested fragments were electrophoresed on 4% agarose gel. The expected fragment includes an uncut Pro allele of 353bp and the Arg allele to be restricted into two fragments of 212bp and 141bp while the heterozygous individuals will have all the three bands. 1/4 samples were repeated for quality control and we recorded >95% concordance. The homozygous Arg samples and heterozygous (Pro/Arg) were sequenced to confirm and validate our PCR-RFLP results.

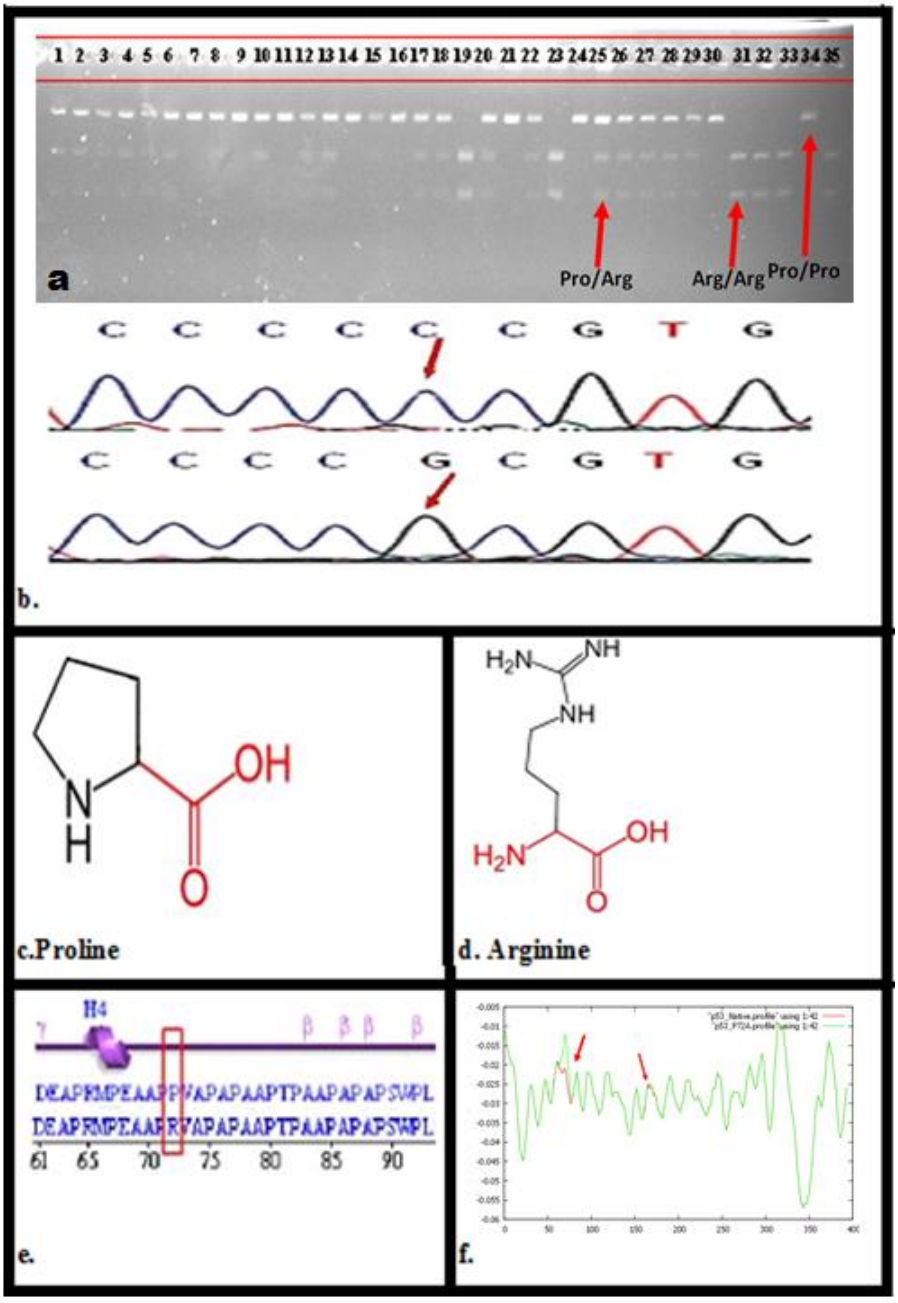

Figure 4.8 represents the pattern of P72R genotyping, while Table 4.10 is representing the genotype frequencies. Statistically significant differences were detected in genotypes distribution between cases and controls (df=2, P<0.001). It was also observed that the allelic frequencies were following Hardy-Weinberg Equilibrium with frequencies of 0.27 (Cases) and 0.78 (Controls). The association between P72R SNP and prostate cancer was evaluated by using un-conditional logistic regression under different genetic models. Highly significant associations were discovered in all

genetic models except the Dominant model (OR=0.77, 95%CI: 0.39-1.52, P=0.460) (Table 4.11).

The mutant residue was annotated at different levels to access its impacts on the phenotype. It can be seen that the polymorphism substitutes a comparatively bigger and a positively charged residue instead of a neutral residue. The mutant residue is comparatively less hydrophobic. Proline, the wild residue, is a very rigid residue and thus induces a special backbone conformation which might be required at this position while the polymorphic residue can disturb this special conformation.

The residue at position 72 in p53 is part of an interpro domain named "P53 Tumor Suppressor Family (IPR002117)", annotated with the GO terms, DNA Binding (GO:0003677) and Sequence-Specific DNA Binding Transcription Factor Activity (GO:0003700) to indicate its functions. It means that, the domain has its functions in Nucleic Acid Binding (GO:0003676) and Nucleic Acid Binding Transcription Factor Activity (GO:0001071). The polymorphism is in the proline rich region interacting with WWOX and HRMT1L2. This substitution which is bigger and less hydrophobic than the wild might cause bumps, changes in hydrophobicity and lead to loss of hydrophobic interactions. The mutation introduces a change in charge which can lead to repulsion of ligands.

The P72R polymorphism is associated with changes in the efficacy, efficiency and functionality of p53 (Murphy, 2006). The C allele (Lacey et al.) induces an enhanced cell-cycle arrest at G1 phase (Thut et al., 1995) while the G allele (Arg) promotes apoptosis (Dumont et al., 2003). The Arg residue is comparatively less stable thermodynamically than Pro residue. Pro promotes p53 induced repair against oxidative damage (Khoo et al., 2009) and has been hypothesized to be under selective pressure (Kiraga et al., 2007). It has also been hypothesized that Arg residue could provide selective advantage to the populations living in comparatively colder regions (Bensaad et al., 2006).

Table 4.10: Genotypes frequencies of TP53 P72R Polymorphism in Prostate cancer in Pakistani population

Genotype	Case (n=146)			Controls (n=107)		
	Number	Percentage	Frequency	Number	Percentage	Frequency
C allele	155	53.1	0.53	60	28.04	0.28
G allele	137	46.9	0.47	154	71.96	0.72
CC	27	18.5	0.19	16	14.95	0.15
CG	101	69.2	0.69	28	26.17	0.26
GG	18	12.3	0.12	63	58.88	0.59

Table 4.11: Association of P72R Polymorphism of TP53 and increased risk of Prostate cancer in Pakistani population.

Genetic Model	OR	Lower 95%CI	Upper 95%CI	Z-stat	P-value
G vs C (Allelic contrast)	0.34	0.24	0.50	-5.55	0.000
GG vs CC (Rare vs Common)	0.17	0.08	0.38	-4.30	0.000
GG vs CG (Rare vs Heterozygotes)	0.08	0.04	0.15	-7.41	0.000
CG vs CC (Heterozygotes vs Common)	2.14	1.01	4.51	1.99	0.046
GG+CG vs CC (Dominant)	0.77	0.39	1.52	-0.74	0.460
GG vs CC+CG (Recessive)	0.10	0.05	0.18	-7.27	0.000
GG/CG/CC (Log additive)	3.54	2.13	5.89	4.86	0.000

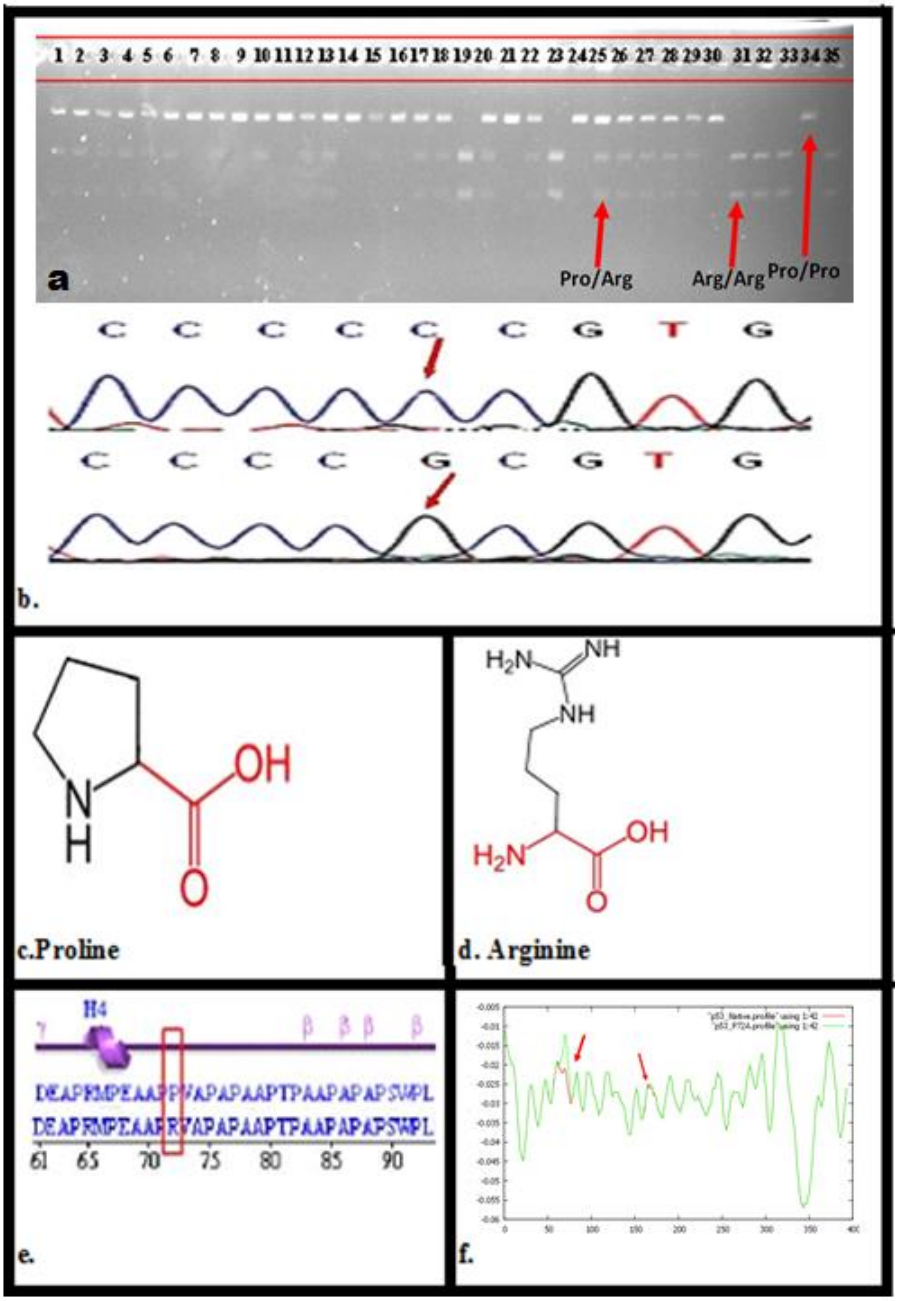

Figure 4.8: Gel electrophoresis, Sequencing and structural analysis of P72R polymorphism.

a. The uncut pro allele has a single band (353bp), Arg allele has restricted into two fragments (212bp and 141bp) by BstU1 restriction enzyme, while the heterozygotes have all three bands (353bp, 212bp and 141bp).

b. Sequence of TP53 exon 4 showing a common variant (rs1042522) at codon 72 (CGC to CCC).

c. & d. Structures of proline and arginine residues.

e. Partial diagram of secondary structure elements alignment of wild and polymorphic sequence.

f. Graphic representation of the changes endorsed by the A72P polymorphism in the p53 protein structure.

4.6 P72R Meta-analysis

Critical literature review of P72R polymorphism of TP53 showed notorious results about its association with prostate cancer. We therefore designed a study to explore this correlation in a systematic review and meta-analysis of published literature from different countries and different ethnic groups. It is hoped that this study will help to better understand the disease development and the biological phenomenon increasing PCa risk.

Literature search was carried out through different online resources including Science direct, Pubmed, Chinese biomedical literature, Human genetic mutation database and Google scholar from 1999-2013 by using the keywords "TP53 Pro72Arg polymorphism OR TP53 P72R polymorphism OR TP53 single nucleotide polymorphism OR TP53 codon 72 polymorphism OR rs1042522 and prostate adenocarcinoma OR prostate cancer OR PCa". Only full text articles were retrieved initially but the inclusion criterion was then narrowed down and articles having full desired information were only included in the study. All the papers were manually scanned for their inclusion on the basis of following parameters, (i) case–control studies with full Author/s, publication year, population and publication country information were selected (Glinsky et al.) studied having information about the number of cases and controls and (iii) genotypes information of both cases and controls. A flow diagram of the collected data was constructed according to the PRISMA statement (

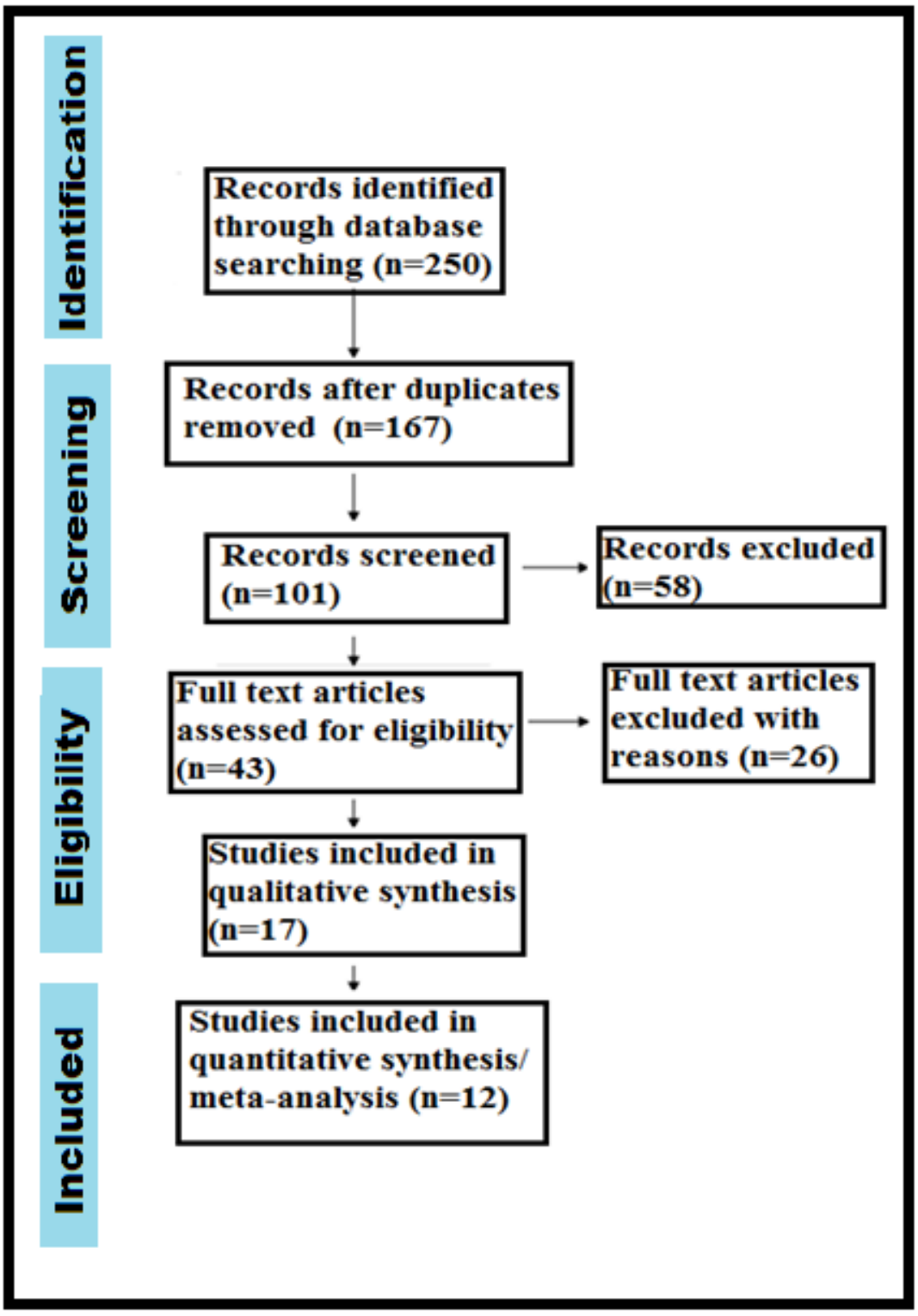

Figure 4.9). The required data was then carefully extracted as whole from each qualified study, their genotypes were analyzed and the Meta-analysis was performed on the basis of both the fixed and random effect models.

Although a very strict and precise inclusion criterion was set, heterogeneity still exists due to a number of probable factors. This inter-study heterogeneity was assessed more precisely by using I-squared, Cochran's Q and Chi2-p statistics to quantitatively assess the percent deviation due to heterogeneity. Results of meta-analysis were graphical represented through Forest plot while publication biasness through the funnel plots on the basis of different genetic models.

A total of 250 relevant journal papers were retrieved in the defined time frame. Only 12 papers were satisfying the required selection criteria and thus data was extracted only from them in addition to our own study. The included studies comprised of a total of 1686 cases and 1888 controls (Table 4.12). When all the data was pooled, a significant association was observed in between the P72R polymorphism and increased risk of the disease in all the genetic models (Figure 4.10). Begg's and Egger's test were used to predict the publication biasness along with funnel plots. The results showed significant heterogeneity among the included studies. Cautious exploration of the results highlighted that population is the most significant causative factor for this heterogeneity. It can also be inferred that the heterogeneity is due to the differential distribution of the genotypes among different populations (Figure 4.11).

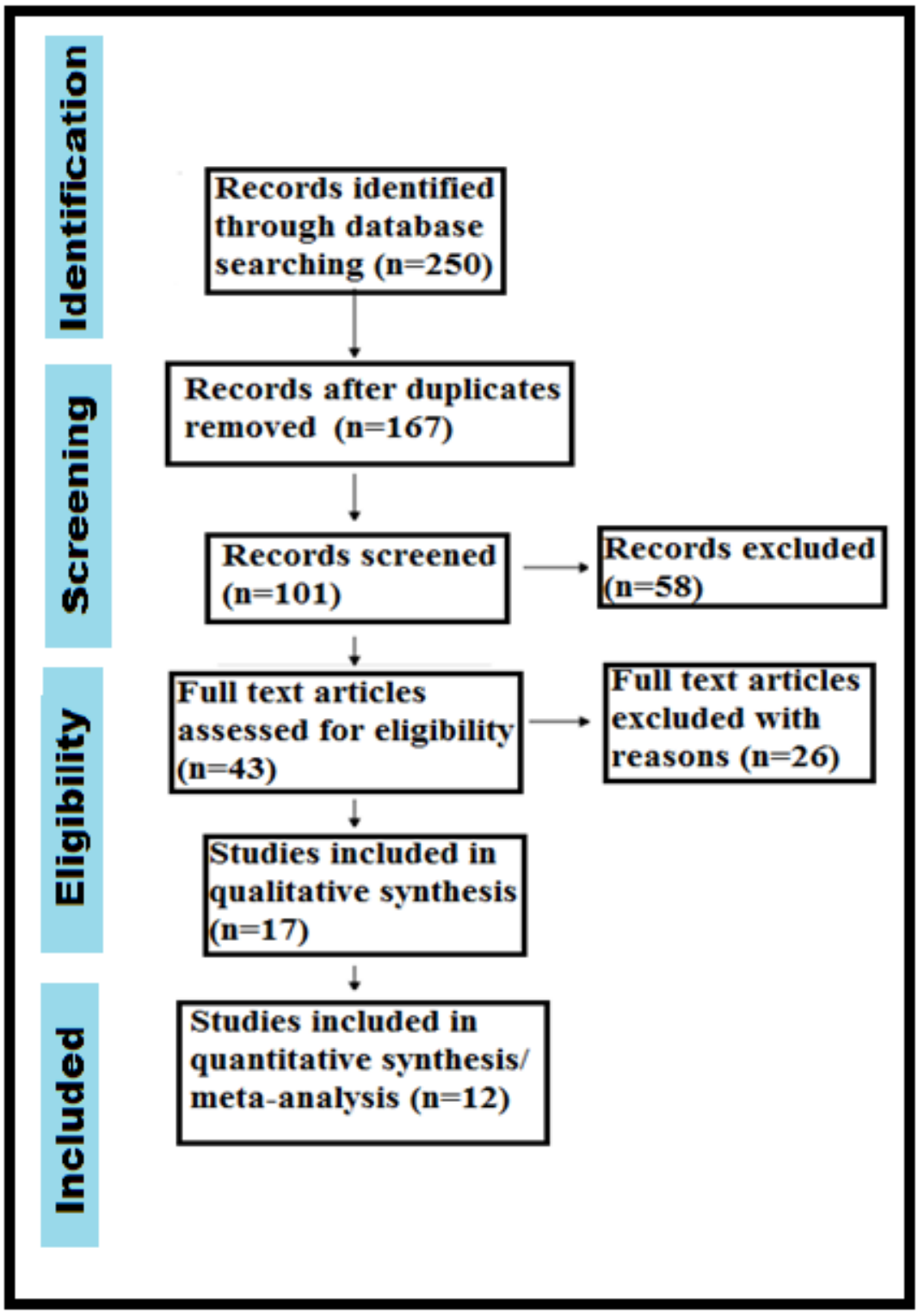

Figure 4.9: **Flow diagram representing stepwise inclusion and exclusion of the studies in the Meta-analysis according to the PRISMA statement.**

Table 4.12: Genotypes details of included studies

Author/s	Population	Cases				Controls			
		No.	CC	GC	GG	No.	CC	GC	GG
Henner et al., 2001	Caucasian	109	66	41	2	146	93	38	15
Suzuki et al., 2003	Japanese.	114	20	46	48	105	7	57	41
Huang et al., 2004	Taiwanese	200	66	92	42	247	54	109	84
Wu et al., 2004	Taiwanese	92	20	61	11	126	30	53	43
Leiros et al., 2005	Caucasians	39	2	17	20	48	2	23	23
Quinones et al., 2006	Chile (Caucasian)	60	14	24	22	117	13	45	59
Hirata et al., 2007	Japanese	167	22	89	56	167	26	80	61
Xu et al., 2010	southern Chinese	209	41	129	39	268	86	140	42
Ricks-Santi et al., 2010	African descent	245	73	135	37	178	70	86	22
Doosti and Dehkordi. 2011	Iranian	187	15	98	74	185	24	111	50
Rogler et al. 2011	Caucasian	118	9	44	65	194	11	79	104
Our Study, 2013	Pakistani	146	27	101	18	107	16	28	63

Table 4.13: Log Odd Ratios (LOR), Standard Error (SE), Variance (Var) and P-value (P) of the included studies under allelic contrast, GG vs CC and GG vs GC models.

Author/s	Allelic Contrast				GG vs CC				GG vs GC			
	LOR	SE	Var	P	LOR	SE	Var	P	LOR	SE	Var	P
Henner et al., 2001	-0.15	0.22	0.05	0.48	-1.67	0.77	0.59	0.03	-2.09	0.79	0.62	0.01
Suzuki et al., 2003	-0.17	0.20	0.04	0.39	-0.89	0.49	0.24	0.07	0.37	0.29	0.08	0.20
Huang et al., 2004	-0.49	0.14	0.02	0.00	-0.89	0.26	0.07	0.00	-0.52	0.24	0.06	0.03
Wu et al., 2004	-0.40	0.19	0.04	0.04	-0.96	0.44	0.20	0.03	-1.50	0.39	0.15	0.00
Leiros et al., 2005	0.06	0.34	0.12	0.86	-0.14	1.05	1.09	0.89	0.16	0.44	0.20	0.71
Quinones et al., 2006	-0.56	0.23	0.05	0.02	-1.06	0.46	0.21	0.02	-0.36	0.36	0.13	0.31
Hirata et al., 2007	-0.01	0.16	0.03	0.94	0.08	0.34	0.12	0.81	-0.19	0.24	0.06	0.42
Xu et al., 2010	0.31	0.13	0.02	0.02	0.67	0.29	0.09	0.02	0.01	0.25	0.06	0.98
Ricks-Santi et al., 2010	0.26	0.14	0.02	0.07	0.48	0.32	0.10	0.13	0.07	0.30	0.09	0.82
Doosti and Dehkordi. 2011	0.37	0.15	0.02	0.01	0.86	0.38	0.14	0.02	0.52	0.23	0.05	0.02
Rogler et al. 2011	-0.01	0.19	0.04	0.95	-0.27	0.48	0.23	0.57	0.12	0.25	0.06	0.64
Our Study, 2013	-1.07	0.19	0.04	0.00	-1.78	0.41	0.17	0.00	-2.54	0.34	0.12	0.00

Table 4.14: Log Odd Ratios (LOR), Standard Error (SE), Variance (Var) and P-value (P) of the included studies under GC vs CC, Dominant and Recessive models.

Author/s	GC vs CC				Dominant Model				Recessive Model			
	LOR	SE	Var	P	LOR	SE	Var	P	LOR	SE	Var	P
Henner et al., 2001	-0.42	0.28	0.08	0.13	0.13	0.26	0.07	0.61	-1.81	0.76	0.58	0.02
Suzuki et al., 2003	1.26	0.48	0.23	0.01	-1.09	0.46	0.21	0.02	0.13	0.28	0.08	0.65
Huang et al., 2004	0.37	0.23	0.05	0.11	-0.57	0.22	0.05	0.01	-0.66	0.22	0.05	0.00
Wu et al., 2004	-0.55	0.34	0.12	0.11	0.12	0.33	0.11	0.72	-1.34	0.37	0.14	0.00
Leiros et al., 2005	0.30	1.05	1.10	0.77	-0.22	1.02	1.05	0.83	0.13	0.43	0.19	0.75
Quinones et al., 2006	0.70	0.46	0.21	0.13	-0.89	0.42	0.18	0.04	-0.56	0.33	0.11	0.08
Hirata et al., 2007	-0.27	0.33	0.11	0.40	0.20	0.31	0.10	0.53	-0.13	0.23	0.05	0.57
Xu et al., 2010	-0.66	0.23	0.05	0.00	0.66	0.22	0.05	0.00	0.21	0.24	0.06	0.39
Ricks-Santi et al., 2010	-0.41	0.22	0.05	0.06	0.42	0.21	0.04	0.04	0.23	0.29	0.08	0.42
Doosti and Dehkordi. 2011	-0.35	0.36	0.13	0.33	0.54	0.35	0.12	0.12	0.57	0.22	0.05	0.01
Rogler et al. 2011	0.38	0.49	0.24	0.43	-0.32	0.47	0.22	0.50	0.06	0.23	0.05	0.80
Our Study, 2013	-0.76	0.38	0.15	0.05	-0.25	0.34	0.12	0.46	-2.32	0.32	0.10	0.00

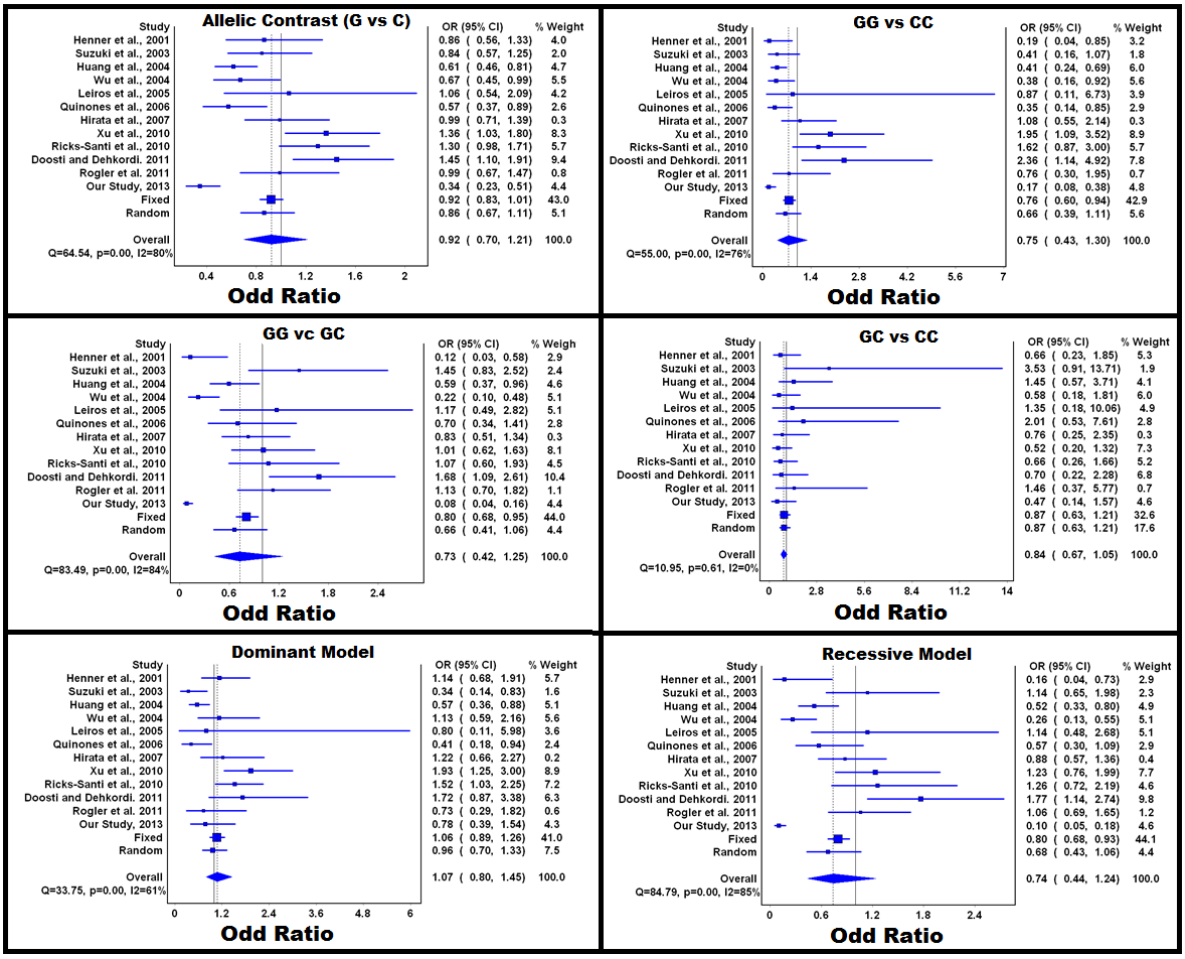

Figure 4.10: Graphical representation of the Meta-analysis results through forest plots in different populations according to different genetic models.

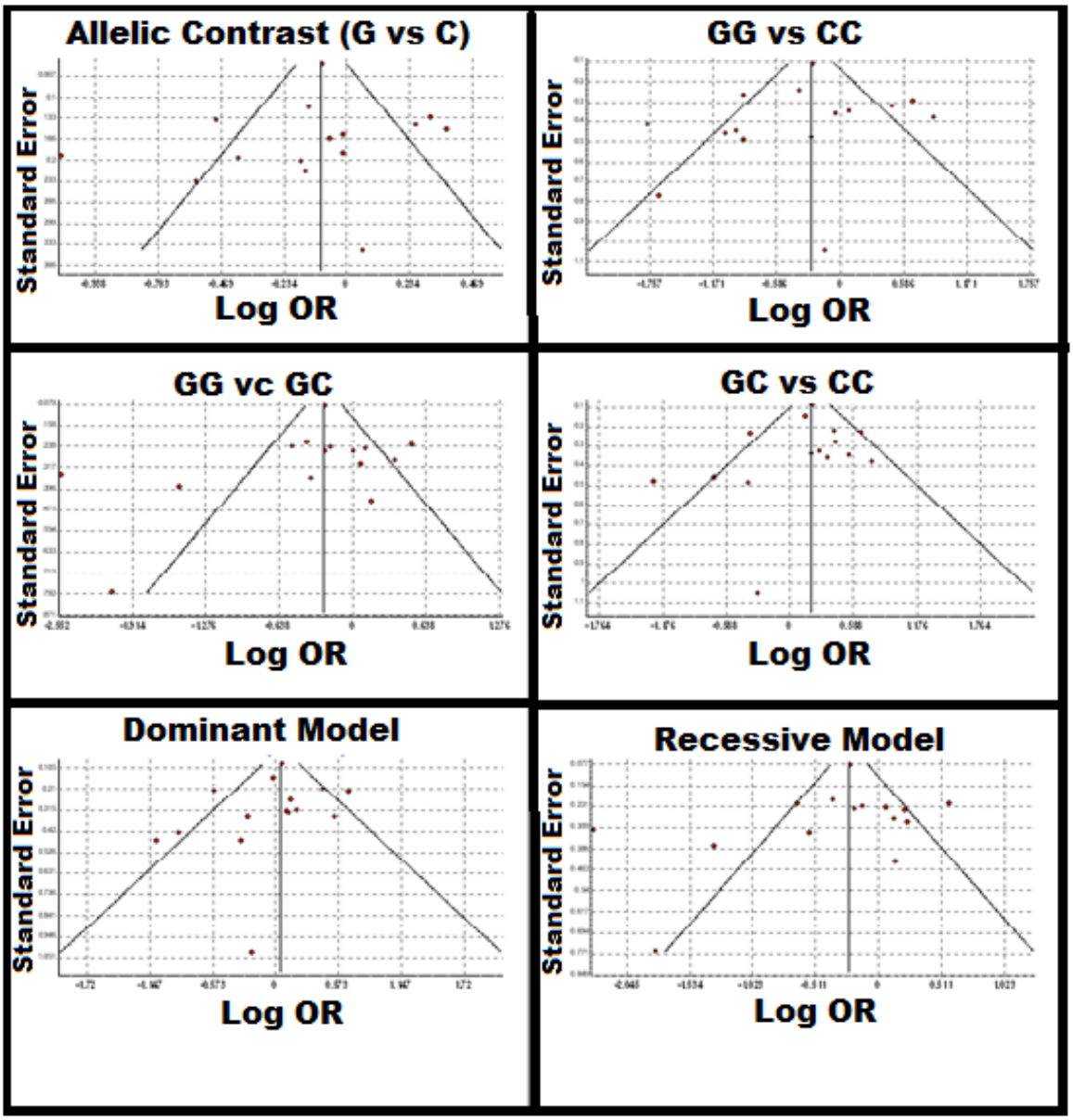

Figure 4.11: Funnel plots representing publication biasness in different populations according to different genotype models.

4.6.1 Ins162G (g.7675075) H179Q

The mutation was confirmed as novel through HGMD and Ensemble. The insertion mutations is in exon-5 position 162 of TP53, results in the substitution of Histidine to Glutamine at position 179 and frame shit downstream. The wild residue was comparatively bigger, basic and Hydrophilic with pK=6.0 while the mutant residue has no charge.

The residue is part of an interpro domain named P53 Tumor Suppressor Family (IPR002117), annotated with the GO terms (GO:0003677) and (GO:0003700) to indicate its function. These annotations indicate that the domain has a function in Nucleic Acid Binding (GO:0003676) and Nucleic Acid Binding Transcription Factor Activity(GO:0001071). The residue has also annotated as part of an interpro domain named P53/runt-Type Transcription Factor, DNA-Binding Domain (IPR012346), an interpro domain named P53, DNA-Binding Domain (IPR011615), P53-Like Transcription Factor and DNA-Binding (IPR008967).

The mutation is positioned in a domain important for binding of other molecules. The mutation could affect this interaction and thereby disturb regulation of the protein by disturbing these contacts. The mutation might affect the domain interactions and thereby disturb signal transfer from binding domain to the activity domain (Figure 4.12). Functionally the mutation was predicted as damaging by SIFT, probably damaging by PolyPhen2, deleterious by PROVEAN and medium by MutationAssessor (**Table 4.9**).

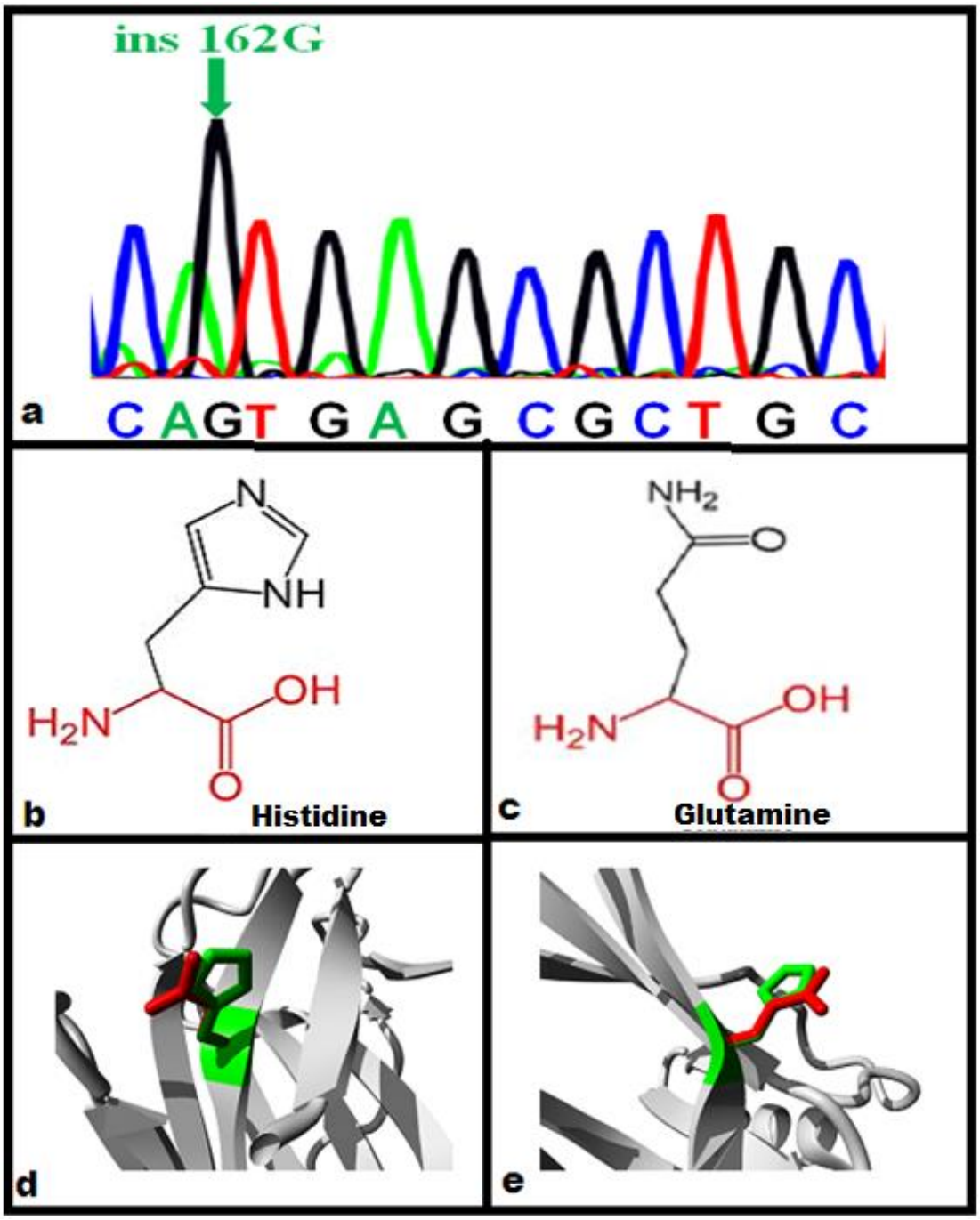

Figure 4.12: ins162G in the sequence and structure of TP53

a. Sequence Chromatogram showing ins 162G in the TP53 exon-5
b. Structure of wild residue Histidine
c. Structure of mutant residue Glutamine.
d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of p53.

4.6.2 C134T (g.7675103) T170M and ins135A (g.7675101) E171R

i. C134T (T170M)

The mutation was confirmed as novel through HGMD and Ensemble. The transition of C to T nucleotide at position 134 in exon-5 of TP53 results in the substitution of Threonine to Methionine at position 170. The wild type residue was buried in the core of the protein. The mutant residue is bigger and probably will not fit. The mutant residue is more hydrophobic than the wild. The mutation will cause loss of hydrogen bonds in the core of the protein and as a result disturb correct folding. The wild type residue is predicted to be located in its preferred secondary structure, β-strand. The mutant residue does not prefer beta-strand confirmation and therefore the local conformation will be slightly destabilized.

The wild type residue occurs often at this position in the sequence, but other residues have also been observed in other homologous sequences. This residue is part of an interpro domain named P53 Tumor Suppressor Family (IPR002117), annotated with the following Gene-Ontology (GO) terms DNA Binding (GO:0003677) and Sequence-Specific DNA Binding Transcription Factor Activity (GO:0003700)to indicate its function. More broadly speaking, these GO annotations indicate that the domain has a function in Nucleic Acid Binding (GO:0003676) and Nucleic Acid Binding Transcription Factor Activity(GO:0001071). This residue is also annotated as part of an interpro domain named P53/runt-Type Transcription Factor, DNA-Binding Domain (IPR012346), interpro domain named P53, DNA Binding Domain (IPR011615).This domain is annotated with the GO terms Transcription Regulatory Region DNA Binding (GO:0044212).

The mutated residue is located in a domain that is important for binding of other molecules and is in contact with a regulatory domain. The mutation could affect this interaction and thereby disturb regulation of the protein and is possible that the mutation disturbs these contacts. The mutation might disturb the interaction between these two domains and as such affect the function of the protein (Figure 4.13). Functionally the mutation was predicted as damaging by SIFT, benign by PolyPhen2, deleterious by PROVEAN and medium by MutationAssessor (**Table 4.9**).

ii. Ins 135A (g.7675101) E171A

The mutation was confirmed as novel through HGMD and Ensemble. This insertion of A at position 135 in exon-5 of TP53 results in the substitution of Glutamic acid to Arginine at position 171 and a frame shift downwards. The mutant residue is bigger than the wild type and is positively charged, while the wild residue was negatively charged and making hydrogen bond with arg-249. The size difference makes it difficult for the mutant to make the same hydrogen bond. The wild residue also making a salt bridge with trp-91, arg-74 and arg-249, the charge differences will disturb these interactions. It was also recorded that the wild residue was much conserved at this position and thus on basis of conservation scores it may be probably damaging.

This residue is part of an interpro domain named P53 Tumour Suppressor Family (IPR002117), annotated with the GO terms i.e. DNA Binding (GO:0003677) and Sequence-Specific DNA Binding Transcription Factor Activity (GO:0003700). These GO annotations indicates that the domain has a function in Nucleic Acid Binding (GO:0003676) and Nucleic Acid Binding Transcription Factor Activity (GO:0001071). In another annotation this residue is part of an interpro domain named P53/runt-Type Transcription Factor, DNA-Binding Domain (IPR012346), P53 DNA-Binding Domain (IPR011615), annotated with the GO term, Transcription Regulatory Region DNA Binding (GO:0044212).

The mutant residue is part of a domain interacting with other molecules and also is in direct contact with a regulatory region. The ins135A may possibly disturb this interaction and thus disturbs the regulation of the protein (Figure 4.13). Functionally the mutation was predicted as damaging by SIFT, benign by PolyPhen2, deleterious by PROVEAN and medium by MutationAssessor (Table 4.9).

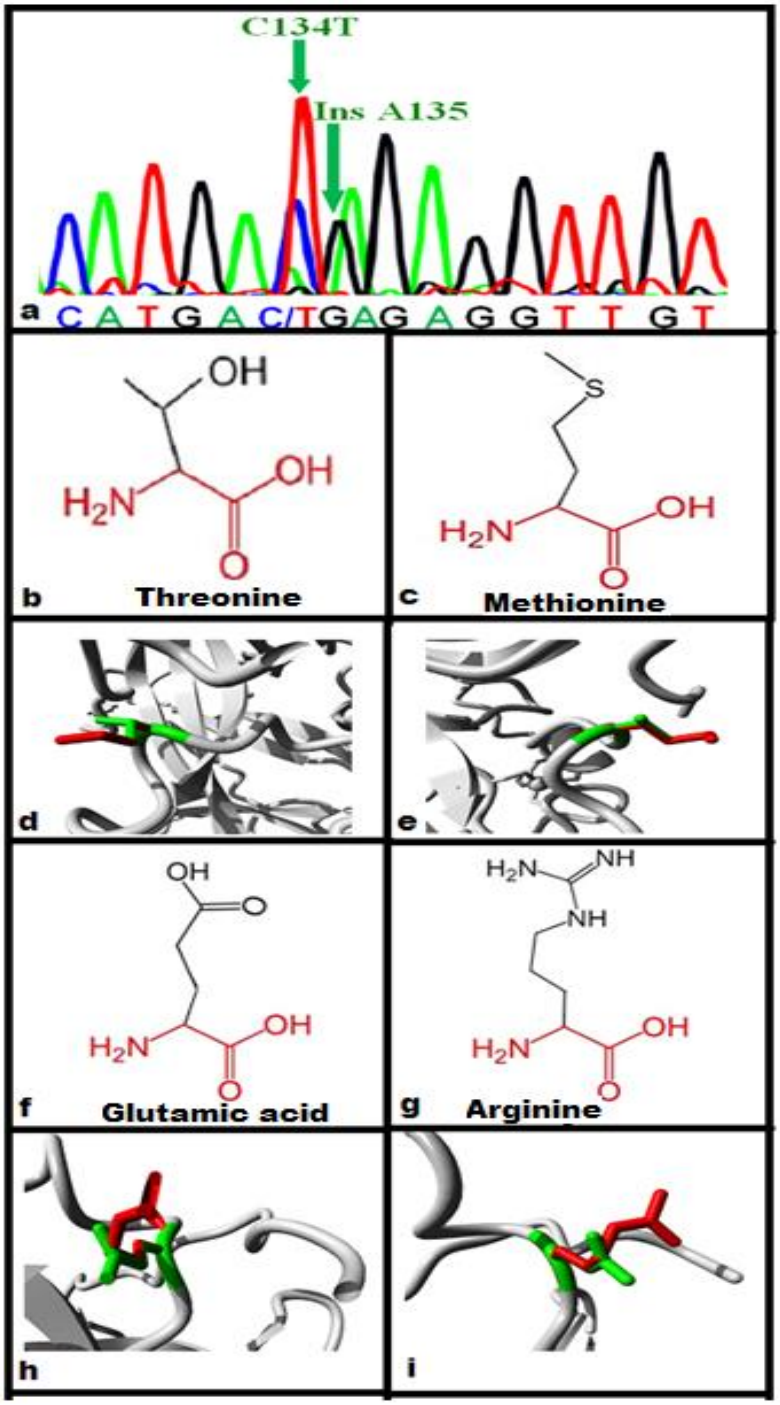

Figure 4.13: C134T and ins 135A in the sequence and structure of TP53

a. Sequence Chromatogram showing ins C134T and ins 135A in the TP53 exon-5

b. Structure of wild residue threonine c. mutant residue Methionine

d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of p53.

f. Structure of wild residue glutamic acid g. mutant residue arginine

h. & i. Comparison of mutant (red) and wild type (green) residues in the 3D structure of p53.

4.6.3 C79G (g.7675158) P152A

The mutation was confirmed as novel through HGMD and Ensemble. The C to G transversion at position 79 in exon-5 of TP53 gene results in the substitution of Proline to Alanine at position 152. Proline is very rigid and hence induces a special backbone conformation. The substituted residue is comparatively smaller and thus can disturb the original special conformation. The wild residue is conserved at this position and based on conservation scores C79G might probably be damaging to the protein.

This residue at position 152 is part of an interpro domain named, P53 Tumor Suppressor Family (IPR002117), annotated with the Gene-Ontology (GO) terms DNA Binding (GO:0003677) and Sequence Specific DNA Binding Transcription Factor Activity (GO:0003700). These GO annotations indicates that the domain has its function in Nucleic Acid Binding (GO:0003676) and Nucleic Acid Binding Transcription Factor Activity (GO:0001071). It is also annotated as part of an interpro domain named P53/runt-Type Transcription Factor, DNA-Binding Domain (IPR012346), and interpro domain named P53, DNA-Binding Domain (IPR011615).

The mutated residue is located in a domain that is important for binding of other molecules and also is in contact with a regulatory domain. The mutation could affect this interaction and thereby disturb regulation of the protein. It is also possible that the mutation disturbs these protein contacts and might affect interaction and thereby disturb signal transfer from binding domain to the activity domain (Figure 4.14). Functionally the mutation was predicted as damaging by SIFT, probably damaging by PolyPhen2, deleterious by PROVEAN and medium by MutationAssessor (**Table 4.9**).

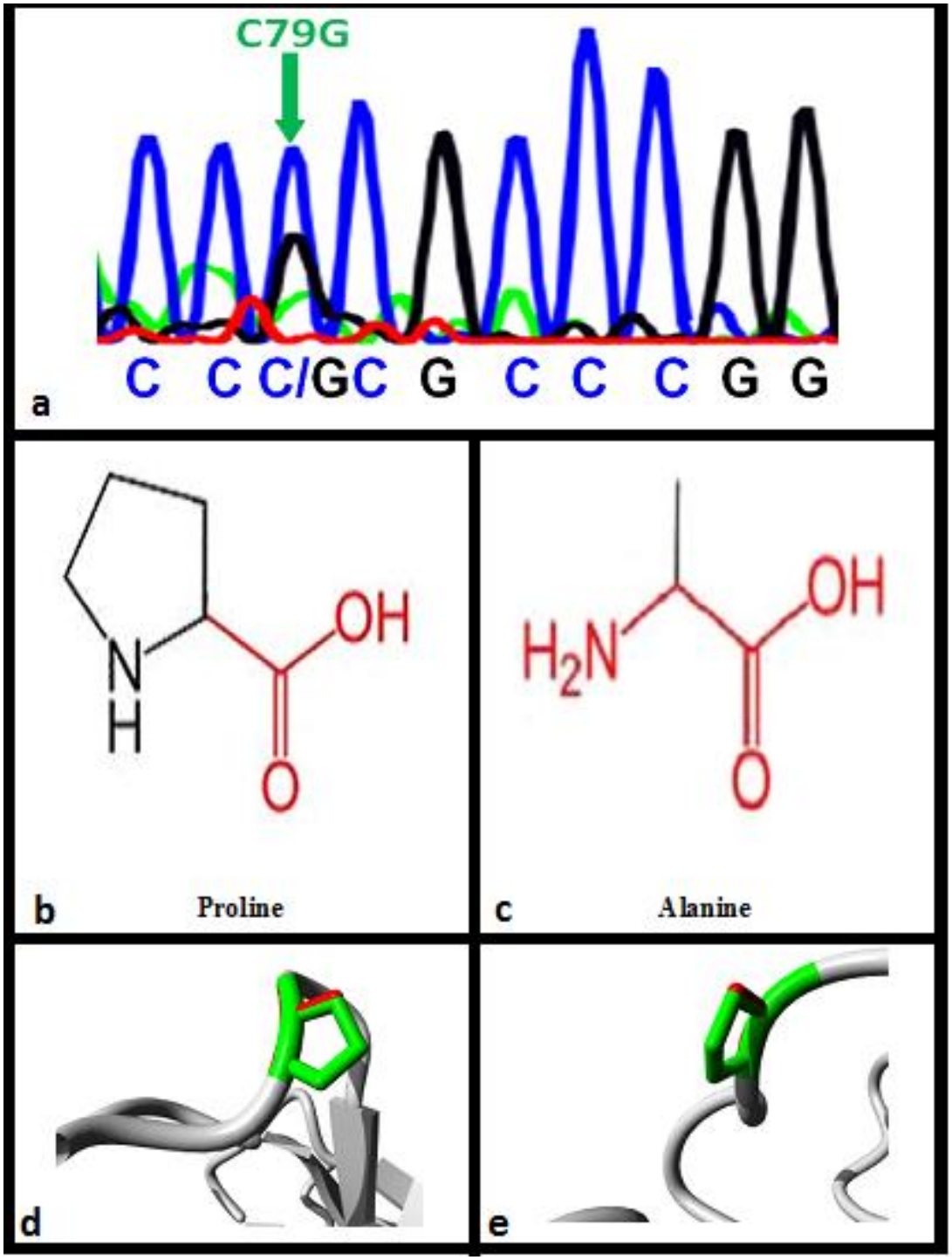

Figure 4.14: Sequence and structural representation of C79G Mutation TP53 exon-5.

a. Sequence Chromatogram showing C79G transversion.
b. Structure of Wild residue Proline.
c. Structure of mutant residue Alanine.
d. & e. Illustrations of comparison in the 3D structure of p53, wild residue is green while mutant is in red.

4.6.4 T49G (g.7674923) V203G

The mutation was confirmed as novel through HGMD and Ensemble. This T to G transversion at position 49 in exon-6 of TP53 gene on the negative strand results in substitution of Valine to Glycine at position 203. The mutant residue is smaller and less hydrophobic than the wild residue. Glycine is very flexible and can disturb the required rigidity of the protein at this position.

This residue is part of an interpro domain named P53 Tumor Suppressor Family (IPR002117), annotated with the Gene-Ontology (GO) terms, DNA Binding (GO:0003677) and Sequence-Specific DNA Binding Transcription Factor Activity (GO:0003700). These GO annotations indicate that the domain has a function in Nucleic Acid Binding (GO:0003676) and Nucleic Acid Binding Transcription Factor Activity (GO:0001071). This residue is also annotated as part of an interpro domain named P53/runt-Type Transcription Factor, DNA-Binding Domain (IPR012346), P53 DNA-Binding Domain (IPR011615) and P53-Like Transcription Factor, DNA-Binding (IPR008967).

The mutated residue is located in a domain that is important for binding of other molecules and is in contact with a regulatory domain. The mutation could affect this interaction and thereby disturb regulation of the protein and might also can affect its interaction and thereby disturb signal transfer from binding domain to the activity domain. Due to the comparatively smaller size of the mutant residue, it will cause an empty space in the core of the protein and also will result in of hydrophobic interactions in the protein core (Figure 4.15). Functionally the mutation was predicted as damaging by SIFT, probably damaging by PolyPhen2, deleterious by PROVEAN and medium by MutationAssessor (**Table 4.9**).

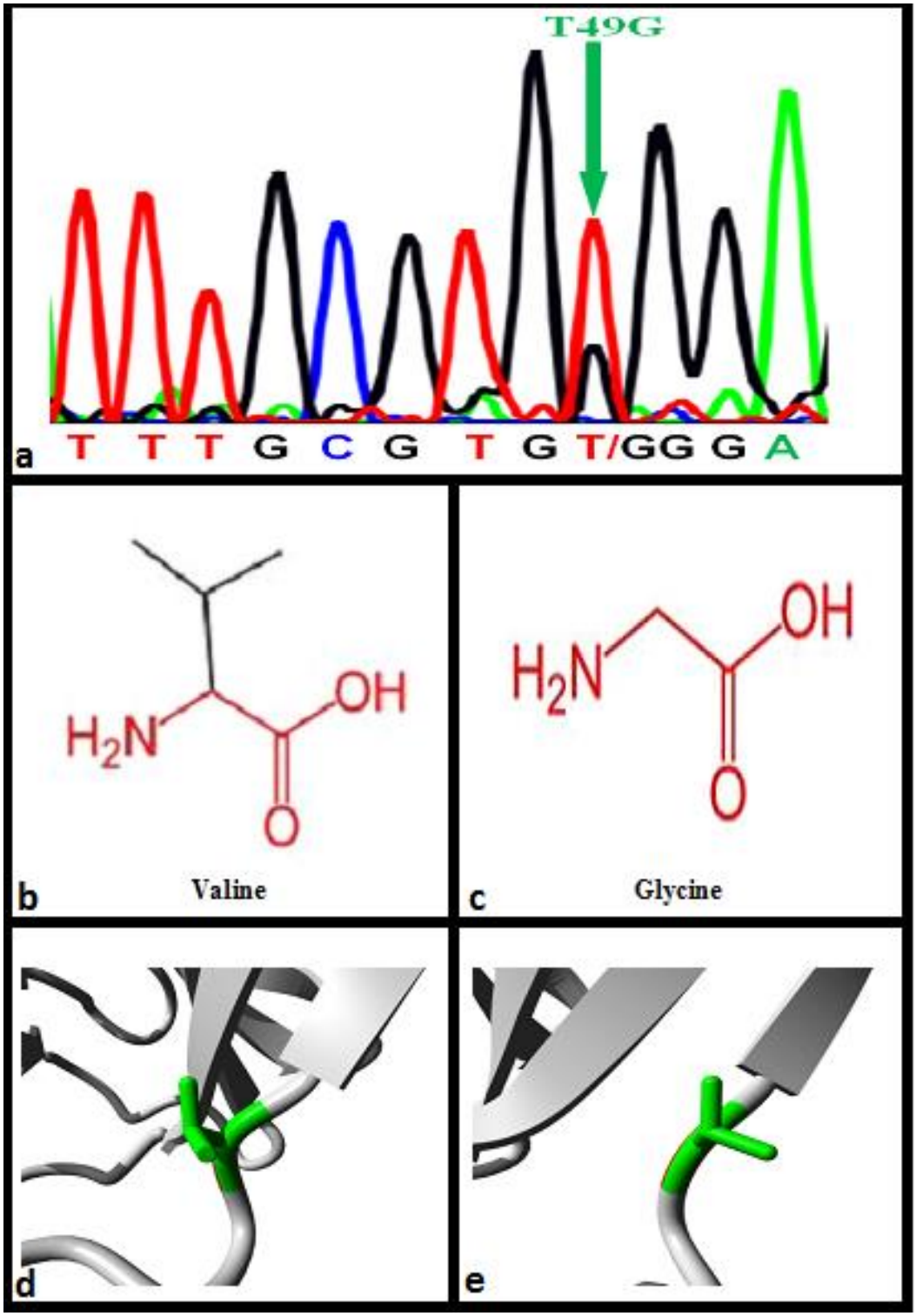

Figure 4.15: Sequence and structural representation of T49G Mutation TP53 exon-6.

a. Sequence Chromatogram showing T49G transversion.
b. Structure of Wild residue valine.
c. Structure of mutant residue glycine.
d. & e. Illustrations of comparison in the 3D structure of p53, wild residue is in green while mutant is in red.

4.6.5 C80G (g.7675153) P153P

This transversion at codon 153 in exon-5 of TP53 gene at position 79 is a silent mutation, confirmed as a novel one from HGMD and Ensembl. The substitution will have no effect on the protein structure and function due to codon degeneracy. The mutation was predicted as neutral and benign by the online resource (**Table 4.9**).

4.6.6 T14C (g.7674958) P191P

This transition at codon 191 in exon-6of TP53 gene at position 14 is also a silent mutation, confirmed as a novel one from HGMD and Ensembl. The substitution will have no effect on the protein structure and function due to codon degeneracy. The mutation was predicted as neutral and benign by the online resource (**Table 4.9**).

4.7 TP53 Intronic Variants

We observed two novel intronic variants in the sampled population presented in

Table 4.15. It was also noted that these intronic variants were recorded both in cases and controls and thus were further analyzed in detail for their possible association with the disease.

The analysis of T/A at g.7675016 reveals that A allele carriers showed significantly positive association with increased risk of PCa in allelic contrast (OR=1.84; 95%CI=1.29-2.62; P = 0.001), TA vs TT(OR=2 .25; 95%CI=1.37-3.69; P = 0.001) and Dominant model (OR=2.13; 95%CI=1.38-3.28; P = 0.001). At the same time, when comparing the case-controls in other models, we found a statistically non-significant association, for AA vs TT (OR = 1.85; CI = 0.92-3.73; P = 0.08), AA vs TA (OR = 0.82; CI = 0.38-1.81; P =0.63), and for the Recessive model (OR = 1.52; CI = 0.77-3.03; P = 0.230).

For the G allele of g.7674991, we found a positive association with increased rick of PCa, with an OR of 1.52 (CI = 1.15-2.01; P=0.00) for allelic contrast, an OR of 2.25 (CI = 1.30-3.88; P=0.00) for TT vs CC model, OR of 2.20 (CI = 1.34-3.62; P=0.00) for TC vs CC model and an OR of 2.40 (CI = 1.51-3.82; P=0.00) for dominant model.

At the same time, we negative association of PCa when compared to controls with an OR of 1.02(CI = 0.64-1.63; P=0.94) for TT vs TC model and an OR of 0.70(CI = 0.47-1.03; P=0.07) for recessive model (Table 4.16, Table 4.17 and Figure 4.16).

Table 4.15: TP53 intronic variants (Reference sequence; NC_000017.11)

S. No	Intron	Chr StartPosition	ChrEnd Position	variant	Status
1.	5	7675016	7675016	g.7675016 T/A	Novel
2.	5	7674991	7674991	g.7674991 T/A	Novel

Table 4.16: Genotypes and their allelic frequencies of TP53 T/A transversion at chromosomal position g.7675016 and C/T transition at chromosomal position g.7674991 in Prostate adenocarcinoma

Alleles	Genotype	Case (n=200)			Controls n=200		
		No.	%age	Frequency	No.	%age	Frequency
g.7675016	T-allele	299	74.75	0.75	338	84.5	0.85
	A-allele	101	25.25	0.25	62	15.5	0.15
	TT	121	60.5	0.60	153	76.5	0.77
	TA	57	28.5	0.29	32	16	0.16
	AA	22	11.0	0.11	15	7.5	0.07
g.7674991	C-allele	175	43.75	0.44	217	54.25	0.54
	T-allele	225	56.25	0.56	183	45.75	0.46
	CC	37	18.50	0.19	67	33.50	0.34
	CT	101	50.50	0.50	83	41.50	0.41
	TT	62	31.00	0.31	50	25.00	0.25

Table 4.17: Association of TP53 T/A transversion at chromosomal position g.7675016 and C/T transition at chromosomal position g.7674991 in and risk of Prostate adenocarcinoma

Alleles	Genetic Model	Odd Ratio	95%CI	Z-stat	P-value
g.7675016	Allelic contrast (A vs T)	1.84	1.29-2.62	3.40	0.001
	Rare vs Common (AA vs TT)	1.85	0.92-3.73	1.73	0.08
	Rare vs Heterozygotes (AA vs TA)	0.82	0.38-1.81	-0.48	0.63
	Heterozygotes vs Common (TA vs TT)	2.25	1.37-3.69	3.22	0.001
	Dominant (AA+TA vs TT)	2.13	1.38-3.28	3.42	0.001
	Recessive (AA vs TT+TA)	1.52	0.77-3.03	1.20	0.230
g.7674991	Allelic contrast (T vs C)	1.52	1.15-2.01	2.96	0.00
	Rare vs Common (TT vs CC)	2.25	1.30-3.88	2.89	0.00
	Rare vs Heterozygotes (TT vs TC)	1.02	0.64-1.63	0.08	0.94
	Heterozygotes vs Common (TC vs CC)	2.20	1.34-3.62	3.13	0.00
	Dominant (TT+TC vs CC)	2.40	1.51-3.82	3.69	0.00
	Recessive (TT vs CC+TC)	0.70	0.47-1.03	-1.80	0.07

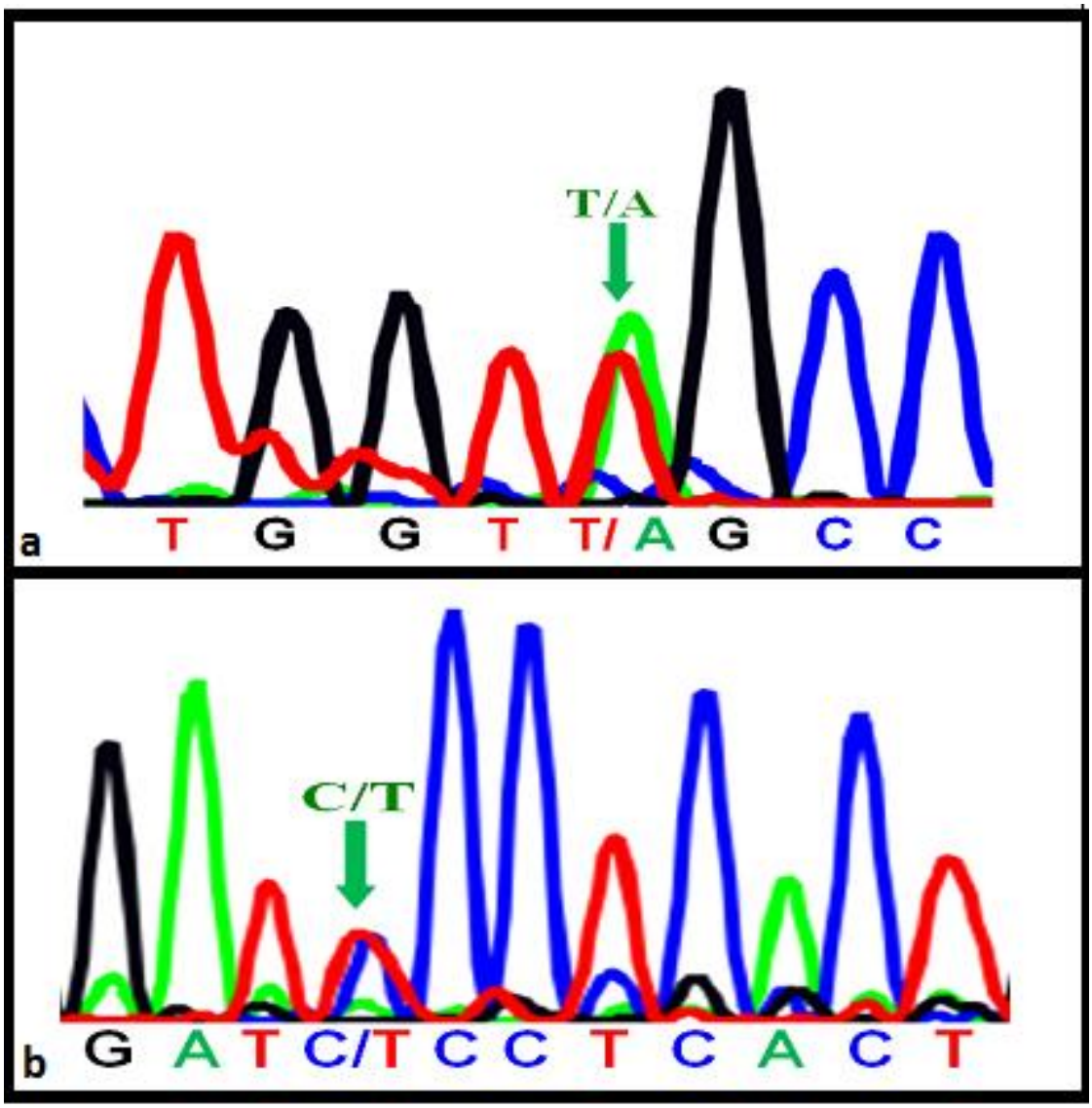

Figure 4.16: TP53 intronic variants.

a. Sequence chromatogram showing TP53 T>A variant in intron-5 at genomic position g.7675016.

b. Sequence chromatogram showing TP53 C>T variant in intron-5 at genomic position g.7674991.

Activation of p53 is initiated through multiple mechanisms including phosphorylation by ATR, ATM and Chk1 etc. MDM2 binds with p53 and targets it for proteasomal degradation (Shi and Gu, 2012). Phosphorylation, p14ARF and USP7 avoid p53-MDM2 binding and leads to an escalation of stable p53 tetramers in the cytoplasm.

According to the literature, p53 has two NES (nuclear export signals), i.e. one at the C-terminus (Stommel et al., 1999) and the other one at the N-terminus transactivation domain (Zhang and Xiong, 2001), (Kudo et al., 1998, Wolff et al., 1997). The NES at the C-terminus has the ability to connect p53 with subcellular localization and nuclear functions (Stommel et al., 1999). Publish data revealed that phosphorylation of the C-terminal serine-392 might stabilize p53 tetramers, while phosphorylation of serine-315 and serine-392 may destabilize the tetramer (Jimenez et al., 1999, Liang and Clarke, 2001, Geyer et al., 2000). Some studies indicated that, MDM2 mediated ubiquitination of p53 enables its export to the cytoplasm (Geyer et al., 2000, Lohrum et al., 2001).

The nuclear export signals at the N-terminus of p53 also been proved to induce its nuclear export. This N-terminal NES is usually active in unstressed cells, but gets inactivated by DNA damage which allow its quick nuclear accumulation (Zhang and Xiong, 2001, (DUMAZ et al., 2001). It was further proposed that, mutations in the N-terminal NES limit its interaction with the export receptor (Kussie et al., 1996).

Keeping in view the vital role of p53 in cancer suppression and prevention, it may be revoked during carcinogenesis. Point mutations deactivated p53 in more than half of the human cancers (see http://www.iarc.fr/p53), majority of which occurring in the DNA binding domain. These mutations either change its conformation or avoid its DNA contact (Hollstein et al., 1994). All of our reported point mutations are in the DNA binding domain of p53 and it is therefore expected that these mutations will result in conformational change and thus will affect its interactions as well. TP53 missense mutations generally repress its functions and may have a dominant negative effect on transactivation of other genes containing p53-specific responsive elements (Ecke et al., 2010).

4.8 PTEN

The PTEN protein is made up of 403 amino acids. Structure and functional annotation of genetic variations is vital to apprehend their diverse impacts. A single base change can even significantly endorse changes in a gene product or some times of multiple gene products. We first confirmed the conservativeness of PTEN across different species phylogenetically through PhylomeDB and TreeFam (Figure 4.17) which confirmed that the gene/protein is conserved across the species evolutionarily.

Molecular level changes can affect the phenotype of cells, tissues and finally the organisms. Therefore the PTEN protein was analyzed in detail for the possible impact of point mutations on its structure, function and folding pattern to explore how they affect the structures and properties of a gene product.

In this study, all the PTEN exons from case-control samples of prostate adenocarcinoma from Pakistani population were analyzed. A total of nine exonic and five intronic variants were identified, out of the nine exonic variants 8 are novel while only one (H93Q) was previously reported (Table 4.18).

Tertiary structure of wild type PTEN protein (Figure 4.18) has already been determined through X-Ray crystallography (PDB id: 1D5R) (Lee et al., 1999) having a resolution of 2.10 Å. It comprises a 179 residue N-terminal domain (residues 7–185) having PTP signature motif and a 166 residue C-terminal domain consisting of a structure analogous to the C2 domain. This domain facilitates the Ca2+ dependent membrane recruitment of multiple signaling proteins including phosphoinositide 3-kinase (PI3K), Cδ1 (PLCδ1) and protein kinase C (PKC) (Rizo and Südhof, 1998). The C2 and phosphatase domains interact through a large interface, consists of several conserved residues which are reported frequently mutated in different cancers.

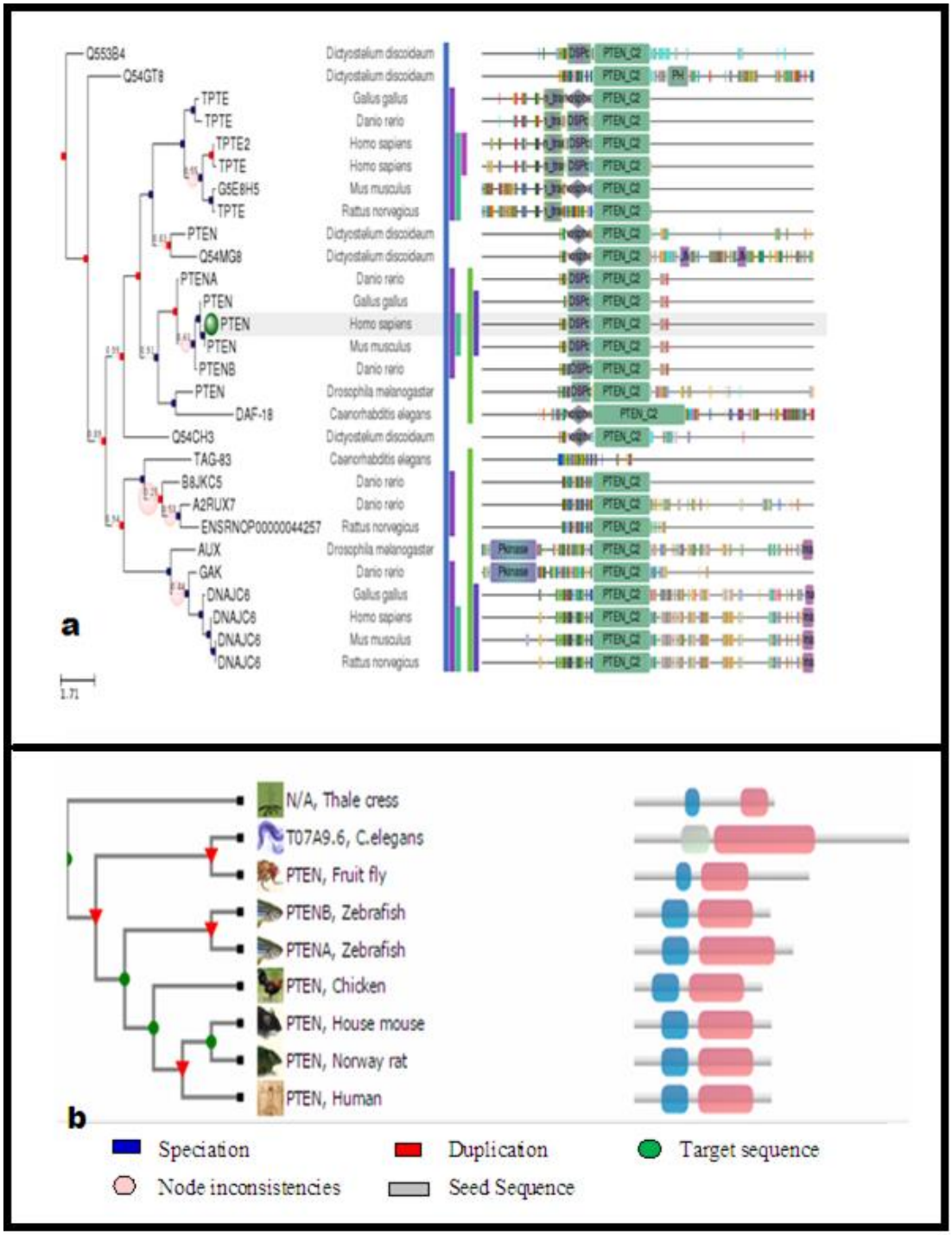

Figure 4.17: Evolutionary history showing conservativeness of PTEN protein.

a. Evolutionary history of PTEN along with pfam domains through PhylomeDB.

b. PTEN gene tree through TreeFam

Table 4.18: PTEN Exonic variants

S. No	Exon	Ref Allele	VarAllele	Chr StartPosition	ChrEnd Position	Residue change	MutType	Reported By
1.	5	G	C	87933015	87933015	A86P	Missense	Novel
2.	5	G	A	87933030	87933030	E91K	Missense	Novel
3.	5	T	G	87933038	87933038	H93Q	Missense	Rodriquez-Escudero et. al., 2011
4.	5	G	C	87933050	87933050	Q97H	Missense	Novel
5.	5	G	T	87933054	87933054	E99*	Nonsense	Novel
6.	5	G	A	87933228	87933228	E157K	Missense	Novel
7.	8	CAC	TTT	87960906	87960908	H272F	Missense	Novel
8.	8	T	C	87960951	87960951	S287P	Missense	Novel
9.	8	G	A	87960954	87960954	E288F	Missense	Novel

In view of the structure and function relation of protein, native and mutant structures were compared to evaluate the level of similarity. It was observed that six mutations (A86P, E91K, H93Q, Q97H, E99* and E157K) are in the C2 domain, while mutations H272F, S287P and E288F are in DSPc domain (Figure 4.17). Secondary structures were predicted for seed and mutated versions of the PTEN proteins and were compared with themselves as well as other model organisms (Figure 4.19). There were observed significant changes. It was also observed that majority of the mutations were at the conserved sites and thus it can be inferred that, these changes may be translated into the higher order conformations and thus were studied in the tertiary structures as well.

The HCXXGXXR signature motif of PTEN makes a P loop (residues 123–130) at the bottom of its active site. The backbone of the pocket are made up of the P loop while its walls from the side chains. The pocket is approximately 8Å deep and has an opening of approximately 5×11Å. Cys-124 and Arg-130 residues of the HCXXGXXR motif are important for catalysis, His-123 and Gly-127 for P loop conformation, Asp-92 acts as an acid to assist protonation of the phenolic oxygen atom (Jia et al., 1995). One of our reported mutations H93Q is very close to Asp-92, which can possibly disrupt the functionality of protein. The P loop sequence of PTEN is unique among known PTPs, having two basic residues i.e. Lys-125 and Lys-128 in its center and is conserved in yeast homologs (Maehama and Dixon, 1998). Results of the functional impacts of our identified variations characterized through automated resources are summarized in Table 4.19.

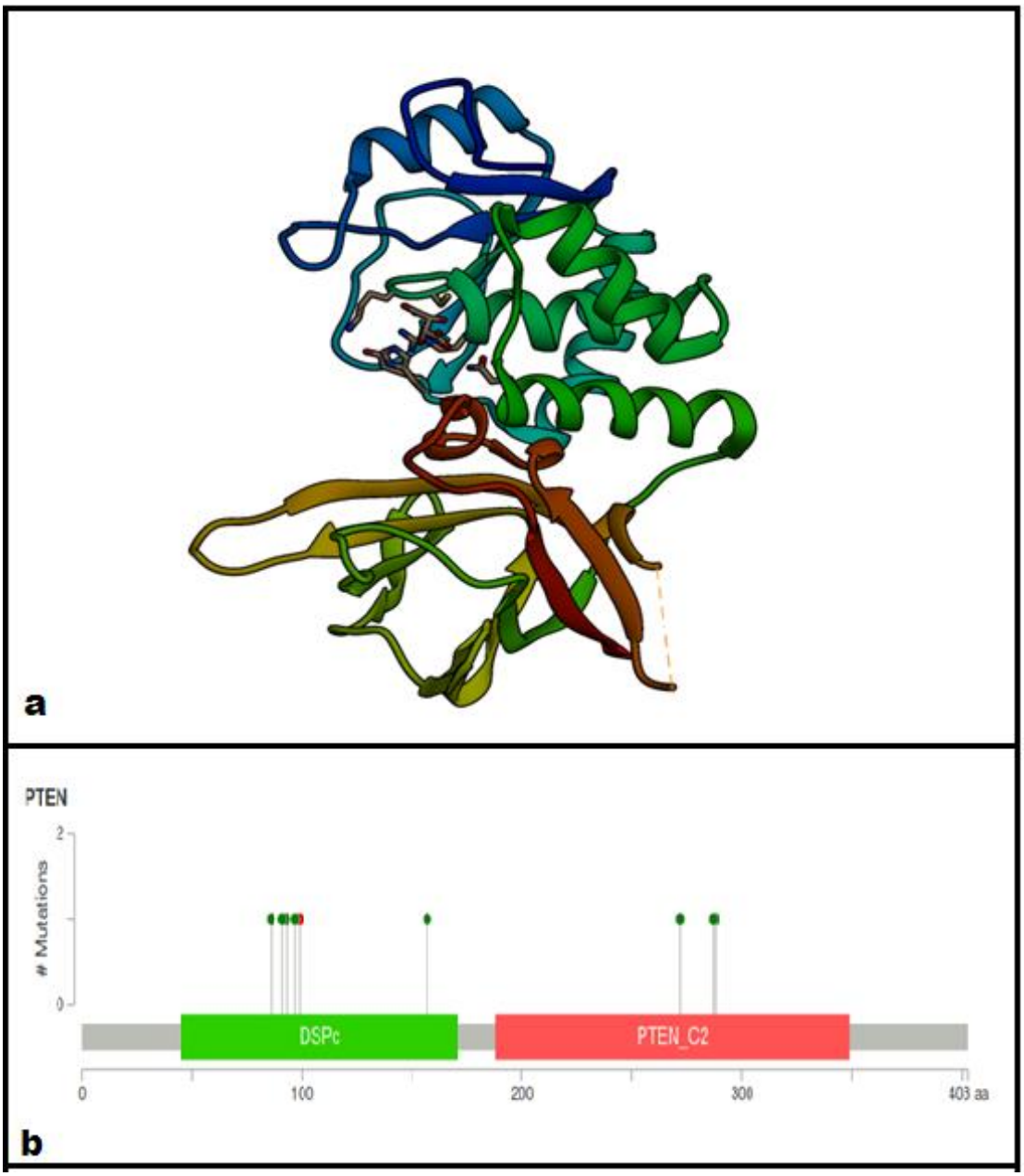

Figure 4.18: Structure of PTEN protein.

a. Crystal structure of PTEN protein (PDB ID: 1D5R) bound with L(+)-tartrate molecule.

b. Lollipop plot of PTEN mutations of exonic regions in the sampled population, six mutations are in the DSCP domain while three are in the PTEN_C2 domain.

Table 4.19: Status of PTEN Exonic variants confirmed through HGMD, Ensembl along mutants functional impact analysis through different servers

S. No	Residue change	Reported By	SIFT	PROVEAN	Mutation Assessor	PolyPhen2
1.	**A86P**	Novel	Damaging	Deleterious	Medium	Probably Damaging
2.	**E91K**	Novel	Tolerated	neutral	Low	Probably Damaging
3.	**H93Q**	Rodriquez-Escudero et. al., 2011	Damaging	Deleterious	High	Probably Damaging
4.	**Q97H**	Novel	Tolerated	neutral	Low	Probably Damaging
5.	**E99***	Novel	Truncated protein			
6.	**E157K**	Novel	Tolerated	neutral	Low	Benign
7.	**H272F**	Novel	Damaging	Deleterious	Low	Probably Damaging
8.	**S287P**	Novel	Tolerated	neutral	Low	Benign
9.	**E288F**	Novel	Damaging	Deleterious	Low	Possibly Damaging

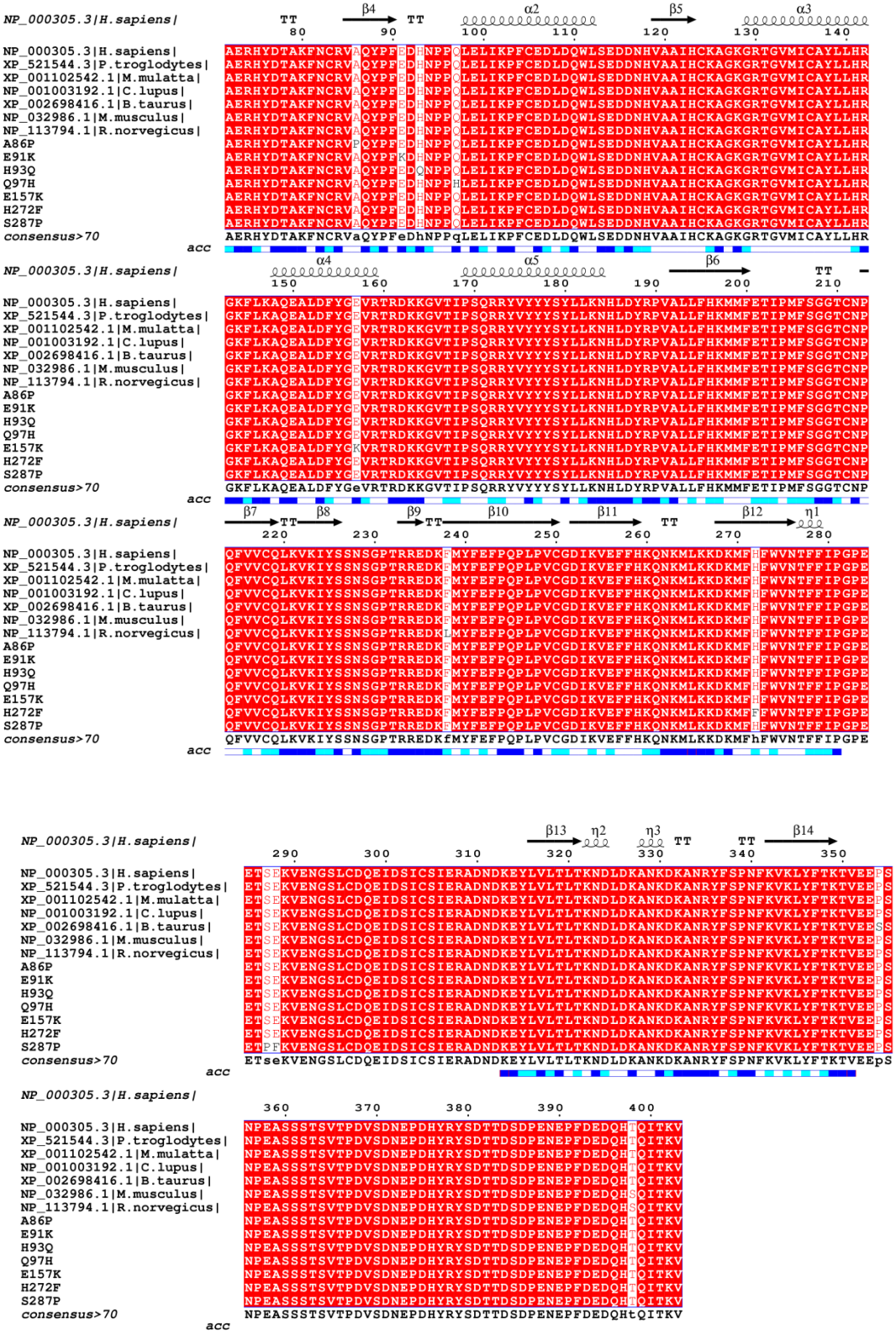

Figure 4.19: Secondary structure analysis of wild and mutant PTEN proteins along with some other species (mutations in the exonic regions only)

Each variant was studied independently in detail for its possible structural and functional impacts as follow,

4.8.1 G3C (g. 87933015) A86P

The transversion of G to C nucleotide at position-3 (exon-5) in PTEN gene results in the substitution of Alanine to Proline at position-86. The mutation was confirmed as a novel one from both the HGMD and Ensembl. The mutant residue is bigger than the wild type and is located within a domain, annotated in UniProt as "Phosphatase tensin-type". It introduces a residue with different properties, able to disturb the domain. The wild residue (alanine) was smaller and buried while the mutant is bigger and will not fit properly. Proline is a very rigid residue and can disturb the required flexibility at this position.

This residue is part of an interpro domain named Dual Specificity Phosphatase and Catalytic Domain (IPR000340), annotated with the GO term, Protein Tyrosine/serine/threonine Phosphatase Activity (GO:0008138). The GO annotation indicates that, the domain has a function in Hydrolase Activity (GO:0016787). The residue is also annotated as part of an interpro domain named Protein-Tyrosine Phosphatase-Like (IPR029021), Tensin Phosphatase, Lipid Phosphatase Domain (IPR029023), Protein-Tyrosine Phosphatase (IPR003595), Pdz Domain Binding (GO:0030165), Magnesium ion Binding (GO:0000287), Phosphatidylinositol-3,4,5-Trisphosphate 3-Phosphatase Activity (GO:0016314) and Protein Tyrosine Phosphatase Activity (GO:0004725). These GO annotations indicate that the domain has a function in Protein Binding (GO:0005515), ion Binding (GO:0043167) and Hydrolase Activity (GO:0016787).

The mutation is in a domain that is central to the protein activity and is also in contact with another domain. The mutation can affect this interaction and thus affect the protein function thereby affecting its catalytic activity and signal transduction between the domains (Figure 4.20). Only this residue was found at this position, functionally the mutation was predicted as Damaging by SIFT, Probably damaging by PolyPhen2, Deleterious by PROVEAN and medium by MutationAssessor (Table 4.19).

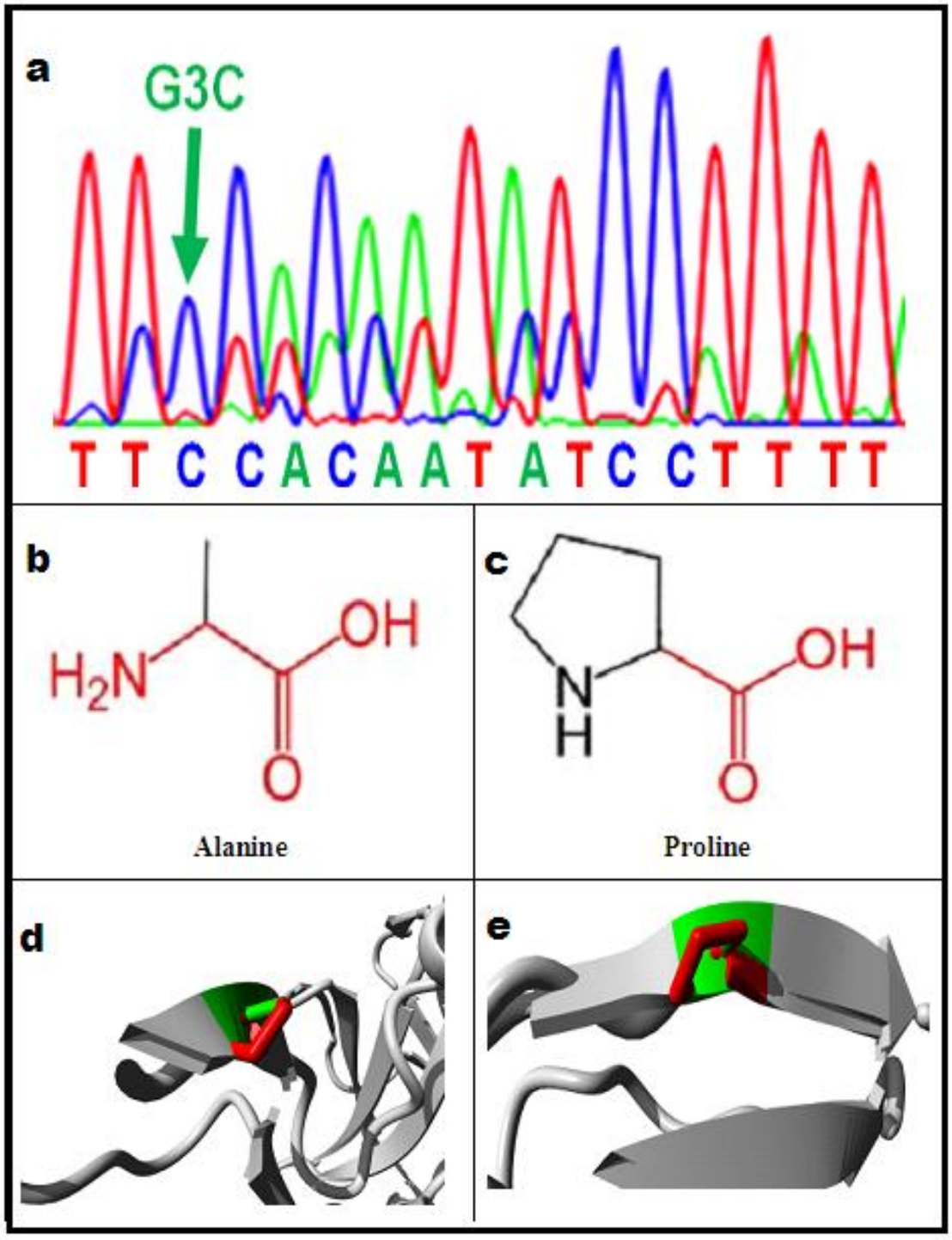

Figure 4.20: Sequence and structure analysis of PTEN A86P mutation

a. Sequence chromatogram showing G3C transversion in the gene sequence.
b. Wild residue alanine.
c. Mutant residue Proline.
d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of PTEN.

4.8.2 G18A (g.7933030) E91K

This transition of G to A nucleotide at codon 91 at position 18 in the exon-5 of PTEN gene results in the substitution of Glutamic acid to Lysine. The mutation was confirmed as a novel one from both the HGMD and Ensembl.

The mutant residue is bigger than the wild residue. The wild type residue was negatively charged, the mutant residue is positively charged. The mutant amino acid is not in direct contact with a ligand, however the mutation could affect the local stability which in turn could affect the ligand contacts made by one of the neighboring residues. The size difference between the wild and mutant residue disturb the position of residue and the new residue is unable to make a hydrogen bond with Asp-94. The mutation also disturbs the salt bridge with Lys-221 and 254 because of charge difference. These differences may possibly abolish contacts with other molecules.

Neither the mutant nor another analogous residue was witnessed at this position in other homologous sequences and thus this mutation is probably damaging to the protein (Figure 4.21). The mutation is located within a domain, annotated in UniProt as "Phosphatase tensin-type" and introduces a residue with totally different properties. The domain is annotated with the GO term, Protein Tyrosine/serine/threonine Phosphatase Activity (GO:0008138). The GO annotations indicates that the domain has a function in Hydrolase Activity (GO:0016787). The residue is also annotated as part of an interpro domain named Protein-Tyrosine Phosphatase-Like (IPR029021), Lipid Phosphatase Domain (IPR029023), an interpro domain named Bi-functional Phosphatidylinositol Tris-phosphate Phosphatase/dual Specificity Phosphatase (IPR017361), annotated with GO terms, Pdz Domain Binding (GO:0030165), Magnesium ion Binding (GO:0000287), Phosphatidylinositol-3,4-Bisphosphate 3-Phosphatase Activity (GO:0051800), Phosphatidylinositol-3,4,5-Trisphosphate 3-Phosphatase Activity (GO:0016314) and Protein Tyrosine Phosphatase Activity (GO:0004725).

Functionally the mutation was predicted as Tolerated by SIFT, Probably damaging by PolyPhen2, neutral by PROVEAN and low by MutationAssessor (Table 4.19).

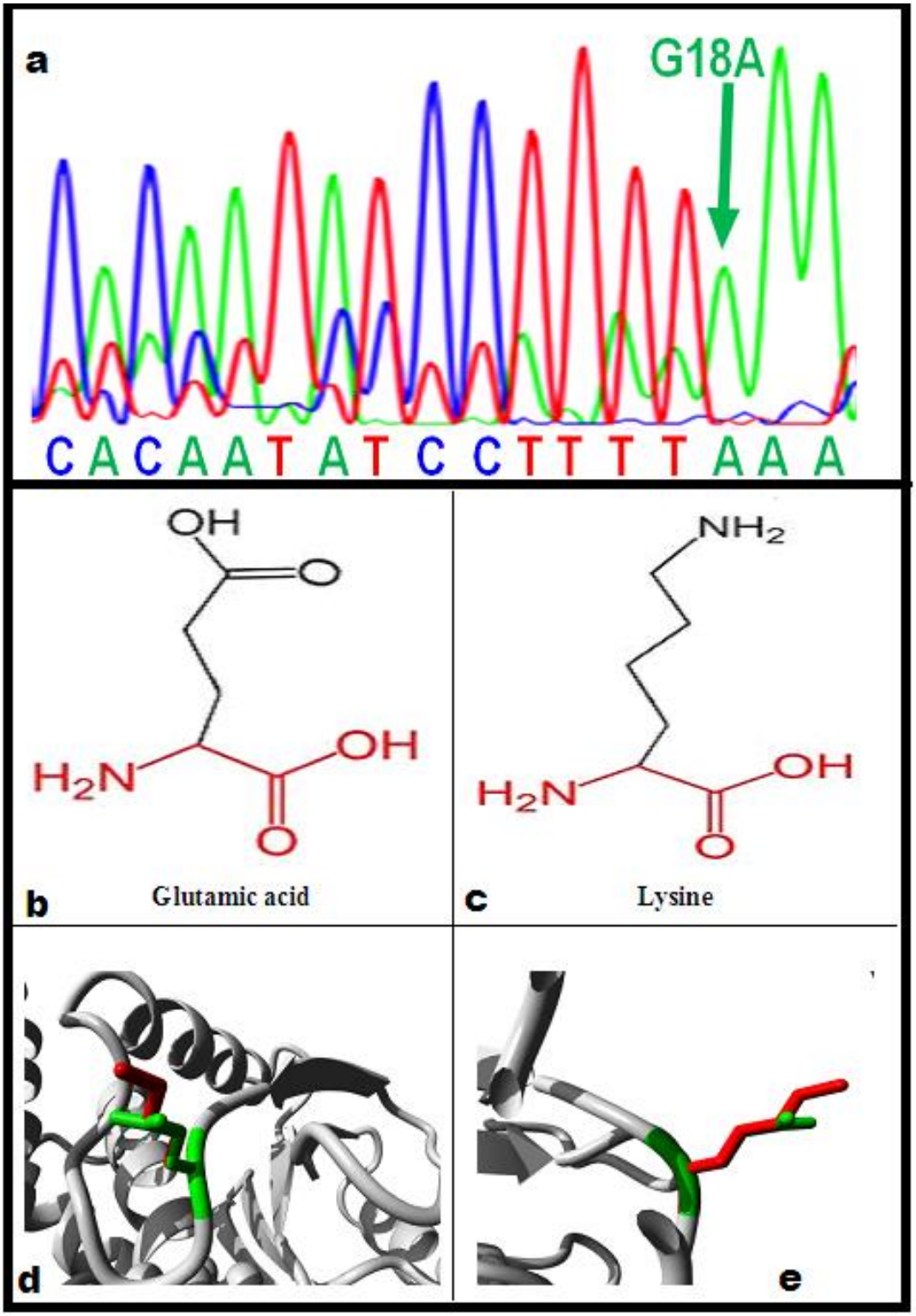

Figure 4.21: Sequence and structure analysis of PTEN E91K mutation

a. Sequence chromatogram showing G18A transition in the gene sequence.
b. Wild residue Glutamic acid c. Mutant residue Lysine.
d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of PTEN protein.

4.8.3 T26G (g.87933038) H93Q

The T to G transversion at codon 93 in exon-5 at position 26 of PTEN gene results in the substitution of Histidine to Glutamine. The mutation is a previously reported one (Rodríguez-Escudero et al., 2011). The mutant residue is smaller and has different properties than the wild residue. The wild interacts with a ligand (TLA), the difference in properties can abolish this interaction which is important for the protein function. The residue at this position is conserved and probably be damaging to the protein.

The mutation is located within a domain, annotated in UniProt as "Phosphatase tensin-type". This residue is part of an interpro domain named "Dual specificity phosphatase, catalytic domain" (IPR000340), annotated with the Gene-Ontology term, Protein Tyrosine/serine/threonine Phosphatase Activity (GO:0008138). It indicates that the domain has a function in Hydrolase Activity (GO:0016787). This residue is also annotated as part of an interpro domain named, Protein-Tyrosine Phosphatase-Like (IPR029021), an interpro domain named Tensin Phosphatase, Lipid Phosphatase Domain (IPR029023), Bi-functional Phosphatidylinositol Tris-phosphate Phosphatase (IPR017361), Pdz Domain Binding (GO:0030165), Magnesium ion Binding (GO:0000287), Phosphatidylinositol-3,4-Bisphosphate 3-Phosphatase Activity (GO:0051800) and Protein Tyrosine Phosphatase Activity (GO:0004725). These GO annotations indicate that the domain has a function in Protein Binding (GO:0005515), ion Binding (GO:0043167) and Hydrolase Activity (GO:0016787).

The residue is positioned on the surface of the protein in a domain required for its main activity. It may be in contact with other molecules or domains, the smaller size of the mutant residue will cause a possible loss of external interactions (Figure 4.22). Functionally the mutation was predicted as Damaging by SIFT, Probably damaging by PolyPhen2, Deleterious by PROVEAN and High by MutationAssessor (Table 4.19).

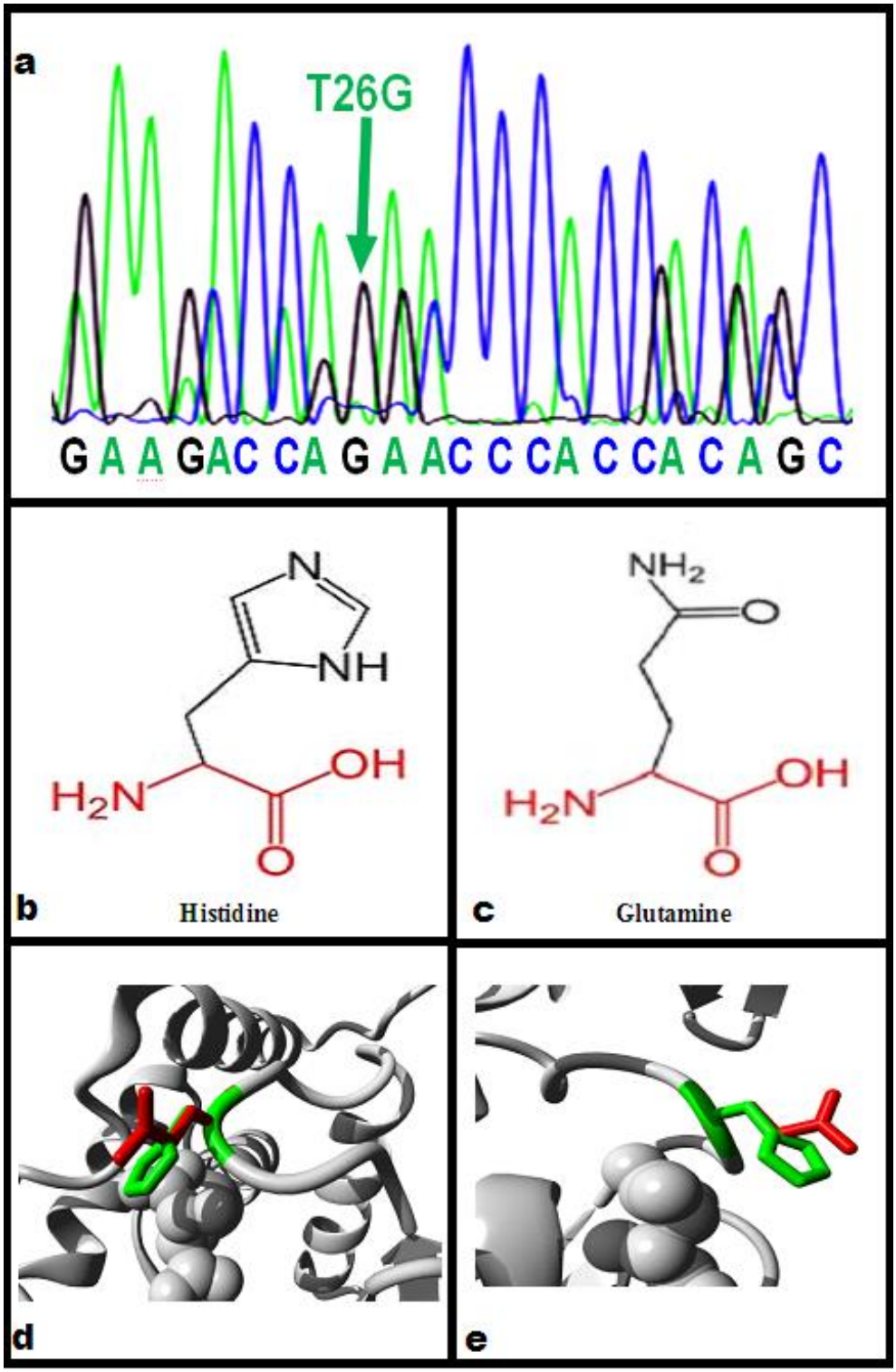

Figure 4.22: **Sequence and structure analysis of PTEN H93Q mutation**

a. Sequence chromatogram showing T26G transversion in the gene sequence.

b. Wild residue Histidine c. Mutant residue Glutamine.

d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of PTEN protein.

4.8.4 G38C (g.87933050) Q97H

The transition of G to C nucleotide at codon 97 at position 38 in the exon-5 of PTEN gene results in the substitution of Glutamine to Histidine. The mutation was confirmed as a novel one from both the HGMD and Ensembl. The mutant residue is bigger and is positioned in a domain, annotated in UniProt as Phosphatase tensin-type. The genetic change introduces a residue with different properties able to affect the domain and its function. This domain is actually important for the main activity of the protein. It is found to be in contact with another domain known to be involved in binding. The mutation can affect this interaction and thereby disrupts the functionality and signal transduction between the domains (Figure 4.23).

As the mutation is in codon 97, which is part of an interpro domain named Dual Specificity Phosphatase, Catalytic Domain (IPR000340), annotated with the GO term, Protein Tyrosine/serine/threonine Phosphatase Activity (GO:0008138) and has a function in Hydrolase Activity (GO:0016787). The domain is also annotated as Protein-Tyrosine Phosphatase-Like (IPR029021), Tensin Phosphatase, Lipid Phosphatase Domain (IPR029023), Protein-Tyrosine Phosphatase, Catalytic (IPR003595), Bi-functional Phosphatidylinositol Tris-phosphate Phosphatase/dual Specificity Phosphatase PTEN (IPR017361), also annotated with the GO terms, GO:0030165 (Pdz Domain Binding), GO:0000287 (Magnesium ion Binding), GO:0051800 (Phosphatidylinositol-3,4-Bisphosphate 3-Phosphatase Activity), GO:0016314 (Phosphatidylinositol-3,4,5-Trisphosphate 3-Phosphatase Activity) and GO:0004725 (Protein Tyrosine Phosphatase Activity) to indicate its function.

The mutation was found conserved at this position and is probably damaging to the protein. Functionally the mutation was predicted as Damaging by SIFT, Probably damaging by PolyPhen2, Deleterious by PROVEAN and High by MutationAssessor (Table 4.19).

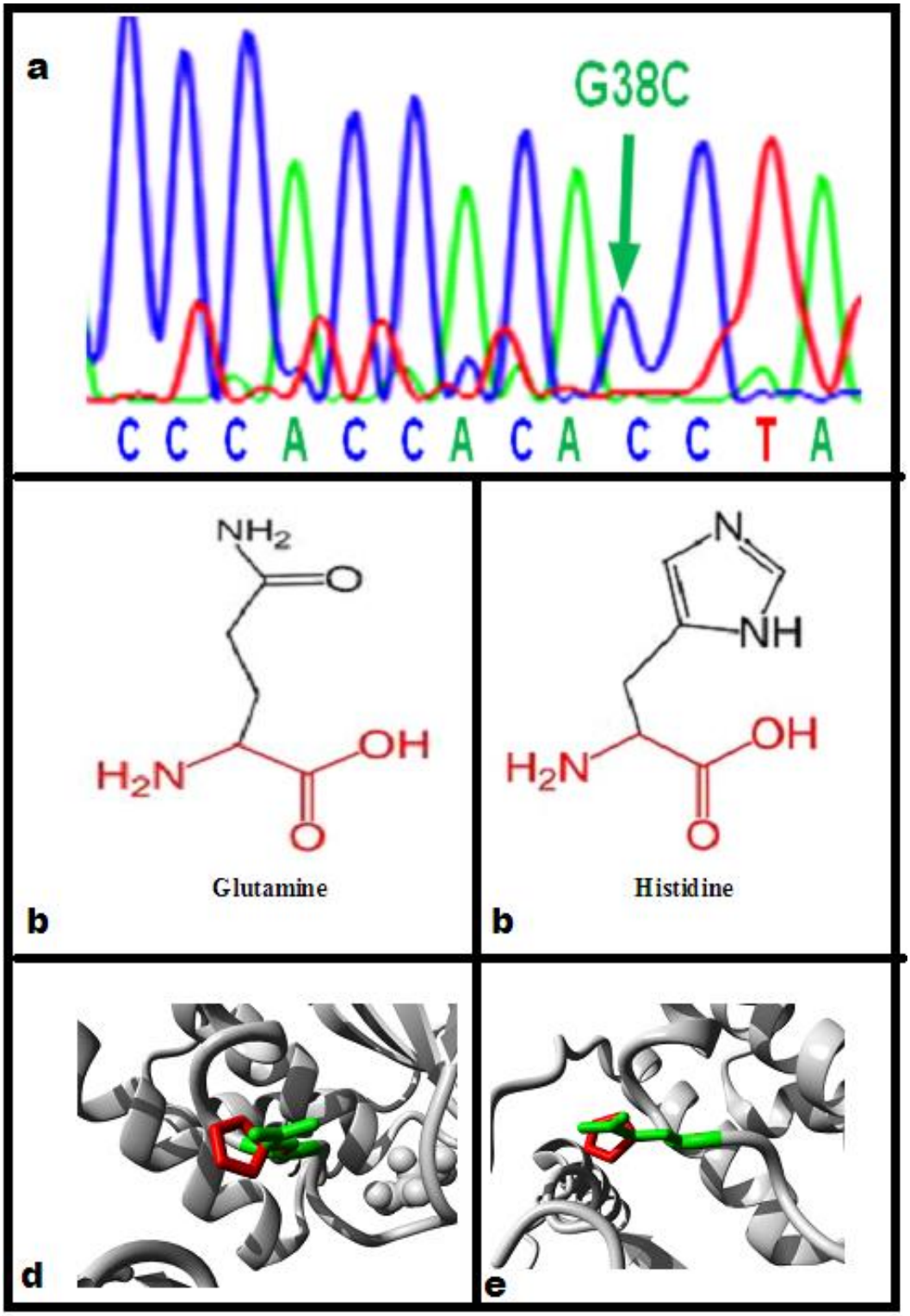

Figure 4.23: Sequence and structure analysis of PTEN Q97H mutation

a. Sequence chromatogram showing G38C transition in the gene sequence.

b. Wild residue Glutamine c. Mutant residue Histidine.

d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of PTEN protein.

4.8.5 G42T (g.87933054) E99X

The transversion of G42T in exon-5 of PTEN gene results in an introduction of stop codon at position 99 of the polypeptide. The mutation was confirmed as a novel one from both the HGMD and Ensembl. The mutation will yield a truncated polypeptide. The PTEN protein has 403 residues and contains a 179 residue N-terminal domain and a 166 residue C-terminus domain. The stop codon is in the DSPc domain so protein will lose part of this domain and a complete C2-domain, thus the truncated polypeptides will lose its function. So it can be concluded that the mutation will have a loss of function affect, but needs further characterization in the wet lab (Figure 4.24).

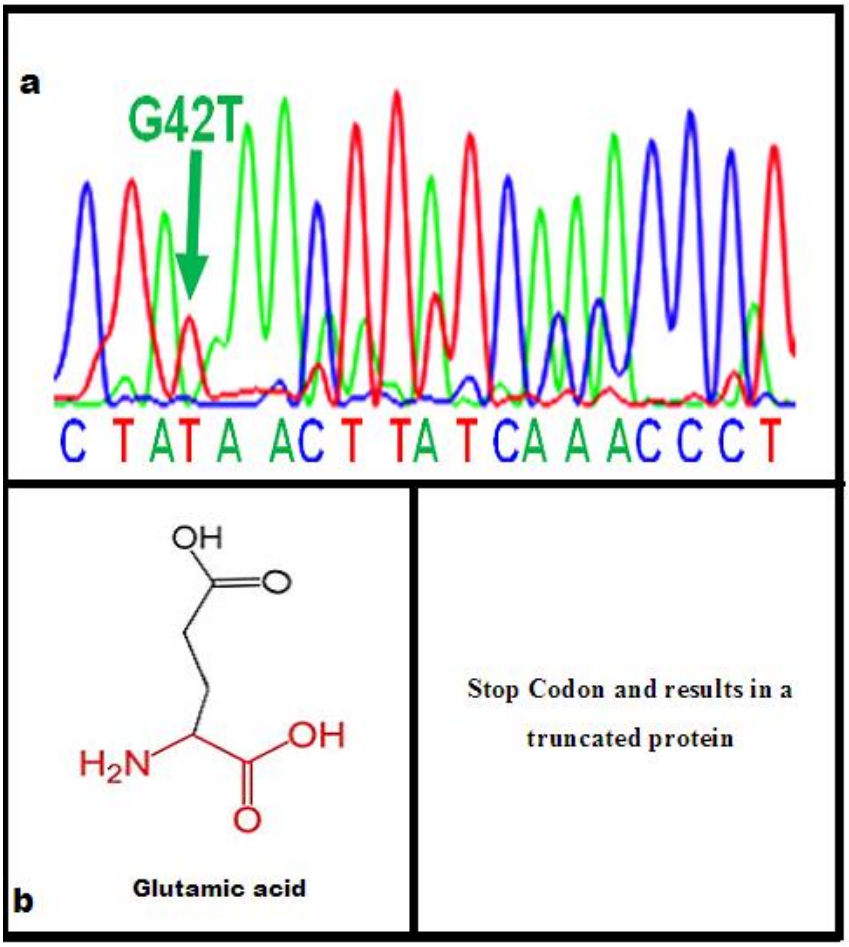

Figure 4.24: Sequence and structure analysis of PTEN E99X mutation

a. Sequence chromatogram showing G42T transversion in the gene sequence.
b. Wild residue Glutamic acid. .

4.8.6 G217A (g.87933228) E157K

G to C nucleotide transition of at codon 157 (position 217) in exon-5 of PTEN gene results in the substitution of Glutamic acid to Lysine. The mutation was confirmed as a novel one from both the HGMD and Ensembl. The mutant residue is bigger and is positively charged while the wild residue was negatively charged. This difference of charge can possibly disturb interactions with other molecules. Furthermore, the salt bridge with Arg-14 and Lys-163 formed by the wild residue also gets disturbed by this change.

The mutation occurred within a domain, annotated in UniProt as "Phosphatase tensin-type", termed in interpro as Dual Specificity Phosphatase, Catalytic Domain (IPR000340) and annotated with the Gene-Ontology term, Protein Tyrosine/serine/threonine Phosphatase Activity (GO:0008138) to indicate its function. The GO annotation indicates that the domain has a function in Hydrolase Activity (GO:0016787). The domain is also termed by interpro as Protein-Tyrosine Phosphatase-Like (IPR029021), Tensin Phosphatase, Lipid Phosphatase Domain (IPR029023), Protein-Tyrosine Phosphatase, Catalytic (IPR003595) and Bi-functional Phosphatidylinositol Tris-phosphate Phosphatase/dual Specificity Phosphatase PTEN (IPR017361), Pdz Domain Binding (GO:0030165), Magnesium ion Binding (GO:0000287), Phosphatidylinositol-3,4-Bisphosphate 3-Phosphatase Activity (GO:0051800), Phosphatidylinositol-3,4,5-Trisphosphate 3-Phosphatase Activity (GO:0016314), Protein Tyrosine Phosphatase Activity (GO:0004725) and Protein Binding (GO:0005515).

The mutation is in a domain, important for the main activity (Figure 4.25). The wild residue is highly conserved at this position. Functionally the mutation was predicted as damaging by SIFT, probably damaging by PolyPhen2, deleterious by PROVEAN and low by MutationAssessor (Table 4.19).

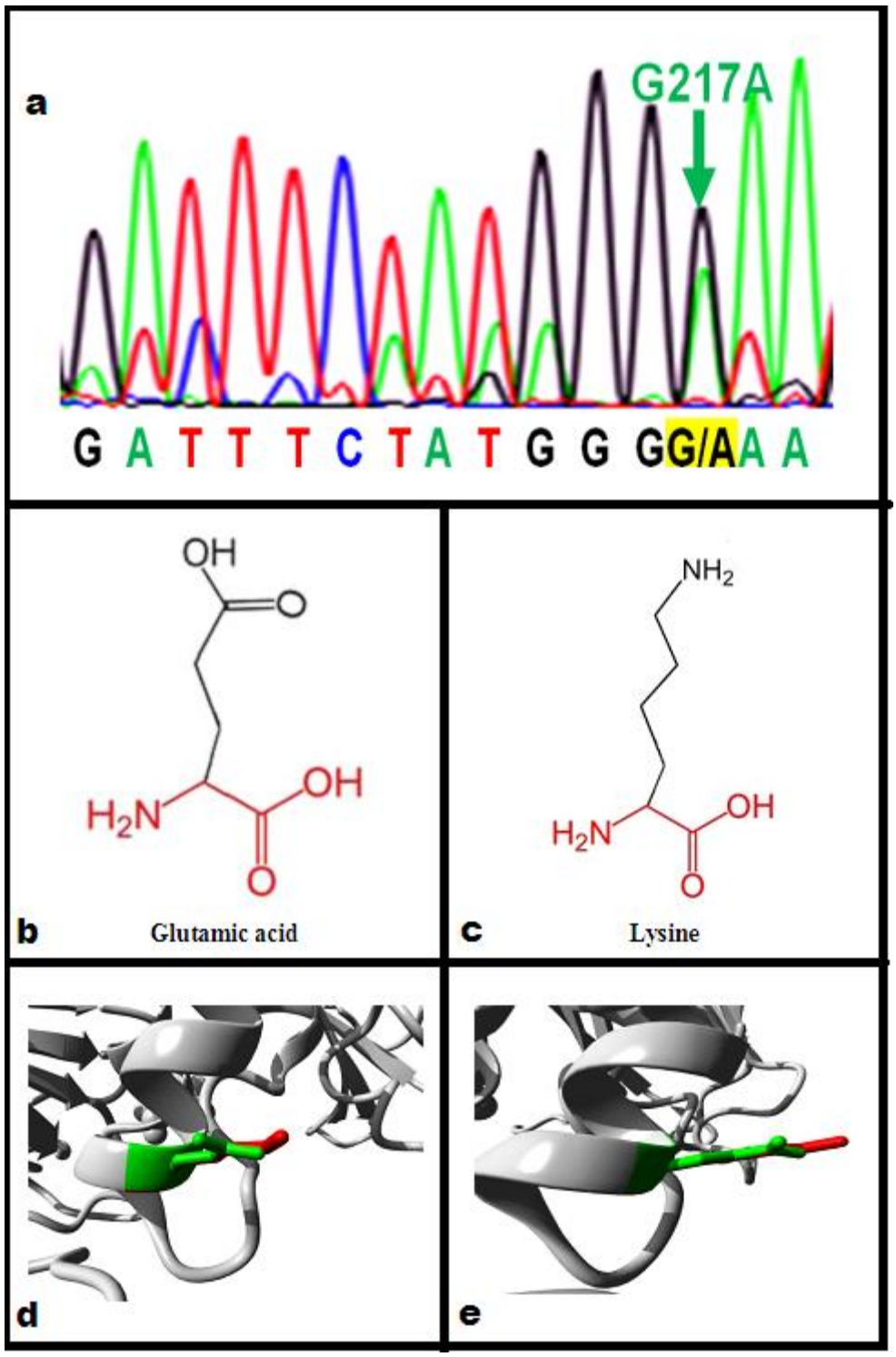

Figure 4.25: Sequence and structure analysis of PTEN E157K mutation

a. Sequence chromatogram showing G217A transition in the gene sequence.
b. Wild residue Glutamic acid c. Mutant residue Lysine.
d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of PTEN protein.

4.8.7 CAC13-15TTT (g.87960906-87960908) H272F

This mutation of codon 272 in exon-8 of PTEN at position 13-15 results in the substitution of Histidine to Phenylalanine. The mutation was confirmed as a novel one from both the HGMD and Ensembl. The mutant residue is comparatively bigger and more hydrophobic. The substitutions are positioned within a domain, annotated in UniProt as "C2 tensin-type" and introduces a residue with different properties. The wild residue was found highly conserved at this position and thus it can be inferred that this change can potentially disturb this domain and is probably damaging.

This residue is part of an interpro domain named Tensin Phosphatase, C2 Domain (IPR014020), interpro domain named Bi-functional Phosphatidylinositol Tris-phosphate Phosphatase/dual Specificity Phosphatase PTEN (IPR017361), annotated with the Gene-Ontology (GO) terms “GO:0030165 (Pdz Domain Binding), GO:0000287 (Magnesium ion Binding), GO:0051800 (Phosphatidylinositol-3,4-Bisphosphate 3-Phosphatase Activity), GO:0016314 (Phosphatidylinositol-3,4,5-Trisphosphate 3-Phosphatase Activity), GO:0004725 (Protein Tyrosine Phosphatase Activity)” to indicate its functions as Protein Binding (GO:0005515), ion Binding (GO:0043167) and Hydrolase Activity (GO:0016787).

Codon 272 is part of a domain vital for the PTEN main activity, thus it is probable that the differences in size and hydrophobicity will lead to disturbance in the core structure. This change is also able to abolish hydrogen bonds in the protein core of and as a result disturb correct folding and thereby affect its catalytic activity (Figure 4.26). Functionally the mutation was predicted as Tolerated by SIFT, Benign by PolyPhen2, Neutral by PROVEAN and Low by MutationAssessor (Table 4.19).

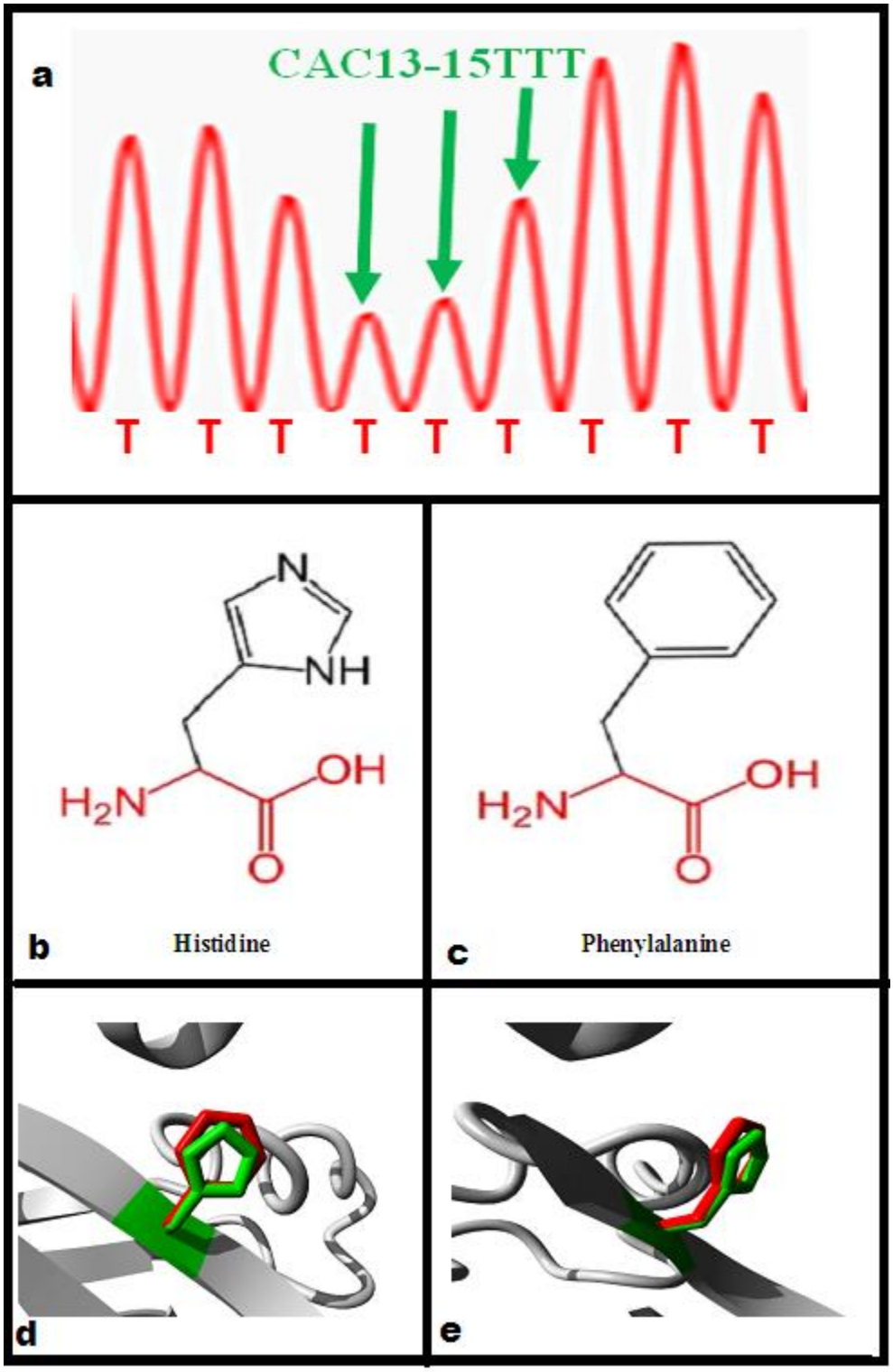

Figure 4.26: Sequence and structure analysis of PTEN H272F mutation

a. Sequence chromatogram showing CAC13-15TTTmutation in the gene sequence.

b. Wild residue Histidine c. Mutant residue Phenylalanine.

d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of PTEN protein.

4.8.8 T58C (g.87960951) S287P and G61A (g.87960954) E288F

i. **T58C (S287P)**

This transition mutation of codon 287 in exon-8 of PTEN at position 58 results in the substitution of Serine to Proline. The mutation was confirmed as novel from HGMD and Ensembl. The mutant residue is bigger and more hydrophobic than the wild residue which was making a hydrogen bond with the Pro-283. Due to larger size and difference in hydrophobicity, the mutant residue is not in the accurate distance and position to make similar bond.

Ser-287 is positioned in a domain, annotated in UniProt as "C2 tensin-type". The mutation leads to a residue with different properties, which can disrupt this domain and suppress its function. The wild residue favors α-helix in the 3D-structure, while the mutant residue (Proline) tends to disrupt this conformation if not present at one of the first three positions (Figure 4.27). If the helix gets disturbed, this might have severe structural impacts.

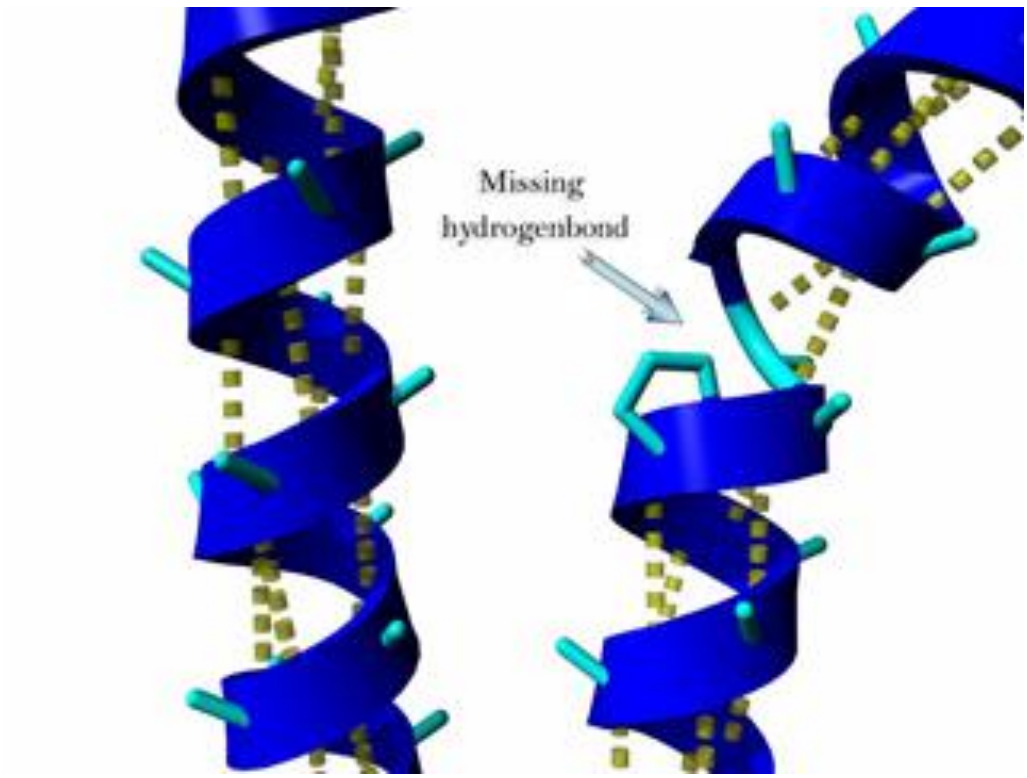

Figure 4.27: α-helix showing disruption of hydrogen bond due to mutation

This residue (Ser-28) is part of an interpro domain named Tensin Phosphatase, C2 Domain (IPR014020), annotated as Bi-functional Phosphatidylinositol Tris-phosphate Phosphatase/dual Specificity Phosphatase PTEN (IPR017361). This domain is annotated with the Gene-Ontology (GO) term, Pdz Domain Binding (GO:0030165), Magnesium ion Binding (GO:0000287), Phosphatidylinositol-3,4-Bisphosphate 3-Phosphatase Activity (GO:0051800), Phosphatidylinositol-3,4,5-Trisphosphate 3-Phosphatase Activity (GO:0016314), Protein Tyrosine Phosphatase Activity (GO:0004725). These annotations demonstrate its function in Protein Binding (GO:0005515), ion Binding (GO:0043167) and Hydrolase Activity (GO:0016787).

The mutated residue is based in an important domain on its surface, thus it can be suggested that this change can interrupt its interactions (Figure 4.28). The wild type residue is conserved at this position. Functionally the mutation was predicted as tolerated by SIFT, benign by PolyPhen2, neutral by PROVEAN and low by MutationAssessor (Table 4.19).

ii. **G61A (E288F)**

The mutation was confirmed as a novel one from both the HGMD and Ensembl. This transition of G to A at codon 288 in exon-8 of PTEN results in the substitution of Glutamic acid to Phenylalanine. The mutant residue is bigger, more hydrophobic and neutral, while the than the wild residue was negatively charged. The mutation is located within a domain, annotated in UniProt as "C2 tensin-type" and introduces a residue with totally different properties. These differences can potentially interrupt this domain structurally and functionally.

Glu-288 is part of an interpro domain named, C2 Domain (IPR014020) and Bi-functional Phosphatidylinositol Tris-phosphate Phosphatase/dual Specificity Phosphatase PTEN (IPR017361), annotated with the GO term, Pdz Domain Binding (GO:0030165), Magnesium ion Binding (GO:0000287), Phosphatidylinositol-3,4-Bisphosphate 3-Phosphatase Activity (GO:0051800), Phosphatidylinositol-3,4,5-Trisphosphate 3-Phosphatase Activity (GO:0016314) and Protein Tyrosine Phosphatase Activity (GO:0004725). These annotations demonstrate

its functions in Protein Binding (GO:0005515), ion Binding (GO:0043167) and Hydrolase Activity (GO:0016787).

Glu-288 is located on the surface of the domain and its interactions might be affected by this mutation. The wild residue loses its charge which can cause loss of interactions with other molecules (Figure 4.28).

The wild residue is not very conserved at this position. Functionally the mutation was predicted as damaging by SIFT, probably damaging by PolyPhen2, deleterious by PROVEAN and low by MutationAssessor (Table 4.19).

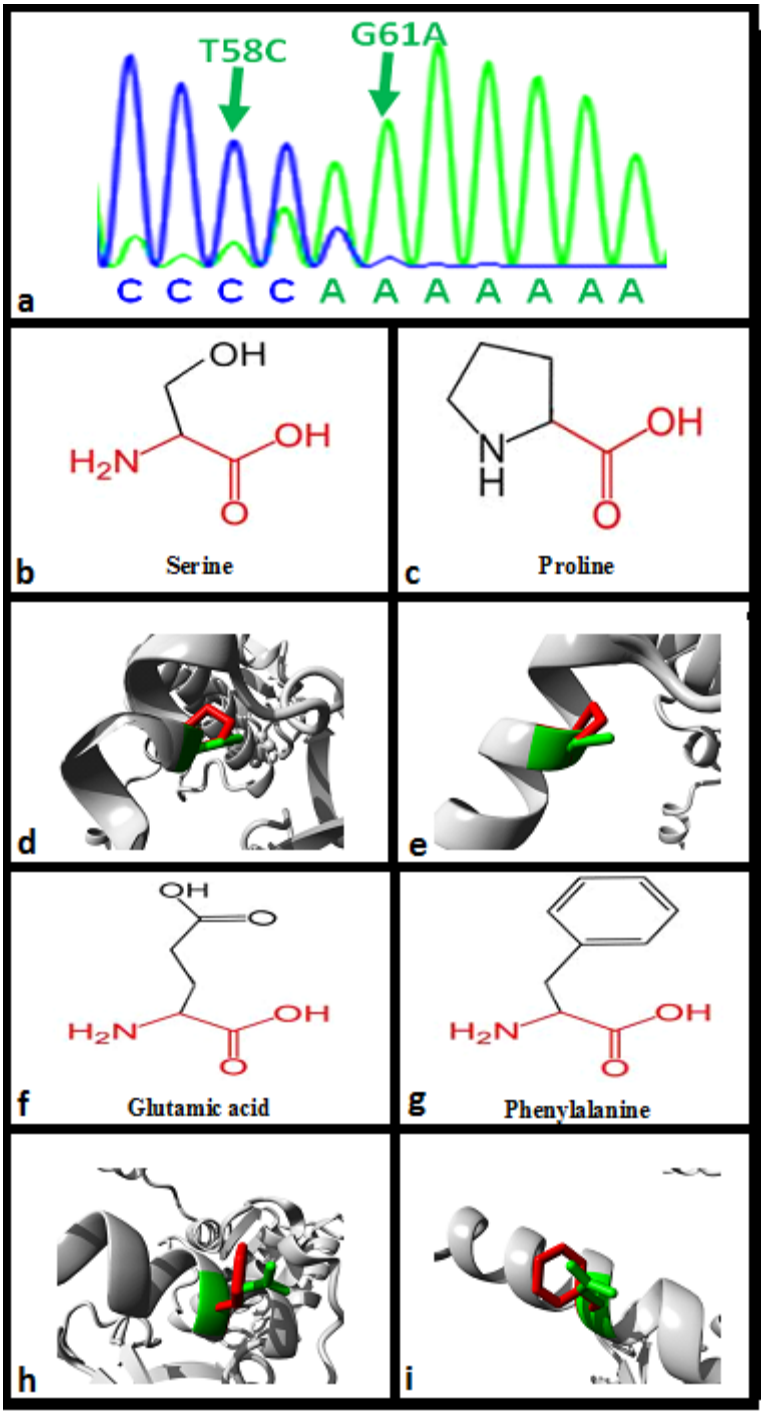

Figure 4.28: Sequence and structure analysis of PTEN S287Pand E288F mutations

a. Sequence chromatogram showing T58C and G61A mutations in the gene sequence.
b. Wild residue Serine c. Mutant residue Proline.
d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of PTEN.
f. Wild Glutamic acid g. Mutant residue Phenylalanine.
e. & i. Comparison of mutant (red) and wild type (green) residues in the 3D structure of PTEN

4.9 PTEN Intronic Variants

We observed four novel intronic variants in PTEN along with a number of exonic variants (Table 4.20). The intronic variants were recorded in both cases and controls and thus were evaluated for their possible association with prostate cancer using unconditional logistic regression under different genetic models.

In our study population, we found no association between prostate cancer and the g.87891382 G>A variant in all models, AA vs TT (OR = 1.85; CI = 0.91-3.76; P = 0.09), AA vs TA (OR = 1.96; CI = 0.95-4.08; P = 0.07), TA vs TT (OR = 0.94; CI = 0.62-1.44; P = 0.79), Dominant model (OR = 1.08; CI = 0.73-1.61; P = 0.69) and Recessive model (OR =1.90; CI = 0.96-3.77; P = 0.07) except allelic contrast (OR = 1.87; CI = 1.36-2.58; P = 0.00).

The T allele of variant at g.87932999 had an association with increased risk of the disease in all models except TT vs CC (OR = 0.94; CI = 0.53-1.67; P=0.83). Highest OR was recorded for TC vs CC (OR = 4.14; CI = 2.60-6.59; P=0.00) and Dominant model (OR = 4.08; CI = 2.63-6.32; P=0.00). Our analyses also revealed positive correlation for the Recessive model (OR = 1.73; CI = 1.01-2.95; P=0.05) and allelic contrast (OR = 2.26; CI = 1.69-3.01; P=0.00).

The analysis of T>C variant at g.87960832 showed that, C allele carriers are significantly associated with the disease risk in Dominant model (OR = 1.53; CI = 1.03-2.27; P=0.04) and allelic contrast (OR = 1.38; CI = 1.03-1.85; P=0.03). Negative association was observed when compared through other models.

For the G allele of A>G variant at g.87960772, we found significant association for all the models except TC vs CC (OR = 0.83; CI = 0.45-1.54; P=0.56) (Table 4.21 and Table 4.22).

Table 4.20: PTEN intronic variants (Reference sequence; NC_000010.10)

S. No	Intron	Chr StartPosition	ChrEnd Position	Variant	Status
1.	1	87891382	87891382	g.87891382G>A	Novel
2.	4	87933000	87932999	g.87932999C>T	Novel
3.	7	87960832	87960832	g.87960832T>C	Novel
4.	7	87960772	87960772	g.87960772A>G	Novel

Table 4.21: Genotypes and their allelic frequencies of PTEN G/A transition at chromosomal position g.87891382, C/T transition at g.87932999, T/C transition at g.87960832 and A/G transition at g.87960772 in Prostate adenocarcinoma

Alleles	Genotype	Cases (n=200)			Controls n=200		
		No.	%age	Frequency	No.	%age	Frequency
g.87891382	G-allele	180	45.00	0.45	295.00	73.75	0.74
	A-allele	120	30.00	0.30	105.00	26.25	0.26
	GG	105	26.25	0.26	109.00	27.25	0.27
	GA	70	17.50	0.18	77.00	19.25	0.19
	AA	25	6.25	0.06	14.00	3.50	0.04
g.87932999	C-allele	201	50.25	0.50	278.00	69.50	0.70
	T-allele	199	49.75	0.50	122.00	30.50	0.30
	CC	42	10.50	0.11	104.00	26.00	0.26
	CT	117	29.25	0.29	70.00	17.50	0.17
	TT	41	10.25	0.10	26.00	6.50	0.07
g.87960832	T-allele	248	62.00	0.62	277.00	69.25	0.69
	C-allele	152	38.00	0.38	123.00	30.75	0.31
	TT	80	20.00	0.20	101.00	25.25	0.25
	TC	88	22.00	0.22	75.00	18.75	0.19
	CC	32	8.00	0.08	24.00	6.00	0.06
g.87960772	A-allele	235	58.75	0.59	334.00	83.50	0.84
	G-allele	165	41.25	0.41	66.00	16.50	0.17
	AA	108	27.00	0.27	151.00	37.75	0.38
	AG	19	4.75	0.05	32.00	8.00	0.08
	GG	73	18.25	0.18	17.00	4.25	0.04

Table 4.22: Association of PTEN G/A transition at chromosomal position g.87891382, C/T transition at g.87932999, T/C transition at g.87960832 and A/G transition at g.87960772 in Prostate adenocarcinoma

Alleles	Genetic Model	OR	LCI	UCI	Z-stat	P-value
g.87891382	Allelic contrast (A vs G)	1.87	1.36	2.58	3.83	0.00
	Rare vs Common (AA vs GG)	1.85	0.91	3.76	1.71	0.09
	Rare vs Heterozygotes (AA vs GA)	1.96	0.95	4.08	1.81	0.07
	Heterozygotes vs Common (GA vs GG)	0.94	0.62	1.44	-0.27	0.79
	Dominant (AA+GA vsGG)	1.08	0.73	1.61	0.40	0.69
	Recessive (AA vs GG+GA)	1.90	0.96	3.77	1.83	0.07
g.87933000	Allelic contrast (T vs C)	2.26	1.69	3.01	5.51	0.00
	Rare vs Common (TT vs CC)	3.90	2.13	7.17	4.39	0.00
	Rare vs Heterozygotes (TT vs TC)	0.94	0.53	1.67	-0.20	0.84
	Heterozygotes vs Common (TC vs CC)	4.14	2.60	6.59	5.99	0.00
	Dominant (TT+TC vs CC)	4.08	2.63	6.32	6.27	0.00
	Recessive (TT vs CC+TC)	1.73	1.01	2.95	1.99	0.05
g.87960831	Allelic contrast (C vs T)	1.38	1.03	1.85	2.16	0.03
	Rare vs Common (CC vs TT)	1.68	0.92	3.08	1.69	0.09
	Rare vs Heterozygotes (CC vs TC)	1.14	0.62	2.10	0.41	0.68
	Heterozygotes vs Common (CT vs TT)	1.48	0.97	2.27	1.81	0.07
	Dominant (CC+TC vs TT)	1.53	1.03	2.27	2.11	0.04
	Recessive (CC vs TT+TC)	1.40	0.79	2.47	1.15	0.25

g.87960772	Allelic contrast (G vs A)	3.55	2.55	4.95	7.52	0.00
	Rare vs Common (GG vs AA)	6.00	3.35	10.75	6.03	0.00
	Rare vs Heterozygotes (GG vs AG)	7.23	3.33	15.70	5.00	0.00
	Heterozygotes vs Common (AG vs AA)	0.83	0.45	1.54	-0.59	0.56
	Dominant (GG+AG vs AA)	2.63	1.72	4.02	4.44	0.00
	Recessive (GG vs AA+AG)	6.19	3.48	10.99	6.22	0.00

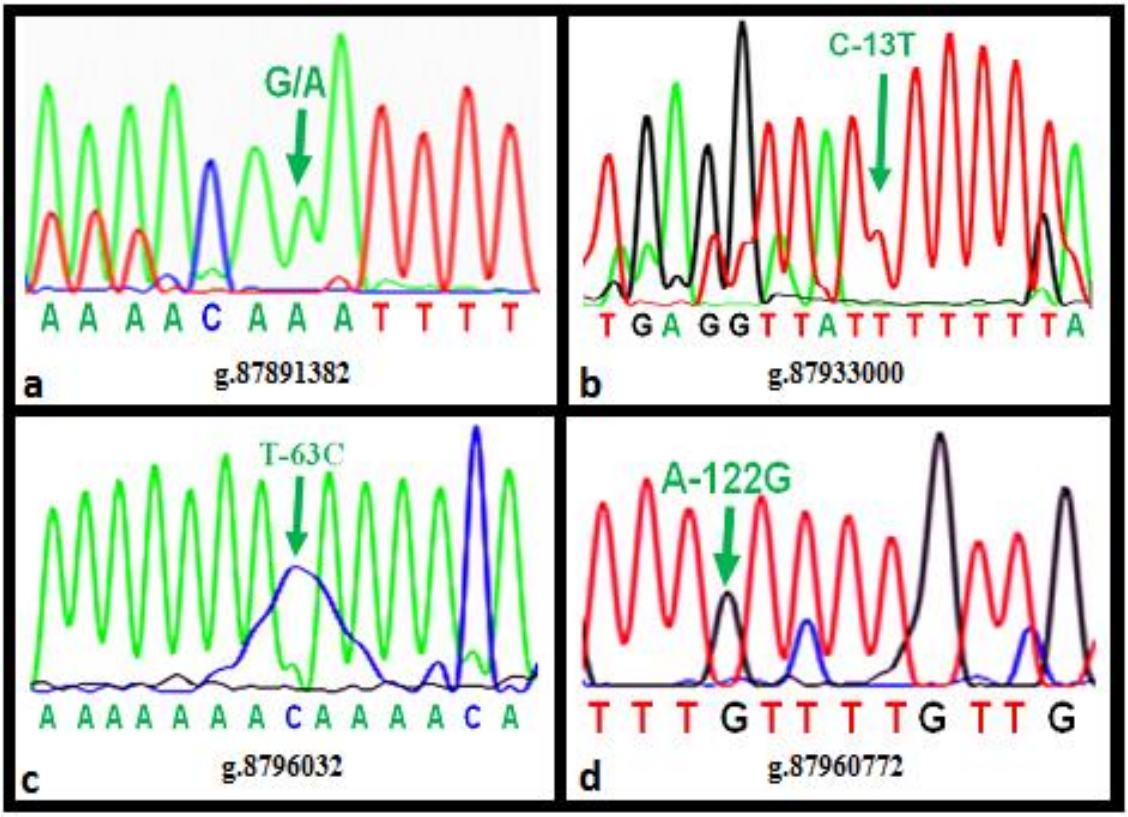

Figure 4.29: Sequencing results of PTEN gene intronic variants.

a. Sequence chromatogram showing G>A transition at genomic position g.87891382.
b. Sequence chromatogram showing C>T transition at genomic position g.87933000.
c. Sequence chromatogram showing T>C transition at genomic position g.8796032.
d. Sequence chromatogram showing A>G transition at genomic position g.87960772.

The PTEN phosphatase is a vital regulatory hub of PI3K-mediated signaling (Figure 4.30 and Figure 4.31) and expressing multiple isoforms of PI3K and PIP3-dependent downstream kinases in mammals. PTEN is a unique tumor suppressor that encodes a common repressor of both the pathways and thus has been found frequently mutated in tumors and hereditary cancer syndromes (Leslie and Downes, 2004, Salmena et al., 2008, Leslie and Foti, 2011). Some of the publications have pointed towards its functions independent of its lipid phosphatase activity (Planchon et al., 2008, Song et al., 2011) but its tumor suppressor property is mainly due to its catalytic activity.

The P-loop of PTEN includes catalytic Cys residue, sensitive to mutations and constitutes a tumor associated mutations hot spot (Bonneau and Longy, 2000, Eng, 2003). The TI and WPD loops are more permissive to small genetic changes. WPD loop harbors residues like Asp-92 and His-93, critical for its function. In our study we have reported novel mutations like E91K and H93Q. E91K is very close to Asp-92 (Denu et al., 1995). Both the residues (Asp-92 and His-93) have very low tolerance to genetic alterations and thus their substitution yields an inactive protein (Denu et al., 1995, Xiao et al., 2007).

In functional impact analysis, we observed that A86P, H93Q, E99X, H272F and E288F are damaging and E91K, Q97H, E157K and S287P are tolerated. A large number of mutations related to human tumors have already been reported in PTEN (Ali et al., 1999), some of them targeting its exon–intron junctions and promoter region. Majority of these mutations affect its expression level and cause functional deficiency (Zhou et al., 2003). Some of its mutations in coding region are nonsense or frame-shifts, generating unstable and truncated products (Georgescu et al., 2000). We also have reported one novel mutation at codon 99, which creates a stop signal and results in a truncated polypeptide. Functional characterization of some of these mutations have revealed that, the resulting residue substitutions produce PTEN proteins with compromised intrinsic catalytic activity (Kato et al., 2000).

Our findings show a relatively strong association between prostate cancer and the majority of our observed mutations at the PTEN gene. It is recommended that further

studies are required to verify the functional impacts endorsed by our reported mutations on PTEN activity.

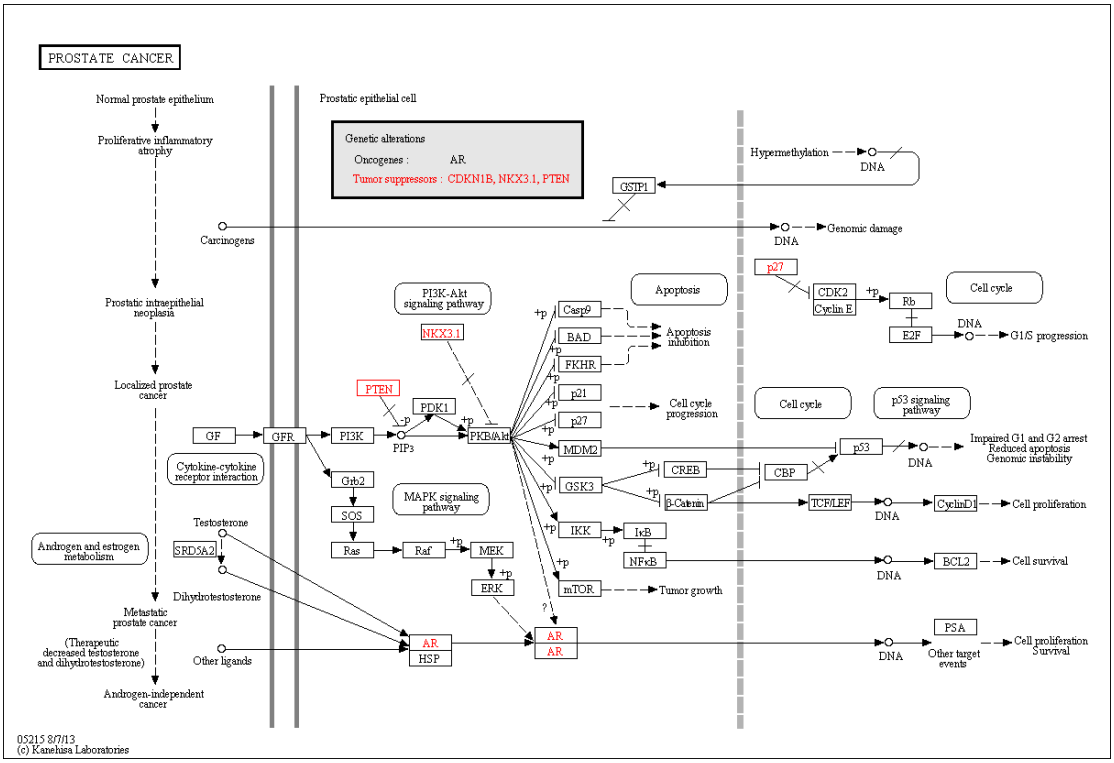

Figure 4.30: Prostate cancer - Reference pathway from KEGG.

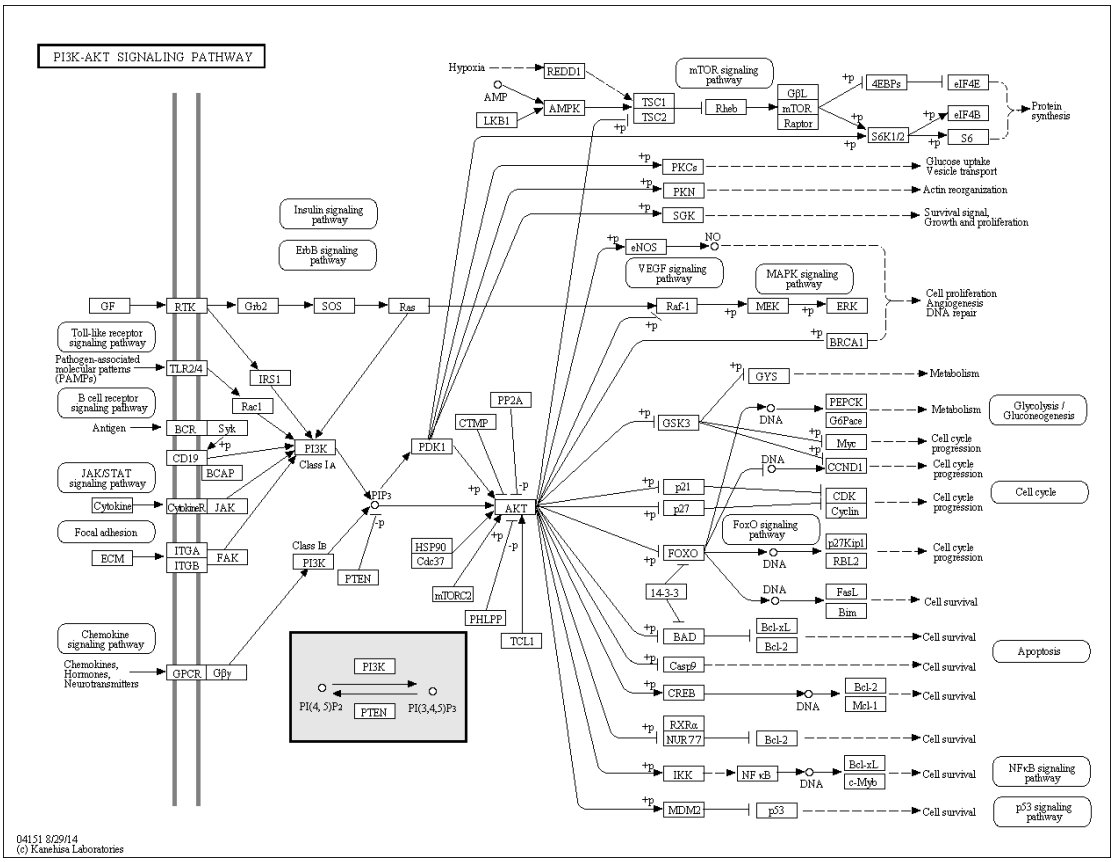

Figure 4.31: PI3K-Akt signaling pathway from KEGG.

The phosphatidylinositol 3'–kinase (PI3K)-Akt signaling pathway is activated by many types of cellular stimuli or toxic insults and regulates fundamental cellular functions such as transcription, translation, proliferation, growth, and survival. The binding of growth factors to their receptor tyrosine kinase (RTK) or G protein-coupled receptors (GPCR) stimulates class Ia and Ib PI3K isoforms, respectively. PI3K catalyzes the production of phosphatidylinositol-3,4,5-triphosphate (PIP3) at the cell membrane. PIP3 in turn serves as a second messenger that helps to activate Akt. Once active, Akt can control key cellular processes by phosphorylating substrates involved in apoptosis, protein synthesis, metabolism, and cell cycle.

4.10 AR (Androgen receptor) gene

AR gene is located on the X chromosome and is also known as the dihydrotestosterone receptor. It belongs to the class of nuclear receptors known as activated class-I steroid receptors. This class also includes glucocorticoid receptor, mineralocorticoid receptor and progesterone receptor, which recognize AREs (canonical androgen response elements). These AREs are inverted repeats of 5'-TGTTCT-3'. Androgen receptor consists of N-terminal and C-terminal activation domains, a ligand binding domain and a poly-glutamine tract. The N-terminal and C-terminal activation domains have selected activation function-1 (AF-1) and AF-2, (Callewaert et al., 2003).

AR is a part of a ternary complex containing AR, EFCAB6/DJBP and PARK7, interacting with HIPK3 and NR0B2 in the presence of androgen. Its ligand-binding domain interacts with KAT7/HBO1 in the presence of dihydrotestosterone, with EFCAB6/DJBP, NCOA1, NCOA2, NCOA3, NCOA4, PQBP1, PELP1, RBAK, RANBP9, RREB1, SLC30A9, SRA1, SPDEF, TGFB1I1, ZMIZ1/ZIMP10, ZMIZ2/ZMIP7 and ZNF318. Both ZMIZ1/ZIMP10 and ZMIZ2/ZMIP7 promote its transactivation activity. Its N-terminal poly-Gln region binds Ran and results in an enhanced AR-mediated transactivation. This domain interacts with HIP1, PRMT2, RNF6, RNF14, TNK2, TRIM24, TRIM68 and USP26. It also interacts with RANBP10 and enhances dihydrotestosterone-induced AR transcriptional activity.

At the first step, we checked through PhylomeDB4 and TreeFam for the conservativeness of AR across different species. It was observed that the protein is majorly conserved evolutionarily (Figure 4.32). In our study we have observed thirteen point mutations (Table 4.23) in the exonic and two novel alterations in the non-coding region (one in intron-1 and the other in 3' UTR) of AR gene (Table 4.25). Out of the 13 exonic variants, 11 are novel while 2 were previously reported i.e D880Y and S889S. Each of the variant was carefully investigated through different online resources for their possible sequence, structural and functional impact. It was further observed that six mutations are in the Hormone receptor domain while the rest are outside of the domains (Figure 4.32 c)

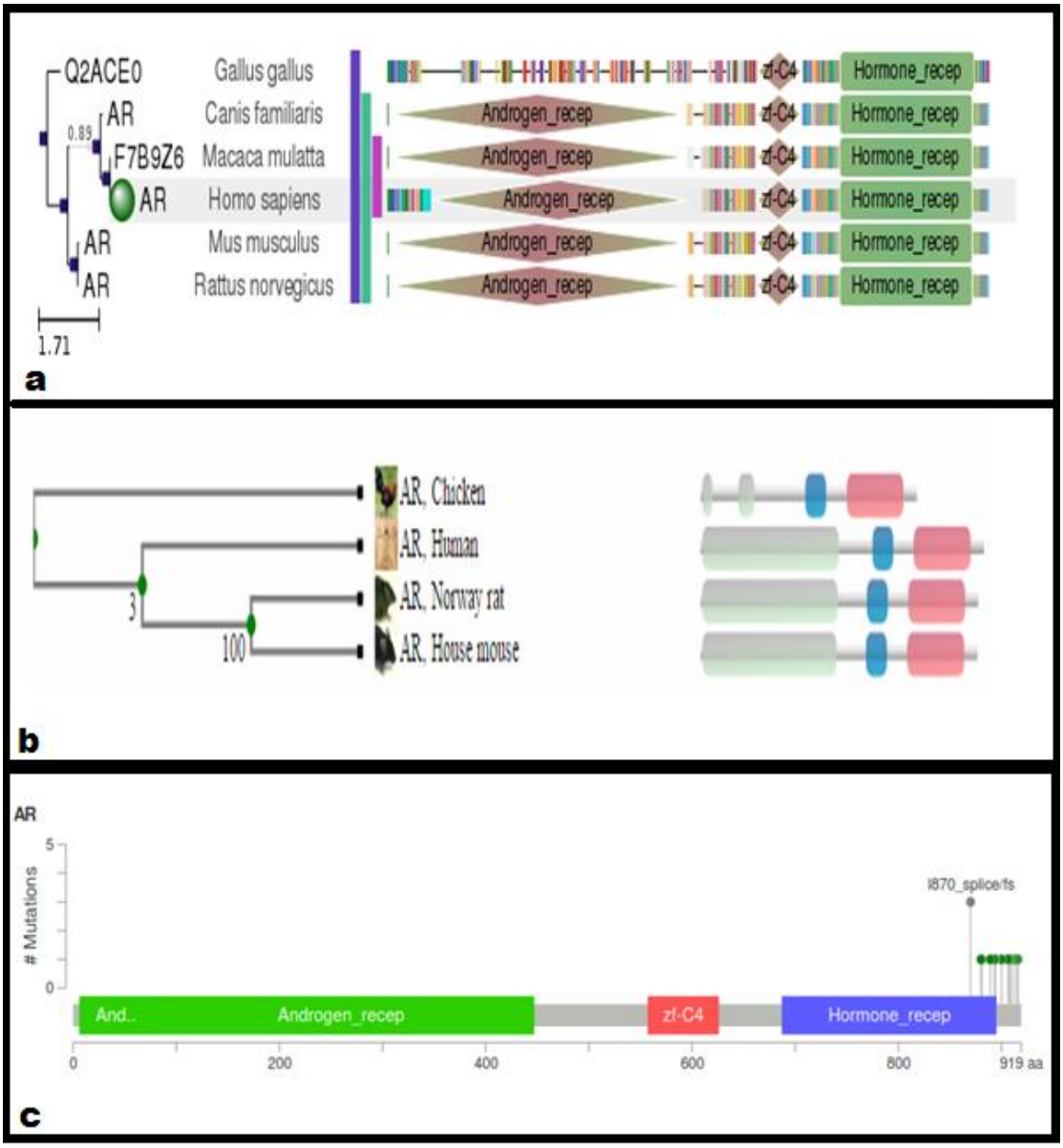

Figure 4.32: Evolutionary history and domains of Androgen receptor.

a. Evolutionary history of PTEN along with pfam domains through PhylomeDB.
b. PTEN gene tree through TreeFam
c. Lolipop plot of AR protein showing domains and mutations.

Table 4.23: AR Exonic variants

S. No	Exon	Ref Allele	VarAllele	Chr StartPos	ChrEnd Pos	Residue change	MutType	Status
1.	8	-	G	67723688	67723688	I870fs	fs sub	Novel
2.	8	TT	G-	67723687	67723688	I870_splice	splice	Novel
3.	8	G	T	67723716	67723716	D880Y	Missense	Reported
4.	8	G	T	67723721	67723721	L881L	Silent	Novel
5.	8	C	T	67723745	67723745	S889S	Silent	Reported
6.	8	GG	AT	67723757	67723758	P893P& E894*	Silent Nonsense	Novel Novel
7.	8	A	G	67723776	67723776	I900V	Missense	Novel
8.	8	C	G	67723804	67723804	S909C	Missense	Novel
9.	8	G	T	67723796	67723796	K906N	Missense	Novel
10.	8	T	C	67723801	67723801	L908P	Missense	Novel
11.	8	TC	GG	67723813	67723814	V912G	Missense	Novel
12.	8	A	T	67723821	67723821	I915F	Missense	Novel
13.	8	A	C	67723825	67723825	Y916S	Missense	Novel

Table 4.24: Status of AR Exonic variants confirmed through HGMD and Ensembl along with mutants functional impact analysis through different servers

S. No	Residue change	Reported By	SIFT	PROVEAN	MutationAssessor	PolyPhen2
1.	I870fs	Novel	Damaging	Deleterious	High	Probably Damaging
2.	I870_splice	Novel	Damaging	Deleterious	High	Probably Damaging
3.	D880Y	Reported	Tolerated	Deleterious	Low	Probably Damaging
4.	L881L	Novel	Tolerated	Neutral	synonymous	Neutral
5.	S889S	Reported	Tolerated	Neutral	synonymous	Neutral
6.	P893P	Novel	Tolerated	Neutral	synonymous	Neutral
7.	E894*	Novel	Damaging	Deleterious	High	Probably Damaging
8.	I900V	Novel	Tolerated	Neutral	Medium	Probably Damaging
9.	S909C	Novel	Damaging	Neutral	Low	Probably Damaging
10.	K906N	Novel	Damaging	Deleterious	Medium	Probably Damaging
11.	L908P	Novel	Damaging	Deleterious	Medium	Probably Damaging
12.	V912G	Novel	Damaging	Neutral	Medium	Probably Damaging
13.	I915F	Novel	Damaging	Neutral	Low	Probably Damaging
14.	Y916S	Novel	Damaging	Deleterious	Low	Probably Damaging

4.10.1 Ins3G (g.67723688) I870M (frame shift)

The mutation was confirmed as a novel one from both the HGMD and Ensembl. The insertion of G nucleotide in codon 870 at position 3 of exon-8 in AR gene will result in a frame shift. The mutation changes ATT codon (I) into ATG (M) and changes the downstream sequence **IARELHQFTF** into **M C E R A A S V H F** and as a result a stop codon will be introduced at position 880 and thus the result is a truncated protein. Functionally the mutation was predicted as damaging by SIFT, probably damaging by PolyPhen2, deleterious by PROVEAN and high by MutationAssessor (Table 4.24 and Figure 4.33).

4.10.1 T2G and del3T (g.67723687-67723688) I870_splice

The mutation was confirmed as a novel from HGMD and Ensembl. This double mutation substitutes T nucleotide into G at position 2 and deletes a T nucleotide normally present at position 3 in exon-8 of AR gene. T2G transversion is at the splice donor site and deletion at position is just next to this donor site. This mutation will probably result in skipping of exon 8 and thus will produce a truncated protein. Functionally the mutation was predicted as damaging by SIFT, probably damaging by PolyPhen2, deleterious by PROVEAN and high by MutationAssessor. It is therefore recommended that the mutant may be characterized in the wet lab for further exploration and validation (Table 4.24 and Figure 4.33).

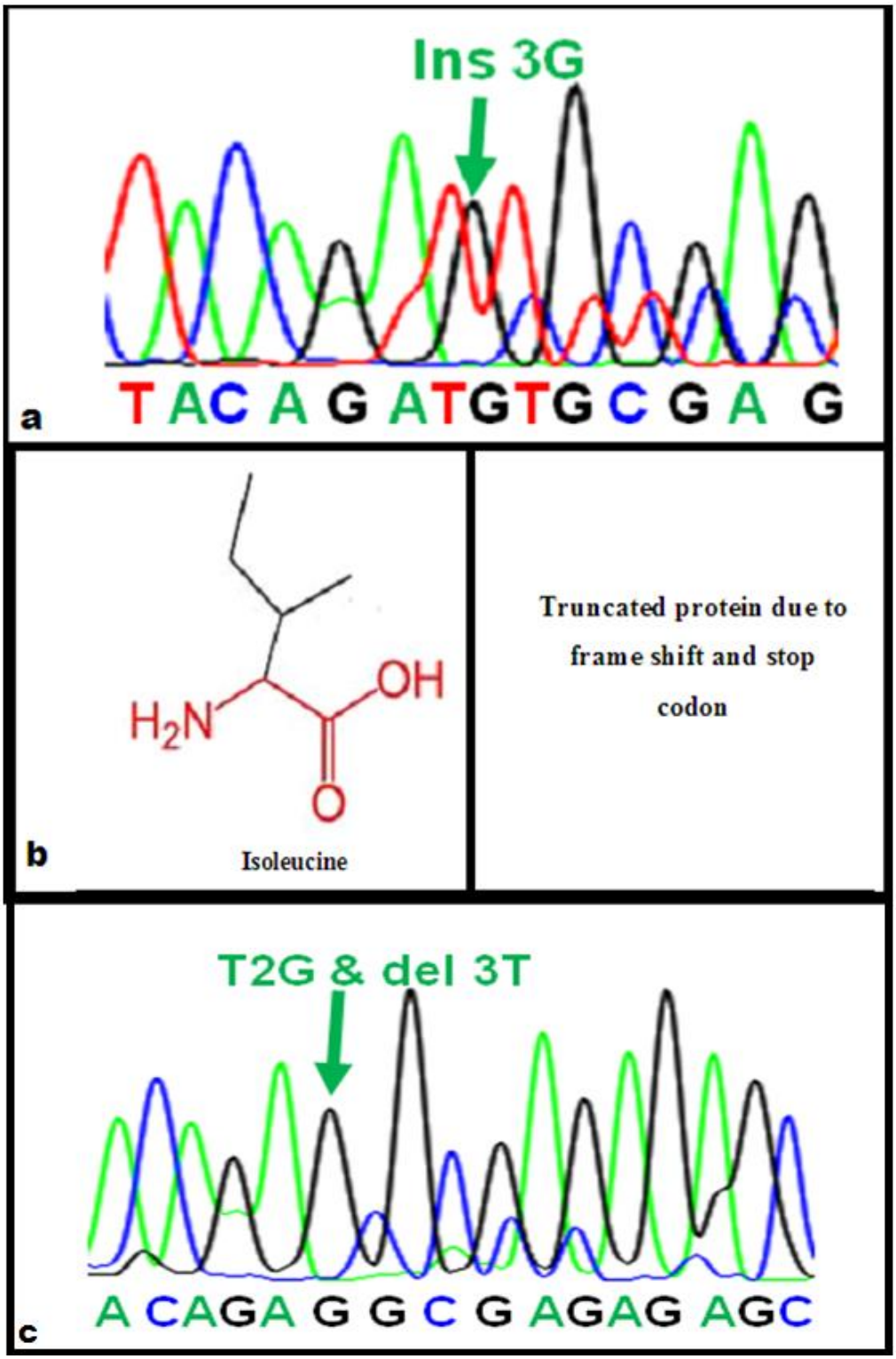

Figure 4.33: Sequence and structure analysis of ARI870M (frame shift) and T2G + del3T mutations

a. Sequence chromatogram showing ins3G mutation in the gene sequence.
b. Wild residue Isoleucine
c. Sequence chromatogram showing T2G + del3T mutation in the gene sequence.

4.10.2 G31T (g.67723716) D880Y

This mutation of codon 880 in exon-8 of AR gene at position 31 results in the substitution of Aspartic acid to Tyrosine. The mutation was confirmed as previously reported. The mutant residue is bigger, neutral and more hydrophobic than the wild residue which was negatively charged. The substituted residue was not in direct contact with a ligand, however this change affects the local stability of the protein which could affect the ligand interactions made by the adjoining residues.

The mutation is located within a region annotated in UniProt as a special region "Interaction with LPXN, KAT7, CCAR1 and Ligand-binding". The differences in amino acid properties can disturb this region. The wild residue interacts with Lys-884 through a salt bridge, the difference in amino acid charges will disturb this interaction and thus its function.

The wild type residue occurs often at this position in the sequence, but other residues have also been observed here. This residue is part of an interpro domain named Nuclear Hormone Receptor, Ligand-Binding (IPR008946), annotated with the Gene-Ontology (GO) terms Steroid Hormone Receptor Activity (GO:0003707) and Sequence-Specific DNA Binding Transcription Factor Activity (GO:0003700). These GO annotations indicate that, the domain has a function in Molecular Transducer Activity (GO:0060089), Receptor Activity (GO:0004872) and Nucleic Acid Binding Transcription Factor Activity(GO:0001071), Nucleic Acid Binding (GO:0003676).

The wild residue loses its charge because of this mutation so it can cause loss of interactions with other molecules (Figure 4.34). Functionally the mutation was predicted as tolerated by SIFT, probably damaging by PolyPhen2, deleterious by PROVEAN and low by MutationAssessor (Table 4.24).

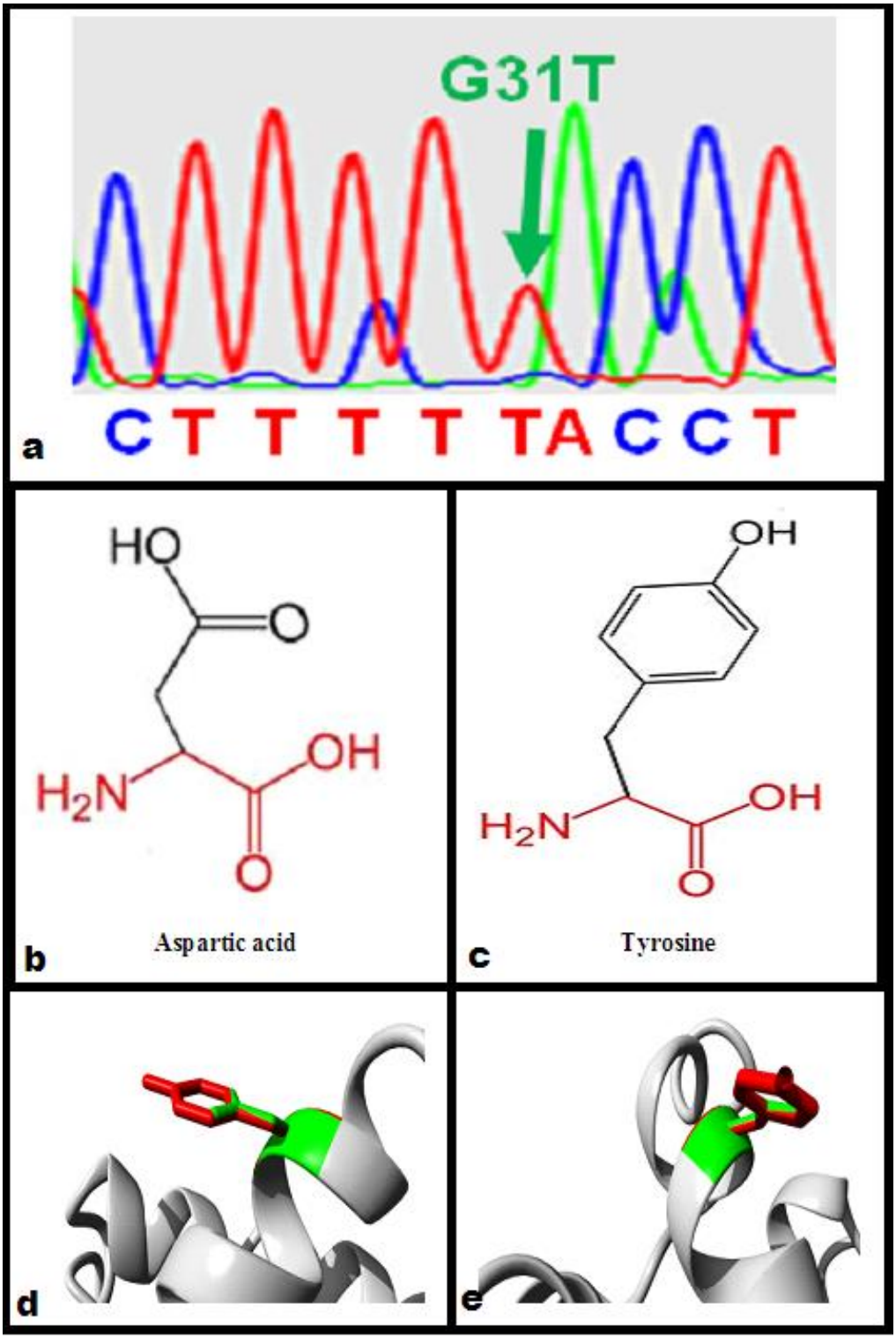

Figure 4.34: Sequence and structure analysis of ARD880Y mutation

a. Sequence chromatogram showing G31T mutation in the gene sequence.
b. Wild residue Aspartic acid c. Mutant residue Tyrosine.
d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of AR protein.

4.10.3 G36T (g.67723721) L881L

This transversion at codon 881 in exon-8 at position 36 of AR is a silent mutation, confirmed as a novel one from HGMD and Ensembl. The mutation was confirmed as novel through HGMD and Ensemble. The substitution will have no effect on the protein structure and function. The mutation was predicted as neutral by the online resource (Table 4.24 and Figure 4.35).

4.10.4 C60T (g.67723745) S889S

This transition in codon 889 of AR gene at position 60 is also of silent nature. It was confirmed as a previously reported one. It is a silent mutation and is already reported (Figure 4.35). The mutation was predicted as neutral by the online resources (Table 4.24).

4.10.5 GG72-73AT (g.67723757-67723758) P893P+E894X

The mutation was confirmed as novel both from HGMD and Ensembl. It results in the transition of CCG to CCA codon number 893 and transversion of GAA to TAA codon number 894 at position 72 and 73 in exon8 of AR gene. Mutation of CCG to CCA codon has no impact on the protein sequence because of the codon degeneracy (both code for Proline), while GAA to TAA introduces a stop codon at position 894 instead of glutamic acid. The result will affect the protein on hormone receptor domain. The mutation was predicted as damaging (Figure 4.35). Functionally the mutation was predicted as damaging by SIFT, probably damaging by PolyPhen2, deleterious by PROVEAN and high by MutationAssessor (Table 4.24).

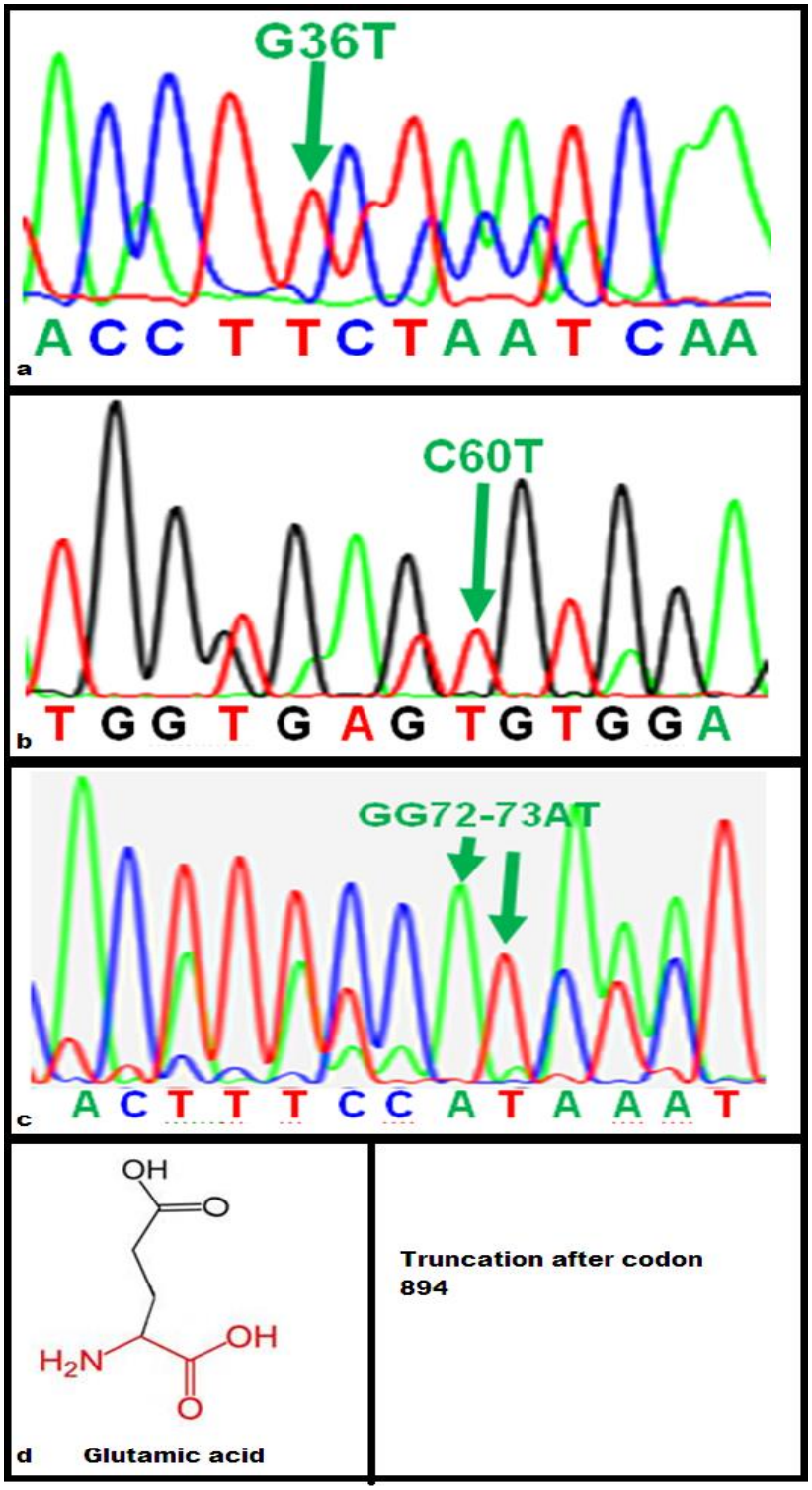

Figure 4.35: Sequence and structure analysis of AR L881L, S889S and P893P+E894X mutation

a. Sequence chromatogram showing G36Ttransversion
b. Sequence chromatogram showing C60Ttransition
c. Sequence chromatogram showing GG72-73ATmutation
d. Wild residue Glutamic acid

4.10.6 A91G (g.66723776) I900V

The mutation was confirmed as novel both from HGMD and Ensembl. It results in transversion of ATC to GTC at codon (900) in exon-8 of AR gene. This mutation results in substitution of Isoleucine to Valine at position 900. The mutant residue is smaller than the wild. According to the PISA-database, containing protein assemblies which are likely to be biologically relevant, the wild residue is involved in a multimer contact. These multimer contacts are lost due to the small size of the mutant residue. The mutation occurs within a region, annotated in UniProt as a special region responsible for interaction with LPXN, KAT7, CCAR1 and Ligand-binding. The differences in amino acid properties can disturb this region and thus its function.

This residue is part of an interpro domain named Nuclear Hormone Receptor and Ligand-Binding (IPR008946), annotated with the Gene-Ontology (GO) terms, Steroid Hormone Receptor Activity (GO:0003707) and Sequence-Specific DNA Binding Transcription Factor Activity (GO:0003700). More broadly speaking, these GO annotations indicate that the domain has a function in Molecular Transducer Activity (GO:0060089), Receptor Activity (GO:0004872), Nucleic Acid Binding Transcription Factor Activity (GO:0001071) and Nucleic Acid Binding (GO:0003676).

The domain having the mutation is vital for the main activity of AR protein but it is possible that, differences between the wild and mutant residue can disturb the core structure and thereby affect its catalytic activity (Figure 4.36). Functionally the mutation was predicted as tolerated by SIFT, probably damaging by PolyPhen2, neutral by PROVEAN and medium by MutationAssessor (Table 4.24).

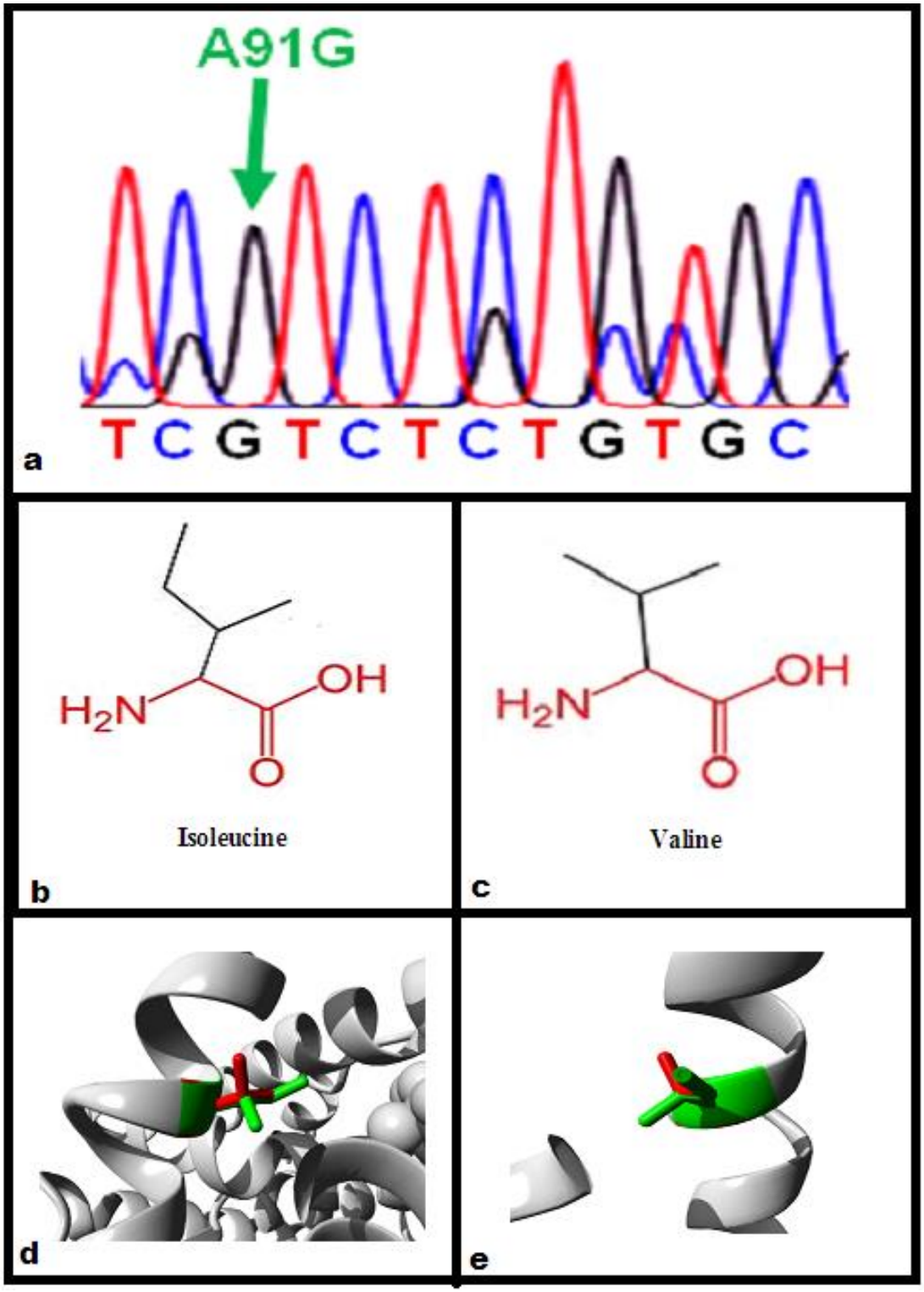

Figure 4.36: Sequence and structure analysis of ARI900V mutation

a. Sequence chromatogram showing A91G mutation in the gene sequence.
b. Wild residue Isoleucine
c. Mutant residue Valine.
d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of Androgen receptor.

4.10.7 G111T (g.67723796) K906N

The mutation was confirmed as novel from both the HGMD and Ensemble. This transversion of G to T at codon 906 (position 111) in exon-8 of AR gene, results in substitution of lysine to aspargine. The mutant residue is smaller and neutral while the wild residue was positively charged and forms a hydrogen bond with the Glu-903. Due to size differences, the mutant residue is no more in suitable distance and position to make the same hydrogen bond.

The mutation is located in a domain annotated in UniProt as a special region responsible for interaction with LPXN, CCAR1, KAT7 and Ligand-binding. The wild residue (lysine) favors α-helix orientation in the structure while the mutant residue (aspargine) does not favor this orientation. The differences in amino acid properties can disturb this region functionally

This residue is part of an interpro domain named Nuclear Hormone Receptor, Ligand-Binding (IPR008946), annotated with the Gene-Ontology (GO) terms, Steroid Hormone Receptor Activity (GO:0003707) and Sequence-Specific DNA Binding Transcription Factor Activity (GO:0003700). These GO annotations indicate that the domain has a function in Molecular Transducer Activity (GO:0060089), Receptor Activity (GO:0004872), Nucleic Acid Binding Transcription Factor Activity (GO:0001071) and Nucleic Acid Binding (GO:0003676).

Charge differences in between wild and mutant residues results in loss of interactions with other molecules. Functionally the mutation was predicted as damaging by SIFT, probably damaging by PolyPhen2, deleterious by PROVEAN and medium by MutationAssessor (Table 4.24 & Figure 4.37).

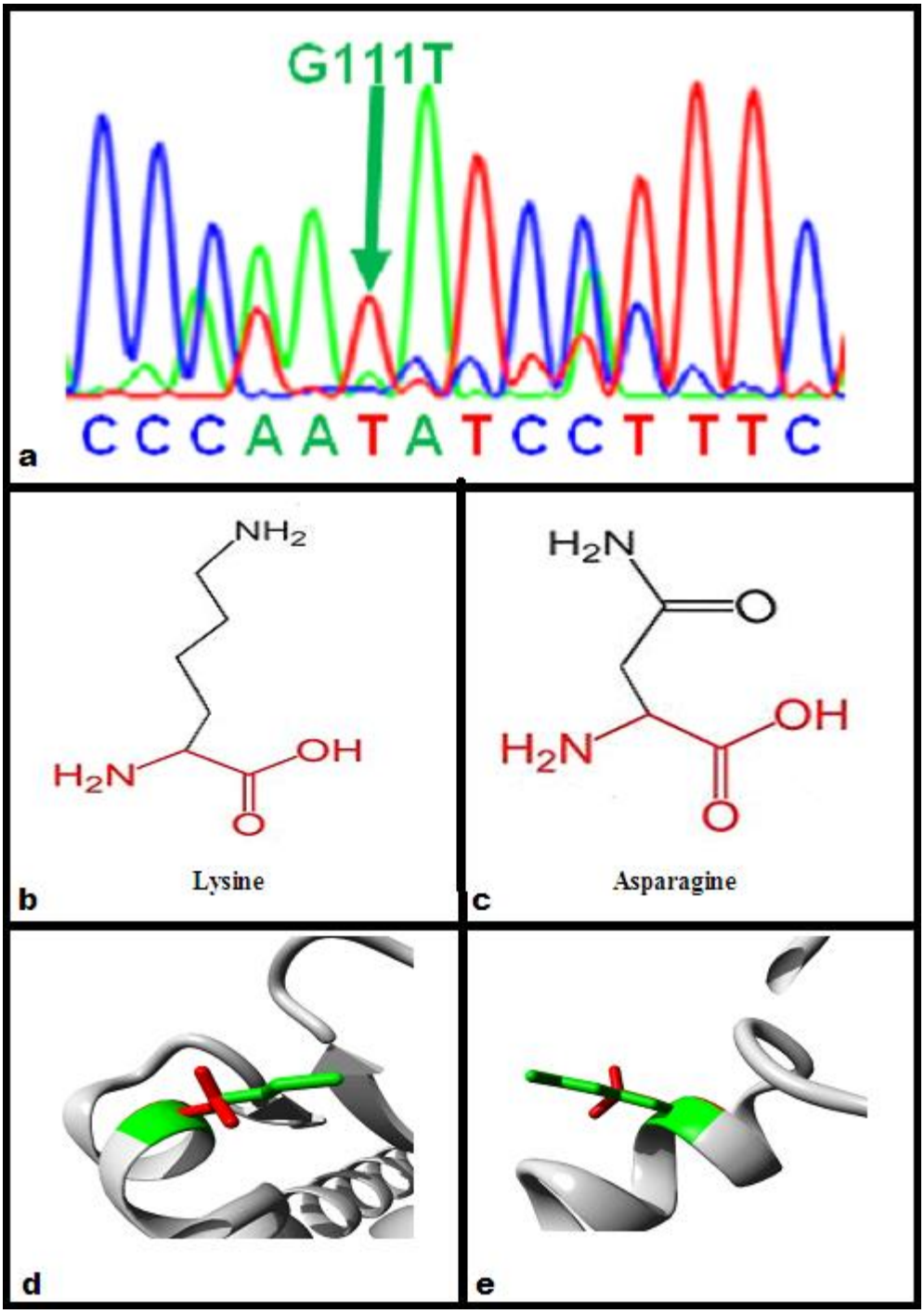

Figure 4.37: Sequence and structure analysis of ARK906N mutation

a. Sequence chromatogram showing G111T mutation in the gene sequence.
b. Wild residue Lysine
c. Mutant residue Asparagine.
d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of Androgen Receptor.

4.10.8 C119G (g.67723804) S909C

The mutation was confirmed as novel from HGMD and Ensemble. The transversion of C to G in codon 909 at position 119 in exon 8 of AR, results in substitution of Serine to Cysteine. The Serine at this position interacts with pro-905 through a hydrogen bond which gets compromised due to change in hydrophobicity, the mutant being more hydrophobic. It is thus possible that by disturbing the interactions, the mutation also affects the protein function by disturbing the signal transfer from binding domain to the activity domain

The mutation is position in a region annotated in UniProt as a special region responsible for interaction with LPXN, KAT7, CCAR1 and Ligand-binding". The differences in residue properties can functionally disturb this region. This residue is part of an interpro domain named Nuclear Hormone Receptor and Ligand-Binding (IPR008946), annotated with the Gene-Ontology (GO) terms Steroid Hormone Receptor Activity (GO:0003707) and Sequence-Specific DNA Binding Transcription Factor Activity (GO:0003700). These GO annotations indicates that the domain has a function in Molecular Transducer Activity (GO:0060089), Receptor Activity (GO:0004872), Nucleic Acid Binding Transcription Factor Activity (GO:0001071) and Nucleic Acid Binding (GO:0003676).

The wild type residue (Serine) is highly conserved at this position in other homologs and thus can be inferred as damaging to the protein. The wild residue prefers a turn in protein orientation, while the mutant prefers to be in another secondary structure and thus slightly destabilized the local conformation (Figure 4.38). Functionally the mutation was predicted as damaging by SIFT, probably damaging by PolyPhen2, neutral by PROVEAN and low by MutationAssessor (Table 4.24).

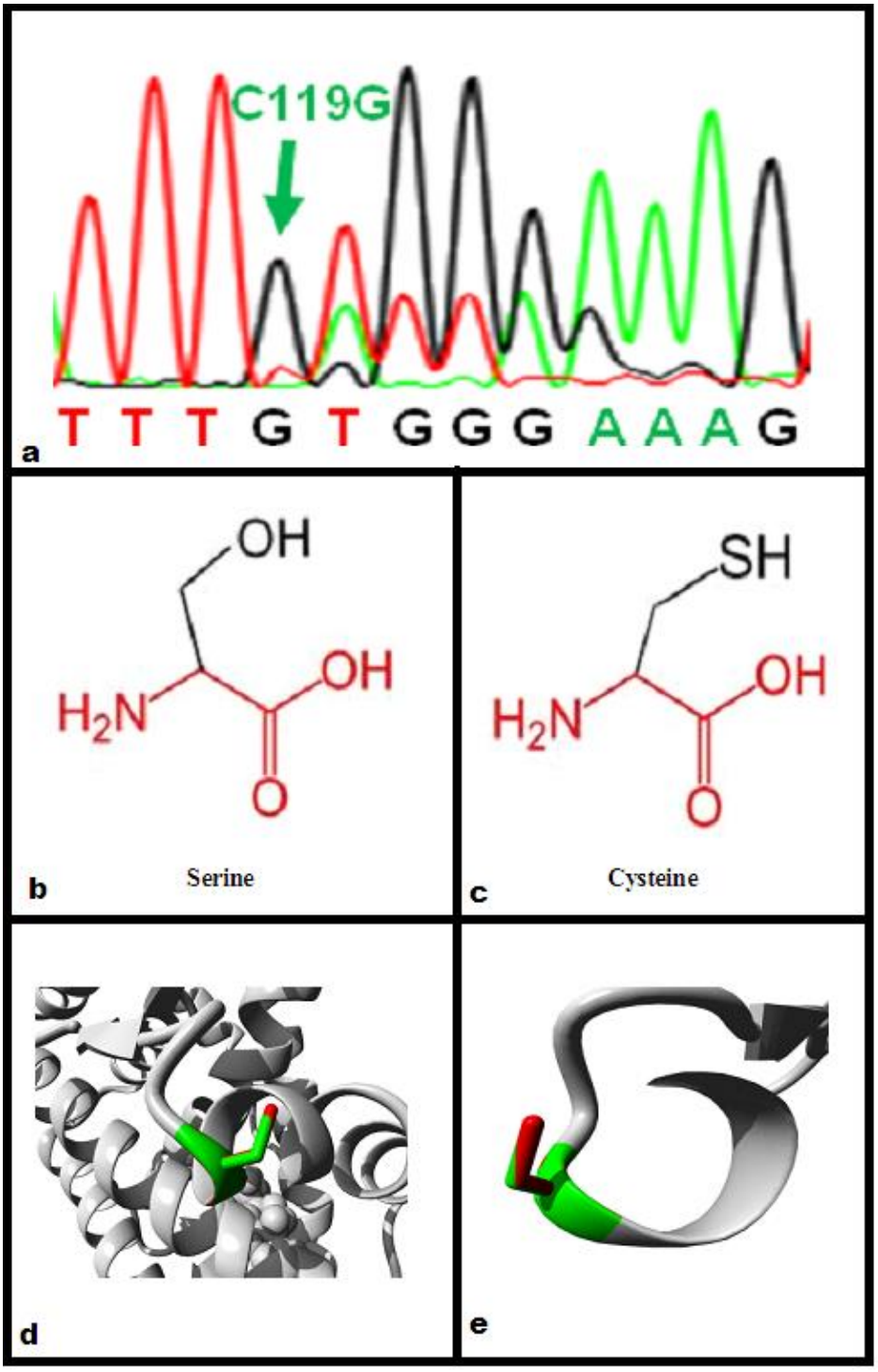

Figure 4.38: Sequence and structure analysis of Androgen Receptor S909C mutation

a. Sequence chromatogram showing C119G mutation in the gene sequence.
b. Wild residue Serine
c. Mutant residue Cysteine.
d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of Androgen Receptor.

4.10.9 T116C (g.67723801) L908P

The mutation was confirmed as novel both from HGMD and Ensemble. The transition of T to C at position 116 in the AR exon-8 at codon 908 results in the substitution of leucine to proline. The mutant residue is smaller and is located within a domain annotated in UniProt as a special region responsible for interaction with LPXN, KAT7 and Ligand-binding. The wild residue favors α-helix in the local conformation, which get disturbed due to differences in amino acid properties (Figure 4.39).

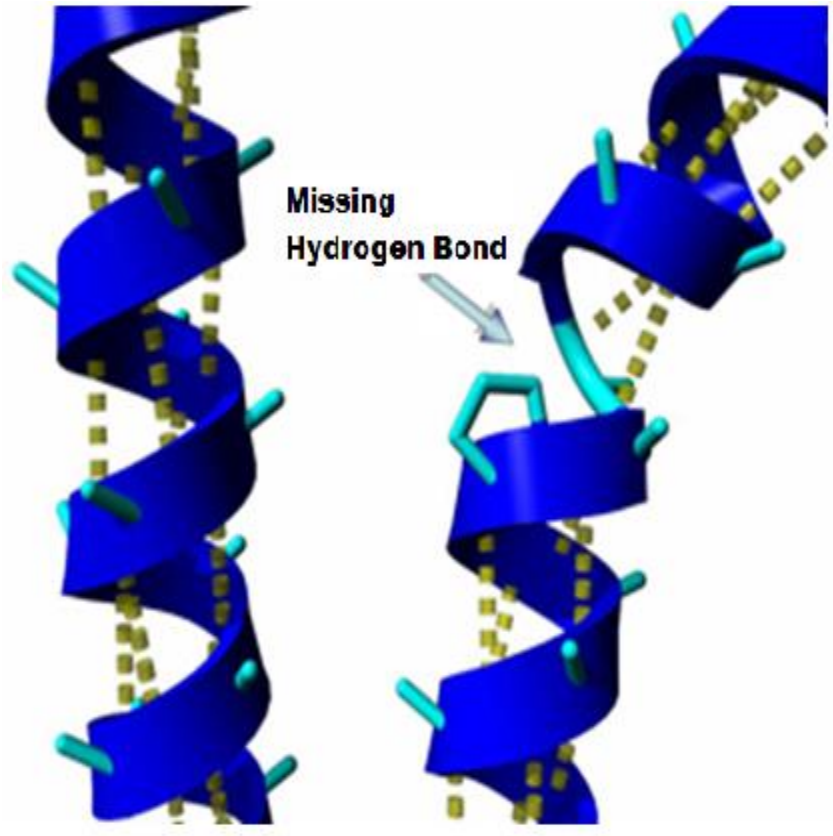

Figure 4.39: Proline disrupts α-helix when not located at one of the first 3 positions of that helix.

A mutation to Phenylalanine has previously been reported at this position, though different from our reported mutation. The wild type residue occurs often at this position in the sequence, but other residues have also been observed in the homologous sequences.

This residue is part of an interpro domain named Nuclear Hormone Receptor, Ligand-Binding (IPR008946), annotated with the Gene-Ontology terms, Steroid Hormone Receptor Activity (GO:0003707) and Sequence-Specific DNA Binding Transcription Factor Activity (GO:0003700), indicating its functions in Molecular Transducer Activity (GO: 0060089), Receptor Activity (GO:0004872), Nucleic Acid Binding Transcription Factor Activity (GO:0001071) and Nucleic Acid Binding (GO:0003676).

Leu-908 is part of an important domain required for the main activity of the Androgen receptor. The mutant residue is comparatively smaller and thus will cause an empty space in the protein core. The differences in properties are able to disturb the core structure of this important domain and thereby affect its catalytic activity. Functionally the mutation was predicted as damaging by SIFT, probably damaging by PolyPhen2, deleterious by PROVEAN and medium by MutationAssessor (Table 4.24 & Figure 4.40).

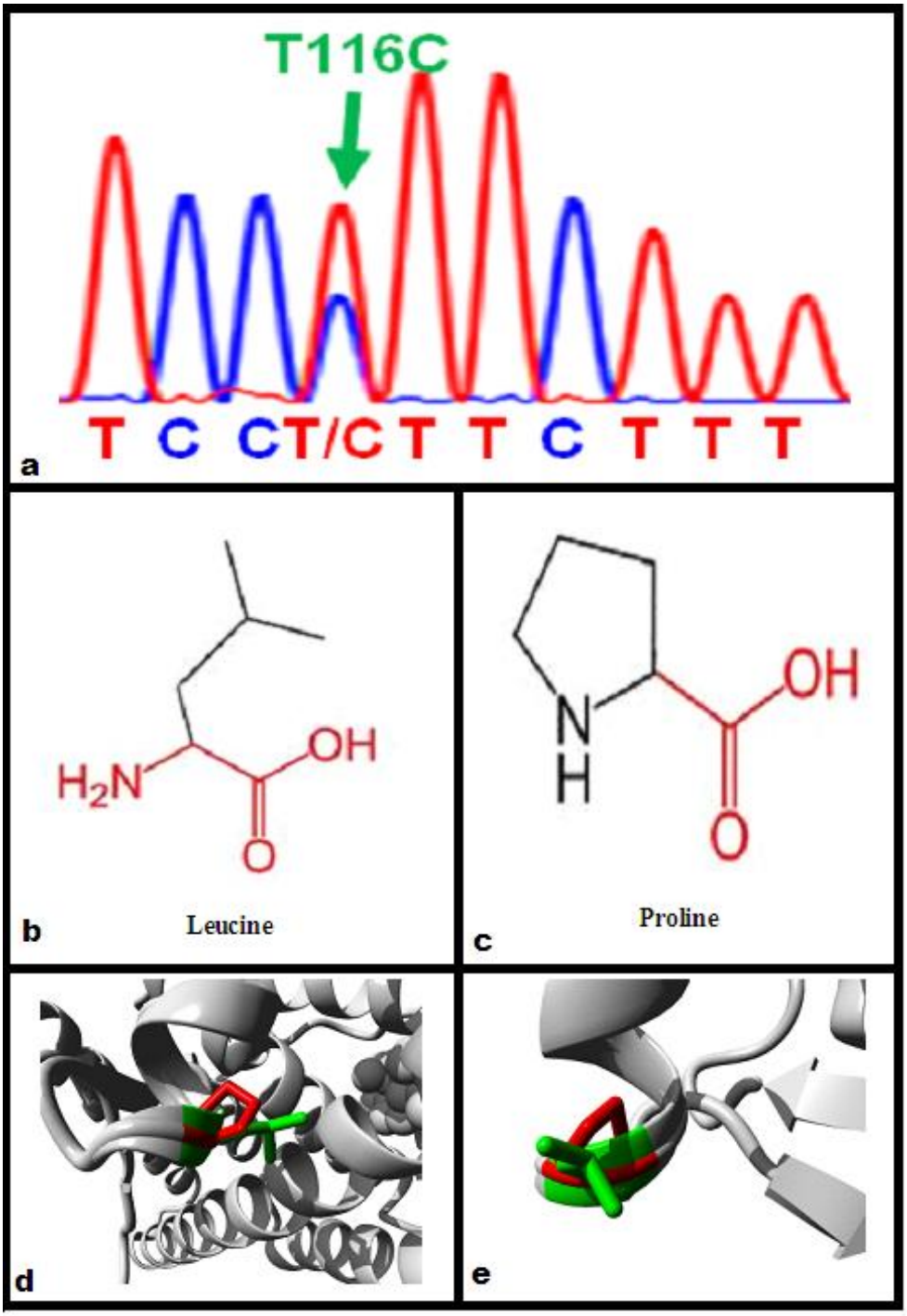

Figure 4.40: Sequence and structure analysis of Androgen Receptor L908P mutation

a. Sequence chromatogram showing T116Cmutation in the gene sequence.
b. Wild residue Leucine
c. Mutant residue Proline.
d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of Androgen Receptor.

4.10.10 TC128-129GG (g.67723813-67723814) V912G

Both the transversions in AR exon-8 at position 128-129 were confirmed as novel at this position and results in substitution of a single amino acids i.e. Valine to Glycine at codon 912. The mutant residue is comparatively smaller and is positioned in a stretch annotated in UniProt as a special region for interaction with LPXN, CCAR1, KAT7 and Ligand-binding, which can be disturbed functionally due to the differences in amino acid properties.

The transversion of TC128-129GG introduces a very flexible amino acid (glycine), which can disturb the essential rigidity at this position. A mutation to "Leucine" has previously been reported at this position. The wild type residue is not very conserved at this position in other homologous sequences. It is also possible that the mutation disturbs the protein interactions and so affects its function by disturbing the signal transfer from binding domain to the activity domain.

This residue is part of an interpro domain named Nuclear Hormone Receptor, Ligand-Binding (IPR008946), annotated with Gene-Ontology (GO) terms Steroid Hormone Receptor Activity (GO:0003707) and Sequence-Specific DNA Binding Transcription Factor Activity (GO:0003700). These GO annotations indicates that the domain has a function in Molecular Transducer Activity (GO:0060089), Receptor Activity (GO:0004872), Nucleic Acid Binding Transcription Factor Activity (GO:0001071) and Nucleic Acid Binding (GO:0003676).

The mutation will cause loss of hydrophobic interactions in the core of the protein due to the differences in hydrophobicity of wild and mutant residues (Figure 4.41). Functionally the mutation was predicted as damaging by SIFT, probably damaging by PolyPhen2, neutral by PROVEAN and medium by MutationAssessor (Table 4.24).

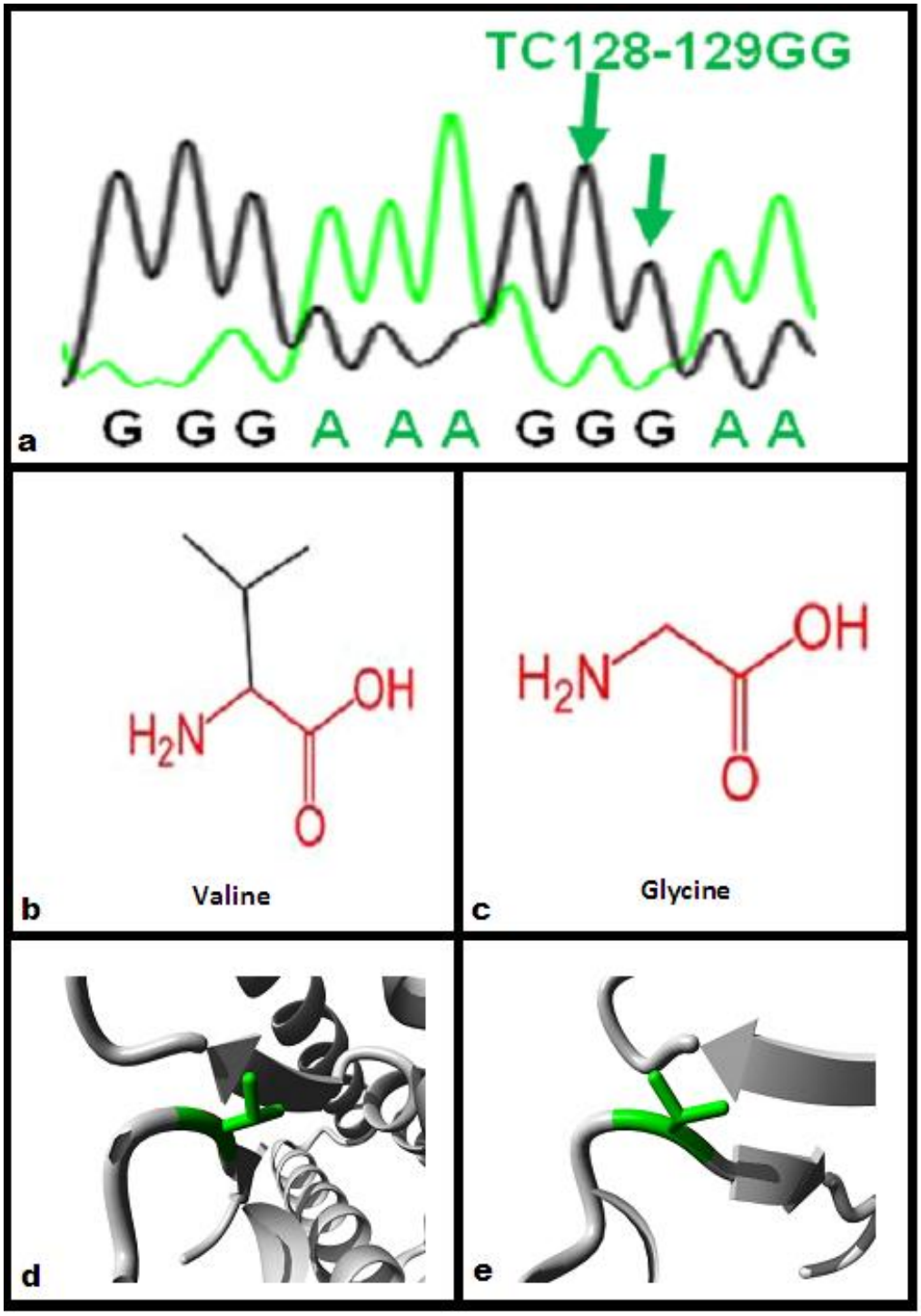

Figure 4.41: Sequence and structure analysis of Androgen Receptor V912G mutation

a. Sequence chromatogram showing TC128-129GG mutation in the gene sequence.
b. Wild residue Valine
c. Mutant residue Glycine.
d. & e. Comparison of mutant (red) and wild type (green) residues in the 3D structure of Androgen Receptor.

4.10.11 A136T and A140C (I915F & Y916S)

i. A136T (g.67723821) I915F

The transition of A to T in codon 915 at AR exon-8 results in substitution of isoleucine to phenylalanine at position 915. It was confirmed as a novel mutation from both HGMD and Ensemble. The mutant residue is comparatively bigger and is located in a region, annotated in UniProt as a special region for "Interaction with LPXN, KAT7, CCAR1 and Ligand-binding". The differences in amino acid properties can disturb this region functionally. The wild type residue occurs often at this position in the sequence, but other residues have also been observed here.

Isoleucine, the wild residue, is part of an interpro domain named Nuclear Hormone Receptor and Ligand-Binding (IPR008946) annotated with the Gene-Ontology (GO) terms Steroid Hormone Receptor Activity (GO:0003707) and Sequence-Specific DNA Binding Transcription Factor Activity (GO:0003700). More broadly speaking, these GO annotations indicates that the domain has a function in Molecular Transducer Activity (GO:0060089), Receptor Activity (GO:0004872), Nucleic Acid Binding Transcription Factor Activity (GO:0001071) and Nucleic Acid Binding (GO:0003676).

This residue is positioned on the surface and was not found to be in contact with other domains (Figure 4.42). Functionally the mutation was predicted as damaging by SIFT, probably damaging by PolyPhen2, neutral by PROVEAN and low by MutationAssessor (Table 4.24).

ii. A140C (g.67723825) Y916S

This transversion of A to C at position 140 in exon-8 (TAT to TCT) results in substitution of tyrosine to serine at codon 916. The mutation is novel as confirmed through HGMD and Ensemble. The mutation introduces a comparatively smaller residue than the wild and is positioned within a region, annotated in UniProt as a special region responsible for interaction with LPXN, KAT7, CCRA1 and Ligand-binding.

The differences in amino acid properties can disturb this region. The original wild type residue is annotated in UniProt as being modified into a "Phosphotyrosine by CSK". This modification can be important for protein function. Mutation into another residue type might disturb this modification and consequently the function of the residue. The wild residue is highly conserved at this position in the sequence.

This residue is part of an interpro domain named Nuclear hormone receptor, ligand-binding" (IPR008946). This domain is annotated with the GO terms sequence-specific DNA binding transcription factor activity (GO: 0003700) and Steroid hormone receptor activity (GO:0003707). These GO annotations indicates that the domain has a function in receptor activity (GO:0004872), nucleic acid binding (GO:0003676), nucleic acid binding transcription factor activity (GO:0001071) and molecular transducer activity (GO:0060089).

The mutated residue is located in a domain that is important for the main activity of the protein. The smaller size of mutant residue may result in loss of external interactions (Figure 4.42). Functionally the mutation was predicted as damaging by SIFT, probably damaging by PolyPhen2, deleterious by PROVEAN and low by MutationAssessor (Table 4.24).

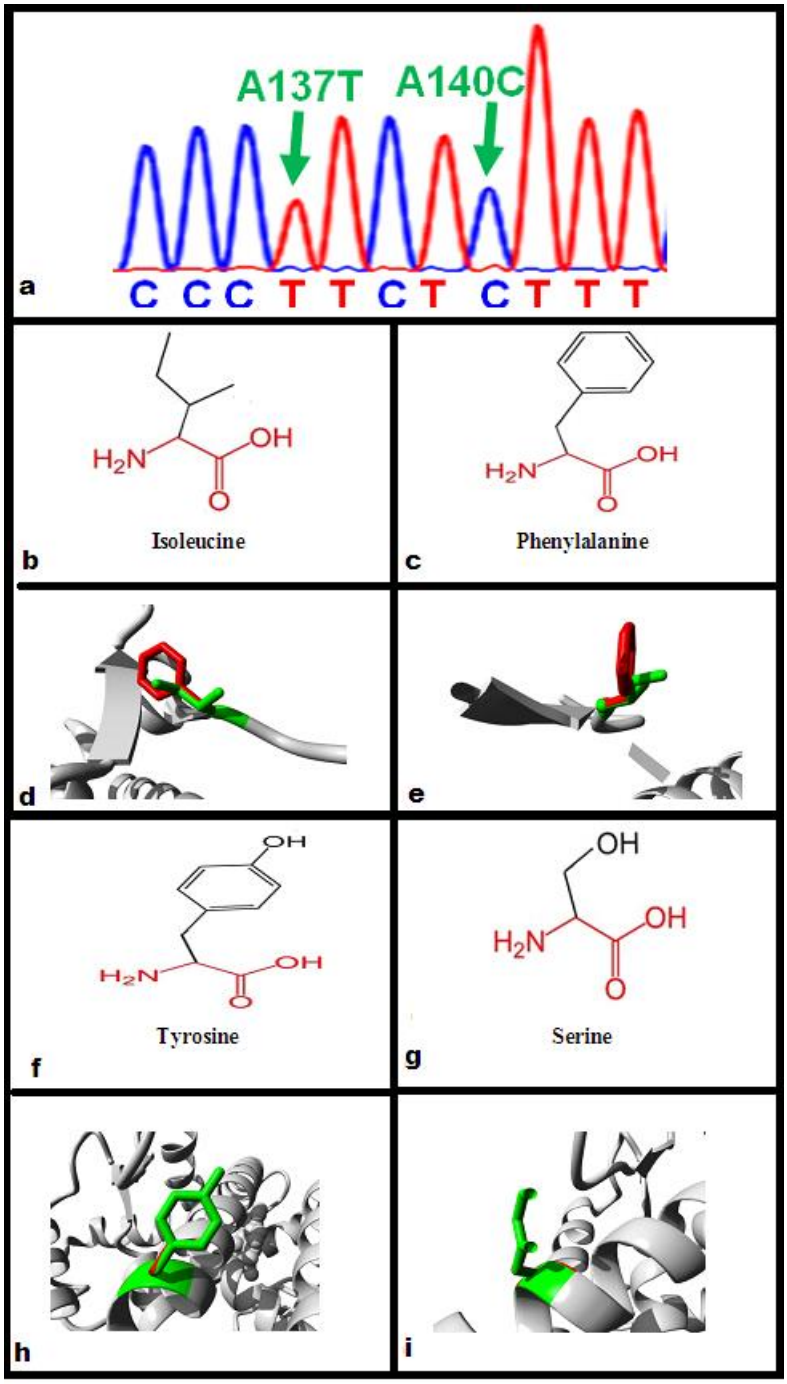

Figure 4.42: Sequence and structure analysis of ARI915F & Y916S mutations.

a. Sequence chromatogram showing A137T + A140C mutations in the AR gene sequence.
b. Wild residue Isoleucine
c. c. Mutant residue Phenylalanine.
d. & e. Comparison of mutant Phenylalanine (red) and wild type Isoleucine (green) residues in the 3D structure of Androgen Receptor.
f. Wild residue Tyrosine g. Mutant residue Serine
h. & i. Comparison of mutant serine (red) and wild type Tyrosine (green) residues in the 3D structure of Androgen Receptor.

4.11 AR Intronic variants

In present study, we observed two novel variants in the non-coding regions of AR gene. One of the two variant (T>A) is in intron-1 at g.67637091 while the other one T>C was in the 3' UTR region at g. 67724021 (Table 4.25). Both the variants were recorded both in cases and control and thus were analyzed in detail to evaluate their possible association with PCa risk. Genotypes distribution and allelic frequency distributions are listed in Table 4.26 while genotypes associations are in Table 4.27. P-values of less than 0.05 were considered statistically significant.

Analysis of the T>A variant at g.67637091 showed a significant difference between PCa patients and control individuals (Table 4.26). The AA genotype was significantly more common in the cases (11.75%) than in the control individuals (5.25%), which further implied that A allele contributes positively towards increased PCA risk. It was also found that, the homozygous AA participants had 3.68-fold higher risk of prostate cancer. The carriers of A allele were found significantly associated with increased risk of PCa in all the models with highest OR of 3.68 (95%CI 2.02-6.71; P = 0.00) in the AA vs TT model except AA vs TA model (OR = 1.83; CI=1.0-3.34; P = 0.05) (Table 4.27 and Figure 4.43 a).

T>C variant at g.67724021 also showed a significant difference between cases and controls (Table 4.26). The C allele was significantly more common in the cases (47%) than in the controls (32%). T>C variant was also found significantly associated with increased risk of PCa in all the models with highest OR of 3.74 (95%CI =1.98-7.07; P = 0.00) in the CC vs TT model (Table 4.28 and Figure 4.43 b).

Table 4.25: AR intronic variants (Reference sequence; NC_000023.11)

S. No	Intron	ChrStartPosition	ChrEnd Position	variant	Status
1.	1	67637091	67637091	g.67637091 T/A	Novel
2.	3' UTR	67724021	67724021	g. 67724021 T/C	Novel

Table 4.26: Genotypes and their allelic frequencies of AR T/A transversion at chromosomal position g.67637091 and T/C transition at 3' UTR g.67724021in Prostate adenocarcinoma

Alleles	Genotype	Cases (n=200)			Controls n=200		
		No.	%age	Frequency	No.	%age	Frequency
g.67637091	T-allele	218	54.50	0.55	286.00	71.50	0.72
	A-allele	182	45.50	0.46	114.00	28.50	0.29
	TT	65	16.25	0.16	107.00	26.75	0.27
	TA	88	22.00	0.22	72.00	18.00	0.18
	AA	47	11.75	0.12	21.00	5.25	0.05
g.67724021	T-allele	212	53.00	0.53	272.00	68.00	0.68
	C-allele	188	47.00	0.47	128.00	32.00	0.32
	TT	55	13.75	0.14	91.00	22.75	0.23
	TC	102	25.50	0.26	90.00	22.50	0.23
	CC	43	10.75	0.11	19.00	4.75	0.05

Table 4.27: Association of AR T/A transversion at chromosomal position g.67637091 and T/C transition at 3' UTR g.67724021in Prostate adenocarcinoma

Alleles	Genetic Model	OR	LCI	UCI	Z-stat	P-value
g.67637091	Allelic contrast (A vs T)	2.09	1.56	2.81	4.95	0.00
	Rare vs Common (AA vs TT)	3.68	2.02	6.71	4.26	0.00
	Rare vs Heterozygotes (AA vs TA)	1.83	1.00	3.34	1.97	0.05
	Heterozygotes vs Common (TA vs TT)	2.01	1.30	3.12	3.13	0.00
	Dominant (AA+TA vs TT)	2.39	1.59	3.59	4.21	0.00
	Recessive (AA vs TT+TA)	2.62	1.50	4.57	3.38	0.00
g.67724021	Allelic contrast (C vs T)	1.88	1.41	2.51	4.32	0.00
	Rare vs Common (CC vs TT)	3.74	1.98	7.07	4.07	0.00
	Rare vs Heterozygotes (CC vs TC)	2.00	1.09	3.67	2.22	0.03
	Heterozygotes vs Common (TC vs TT)	1.88	1.21	2.91	2.81	0.00
	Dominant (CC+TC vs TT)	2.20	1.45	3.34	3.71	0.00
	Recessive (CC vs TT+TC)	2.61	1.46	4.66	3.24	0.00

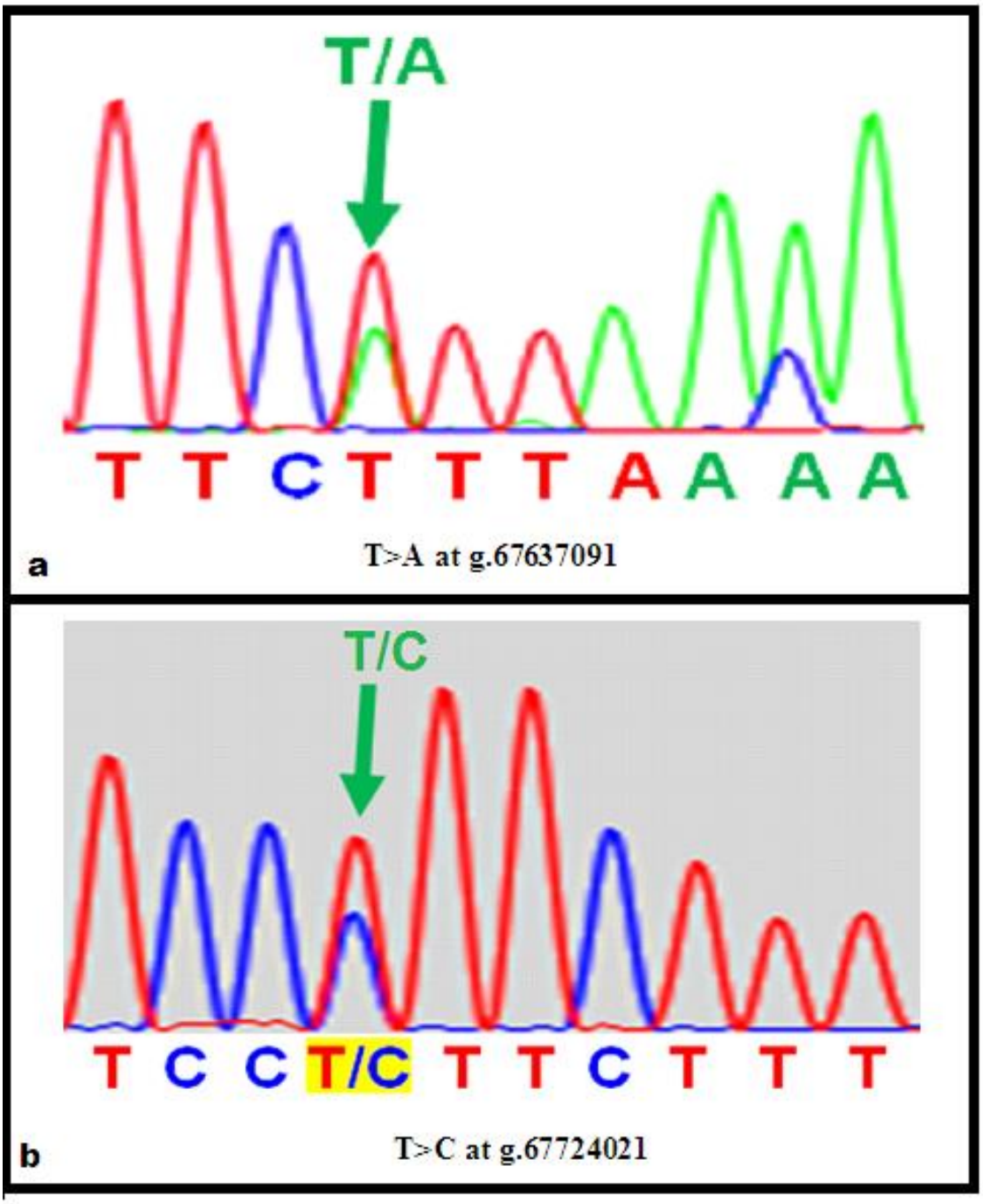

Figure 4.43: Sequence chromatograms of AR gene variants in the non-coding region.

a. Sequence chromatograms showing T>A transversion in intron-1 at g.67637091.
b. Sequence chromatograms showing T>C transition in 3'UTR at g.67724021.

In metastatic androgen-independent prostate cancer, a group of researchers reported that, 100% of metastatic cells carried an AR gene with 4 mutations resulting in the following amino acid substitutions: ser647-to-asn, gly724-to-asp, leu880-to-gln, and ala896-to-thr. They also found a GCC-to-ACC and CAA-to-CGA transition in the AR gene in metastatic cells of prostate cancer in 1 out of 10 patients studied. The nucleotide substitution resulted in a ala721-to-thr and gln902-to-arg amino acid change respectively (Taplin et al., 1995).

In 6 of 24 specimens of prostatic tissue derived from transurethral resections in patients with metastatic prostate cancer, thr877-to-ala mutation in the hormone-binding domain of the AR gene has also been reported. The same mutation had been reported previously in a metastatic prostatic cancer cell line where it conferred on the androgen receptor an altered ligand-binding specificity that was stimulated by estrogens, progestagens, and anti-androgens. It was suggested that the codon 877 mutant AR with altered ligand binding may provide selective growth advantage in the genesis of a subset of advanced prostate cancer. The stimulatory effect of the usual therapeutic agents on the codon 877 mutant AR may contribute to treatment-refractory disease (Gaddipati et al., 1994).

Another goup of researchers reported a somatic AR mutation by study of genomic DNA by PCR followed by denaturing gradient gel electrophoresis (DGGE) in 1 of 26 specimens of untreated organ-confined stage B prostate cancer. Sequencing revealed a G-to-A substitution in exon E, changing valine to methionine at codon 730. An abundance of the mutated fragment indicated its presence in cells with a growth advantage (Newmark et al., 1992).

The identification of key molecular alterations in prostate-cancer cells implicates carcinogen defenses (GSTP1), growth-factor-signaling pathways (NKX3.1, PTEN, and p27), and androgens (AR) as critical determinants of the phenotype of prostate-cancer cells. Glutathione S-transferases (GSTP1) are detoxifying enzymes. Inadequate levels of PTEN and NKX3.1 lead to a reduction in p27 levels and to increased proliferation and decreased apoptosis. Androgen receptor (AR) is a transcription factor that is normally activated by its androgen ligand. During androgen withdrawal therapy, the AR signal transduction pathway also could be activated by amplification

of the AR gene, by AR gene mutations, or by altered activity of AR co-activators. Through these mechanisms, tumor cells lead to the emergence of androgen-independent prostate cancer.

4.12 Analysis of Socio-demographic and clinical factors

Clinical data collected along with the blood samples were analyzed through different statistical techniques to get valuable information about the disease. The clinical and biochemical data including Socio-demographic characteristics and clinical features was analyzed from different angles in detail to evaluate the possible association of risk factors with prostate adenocarcinoma through R version 2.1.1. Descriptive characteristics of the 680 prostate cancer cases and the 500 controls are summarized in Table 4.28.

In prostate patients, mean age was 68.9± 9.0 ranging from 40-90 years while in controls the mean age was 60.3 ± 13.2 ranging from 39-100 years. Mean for BMI, PSA, Systolic and Diastolic BP were 25.9±4.0, 11.3±17.5, 128.5±20.6 and 80.1±10.7 respectively while in controls it were 24.7±4.9, 1.7±1.2, 121.8±19.6 and 80.1±11.5 respectively (Table 4.28).

4.12.1 Age

Age is one among the three established risk factors for prostate cancer (Gann, 2002). Increasing age further enhance the disease risk. In prostate patients, mean age was 68.9± 9.0 ranging from 40-90years while in controls the mean age was 60.3 ± 13.2 ranging from 39-100 years. Majority of the cases are distributed in ages ranges 60-90 years of ages (Table 4.28). It was further observed that none of the cases has age below 40 years in the sampled population, and only 2.2 percent of the cases are in the range of 40-50 years. Variation distribution of cases in different age groups was clearly observed and thus to systematically observe the pattern of distribution of cases in age groups and also to highlight the impact of age on prostate caser risk, we categorized the age of variable into six categories i.e. upto 50 years, 51-59.9, 60-69.9, 70-79.9, 80-89.9 and 90 or above. In our study, we further observed that most of the cases lie in the category of 60-69.9 (39.6%) and 70-79.9 (34.4%). It was further observed that only 0.4% of the cases are in the category of age s above 90 years while

controls are 1.6% in this category. It is hypothesized that due to the disease and other allied complication the prostate cancer patients usually do not survive upto the age of above 90 years.

In Uni-factorial analysis, age groups of 60s (OR=3.62, 95% CI=2.69-4.79), 70s (OR=6.76, 95% CI=4.65-9/83) and 80s (OR=1.75, 95% CI=1.19-2.60) showed positive association with prostate cancer (Table 4.29). Similar results were observed for the same age groups in multi-factorial analysis (Table 4.30). Prostate cancer is one of the most common types of malignancy and its frequency is correlated with advancing age (Jemal et al., 2010). Our results were also supported by other studies who reported that, 64% of the new cases in United States were diagnosed in men older than 65 years of age and 23% in men older than 75 years (Heinzer and Steuber, 2009). Our results were also confirmed by (Gann, 2002) who reviewed that, age-specific risk of prostate cancer incidence sharply rise after the age of 55 and peaks at 70–74 years of age. Same results were reported by the American Cancer Society (2015). According to their report, age significantly increases the risk of prostate cancer and is infrequently been diagnosed in younger men having age less than 40 years. Its incidence rises sharply with each decade thereafter. They further showed that, the probability disease diagnosis is 1/304 in younger having age less than 49 years, 1/ 44 for men aged 50-59 years, 1/16 for men aged 60-69 years and 1/9 for men aged equal to or greater than 70 years, with an overall lifetime risk of 1/7.

Table 4.28: Socio-demographic characteristics and clinical features of the sampled population.

	cases					Controls				
Factors	**No**	**%age**	**Mean**	**SD**	**Range**	**No**	**%age**	**Mean**	**SD**	**Range**
Age	680		68.9	9.0	40-90	500		60.3	13.2	39-100
BMI			25.9	4.0	16.50-45.9			24.7	4.9	15.2-55.8
PSA			11.3	17.5	0.25-113			1.7	1.2	0.01-8.78
Cystolic BP			128.5	20.6	70-190			121.8	19.6	80-210
Diastolic BP			80.1	10.7	40-130			83.2	11.5	40-120
Marital Status										
Married	644	94.7				484	96.8			
Un-married	36	5.3				16	3.2			
BMI Status										
Underweight	11	1.6				10	2			
Normal	239	35.15				354	70.8			
Overweight	348	51.18				84	16.8			
Obese	82	12.1				52	10.4			
BP status										
Hypotension	8	1.2				10	1.8			
Normal	421	61.9				332	66.4			
Pre-Hypertension	95	14.0				77	15.4			
Hypertension	156	22.9				82	16.4			
No. of kids										
0	43	6.3				33	6.6			
1	35	5.1				28	5.6			
2	122	17.9				87	17.4			
3	118	17.4				93	18.6			
4 or 4+	362	53.2				259	51.8			
Smoking										
Yes	457	67.21				180	36			
No	223	32.79				320	64			
Physical Activity										
Active	258	37.9				389	77.8			
Disturb	303	44.6				94	18.8			
Usually at bed	73	10.7				10	2.0			
At bed	46	6.8				7	1.4			

Surgery Status										
Prostectomy	152	22.4				NA	NA			
No	528	77.7				NA	NA			
Disease Stage										
Local	612	90				NA	NA			
Locally Advanced	10	1.5				NA	NA			
Metastatic	58	8.5				NA	NA			
Other Complications										
Diabetes	130	19.1				72	14.4			
Hypertensive / pre-hypertensive	251	36.9				159	31.8			
Diabetic + Hypertensive	52	7.7				33	6.6			
Cardiovascular	41	6				14	2.8			
Others	126	18.5				99	19.8			
No	425	62.5				337	67.4			
Family History										
Yes	276	40.6				16	3.2			
No	308	45.3				437	87.4			
Don’t Know	96	14.1				47	9.4			
Dairy Products Consumption										
Low	51	7.5				309	61.8			
Medium	181	26.6				55	31			
High	209	30.7				26	5			
Very High	239	35.2				11	2.2			

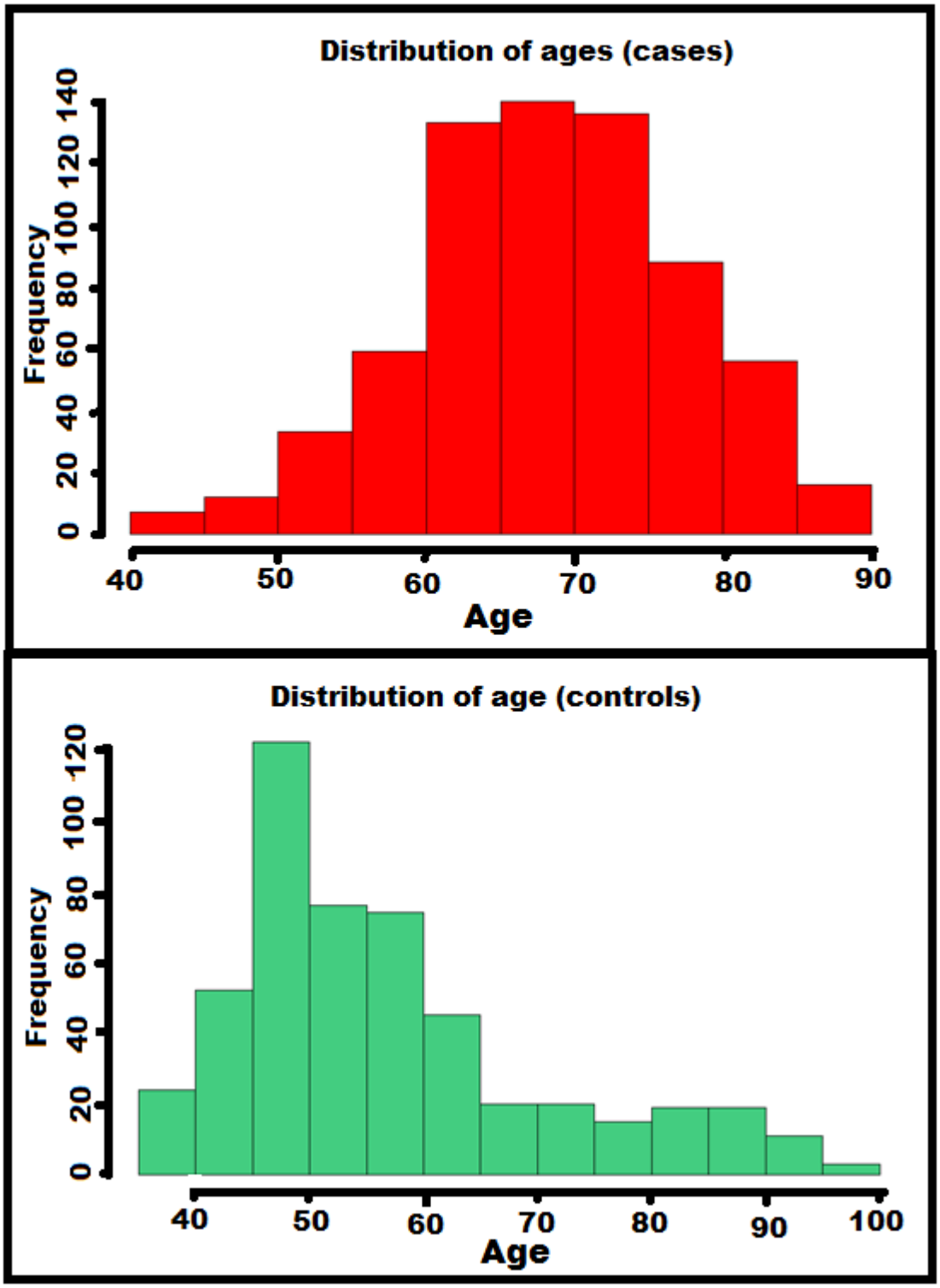

Figure 4.44: Distribution of ages in cases and controls

4.12.2 BMI

Excess bodyweight (BMI) of 25 to 29·9 kg/m²) or obese (BMI of 30 kg/m² or greater), is increasingly recognized as an important risk factor for some common cancers (Renehan et al., 2008), including prostate cancer. Though obesity is an established risk factor for many cancers but the association is not clear with total prostate cancer risk (Allott et al., 2013).

We observed in our study an increase in prostate cancer risk with the increasing BMI. In our studied population, 51.2% of the cases are in the overweight and 12.1% in the obese category (Table 4.28). In present study, it was observed that individuals BMI deviating from the normal range increases the risk of prostate cancer. A positive correlation was observed in underweight, obese and overweight individuals both in uni-variate (Table 4.29) and multi-variate (Table 4.30) analysis. Our results have been supported by many other studies in different ethnic groups (Rodriguez et al., 2001, MacInnis et al., 2003, Gong et al., 2006). Some researchers reported positive association of prostate cancer risk and BMI, specifically for those having BMI ≥30 versus 18.5–24.99 kg/m2 (Engeland et al., 2003). A meta-analysis of prospective cohort studies, including more than 2,000,000 men worldwide, found that obese men are at increased risk of advanced prostate cancer, but lower risk of localized prostate cancer (Discacciati et al., 2012). Multiple reasons for this association have been postulated. Another study reported that obesity is associated with decreased risk of low-grade and increased risk of high-grade prostate cancer. These data provide further support to the hypothesis that obesity is associated with aggressive prostate cancer (Allott et al., 2013) (Figure 4.45).

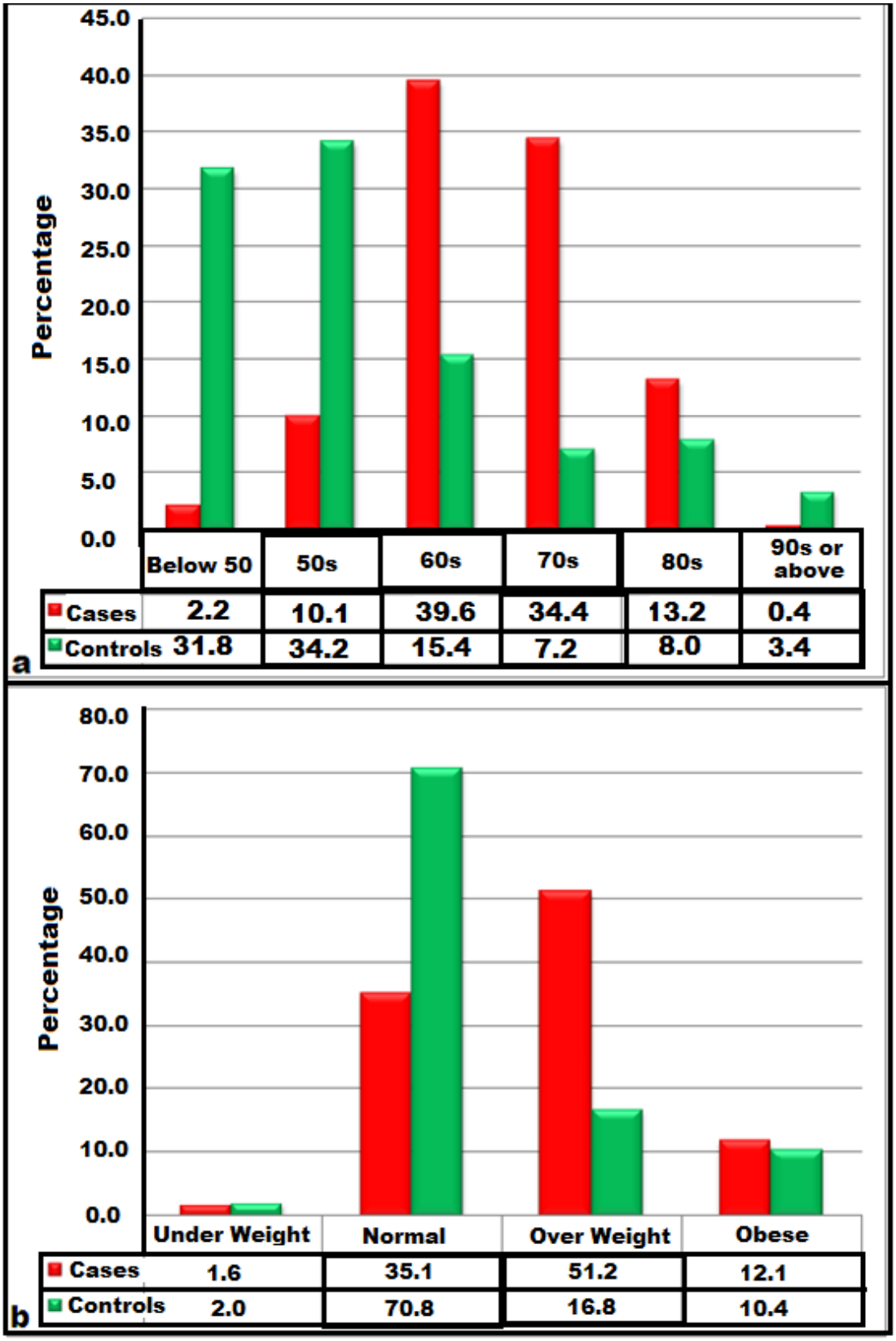

Figure 4.45: Comparison of a. age and b. BMI between cases and controls

4.12.3 PSA

Prostate specific antigen (PSA) is an androgen regulated serine protease of the kallikrein family produced by secretory epithelial cells of the prostate (Stamey et al., 1987) and is the most commonly used biomarker for prostate cancer screening.

On the basis of published facts, we designed our study to focus on the impact and association of PSA and prostate cancer risk. In our studied population, 73.24% of the cases has PSA level >4.0 ng/mL while 5% of the controls also has PSA level >4.0 ng/mL. It was also observed that 26.76% of the cases has PSA level <4.0 ng/mL while in controls, 95% of sampled individuals has PSA level <4.0 ng/mL (Table 4.28, Table 4.29, Table 4.30 and Figure 4.46).

Results from the large Prostate Cancer Prevention Trial (PCPT) showed that only 25–30% of patients with elevated PSA levels (>4.0 ng/ml) were confirmed with PCA on biopsy while approximately 15% patients having PSA levels <4.0 ng/ml were also having PCa on biopsy (Thompson et al., 2005). It suggests that, the predictive power of PSA as a prostate cancer marker is not optimal, which is mainly due to its PCa non-specific nature. PSA levels are also subjective to age, ethnicity, prostate size inflammation or prostatitis (Gsur et al., 2002, Schatzl et al., 2005). It was also reported that men with age ≥60 years and having a PSA level below 1ng/mL, unlikely develop clinically significant prostate cancer and having only 0.2% risk of prostate cancer related deaths (Vickers et al., 2010). Similarly, Baltimore Longitudinal Study of Aging reported that, no men with age ranges from75-80 years and having PSA level <3.0 ng/mL died of prostate cancer (Schaeffer et al., 2009).

4.12.4 Family History

Family history is one of established risk factors for prostate cancer (Gann, 2002) and familial clustering has been frequently reported (Stanford and Ostrander, 2001; Matikaine et al., 2001). In the present study we observed that family history is directly correlated with increased risk of prostate cancer (OR=20.67, 95% CI=1.74-34.80) (Table 4.29 and Table 4.30).

Several case-control and cohort studies from different populations have shown that family history is a major risk factor in PCa. Moreover, 5-10% of the cases are reported primarily due to high-risk inherited genetic factors or PCa susceptibility. (Gronberg, Damber and Damber, 1996) (Figure 4.46).

4.12.5 Marital Status

Social support benefits positively extend to cancer outcomes, which has already been demonstrated in different studies (Osborne et al., 2005, Baine et al., 2011), but still the influence of the marital status and its potential link with prostate cancer is a matter of discussion (Berkson, 1962, Rendall et al., 2011). We examined the impact of marital status if any on prostate cancer risk in the present study. We classified this variable as either married or unmarried. We observed non-significant difference in between married and un-married individuals for their cancer risk in our sampled population in uni-variate analysis (OR=1.69, 95% CI=0.94-3.16), but positive association was observed (OR=2.34, 95% CI=1.07-5.30) for un-married individuals in multi-variate analysis (Table 4.29 and Table 4.30).

Unmarried patients including widowed are at comparatively higher risk of metastatic cancers and have less survival rates than married patients (Aizer et al., 2013). Marriage also has additional potential etiologies. Psychologically cancer diagnosis may result in more distress (Kaiser et al., 2010). Married patients display less anxiety, depression and distress than unmarried cases (Goldzweig et al., 2009). It can thus be inferred that, depression may be a mediator between the marital status and disease susceptibility. Marital status also has a significant protective effect against other metastatic cancers (Inverso et al., 2014).

Some studies also had reported non-significant association in between marital status and prostate cancer risk (Brusselaers et al., 2014). We further investigated the married individuals for the number of kids and its possible association with enhanced PCa risk and the results are illustrated in Figure 4.47, Table 4.29 and Table 4.30.

4.12.6 Smoking

According to the International Agency for Research on Cancer (IARC), smoking is positively associated with multiple cancers (Secretan et al., 2009). Although smoking is correlated with a number of solid tumors but its association with prostate cancer is still debatable. Some authors have published that prostate cancer is not tobacco related (Colditz, 1996). Cigarette smoke contains a number of carcinogens including *N*-nitroso compounds and it may also have a hormonal basis. Elevated level of androsterone and testosterone have also been reported in smoker patients, which can potentially increase the disease risk and also contribute to cancer progression (Huncharek et al., 2010).

We therefore considered this variable in our case–control study to evaluate the role of smoking and prostate cancer risk. In our study, we observed that 67.21% of the cases are smokers while in the control population only 36% are smokers. We found a positive association in between smoking and prostate cancer (OR=3.64, 95% CI=2.86-4.65) in our study population (Figure 4.48, Table 4.29 and Table 4.30).

The possible association between smoking and PCa was studied in a meta-analysis, which showed a positive association in the pooled data (Huncharek et al., 2010). In another large cohort prospective study it was reported that current and former smokers had decreased prostate cancer overall risk (Watters et al., 2009). A systematic review based on 36 prospective cohort studies showed no association between prostate cancer incidence and smoking (Hickey et al., 2001). These studies provide evidence that smoking is not an associated risk factor for prostate cancer. In addition, IACR also does not consider prostate cancer to be tobacco related.

4.12.7 Physical Activity

Researchers are working to gather evidence on the potential benefits of physical activity as a primary means for cancer prevention. A good number of epidemiological studies have been conducted on physical activity and its association with prostate cancer risk (Friedenreich and Orenstein, 2002) but this association is less consistent as compared to its association with either breast or colon cancer. We therefore examined the influence of physical activity on prostate cancer risk in Pakistani population

(Figure 4.48) and observed that lack of healthy physical activities result in an increase of PCa risk (OR=3.47, 95% CI=2.65-4.55 for disturbed cases, OR=5.93, 95% CI=3.01-11.53 for cases usually at bed and OR=5.11, 95% CI=2.29-11.42 for at bed cases) (Figure 4.48; Table 4.29 and Table 4.30).

Studies like (Vena et al., 1987, Paffenbarger et al., 1987, Yu et al., 1988, Brownson et al., 1991, Hsing et al., 1994, Thune and Lund, 1994, Andersson et al., 1995, Steenland et al., 1995), (Oliveria et al., 1996), (Hartman et al., 1998), (Clarke and Whittemore, 2000), (Bairati et al., 2000) reported that physically active individuals were at reduced risk of prostate cancer while other studies (SEVERSON et al., 1989), (Dosemeci et al., 1993), (Pukkala et al., 2000), (Liu et al., 2000), (Lee et al., 2001), (Whittemore et al., 1995) has reported no association in between physical activity and PCa risk.

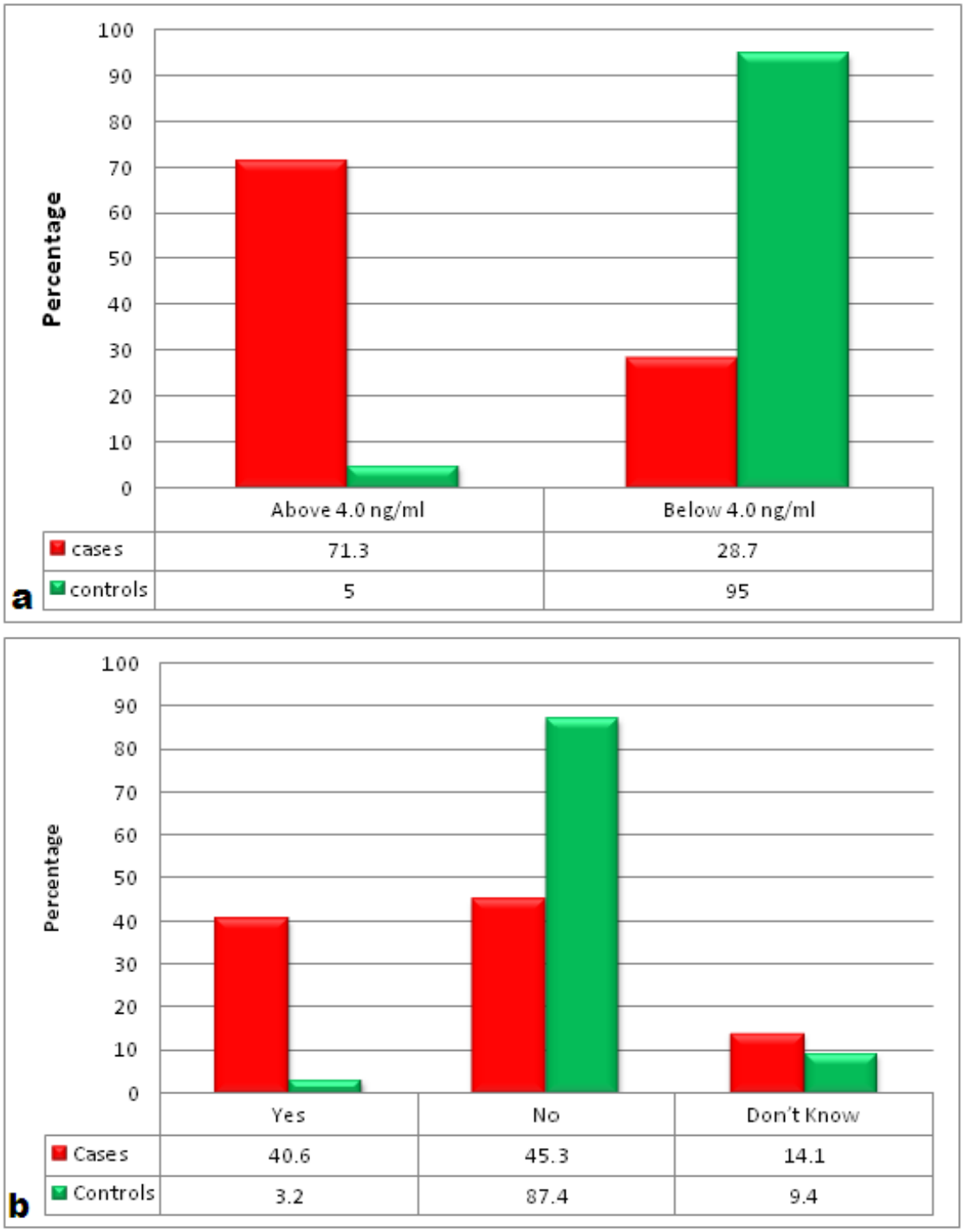

Figure 4.46: Comparison of a. PSA and b. Family history between cases and controls

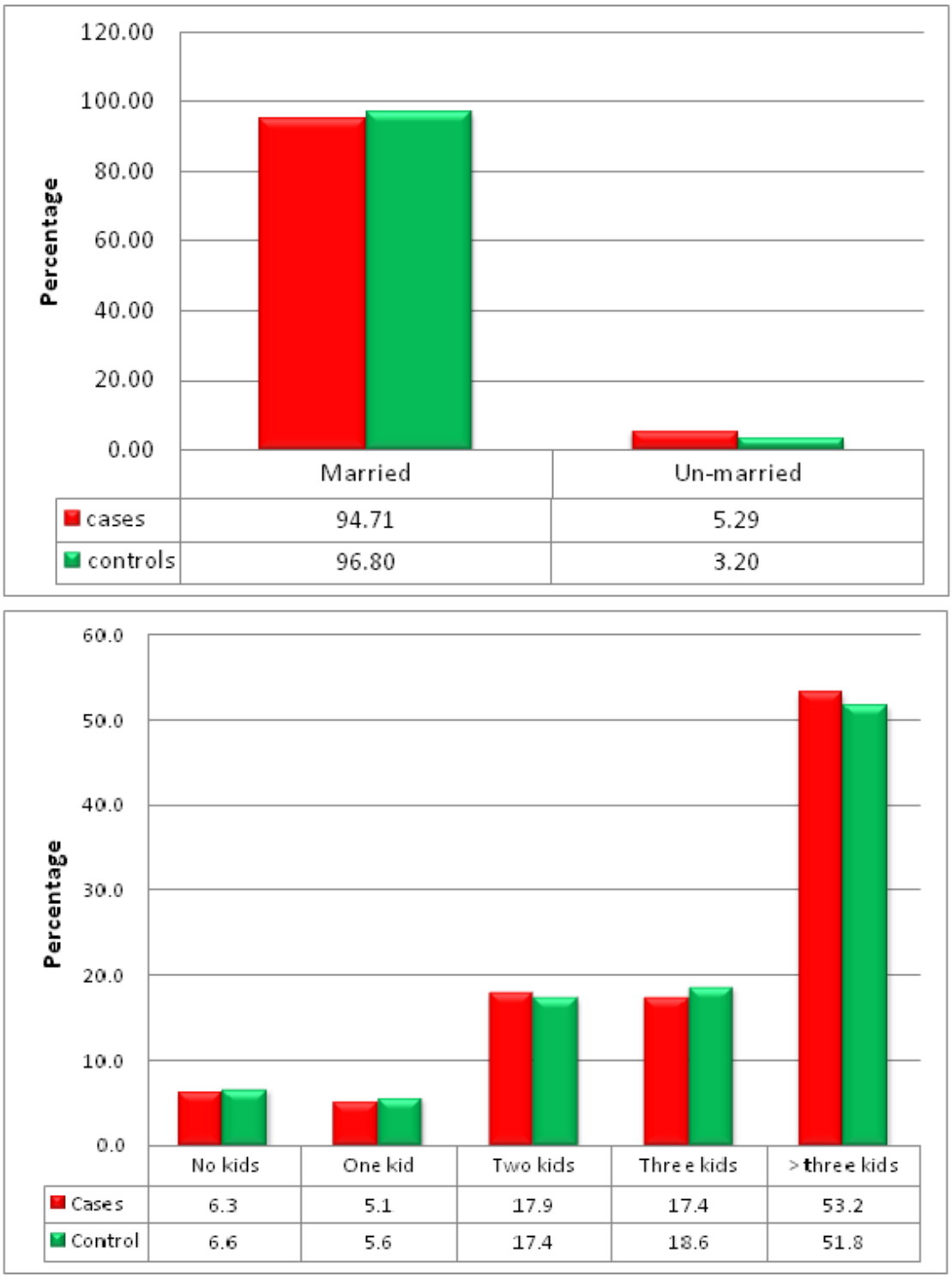

Figure 4.47: Comparisons of a. marital status and b. Number of kids between cases and controls

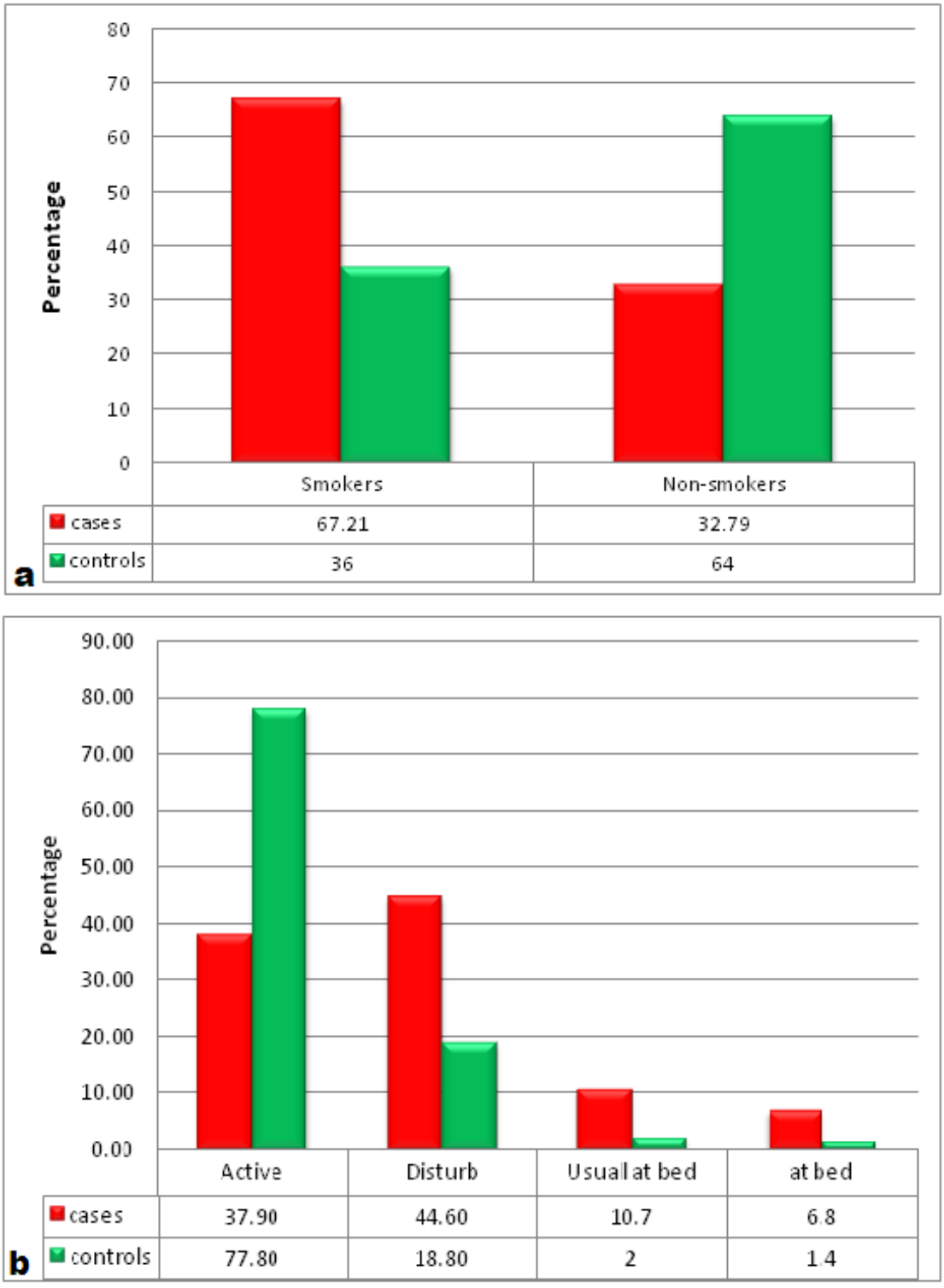

Figure 4.48: Comparison of a. Smoking or other tobacco usage and b. Physical activity in between cases and controls

4.12.8 Surgery Status

Radical prostatectomy is one of the available options to treat clinically localized prostate cancer. It may also be preferred by many patients on the basis of cancer risk, age, preferences and comorbidity (Cooperberg et al., 2012). In the initial period of PSA testing in patients with localized PCa, surgery did not significantly altered the overall prostate cancer mortality (Wilt et al., 2012) but nowadays its reported associated with enhanced survival. Patients with extra-capsular tumor growth may benefit from adjuvant local or systemic treatment (Bill-Axelson et al., 2008). In our study population, only 22.7% of the cases carried out surgery (Table 4.28).

4.12.9 Other Complications

We also investigated in the study, the impact of other allied complications on prostate cancer risk, which includes the following (Figure 4.49).

Diabetes: Both diabetes and cancer are globally widespread diseases (Giovannucci et al., 2010). Overall, cancer patients may also have diabetes (8–18%) and the incidence vary with tumor site (Ko and Chaudhry, 2002). Diabetes mellitus Type-2 is basically a metabolic disorder, positively correlated with elevated cancer risk (Everhart and Wright, 1995). We recorded a positive association for increased risk of PCa in non-diabetic patients in our study (OR=1.41, 95% CI=1.03-1.93) (Table 4.29 and Table 4.30).

It has also been reported that, PCa risk decreases with increasing time since the diabetes mellitus diagnosis. It suggests that hormonal changes due to diabetes mellitus creates a less oncogenic/carcinogenic environment which becomes more prominent with the passage of time (Calton et al., 2007), (Kasper et al., 2009), Gong (Gong et al., 2006), (Rodriguez et al., 2005), (Bonovas et al., 2004), (Giovannucci et al., 1998).

Blood Pressure: High blood pressure level has also been reported associated with an increased cancer risk (Grossman et al., 2002). As compared to other cancers, only limited data is available on the association between BP and PCa risk (Davis et al., 2008, Martin et al., 2011).

In our study, we observed that there was no clear cut difference in the overall systolic and diastolic BP in between cases and controls, but some levels of differences were observed when blood pressure level was categorized into hypotension, normal, pre-hypertension and hypertension (Figure 4.49 b). Hypertension was found associated to the increased PCa risk only in uni-factorial study (OR=1.52, 95% CI=1.13-2.04) (Table 4.29 and Table 4.30). Other studies reported no evidence of a positive association of BP with subsequent prostate cancer (Friedman, 1997, Fitzpatrick et al., 2001).

Cardiovascular diseases: Fatal and non-fatal relative and absolute risks of different types of CVDs, including heart failure, acute MI, ischemic heart disease, arrhythmia, and stroke were investigated in a population based cohort study. They detected an enhanced relative risks CVDs in PCa patients, especially those treated with endocrine treatment (Van Hemelrijck et al., 2010). Our results are in line with the previous results (Table 4.29 and Table 4.30).

4.12.1 Geographical and ethnical distribution of the samples

Samples were collected from different parts of the country and we tried our level best to collect samples from all possible areas across Pakistan and also from the different ethnic groups of those areas (Figure 4.50). We assume that due to the limitations of resources, the number of samples are not very large but are enough so for the systematic evaluation of PCa in different regions and different ethnic groups of Pakistan. We also studied the distribution of cases in the sampled areas (provinces) with respect to sub-ethnic groups across Pakistan (Figure 4.51 and Figure 4.52) and also the distribution of each sub-ethnic group across Pakistan (Figure 4.53 and Figure 4.54).

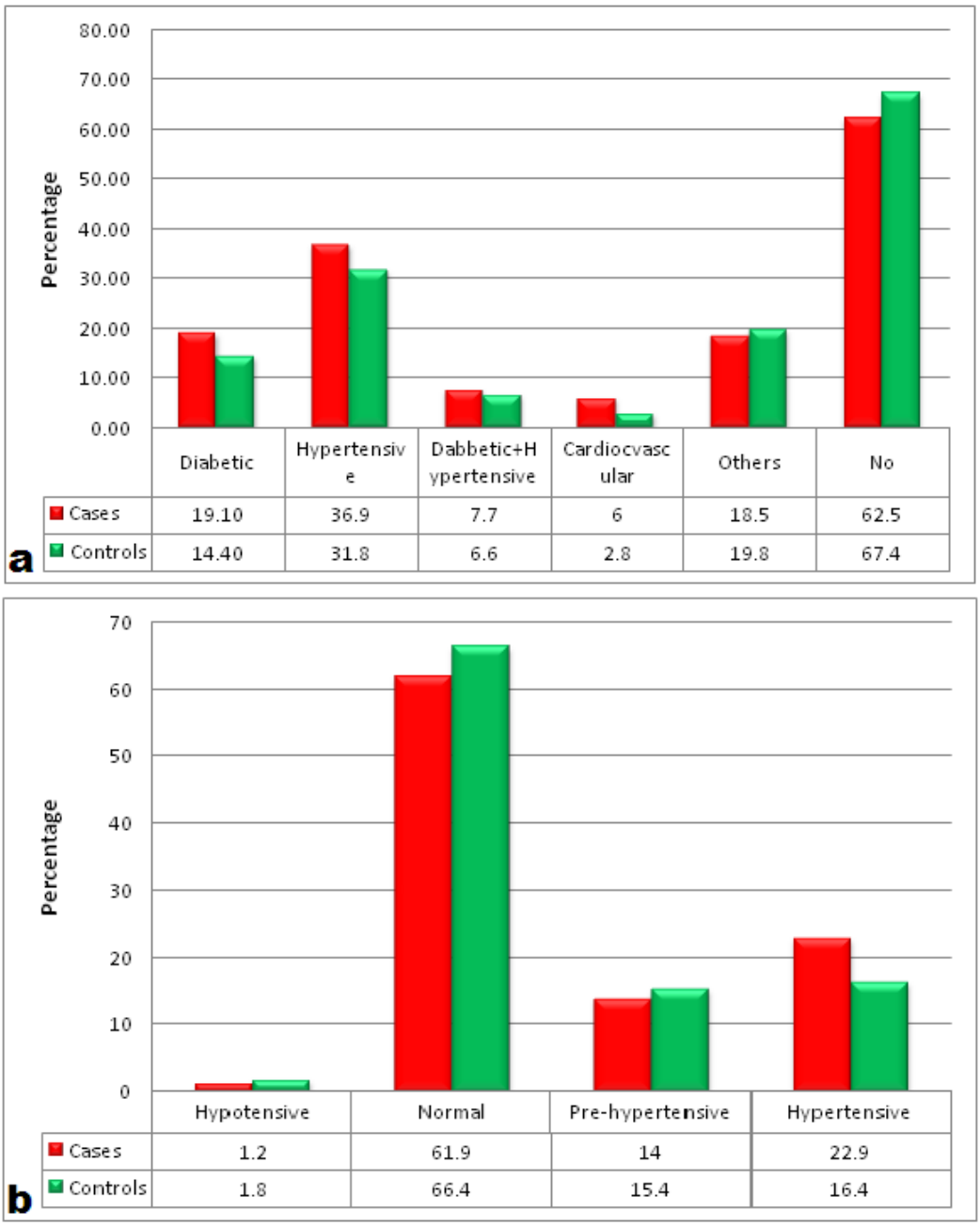

Figure 4.49: Comparison of complications other than PCa in between cases and controls

a. Comparisons of other complications in between cases and controls

b. Comparisons of BP status in between cases and controls

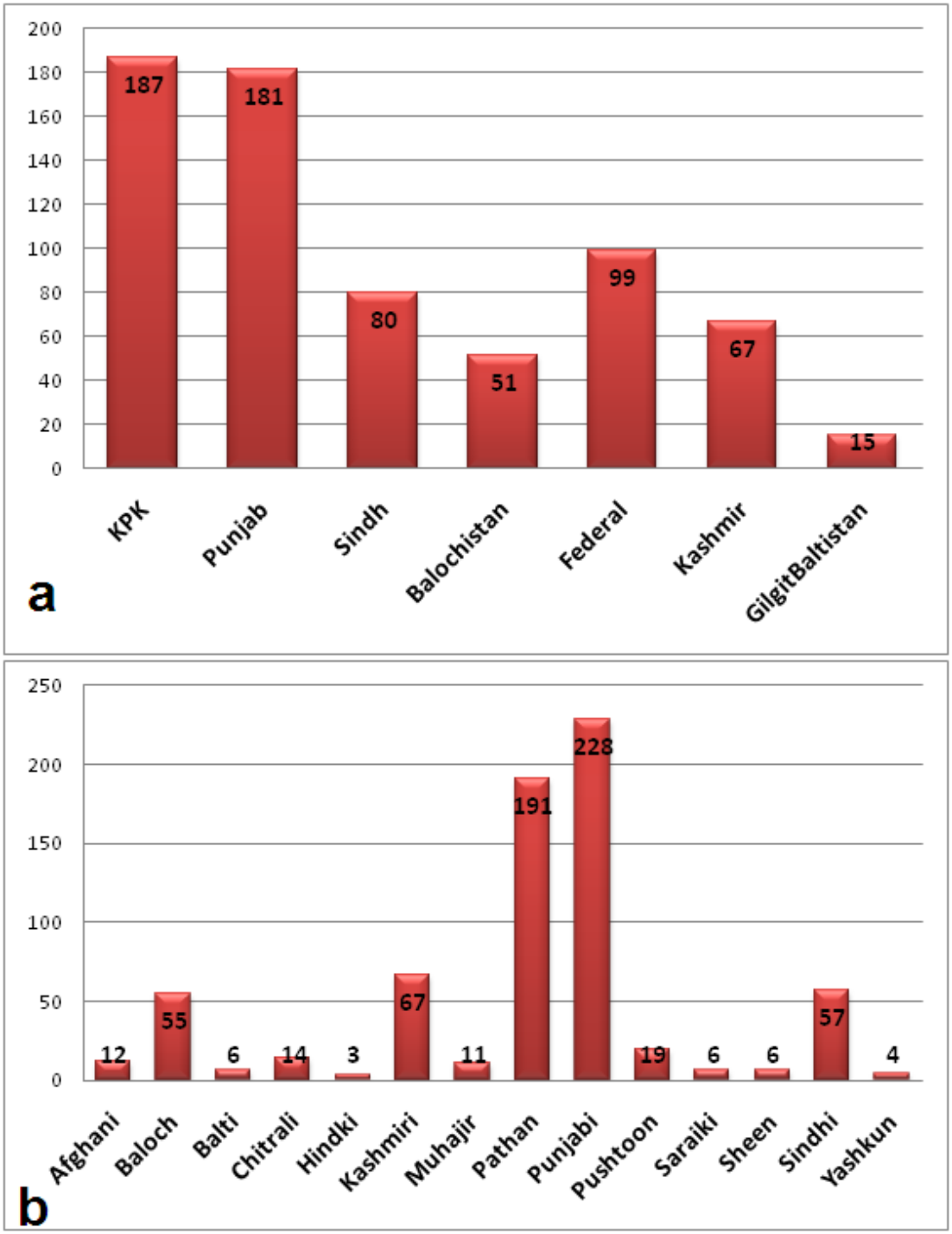

Figure 4.50: Distribution of cases on the basis of a. geographical distribution and b. ethnicity in the selected samples across Pakistan

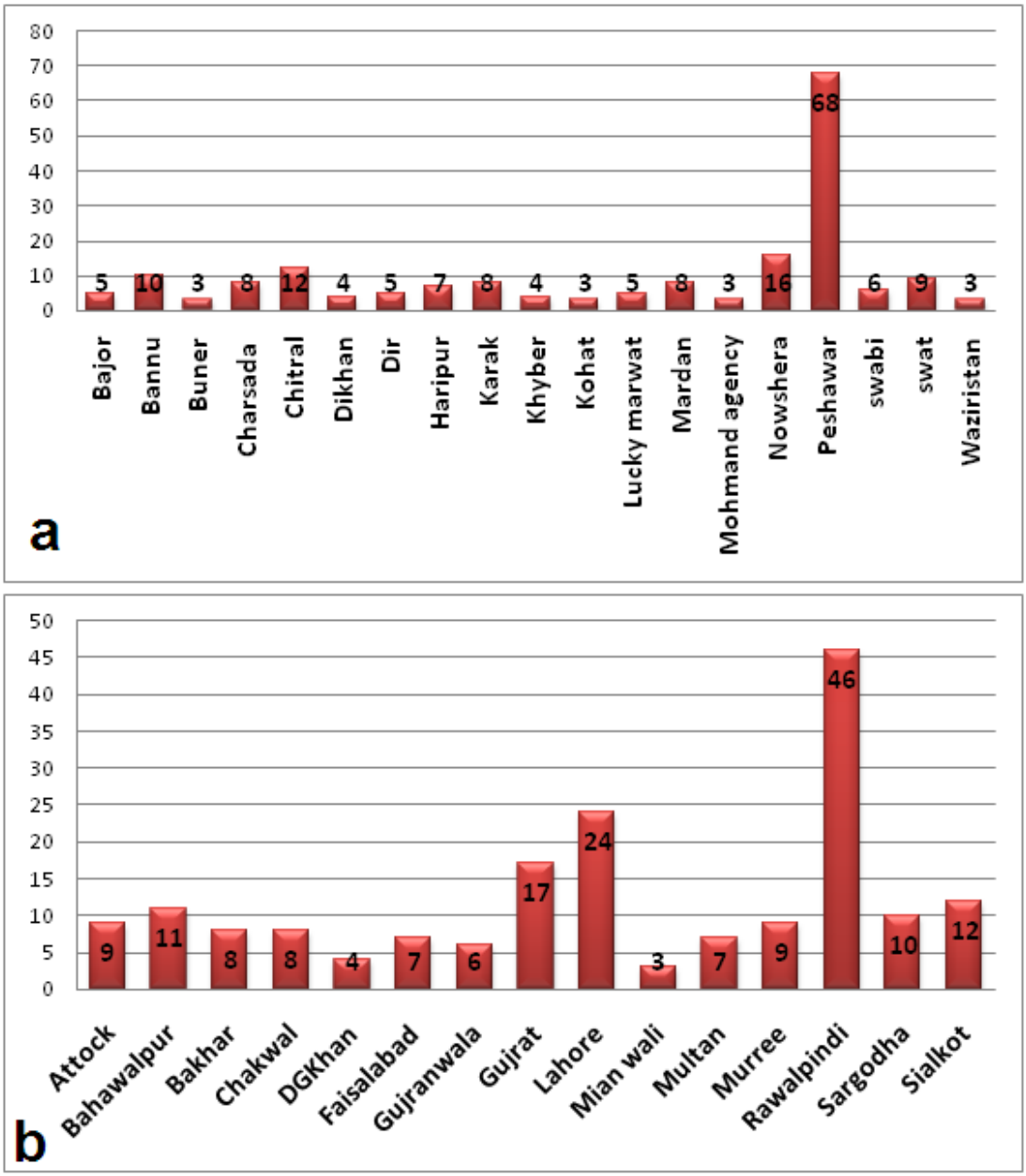

Figure 4.51: Distribution of cases in the selected samples across a. KPK and b. Punjab

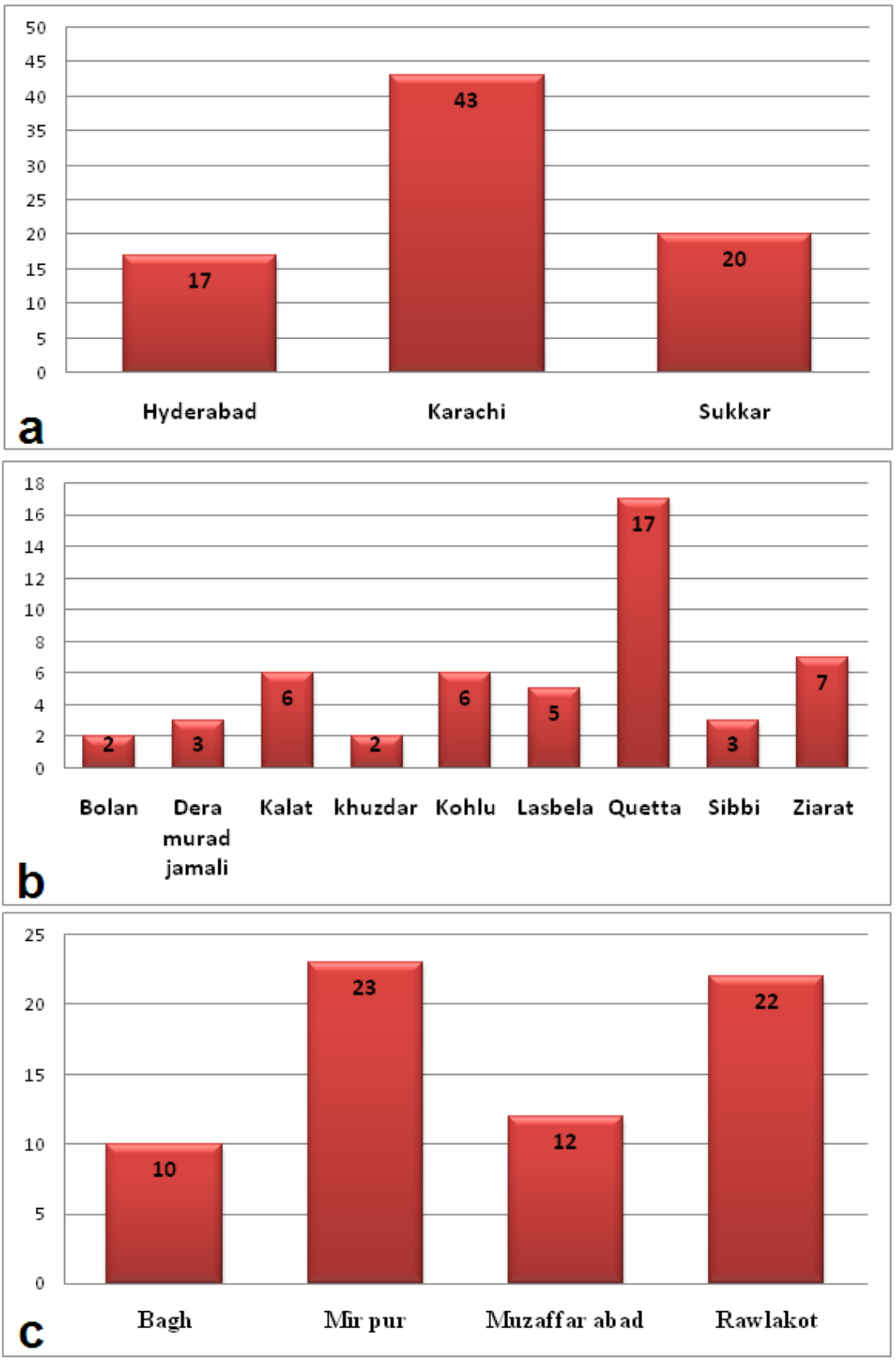

Figure 4.52: Distribution of cases in the selected samples across a. Sindh, b. Balochistan and c. Kashmir

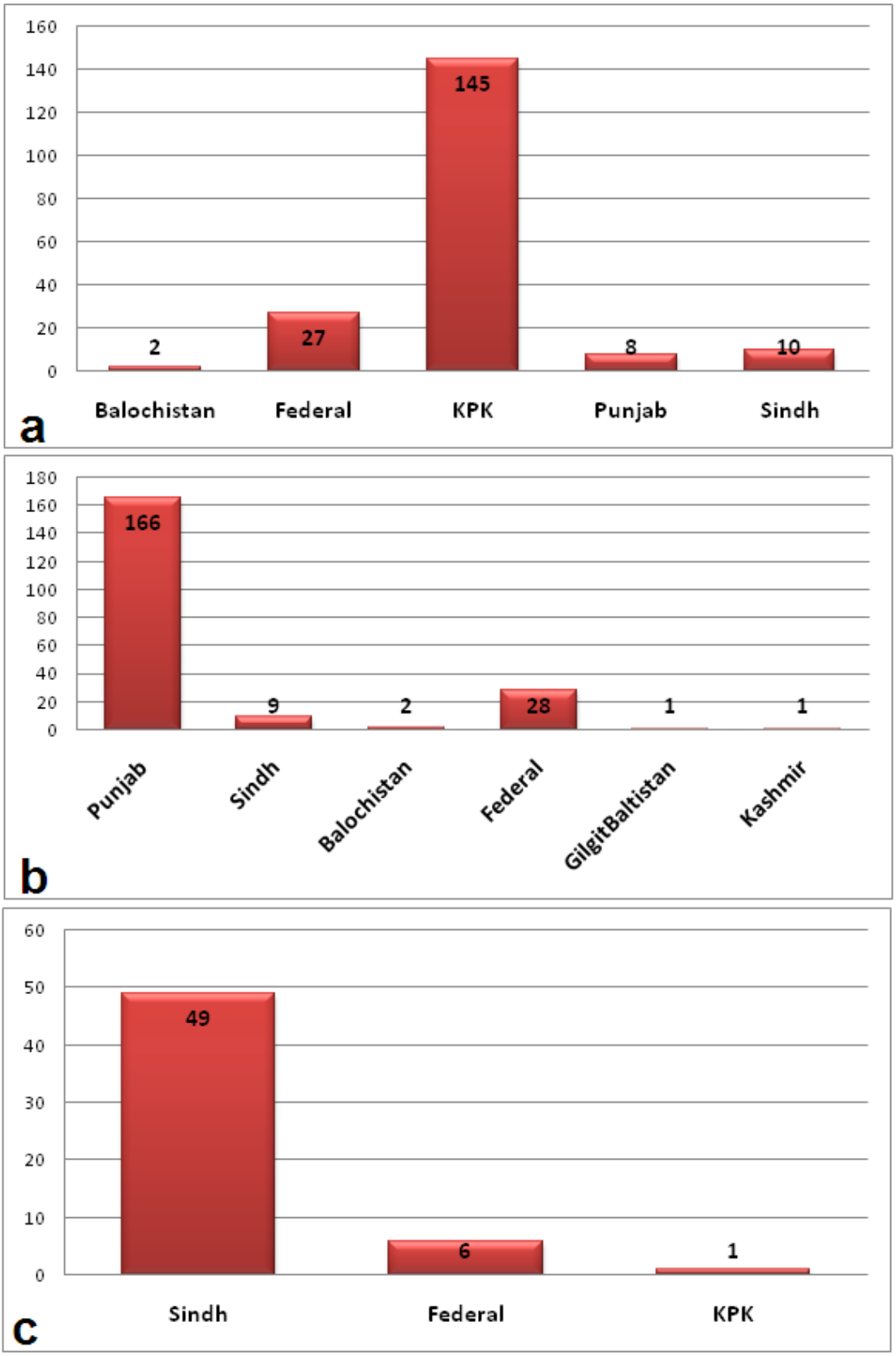

Figure 4.53: Distribution of a. Pathans, b. Punjabis and c. Sindhis in the cases across Pakistan

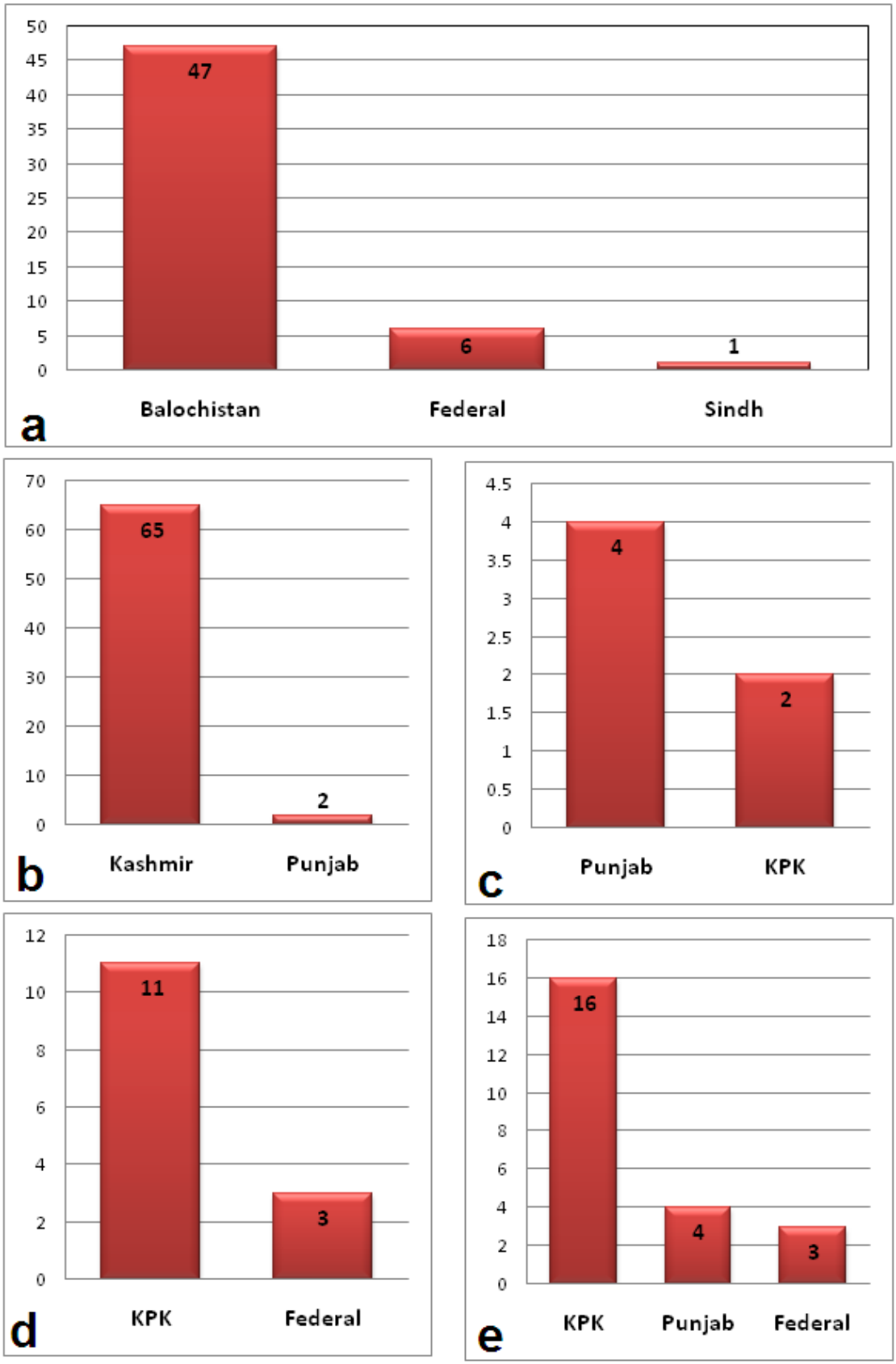

Figure 4.54: Distribution of a. Balochis, b. Kashmiris, c. Sariakis, d. Chitralis and e. Pushtoons in the cases across Pakistan

Table 4.29: Uni-factorial analysis of socio-demographic factors.

S. No	Risk Factor	OR(95%CI)	Z	P
Age groups.				
1.	Young	0.48(0.02-0.08)	-10.89	<0.000
2.	50s	1		
3.	60s	3.60(2.69-4.79)	8.73	<0.000
4.	70s	6.76(4.65-9.83)	10.01	<0.000
5.	80s	1.75(1.19-2.60)	2.81	<0.000
6.	90s/90+	0.13(0.04-0.43)	-3.29	0.000
Smoking				
1.	Yes	3.64(2.86- 4.65)	10.43	<0.000
2.	No	1		
PSA				
1.	>4.0 ng/ml	1		
2.	<4.0 ng/ml	0.02(0.01-0.03)	-17.37	<0.000
Diabetes				
1.	Yes	1		
2.	No	1.41(1.03-1.93)	2.13	0.033
CVD				
1.	Yes	2.23(1.20-4.13)	2.54	0.011
2.	No	1		
Diabetes + Hypertension				
1.	Yes	1.17(0.75-1.84)	0.69	0.492
2.	No	1		
Other disorders				
1.	Yes	0.92(0.69-1.24)	-0.55	0.583
2.	No	1		
No other disorder				
1.	Yes	0.81(0.63-1.03)	-1.74	0.082
2.	No			
BMI				
1.	Under weight	1.63(0.68-3.97)	1.097	0.273
2.	Normal	1		

3.	Obese	2.34(1.60- 3.45)	4.327	<0.000
4.	Over weight	6.14(4.61- 8.23)	12.291	<0.000
Marital status				
1.	Married	1		
2.	Un-married	1.69(0.94-3.16)	1.71	0.086
No of kids				
1.	No kids	0.93(0.58-1.52)	-0.286	0.775
2.	One kid	0.89(0.53- 1.52)	-0.419	0.675
3.	Two kids	1.0(0.73- 1.38)	0.020	0.984
4.	Three kids	0.91(0.66- 1.25)	-0.602	0.547
5.	Four/Over	1		
Dairy products consumption.				
1.	Low	0.02(0.01-0.03)	-15.147	<0.000
2.	Medium	0.15(0.09-0.23)	-8.158	<0.000
3.	High	1		
4.	Very High	2.70(1.34- 5.83)	2.674	0.0075
Physical activity				
1.	Active	1		
2.	Disturb	3.47(2.65-4.55)	9.02	<0.000
3.	Usually at bed	5.93(3.01-11.53)	5.18	<0.000
4.	At bed	5.11(2.29-11.42)	3.98	<0.000
Family history				
1.	Yes	20.67(12.74-34.80)	11.39	<0.000
2.	No	0.12(0.09-0.16)	-13.69	<0.000
3.	Don’t know	1		
Blood pressure				
1.	Hypotension	1		
2.	Normal	0.82(0.65-1.05)	-1.59	0.113
3.	Pre-hypertension	0.89(0.644-1.24)	-0.69	0.492
4.	Hypertension	1.52(1.13-2.04)	2.756	0.006

Table 4.30: Multi-factorial analysis of socio-demographic factors.

S. No	Risk Factor	OR(95%CI)	Z-stat	P-Value
Age groups.				
1.	Young	0.29(0.15-0.54)	-3.792	0.000149
2.	50s	1		
3.	60s	5.49(4.96-9.04)	11.056	<0.000
4.	70s	7.68(5.64-11.72)	11.850	<0.000
5.	80s	3.02(2.25-3.80)	7.486	<0.000
6.	90s/90+	0.42(0.09-1.42)	-1.278	0.201
Smoking				
1.	Yes	3.75(2.94-4.81)	10.51	<0.000
2.	No	1		
PSA				
1.	>4.0 ng/ml	1		
2.	<4.0 ng/ml	0.02(0.01-0.03)	-17.63	<0.000
Diabetes				
1.	Yes	1		
2.	No	1.51(1.03-2.24)	2.09	0.036
CVD				
1.	Yes	2.21(1.21-4.24)	2.50	0.012
2.	No	1		
Diabetes + Hypertension				
1.	Yes	0.8(0.45-1.4)	-0.802	0.422
2.	No	1		
Other disorders				
1.	Yes	0.63(0.42-0.94)	-2.26	0.024
2.	No	1		
No other disorder				
1.	Yes	0.63(0.45-0.87)	-2.768	0.006
2.	No	1		
BMI				
1.	Under weight	1.45(0.49-4.47)	0.657	0.511
2.	Normal	1		

3.	Obese	2.11(1.32-3.41)	3.065	0.002
4.	Over weight	6.62(4.61-9.633)	10.072	<0.000
Marital status				
1.	Married	1		
2.	Un-married	2.34(1.07-5.30)	2.078	0.037688
No of kids				
1.	No kids	0.88(0.44-1.76)	-0.365	0.715
2.	One kid	0.72(0.34-1.54)	-0.851	0.395
3.	Two kids	1.02(0.66-1.59)	0.101	0.919
4.	Three kids	0.96(0.62-1.49)	-0.179	0.858
5.	Four/Over	1		
Dairy products consumption				
1.	Low	0.01(0.001-0.02	-15.18	<0.000
2.	Medium	0.04(0.02-0.07)	-10.03	<0.000
3.	High	1		
4.	Very High	2.89(1.41-6.32)	2.79	0.005
Physical activity				
1.	Active	1		
2.	Disturb	4.81(3.64-6.39)	10.97	<0.000
3.	Usually at bed	8.72(5.68-13.49)	5.50	<0.000
4.	At bed	7.71(4.56-15.87)	6.83	<0.000
PSA				
1.	>4.0 ng/ml			
2.	<4.0 ng/ml			
Family history				
1.	Yes	20.67(12.74-34.80)	11.39	<0.000
2.	No	0.12(0.09-0.16)	-13.69	<0.000
3.	Don’t know	1		
Blood pressure				
1.	Hypotension	1		
2.	Normal	0.31(0.11-0.90	-2.16	0.031
3.	Pre-hypertension	0.51(0.36-0.71)	-3.95	<0.000
4.	Hypertension	0.38(0.25-0.58)	-4.432	<0.000

4.13 Survival analysis

Overall survival analysis tells us that for the 680 cases in study, 411 people were uncensored means they were followed for the entire study duration, until occurrence of event. There was a median survival time of 21 months for the 411 uncencored patients with 95% confidence interval (95% CI) of 18-26 (Table 4.31 and Figure 4.55).

Our next question was to examine the association between different risk factors and their impact on survival length. Intra group differences in survival after adjusting for a potential confounder were confirmed through a proportional hazard model. The hazard of an event is basically the risk of that event, as the time frame vanishes to t = 0.

$$h(t|X) = h_0(t)\, e^{\beta_0+\beta_1*x}$$

H_0(t) is the baseline hazard. Hazard ratio for individuals with X= x vs X = (x+1) as follows,

$$\frac{h(t|X=x+1)}{h(t|X=x)} = \frac{h_0(t)\, e^{\beta_0+\beta_1*(x+1)}}{h_0(t)\, e^{\beta_0+\beta_1*x}} = e^{[\beta_0+\beta_1*(x+1)]-[\beta_0+\beta_1*x]} = e^{\beta_1}$$

We systematically analyses the data for each risk factor and compared the survival curves as follows,

4.13.1 Association between age and length of survival

The data was analyzed for each age group (Young, 50s, 60s, 70s, 80s and 90s/90+). The median follow up was 21, 29, 26, 15, 21 and NA respectively. NA in this case means infinity which is common in survival analysis due to the fact that the data is skewed. In order to determine if there is a statistically significant difference between the survival curves, we perform the log-rank test to test the null hypothesis.

According to log-rank test, there were significant variations among the age groups ($\chi 1^2$ = 21.6 with a p-value = 0.0006 on 5 degree of freedom). To adjust for the covariates we performed the Cox's proportional hazards regression using the coxph() function in R.

When using the Cox regression to perform the test, the results are very similar to the log rank test ($\chi 1^2$ = 20.4 with p-value = 0.0011 on 5 degree of freedom). Age is significantly associated to survival (p-value = 0.0011). The plot (Figure 4.56 a) shows better survival probability (hazard ratio of dying = 0.603) in aged patients but this because we have only three patients in this age group. It was observed that length of survival is proportional to the younger ages, with max observed survival length in young age group, then 50s, 60s and so on. To extend the cox regression to adjust for age, again comparison among the age groups was performed after adjusting for age. Statistically significant differences were observed ($\chi 1^2$ = 20.4 with p-value = 0.0023 on 6 degree of freedom. Significant differences were also recorded within groups. After adjusting for age, younger patients have significantly better survival having 0.66 times the hazard of dying in comparison to the older ones, adjusting for age (HR<1).

4.13.2 Association between BMI and length of survival

The data was also analyzed for each BMI group (Underweight, Normal, Overweight and Obese). The median follow up was 23, 24, 16 and 21 respectively with 95% CI of 13-NA (underweight), 18-32 (normal), 17-29 (over weight) and 23-26 (0bese) (Table 4.31). In order to determine if there is a statistically significant difference between the survival curves, we perform the log-rank test to test the null hypothesis and reject the null hypothesis $\chi 1^2$ = 23.8 with a p-value = <0.000 on 3 degree of freedom). When using the Cox regression, the results are very similar to the log rank test ($\chi 1^2$ = 23.1 with p-value = <0.000 on 3 degree of freedom). BMI is significantly associated to prostate cancer survival (p-value = <0.000). The plot (Figure 4.56) shows comparatively better survival probability (hazard ratio of dying = 0.89) in patients with normal BMI. To extend the cox regression to adjust for age in all the BMI groups, significant differences were observed among the groups with better survival in normal followed by overweight people having 0.89 and 1.50 times the

hazard of dying in comparison to the underweight and obese patients (HR<1 for patients with normal BMI).

4.13.3 Association between Smoking and length of survival

The median length of follow-up for smokers was 21 months in comparison to 24 months in non-smokers with 95% CI of 17-25 and 16-29 respectively. These differences were statistically non-significant ($\chi 1^2$ =13.1, P =0.0003 on 1 degree of freedom). Cox regression was applied and the results closely similar to the log rank test ($\chi 1^2$ = 15.6; p-value < 0.000 on 1 degree of freedom). It is clear from the results that smoking is significantly related to survival (p-value < 0.000), with better survival in non-smokers (HR= 1.56 in smokers). Cox regression was extended to adjust for the covariate age and was thus tested for any difference when comparing smokers to non-smokers after adjusting for age. Again differences were observed (z=3.73; p-value = 0.0002). After adjusting for age, non-smokers have significantly better survival. Smokers have 1.56 times the hazard of dying as compared to non-smokers (Table 4.31 and Figure 4.58 c).

4.13.4 Association between PSA level and length of survival

We observed that, out of the total 680 cases, 495 were having PSA level greater than 4.0ng/ml while the rest have PSA level less than 4.0ng/ml. The median follow up was 24 months for cases having PSA level greater than 4.0 ng/ml with 95% CI=18-29 and 18 months for the cases having PSA level less than 4.0 ng/ml with 95% CI=16-25. Statistically non-significant differences were observed in between length of survival for the cases having PSA level greater than 4.0 ng/ml and those having less 4.0ng/ml ($\chi 1^2$ = 2.7; p-value = 0.102 on 1 degree of freedom). We also tested to confirm if there is a difference in survival functions between the two groups after adjusting for a potential confounder through a proportional hazard model. Cox regression was applied and the results were more or less similar to the log rank test ($\chi 1^2$ = 2.74; p-value = 0.098). PSA level was observed non-significantly related to survival (p-value = 0.098). Cox regression was extended to adjust for the covariate age and again no significant differences were observed (z=1.40; p-value = 0.160) (Table 4.31 and Figure 4.57 c).

4.13.5 Association between Diabetes and length of survival

There were 130 diabetic and 549 non-diabetic patients in our studied population. The median follow up was 14 months for diabetic while 24 for non-diabetic cases with 95% CI of 12-29 and 18-27 respectively. Non-significant differences were observed in between diabetes and length of survival ($\chi 1^2$ = 0.4; p-value = 0.506 on 1 degree of freedom). The results of Cox regression was similar to the log rank test ($\chi 1^2$ = 0.97; p-value = 0.325). It was further observed that though there were non-significant differences in between the two groups but patients with having diabetes have comparatively better survival (HR= 0.883). Cox regression was extended to adjust for the covariate age and was cases with diabetes were again observed having better survival probability, having 0.877 times the hazard of dying in comparison to non-diabetic cases, adjusting for age (HR<1) (Table 4.31 and Figure 4.59 b).

4.13.6 Association between CVD and length of survival

There were only 41 cases with CVD in addition to prostate cancer. The median follow up was 21 months with 95% CI of 19-NA for those with CVD and 29 months with 95% CI of 17-24 for those having no CVD. Significant differences were observed in between the two groups ($\chi 1^2$ = 5.0; p-value = 0.025 on 1 degree of freedom). The results of Cox regression was more or less similar to the log rank test ($\chi 1^2$ = 5.99; p-value = 0.014) with better survival in those having no CVD (HR= 0.546). After adjusting for age, cases without having CVD have significantly better survival having 0.546 times the hazard of dying in comparison to cases with CVD, adjusting for age (HR<1) (Table 4.31).

4.13.7 Association between Marital status and length of survival

The median follow up was 19 months (95% CI=16-19) for married and 19 months (95% CI=7-38) for un-married cases. Non-significant differences were observed in between the two groups ($\chi 1^2$ = 0.0; p-value = 0.964 on 1 degree of freedom). The results of Cox regression produced similar results to the log rank test ($\chi 1^2$ = 0.0; p-value = 0.965 on 1 degree of freedom). After adjusting for age, no clear cut difference was observed in survival for the two groups (z=0.022; p-value=0.980). Un-married

cases have 0.995 times the hazard of dying in comparison to married cases, adjusting for age (HR<1) (Table 4.31 and Figure 4.56 c).

4.13.8 Association between number of kids and length of survival

In our studied population we divided the number of kids into five categories (no kids, one kid, two kids, three kids and four or over four kids). The median follow up was was 19 months (95% CI=7-38), 24 months (95% CI=16-36), 19 months (95% CI=16-34), 16 months (95% CI=16-29) and 16 months (95% CI=12-19) for the cases having no kid, one kid, two kids, three kids and four/over kids respectively. Non-significant differences were observed for the number of kids and length of survival ($\chi 1^2$ = 2.8; p-value = 0.593 on 4 degree of freedom). The results of Cox regression were the the same as the log rank test ($\chi 1^2$ = 1.35; p-value = 0.853). After adjusting for age, When testing the null hypothesis that there is no difference in the survival function among different number of kids, we noticed that we have to accept the null hypothesis ($\chi 1^2$ =6.67, p-value = 0.246), adjusting for age (HR>1) (Table 4.31 and Figure 4.57 a).

4.13.9 Association between Dairy products consumption and length of survival

We categorically divided the cases into Low, Medium, High and very high dairy products consuming groups. The median follow up was 19, 16, 16 and 19 months respectively. In order to determine if there is a statistically significant difference between the survival curves, we perform the log-rank test to test the null hypothesis. According to log-rank test, there were non-significant variations among the dairy products consuming groups ($\chi 1^2$ = 0.8 with a p-value = 0.844 on 3 degree of freedom). When using the Cox regression, the results were very similar to the log rank test ($\chi 1^2$ = 0/79 with p-value = 0.852 on 3 degree of freedom). It means that dairy products consumption is non-significantly associated to survival (p-value = 0.852). Figure 4.57 b shows no clear differences for survival among the four groups. After adjusting for age, comparatively better survival ratio was observed for Low dairy product consumption group (HR=0.852) but again the results are non-significant ($\chi 1^2$ = 5.09 with p-value = 0.279) (Table 4.31).

4.13.10 Association between Prostectomy and length of survival

The median follow up was 29 months (95% CI=18-33) for cases who carried out prostectomy and 18 months (95% CI=16-24) for cases without prostectomy. Statistically significant differences were observed in between the two groups for their length of survival ($\chi 1^2$ = 5.0; p-value = 0.026 on 1 degree of freedom). The results of Cox regression produced similar results ($\chi 1^2$ = 5.39; p-value = 0.020). After adjusting for age, again significant differences were observed in survival for the two groups (z = -2.25; p-value = 0.024). Prostectomy enhances the survival probability and also comparatively increases the length of survival. Patients who carried out prostectomy have 0.753 times the hazard of dying in comparison to those who did not carried out prostectomy, adjusting for age (HR<1) (Table 4.31 and Figure 4.56 c).

4.13.11 Association between Physical activity and length of survival

Physical activity is one among the nine cancer risk factors according to the World Health Organization (Barbaric et al., 2010). Its usefulness in various cancers has previously been reported and its correlation to cancer prevention is the most studied subject (Courneya and Friedenreich, 2007). We evaluated physical activity in relation to survival probability and length of survival in the study population. We divided physically activity into four categories (active, disturb, usually at bed, and at bed). The median follow up was 24 months (95% CI=23-32) for physically active cases, 21 months (95% CI=14-23) for physically disturb cases, 19 months (95% CI=17-NA) for cases physically disturb up to the extent that majority of times remain at bed and 16 months (95% CI=14-NA) for physically too much disturb cases who are at bed. Physical activity was observed significantly associated with PCa mortality and thus affect their length of survival ($\chi 1^2$ = 25.1; p-value <0.000 on 3 degree of freedom). The results of Cox regression produced similar results ($\chi 1^2$ = 25.4; p-value <0.000). After adjusting for age, again significant differences were observed in survival among the physical activity groups (z = 4.989; p-value < 0.000 for physically disturb cases, z = 1.461; p-value < 0.010 for cases usually at bed and z = 0.756; p-value < 0.000 for cases at bed). According to the results, physical activity increases survival probability and survival length also. Physically active patients have the significantly better

survival ratios. Physically disturb cases have 1.254 times, cases usually at bed has 1.386 times and cases at bed has 1.767 times the hazard of dying in comparison to the physical active patients, after adjusting for age (Table 4.31 and Figure 4.58 b). It is now well established that healthy physical activity help to improve quality of life and increases cancer survival by reducing the recurrence risk or slowing cancer progression of cancer (Fong et al., 2012). It has also been proposed that vigorous physical activities positively regulate cell cycle, DNA repair pathways and increase apoptosis (Rundqvist et al., 2013, Soliman et al., 2011).

4.13.12 Association between Family history and length of survival

There were three categories of patients in our study with reference to family history of any type of cancer i.e. patients with having family history of cancer, patients having no family history and patients who do not know about any history of cancer in their families. The median follow up was 18 months (95% CI=16-27) for cases having family history of cancer, 32 months (95% CI=18-34) for cases without family history and 24 months (95% CI=16-30) for cases who do not know about cancer history. Family history was recorded non-significantly associated to survival ($\chi 1^2$ = 2.0; p-value = 0.356 on 2 degree of freedom). The results of Cox regression produced similar results ($\chi 1^2$ = 0.32; p-value = 0.57).

4.13.1 Association between Blood pressure and length of survival

Study population was divided into hypotensive, normal, pre-hypertensive and hypertensive cases. The median follow up was 24 months (95% CI=16-NA) hypotensive cases, 25 months (95% CI=21-30) for normal BP cases, 16 months (95% CI=14-29) for pre-hypertensive cases and 16 months (95% CI=14-29) for hypertensive cases. BP was non-significantly associated with length of survival ($\chi 1^2$ = 3.1; p-value = 0.376 on 3 degree of freedom). The results of Cox regression are similar to the log rank test ($\chi 1^2$ = 3.18; p-value = 0.364). After adjusting for age, again non-significant differences were observed in survival for the BP groups ($\chi 1^2$ = 4.96; p-value = 0.291). It was further noticed that, though BP is non-significantly associated to survival but patients having normal BP has better survival length (Table 4.31 and Figure 4.59 a).

Table 4.31: Survival Analysis on the basis of different risk factors.

S. No	Analysis Type	Records	Events	Median	95% LCI	95% UCI
Overall survival						
1.	Overall	680	411	21	18	26
Survival analysis on the basis of age groups.						
2.	Young	15	11	21	18	NA
3.	50s	69	42	29	18	33
4.	60s	269	168	26	23	32
5.	70s	234	150	15	14	18
6.	80s	90	40	21	14	41
7.	90s/90+	3	1	NA	13	NA
Survival analysis on the basis of smoking.						
1.	Smokers	457	278	21	17	25
2.	Non-smokers	233	133	24	16	29
Survival analysis on the basis of PSA level						
1.	>4.0 ng/ml	495	289	24	18	29
2.	<4.0 ng/ml	185	123	18	16	25
Survival analysis on the basis of Diabetes.						
1.	Diabetic	130	103	14	12	29
2.	Non-diabetic	549	308	24	18	27
Survival analysis on the basis of CVD						
1.	Yes	41	14	21	19	NA
2.	No	639	397	29	16	19
Survival analysis on the basis of BMI.						
1.	Under weight	11	9	23	13	NA
2.	Normal	239	148	24	18	32
3.	Obese	82	53	16	12	26
4.	Over weight	348	202	21	17	29

Survival analysis on the basis of Marital status.						
1.	Married	644	388	19	16	19
2.	Un-married	36	36	19	7	38
Survival analysis on the basis of number of children.						
1.	No kids	43	26	19	7	38
2.	One kid	35	22	24	16	36
3.	Two kids	122	69	19	16	34
4.	Three kids	118	70	16	16	29
5.	Four/Over	362	224	16	12	19
Survival analysis on the basis of Dairy products consumption.						
1.	Low	50	135	19	7	26
2.	Medium	182	30	16	16	19
3.	High	209	120	16	7	29
4.	Very High	239	126	19	16	24
Survival analysis on the basis of Prostectomy.						
1.	No	528	327	18	16	24
2.	Yes	152	84	29	18	33
Survival analysis on the basis of Physical activity						
1.	Active	258	169	24	23	32
2.	Disturb	303	198	21	14	23
3.	Usually at bed	73	73	19	17	NA
4.	At bed	46	46	16	14	NA
Survival analysis on the basis of Family history						
1.	Yes	165	94	18	16	27
2.	No	385	237	32	18	34
3.	Don’t Know	130	80	24	16	30
Survival analysis on the basis of Blood pressure						
1.	Hypotension	8	7	24	16	NA
2.	Normal	421	241	25	21	30
3.	Pre-hypertension	95	63	16	24	29
4.	Hypertension	156	100	16	14	29

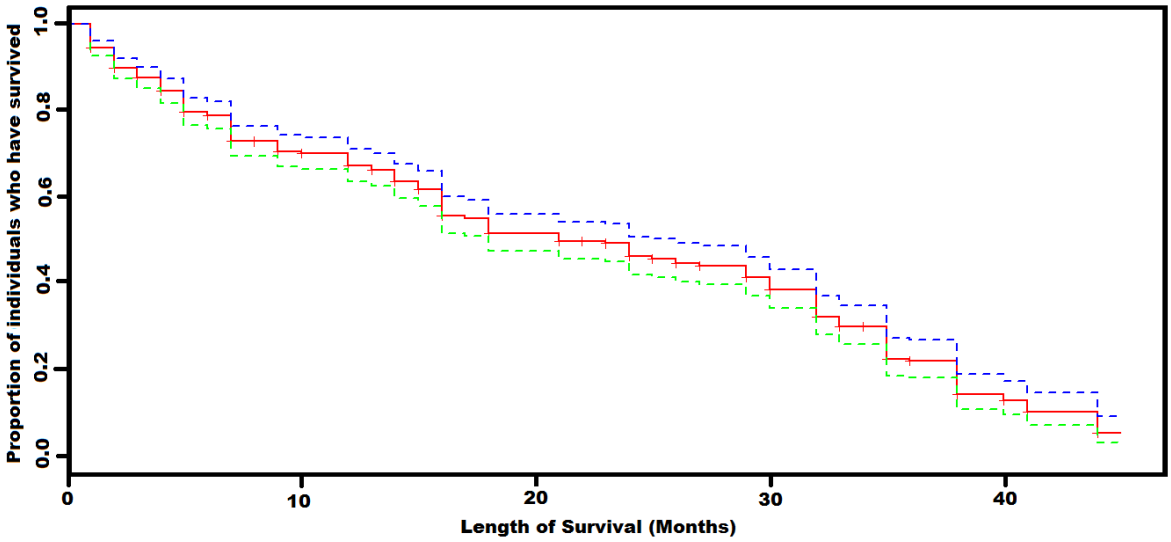

Figure 4.55: Overall Kaplan–Meier survival analysis with 95% CI.
The plot shows the survival curve, the proportion of individual who have survived up for a particular time are shown as a solid red line and the 95% CI as the dashed lines.

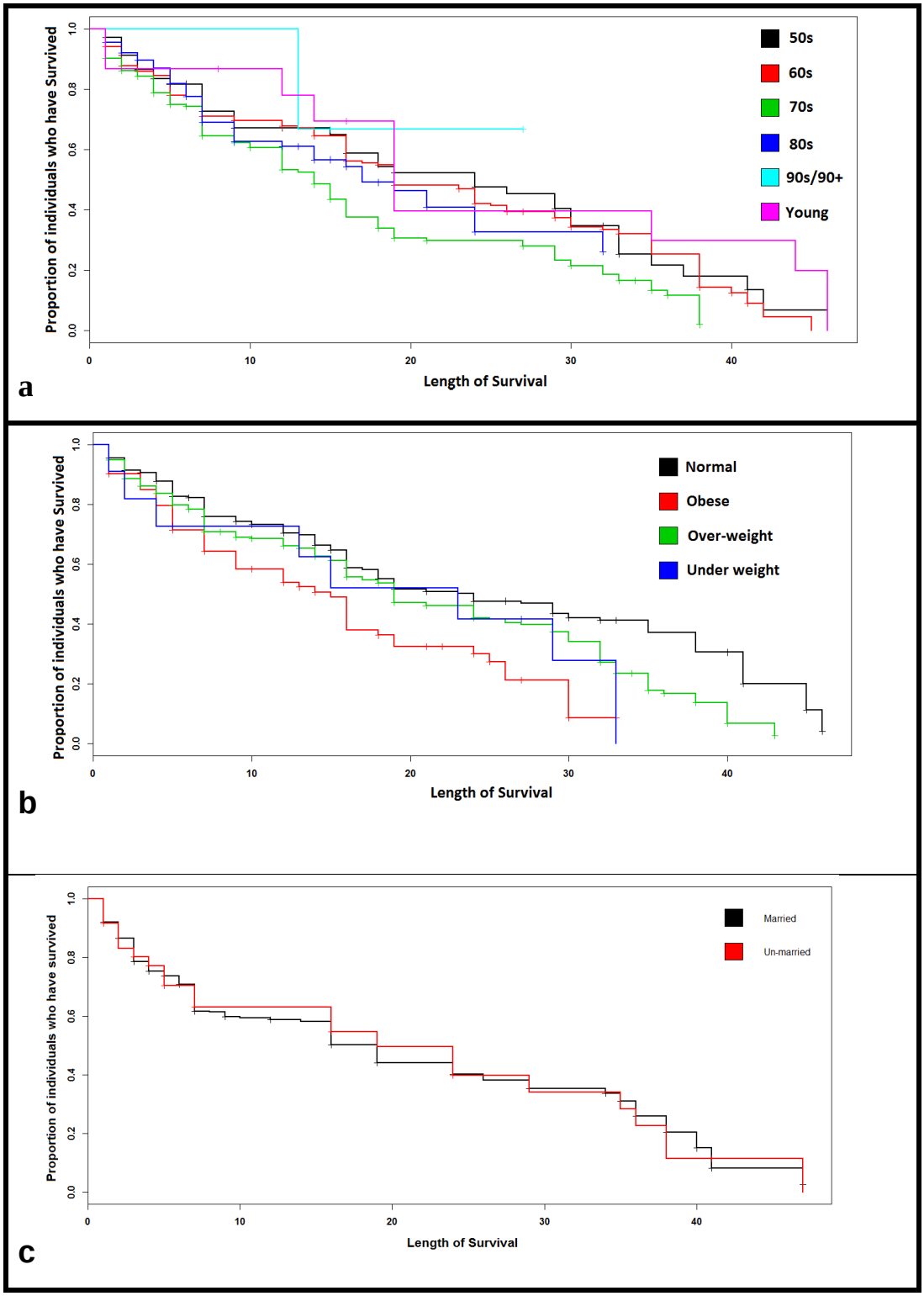

Figure 4.56: Survival curves of Prostate Cancer Patients

a. Kaplan-Meier plot by age groups of PCa patients in Pakistani population.
b. Kaplan-Meier plot by BMI of PCa patients in Pakistani population
c. Kaplan-Meier plot by Marital status of PCa patients in Pakistani population

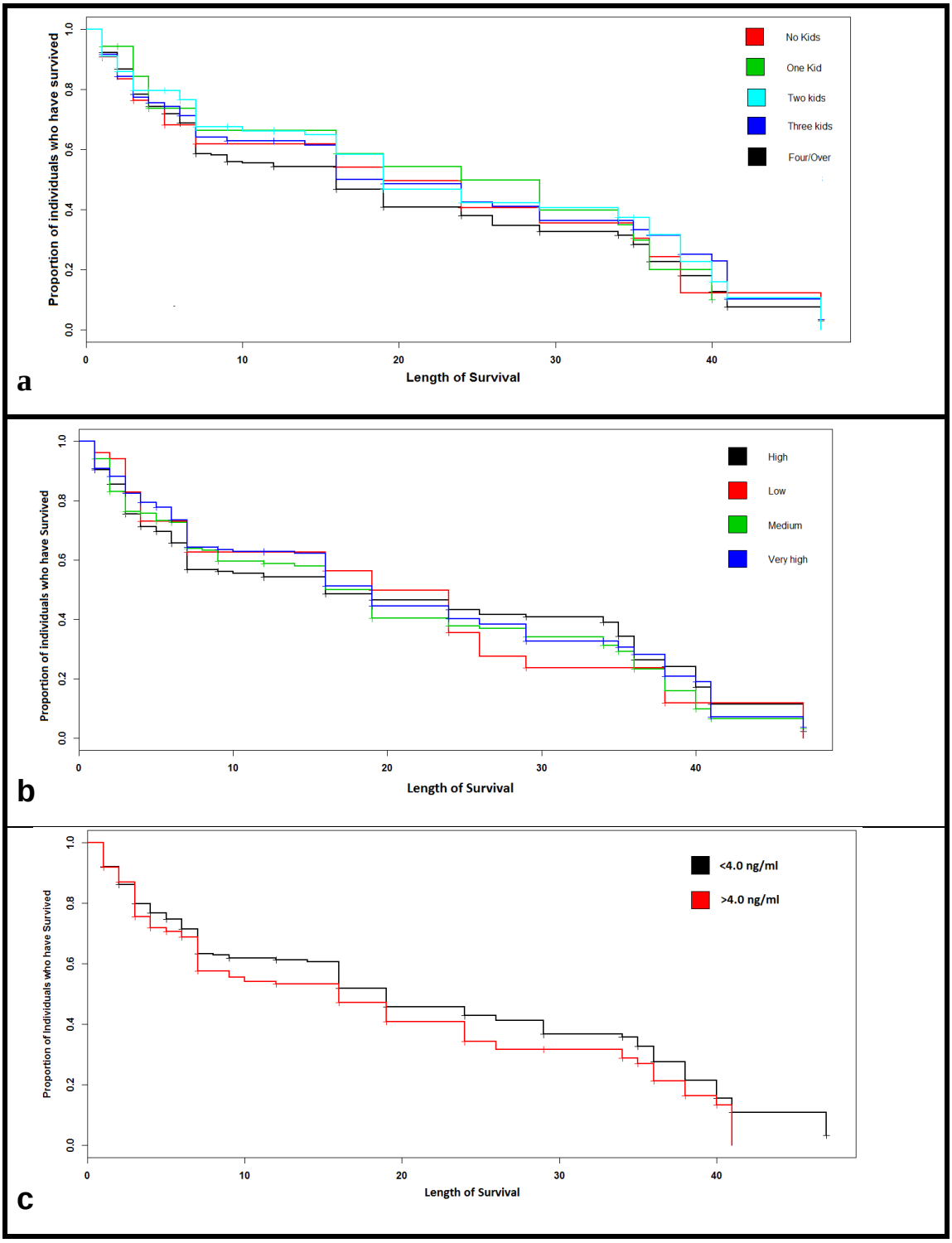

Figure 4.57: Survival curves of Prostate Cancer Patients

a. Kaplan-Meier plot by Number of kids of PCa patients in Pakistani population.
b. Kaplan-Meier plot by Diary products consumption of PCa patients in Pakistani population
c. Kaplan-Meier plot by PSA level of PCa patients in Pakistani population

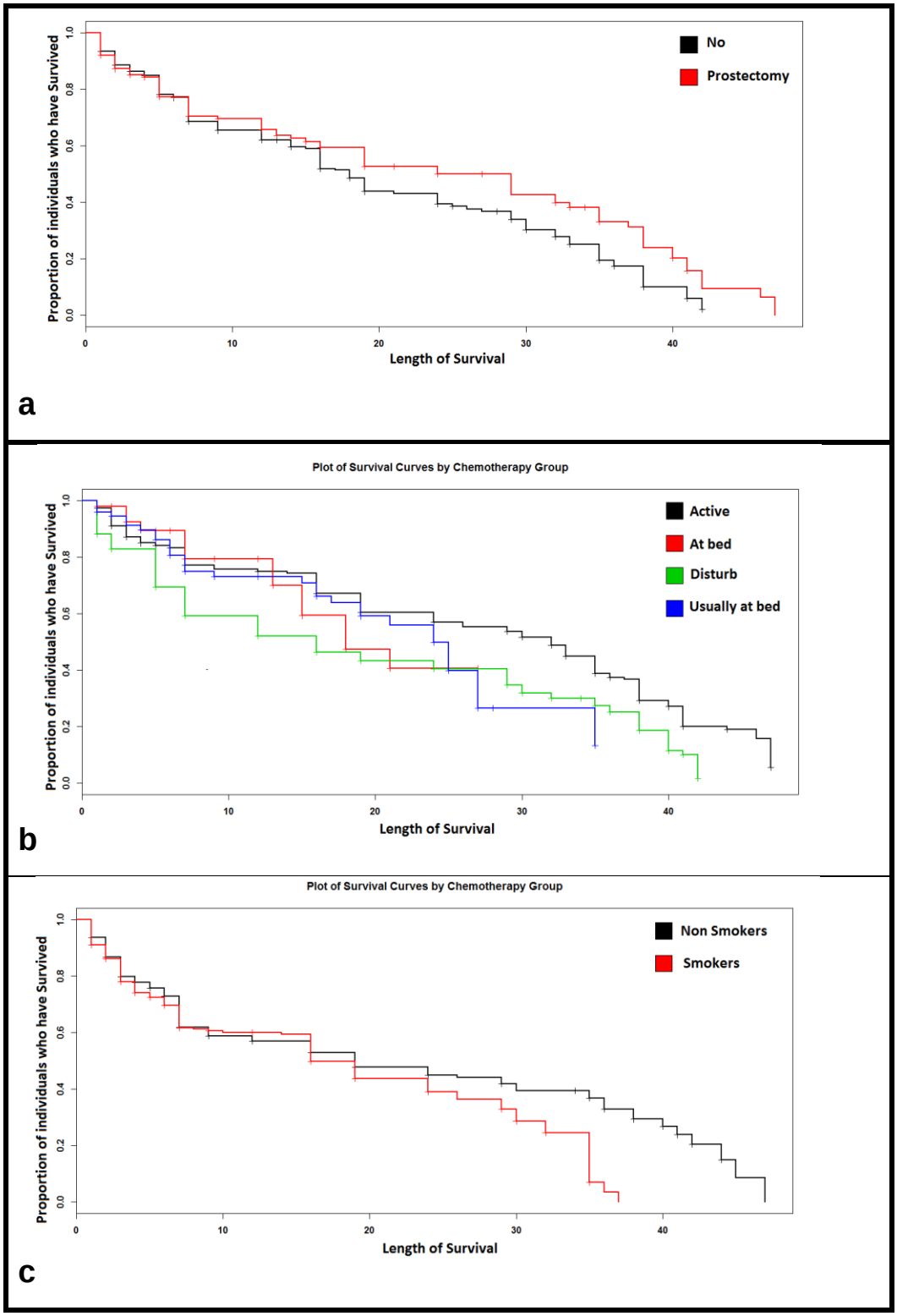

Figure 4.58: Survival curves of Prostate Cancer Patients

a. Kaplan-Meier plot by Prostectomy of PCa patients in Pakistani population.
b. Kaplan-Meier plot by Physical activity of PCa patients in Pakistani population
c. Kaplan-Meier plot by Smoking of PCa patients in Pakistani population

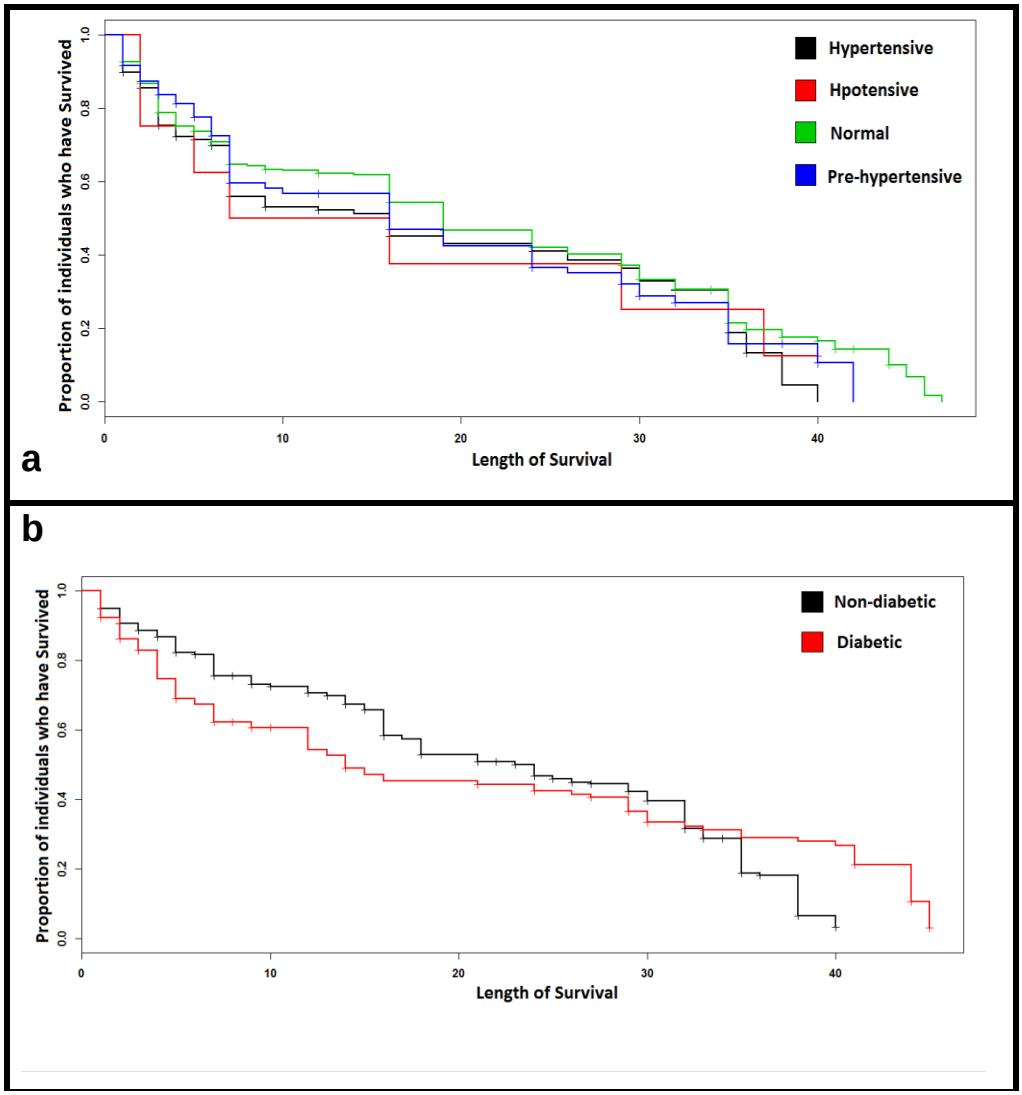

Figure 4.59: Survival curves of Prostate Cancer Patients

a. Kaplan-Meier plot by Blood pressure of PCa patients in Pakistani population.

b. Kaplan-Meier plot by Diabetes of PCa patients in Pakistani population

4.14 PakProstate.com

During this research work, a number of problems were faced related to different domains of the domains of the research work from Literature to data search to genes and sequences etc. So it etc. So it was felt that to avoid hectic queries across different databases a single flat form is form is needed to help the users that supports the information technology needs. This repository repository offers multiple functionalities to the users in different modules including, Literature Literature search, Genes, Drugs, sample data, data upload, simple data analysis and blog option. blog option. User can customize his search on the basis of Date, Journal, Data and Journal and Journal and previous number of days (

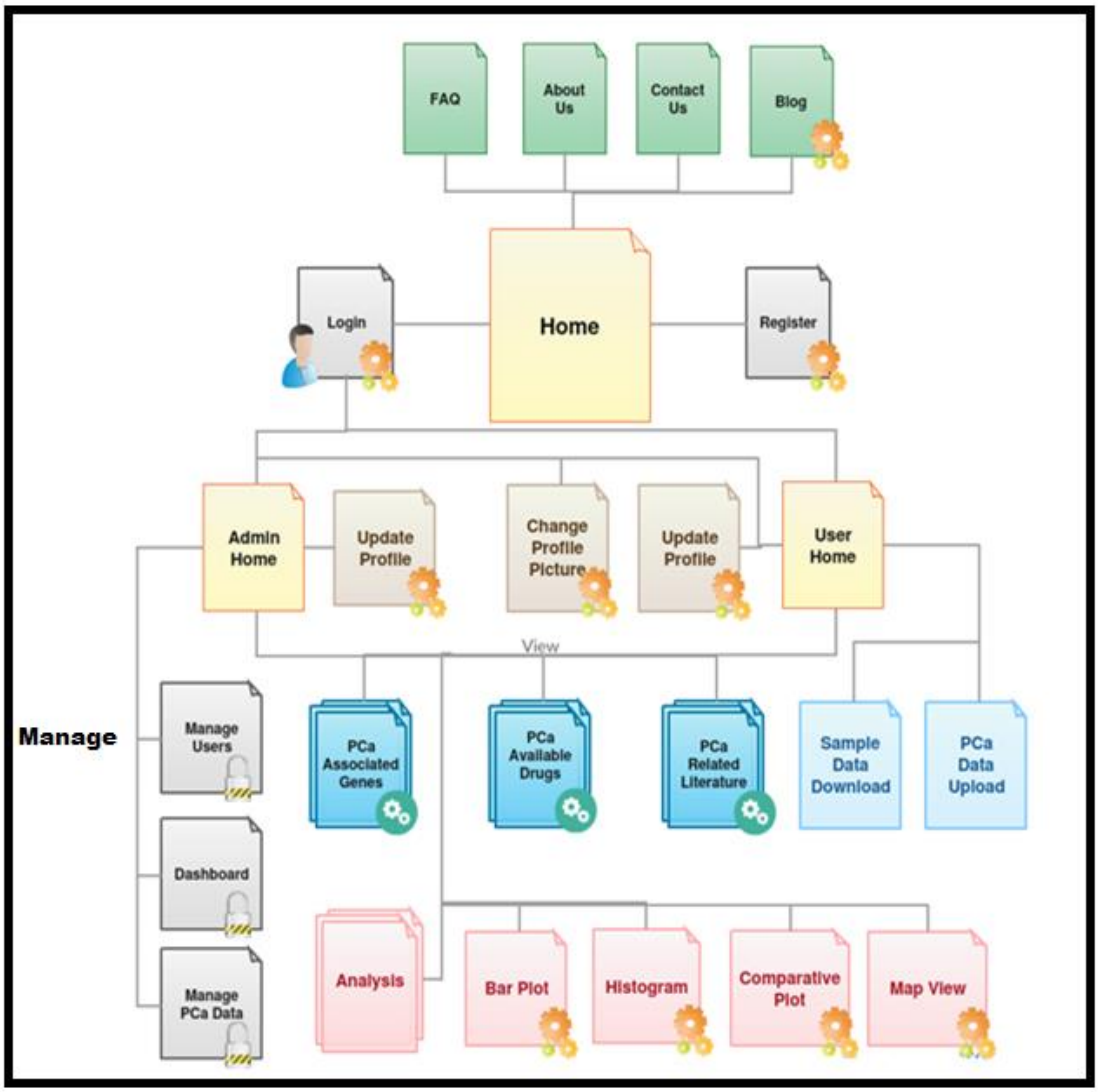

Figure 4.60 to Figure 4.72).

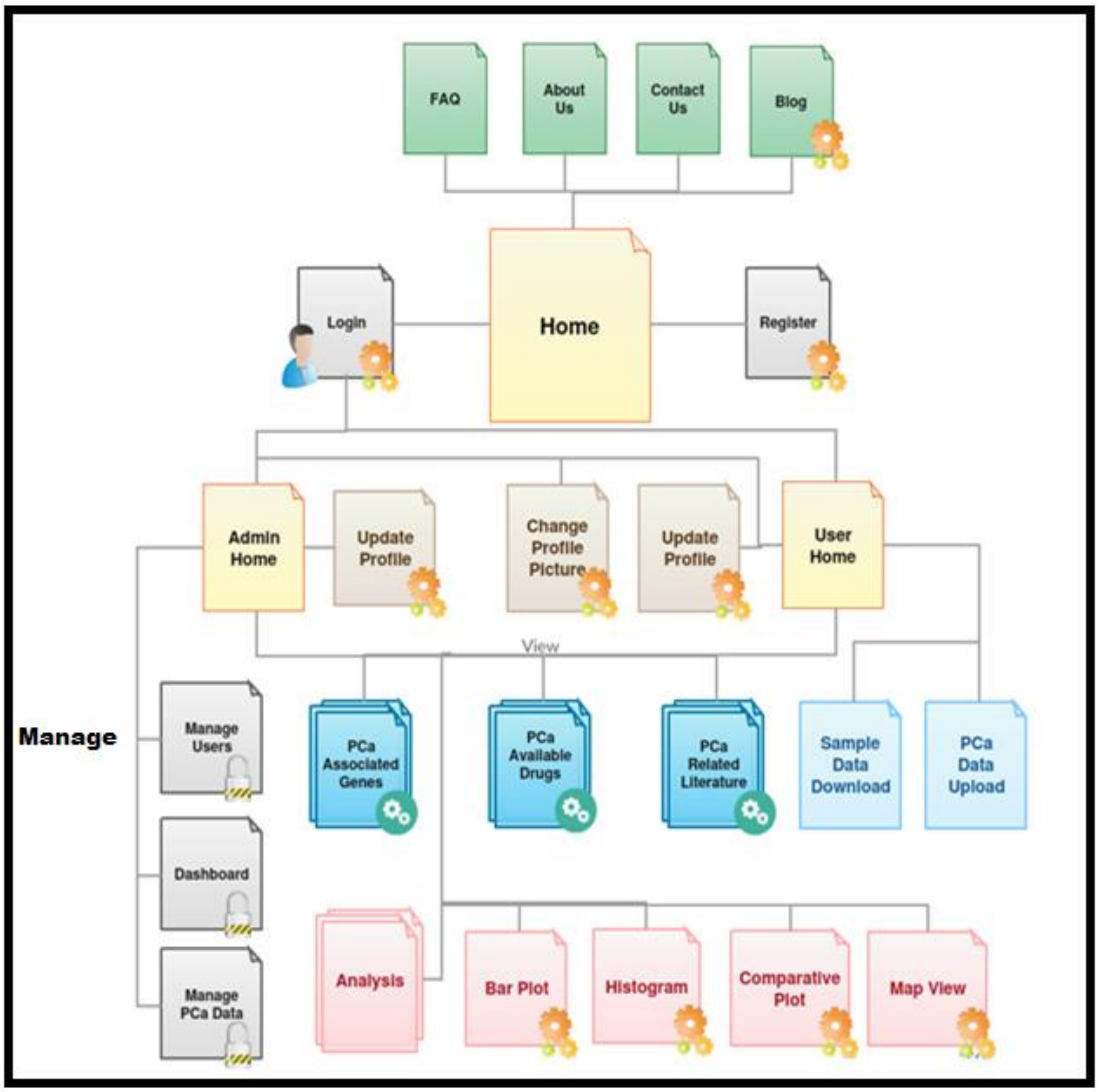

Figure 4.60: Sitemap of pakprostate.com

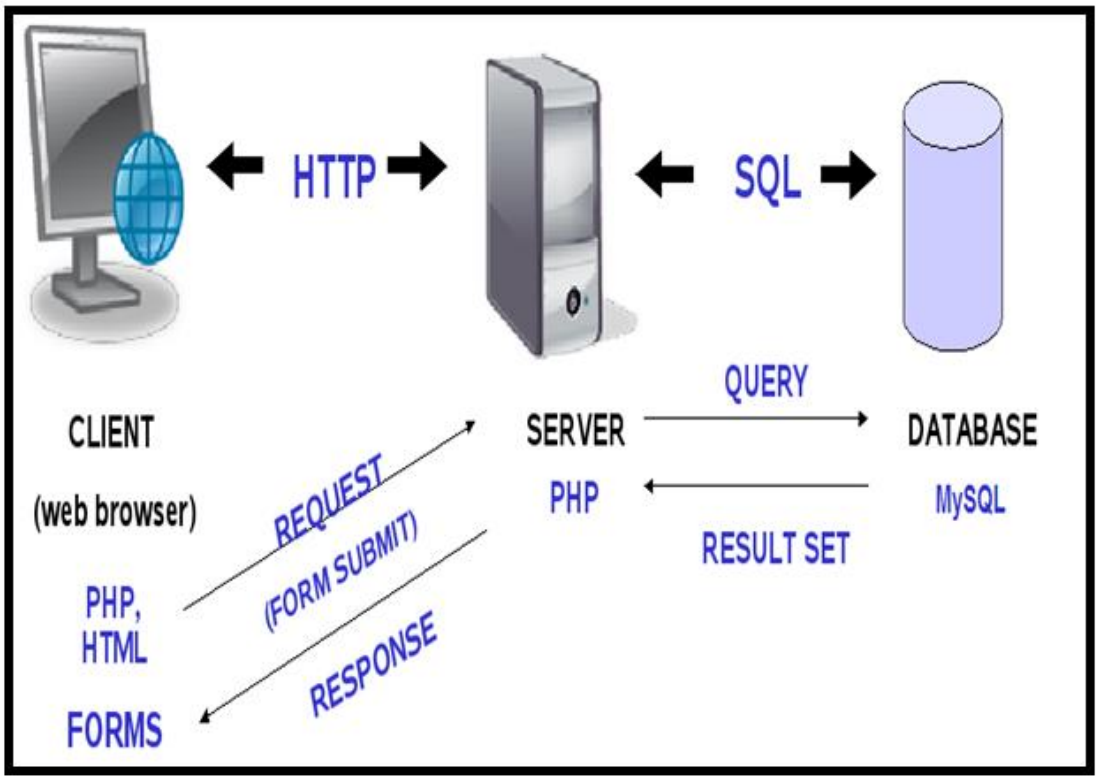

Figure 4.61: Client server diagram of pakprostate.com

Figure 4.62: Snap shot of unregistered view of pakprostate.com

Figure 4.63: Snap shot of registered view of pakprostate.com

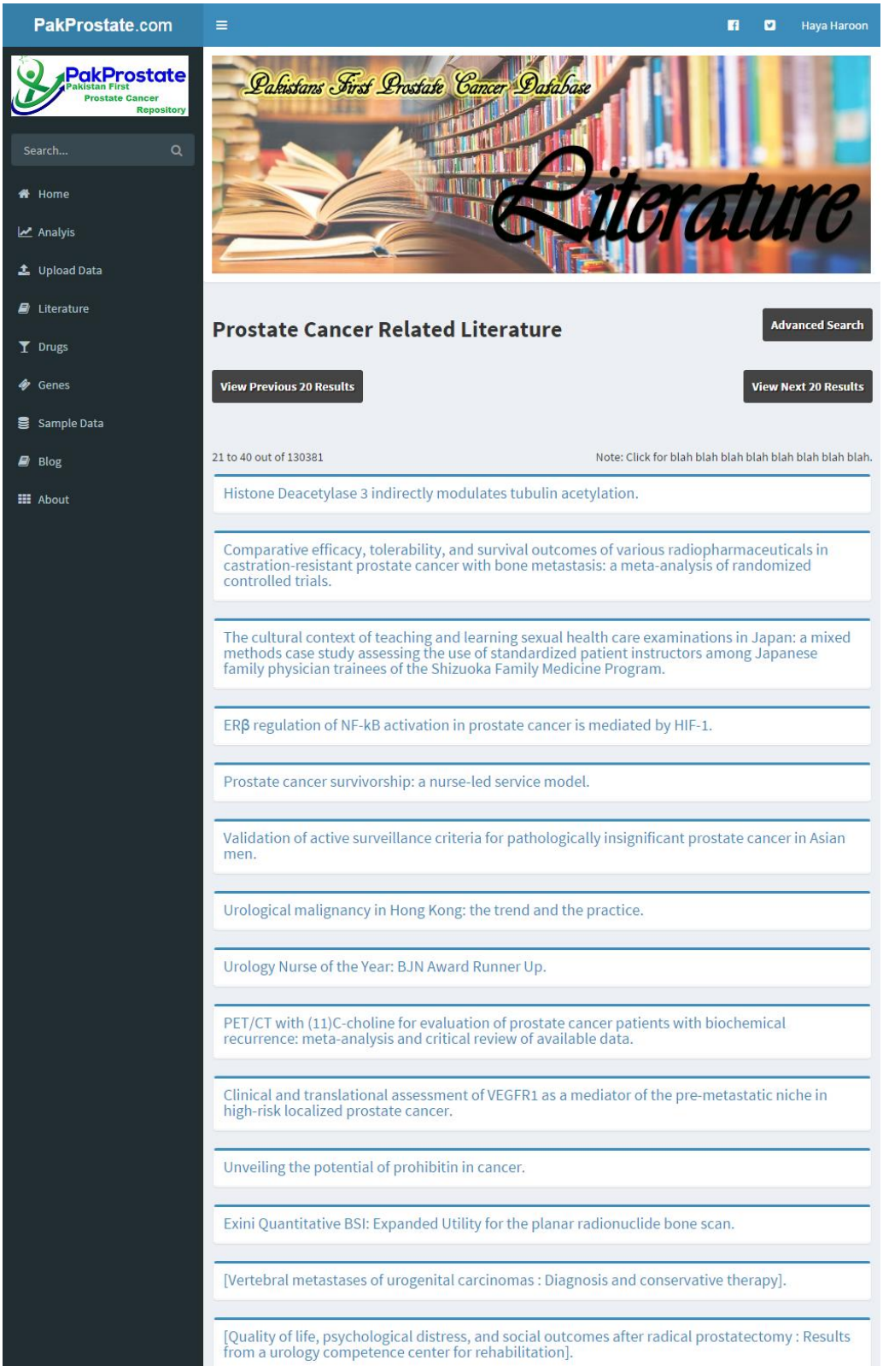

Figure 4.64: Snap shot of registered view showing Literature module.

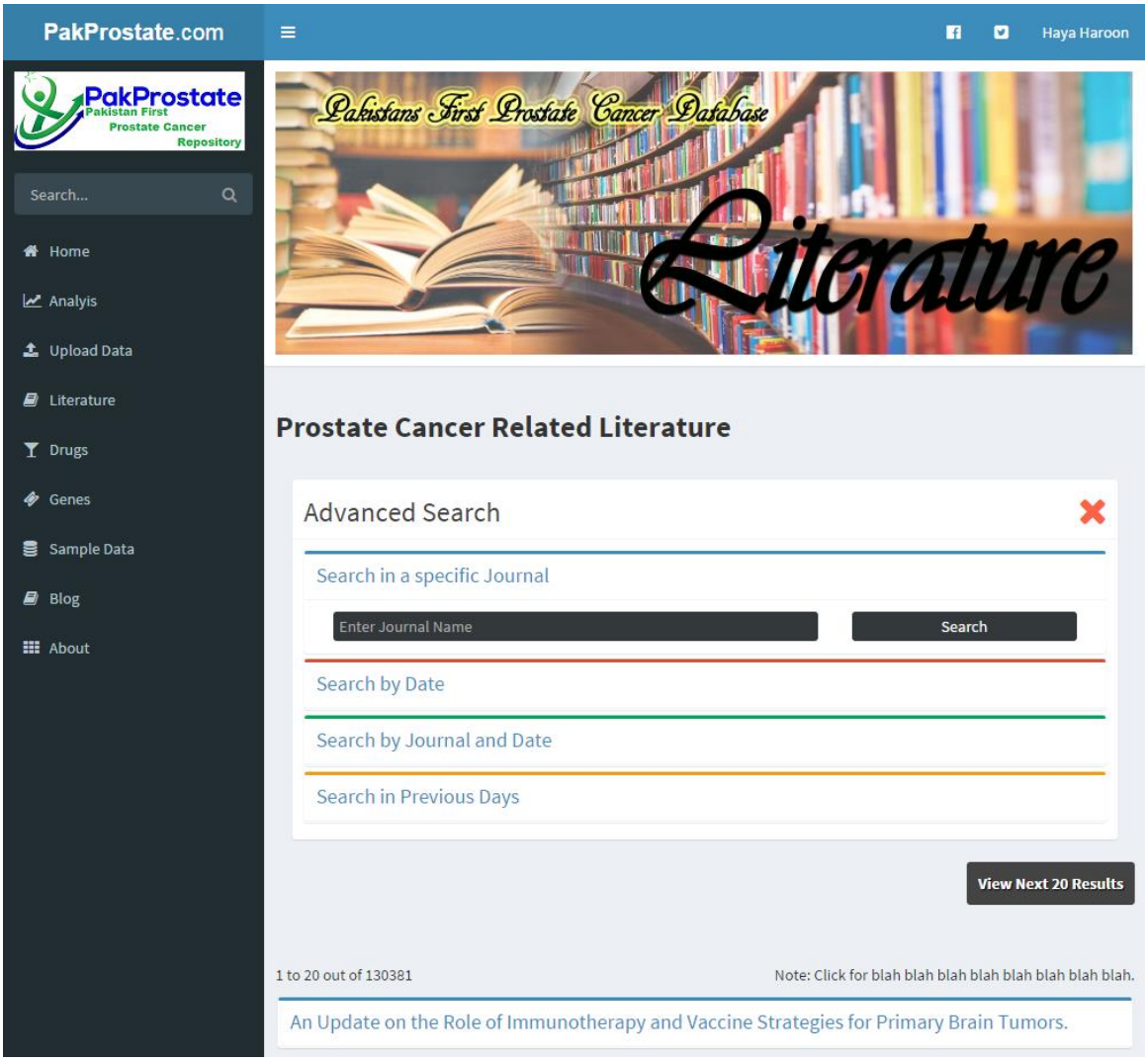

Figure 4.65: Snap shot of registered view showing customized Literature module.

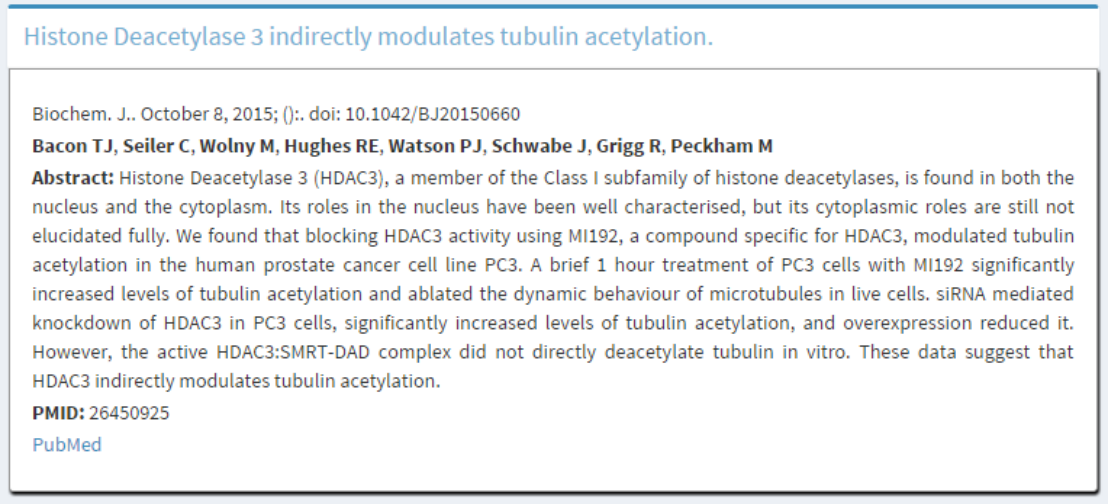

Figure 4.66: Snap shot of registered view showing detail view of Literature results.

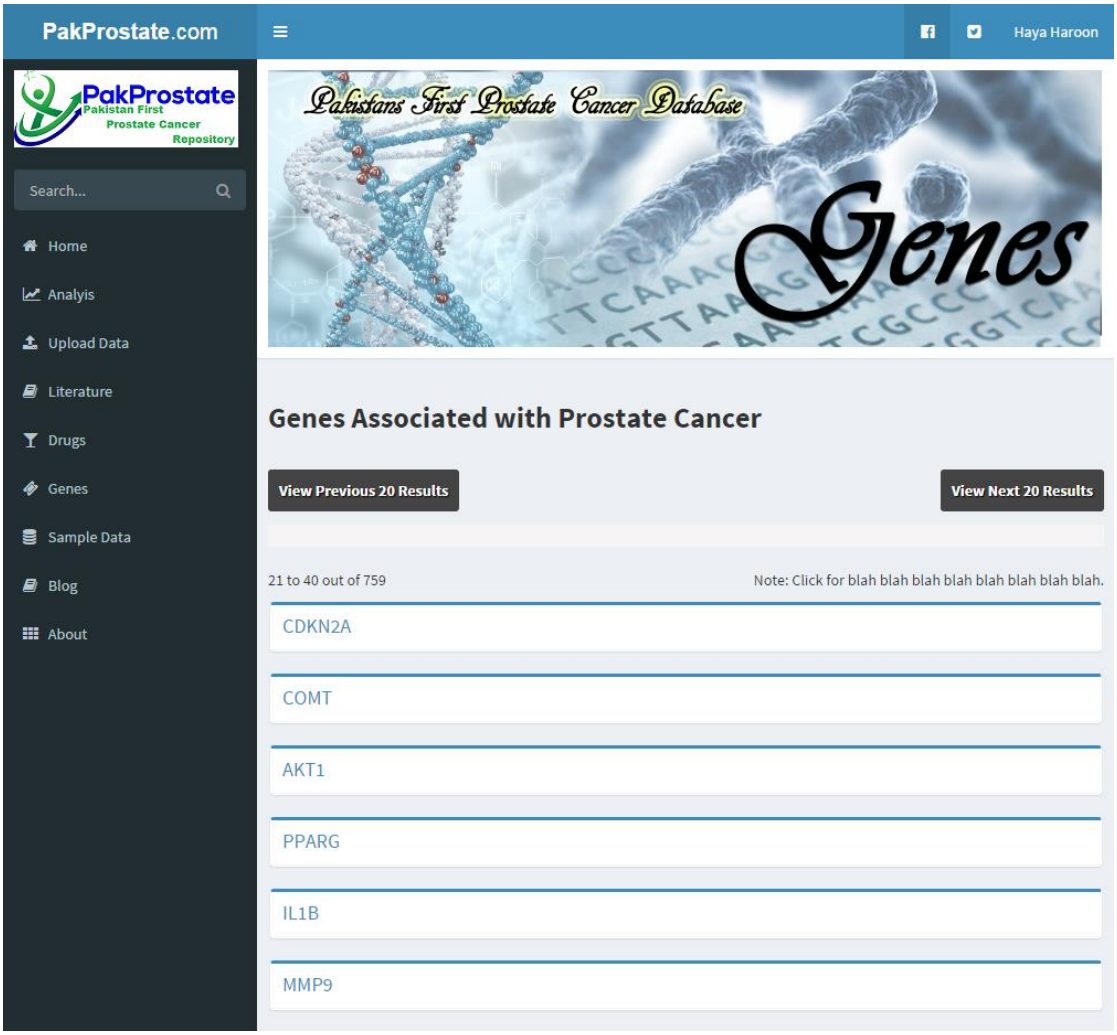

Figure 4.67: Snap shot of registered view showing Genes module.

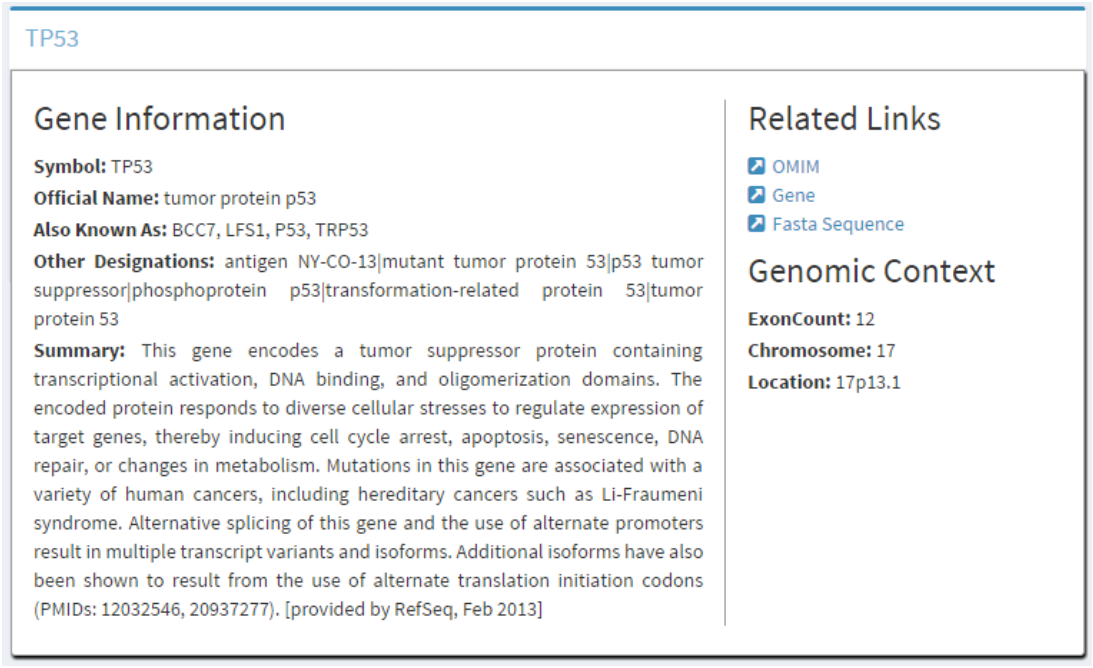

Figure 4.68: Snap shot of registered view showing detail view of Gene results.

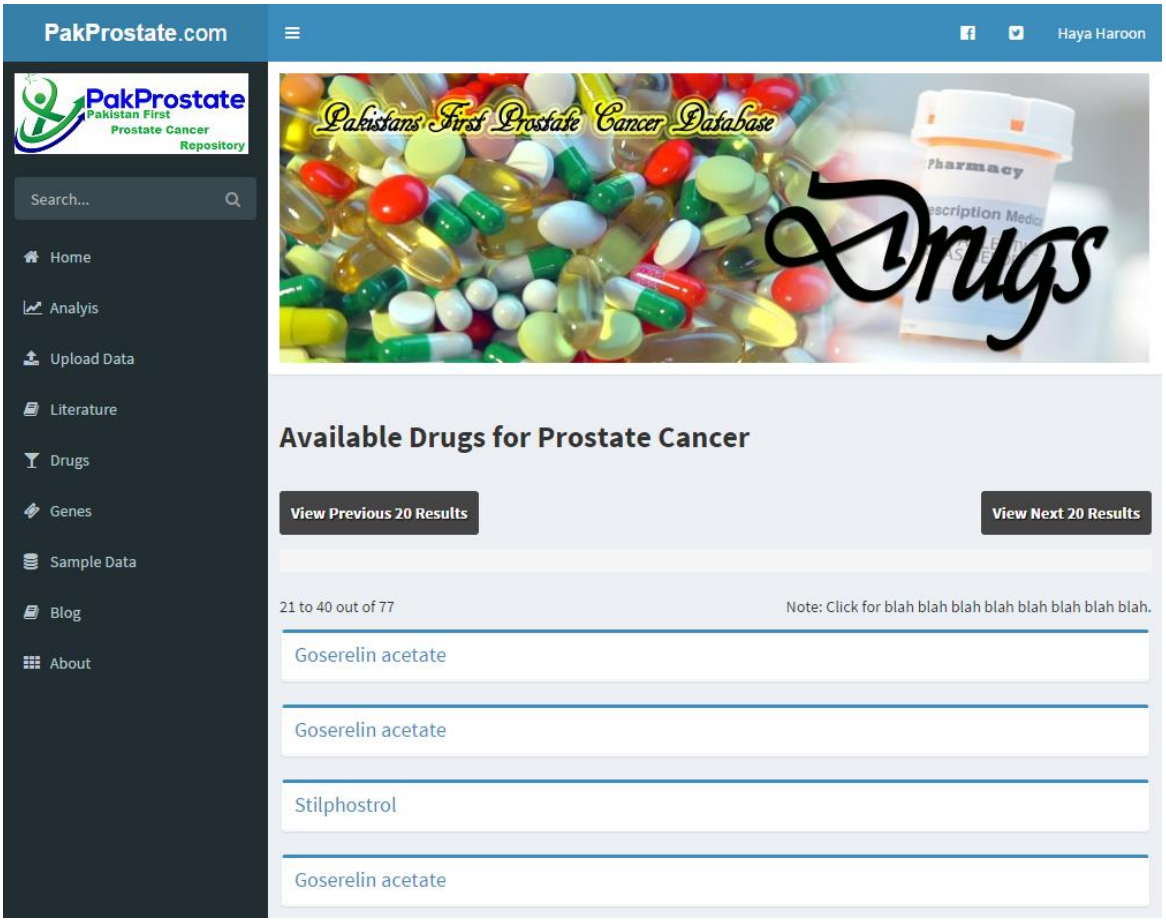

Figure 4.69: Snap shot of registered view showing Drugs module.

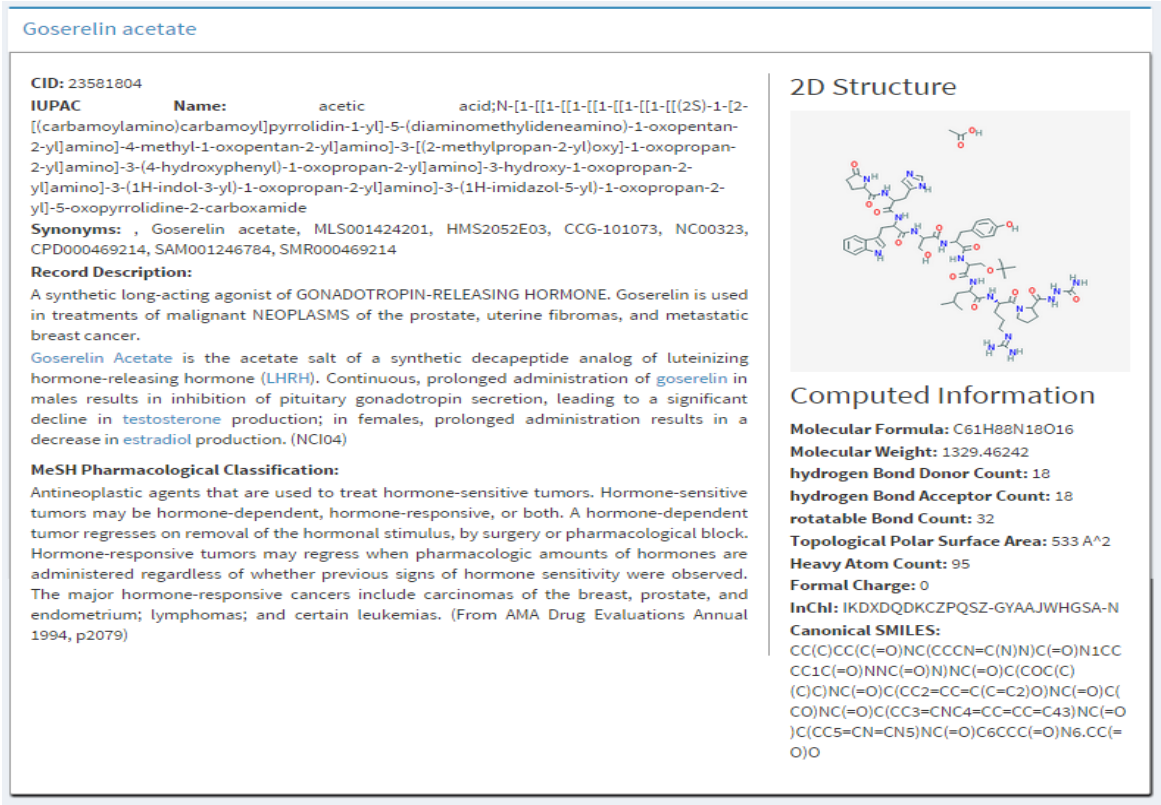

Figure 4.70: Snap shot of registered view showing detail view of drug results.

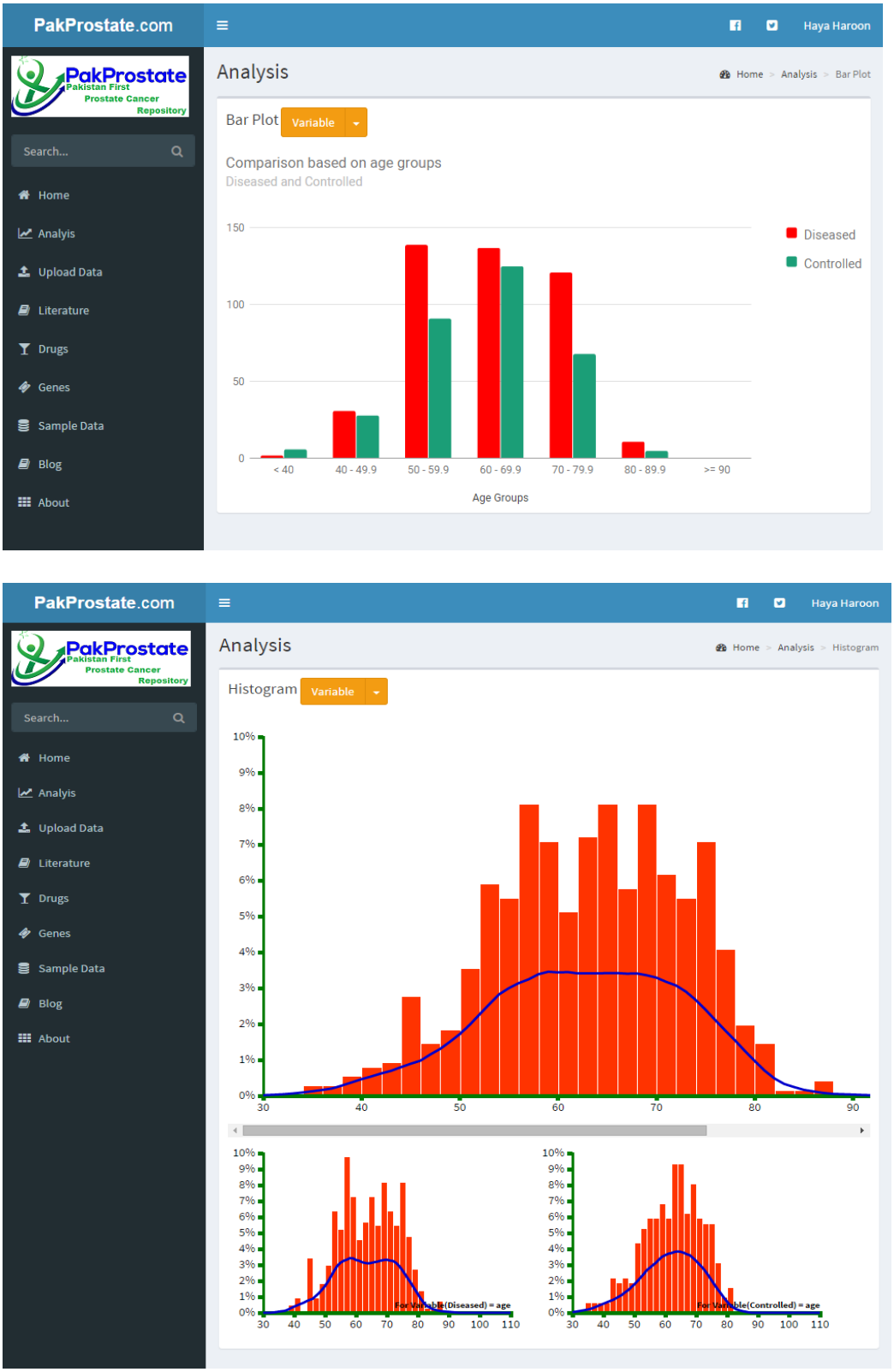

Figure 4.71: Snap shot of registered view showing results through the analysis module.

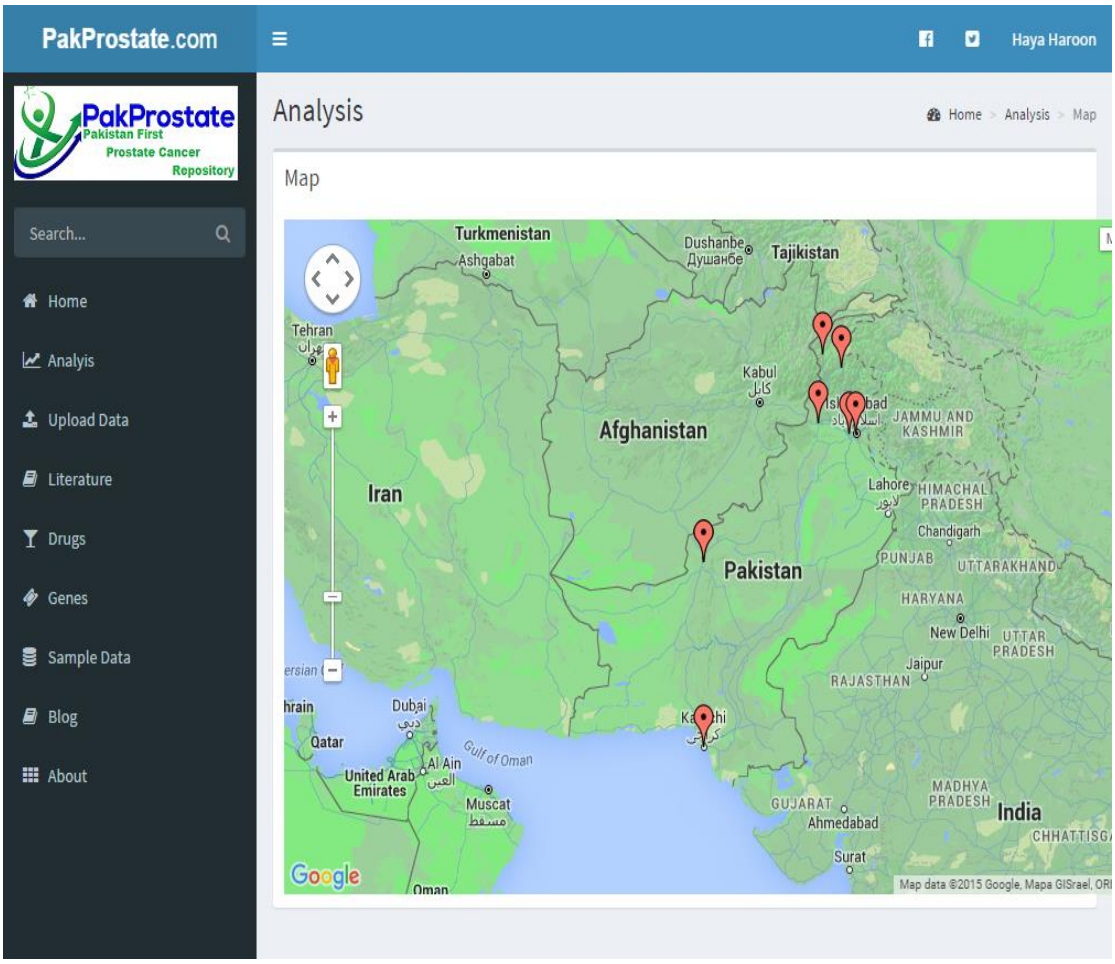

Figure 4.72: Snap shot of Map view.

Chapter 5

CONCLUSIONS

5.1 Summary of Results

All cancers including prostate cancer results due to the accumulation of different types of genetic alterations that cause break down of the respective protein-protein interaction network. TP53, PTEN and AR genes were studied with the help of computational and molecular biology techniques. Study samples consists of 680 histologically confirmed prostate cancer patients and 500 controls. A total of Twenty six novel exonic mutations and six novel intronic variants were detected in the three target genes with the addition of four previously reported exonic alterations. The exonic mutations consist of eight in TP53 (7 novel and one previously reported), nine in PTEN (8 novel and one previously reported) and thirteen in AR (11 novel and two previously reported) gene.

In TP53, we found the g.7676154C>G in exon-4 (p.P72R), a previously reported and seven novel mutations consists of g.7675075insG in exon-5 (p.H179Q), g.7675101insA in exon-5 (p.E171R), g.7675153C>G in exon-5 (p.P153P), g.7674958T>C in exon-6 (p.P191P) g.775103C>T in exon-5 (p.T170M), g.7675158C>G in exon-5 (p.P152A) and g.7674923T>G in exon-6 (p.V203G) mutations. In PTEN, we detected g.87933038T>G in exon-5 (p.H93Q) a previously reported one while eight novel including g.87933015G>C in exon-5 (p.A86P), g.87933030G>A in exon-5 (p.E91K), g.87933050G>C in exon-5 (p.Q97H), g.87933054G>T in exon-5 (p.E99X), g.87933228G>A in exon-5 (p.E157K), g.879960906_87960908CAC>TTT in exon-8 (p.H272F), g.87960951T>C in exon-8 (p.S287P) and g.87960954G>A in exon-8 (p.E288F) mutations. In AR gene, we found the g.67723716G>T in exon-8 (p.D880Y) and g.67723745C>T in exon-8 (p.S889S) previously reported and eleven novel mutations all in exon-8 consisting of g.67723688insG (p.I870fs), g.67723687_67723688TT>Gdel (p.I870_splice), g.67723721G>T (p.L881L), g.67723757_67723758GG>AT (p.P893P & p.E894X), g.67723776A>G (p.I900V), g.67723804C>G (p.S909C), g.67723796G>T (p.K906N),

g.67723801T>C (p.L908P), g.67723813_67723814TC>GG (p.V912G), g.67723821A>T (p.I915F) and g.67723825A>C (p.Y916S).

The intronic variants were further investigated for their association specifically through logistic regression. Both the intronic variants of TP53 (g.7675016T>A and g.7674991T>A) showed significant association with increased PCa risk. Out the four intronic variants of PTEN, only g.87891382G>A showed no correlation with prostate cancer. Similarly, both the intronic variants of AR gene (g.67637091T>A and g.67724021T>C) were also found associated with increased PCa risk.

All the observed exonic mutations were computationally analyzed for their possible structural and functional consequences. 5/8 of TP53 (P152A, T170M, E171R, H179Q and V203G), 7/9 of PTEN (A86P, E91K, H93Q, Q97H, E99X, H272F and E288F) and 11/13 of the AR gene mutations (I870fs, I870_splice, D880Y, E884X, I900V, S909C, K906N, L908P, V912G, I915F and Y916S) were predicted damaging and have functionally impacts.

Age, smoking CVDs, BMI, dairy products consumption, physical activity, family history and hypertension was observed responsible for increasing prostate cancer risk. Overall survival analysis showed a median survival time of 21 months for the PCa patients (95%CI= 18-26). Intra group differences in patient's survival after adjusting for confounder age were confirmed through a proportional hazard model. It was observed that age, BMI, smoking, prostectomy and physical activities are associated with survival probability. It was further noticed that diabetic patients has better survival length as compared to the non-diabetic counterparts. We also developed pakprostate.com, freely available online at the URL www.pakprostate.com. During the course of study to provide users with an integrated platform to search literature, genes, drugs, to have access to download verified dataset, can upload data and also can apply basic stats and visualize data just on click of a button and to avoid hectic queries.

5.2 Future Directions

This thesis is a small contribution to science exploring the study topic from different angels. We have derived a few recommendations for future studies which are as follow,

1. On the basis of small sample size the arguments cannot be fully endorsed and large scale studies are recommended.
2. Despite of the fact that prostate gland is easily accessible to clinical and biochemical examinations, clinicians make significant delays in prostate cancer diagnosis. Early diagnostic methodologies are needed which can potentially enhance survival probabilities.
3. Analysis on the basis of a single gene or single data type to infer disease activities are unable to explore the whole system more realistically. It is therefore recommended to use integrative approaches having multiple genomic data types to enhance the accuracy the diagnostic and prognostic accuracy.
4. Multidisciplinary efforts can yield predictive and prognostic factors that make early diagnosis possible.
5. It is well established that a certain groups of cancer patients get good result from certain types of treatments while not others due to the fact that all cancers are not similar. There is a need to classify prostate cancer patients into different groups on the basis of their genetic data and each distinct group must be treated accordingly.
6. Mathematical modeling can be used to simulate the behavior of various biological systems. Lab experiments are performed in vivo or in vitro are too laborious and costly requiring long execution and preparation time. Mathematical models offer flexibility to vary the experimental parameters as frequently as needed. It is therefore recommended to develop mathematical models for prognosis and mechanism linkages.

I would like that both the patients and doctors may have both the prognostic and predictive methodologies/factors available which can better guide them what treatment option to choose according to their need and support.

Chapter 6

REFERENCES

ADZHUBEI, I. A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V. E., GERASIMOVA, A., BORK, P., KONDRASHOV, A. S. & SUNYAEV, S. R. 2010. A method and server for predicting damaging missense mutations. *Nature methods,* 7**,** 248-249.

AGALLIU, I., GERN, R., LEANZA, S. & BURK, R. D. 2009. Associations of high-grade prostate cancer with BRCA1 and BRCA2 founder mutations. *Clinical cancer research,* 15**,** 1112-1120.

AHMAD, Z., QURESHI, A., IDREES, R. & AFTAB, K. 2009. Prostatic carcinoma: a Pakistani perspective. *Asian Pacific Journal of Cancer Prevention,* 10**,** 323-324.

AIZER, A. A., CHEN, M.-H., MCCARTHY, E. P., MENDU, M. L., KOO, S., WILHITE, T. J., GRAHAM, P. L., CHOUEIRI, T. K., HOFFMAN, K. E. & MARTIN, N. E. 2013. Marital status and survival in patients with cancer. *Journal of Clinical Oncology,* 31**,** 3869-3876.

ALI, I. U., SCHRIML, L. M. & DEAN, M. 1999. Mutational spectra of PTEN/MMAC1 gene: a tumor suppressor with lipid phosphatase activity. *Journal of the national cancer institute,* 91**,** 1922-1932.

ALLOTT, E. H., MASKO, E. M. & FREEDLAND, S. J. 2013. Obesity and prostate cancer: weighing the evidence. *European urology,* 63**,** 800-809.

ANDERSSON, S.-O., BARON, J., WOLK, A., LINDGREN, C., BERGSTRÖM, R. & ADAMI, H.-O. 1995. Early life risk factors for prostate cancer: a population-based case-control study in Sweden. *Cancer Epidemiology Biomarkers & Prevention,* 4**,** 187-192.

AUNE, D., ROSENBLATT, D. A. N., CHAN, D. S., VIEIRA, A. R., VIEIRA, R., GREENWOOD, D. C., VATTEN, L. J. & NORAT, T. 2015. Dairy products, calcium, and prostate cancer risk: a systematic review and meta-analysis of cohort studies. *The American journal of clinical nutrition***,** ajcn. 067157.

AYDIN, Z., SINGH, A., BILMES, J. & NOBLE, W. S. 2011. Learning sparse models for a dynamic Bayesian network classifier of protein secondary structure. *BMC bioinformatics,* 12**,** 154.

BAINE, M., SAHAK, F., LIN, C., CHAKRABORTY, S., LYDEN, E. & BATRA, S. K. 2011. Marital status and survival in pancreatic cancer patients: a SEER based analysis. *PLoS One,* 6**,** e21052.

BAIRATI, I., LAROUCHE, R., MEYER, F., MOORE, L. & FRADET, Y. 2000. Lifetime occupational physical activity and incidental prostate cancer (Canada). *Cancer Causes & Control,* 11**,** 759-764.

BARBARIC, M., BROOKS, E., MOORE, L. & CHEIFETZ, O. 2010. Effects of physical activity on cancer survival: a systematic review. *Physiotherapy Canada,* 62**,** 25-34.

BASHIR, M. N., AHMAD, M. R. & MALIK, A. 2014. Risk factors of prostate cancer: a case-control study in Faisalabad, Pakistan. *Asian Pac J Cancer P,* 15**,** 10237-10240.

BATAGELJ, V. & MRVAR, A. 1998. Pajek-program for large network analysis. *Connections,* 21**,** 47-57.

BENSAAD, K., TSURUTA, A., SELAK, M. A., VIDAL, M. N. C., NAKANO, K., BARTRONS, R., GOTTLIEB, E. & VOUSDEN, K. H. 2006. TIGAR, a p53-inducible regulator of glycolysis and apoptosis. *Cell,* 126**,** 107-120.

BERKSON, J. 1962. Mortality and marital status. Reflections on the derivation of etiology from statistics. *American Journal of Public Health and the Nations Health,* 52**,** 1318-1329.

BHATTACHARYYA, A., EAR, U. S., KOLLER, B. H., WEICHSELBAUM, R. R. & BISHOP, D. K. 2000. The breast cancer susceptibility gene BRCA1 is required for subnuclear assembly of Rad51 and survival following treatment with the DNA cross-linking agent cisplatin. *Journal of Biological Chemistry,* 275**,** 23899-23903.

BILL-AXELSON, A., HOLMBERG, L., FILÉN, F., RUUTU, M., GARMO, H., BUSCH, C., NORDLING, S., HÄGGMAN, M., ANDERSSON, S.-O. & BRATELL, S. 2008. Radical prostatectomy versus watchful waiting in localized prostate cancer: the Scandinavian prostate cancer group-4 randomized trial. *Journal of the National Cancer Institute,* 100**,** 1144-1154.

BONNEAU, D. & LONGY, M. 2000. Mutations of the human PTEN gene. *Human mutation,* 16**,** 109.

BONOVAS, S., FILIOUSSI, K. & TSANTES, A. 2004. Diabetes mellitus and risk of prostate cancer: a meta-analysis. *Diabetologia,* 47**,** 1071-1078.

BOURDON, J.-C. 2007. p53 and its isoforms in cancer. *British journal of cancer,* 97**,** 277-282.

BOURDON, J.-C., FERNANDES, K., MURRAY-ZMIJEWSKI, F., LIU, G., DIOT, A., XIRODIMAS, D. P., SAVILLE, M. K. & LANE, D. P. 2005. p53

isoforms can regulate p53 transcriptional activity. *Genes & development,* 19**,** 2122-2137.

BROWNSON, R. C., CHANG, J. C., DAVIS, J. R. & SMITH, C. A. 1991. Physical activity on the job and cancer in Missouri. *American journal of public health,* 81**,** 639-642.

BRUSSELAERS, N., MATTSSON, F., JOHAR, A., WIKMAN, A., LAGERGREN, P., LAGERGREN, J. & LJUNG, R. 2014. Marital status and survival after oesophageal cancer surgery: a population-based nationwide cohort study in Sweden. *BMJ open,* 4**,** e005418.

CALLEWAERT, L., CHRISTIAENS, V., HAELENS, A., VERRIJDT, G., VERHOEVEN, G. & CLAESSENS, F. 2003. Implications of a polyglutamine tract in the function of the human androgen receptor. *Biochemical and biophysical research communications,* 306**,** 46-52.

CALTON, B. A., CHANG, S. C., WRIGHT, M. E., KIPNIS, V., LAWSON, K., THOMPSON, F. E., SUBAR, A. F., MOUW, T., CAMPBELL, D. S. & HURWITZ, P. 2007. History of diabetes mellitus and subsequent prostate cancer risk in the NIH-AARP Diet and Health Study. *Cancer Causes & Control,* 18**,** 493-503.

CAO, S., LIU, L., YIN, X., WANG, Y., LIU, J. & LU, Z. 2013. Coffee consumption and risk of prostate cancer: a meta-analysis of prospective cohort studies. *Carcinogenesis***,** bgt482.

CAPRIOTTI, E. & MARTI-RENOM, M. A. 2010. Quantifying the relationship between sequence and three-dimensional structure conservation in RNA. *BMC bioinformatics,* 11**,** 322.

CARTER, B., BOVA, G., BEATY, T., STEINBERG, G., CHILDS, B., ISAACS, W. & WALSH, P. 1993. Hereditary prostate cancer: epidemiologic and clinical features. *J Urol,* 150**,** 797-802.

CERHAN, J. R., TORNER, J. C., LYNCH, C. F., RUBENSTEIN, L. M., LEMKE, J. H., COHEN, M. B., LUBAROFF, D. M. & WALLACE, R. B. 1997. Association of smoking, body mass, and physical activity with risk of prostate cancer in the Iowa 65+ Rural Health Study (United States). *Cancer Causes & Control,* 8**,** 229-238.

CHEN, X., XU, H., YUAN, P., FANG, F., HUSS, M., VEGA, V. B., WONG, E., ORLOV, Y. L., ZHANG, W. & JIANG, J. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell,* 133**,** 1106-1117.

CHENG, J., RANDALL, A. Z., SWEREDOSKI, M. J. & BALDI, P. 2005. SCRATCH: a protein structure and structural feature prediction server. *Nucleic acids research,* 33**,** W72-W76.

CHETRAM, M. A., DON-SALU-HEWAGE, A. S. & HINTON, C. V. 2011. ROS enhances CXCR4-mediated functions through inactivation of PTEN in prostate cancer cells. *Biochemical and biophysical research communications,* 410**,** 195-200.

CHI, S.-G., DEVERE WHITE, R. W., MEYERS, F. J., SIDERS, D. B., LEE, F. & GUMERLOCK, P. H. 1994. p53 in prostate cancer: frequent expressed transition mutations. *Journal of the National Cancer Institute,* 86**,** 926-933.

CHOI, Y., SIMS, G. E., MURPHY, S., MILLER, J. R. & CHAN, A. P. 2012. Predicting the functional effect of amino acid substitutions and indels.

CHOSDOL, K., AHUJA, A., RATHORE, A., MISRA, A., MIR, S. H., CHATTOPADHYAY, P. & SINHA, S. 2002. Study of p53 codon 72 polymorphism in various ethnic groups of North India. *Curr Sci,* 82**,** 1253-1255.

CHOURA, M. & REBAI, A. 2011. Structural analysis of hubs in human NR-RTK network. *Biology direct,* 6**,** 49.

CLARKE, G. & WHITTEMORE, A. S. 2000. Prostate cancer risk in relation to anthropometry and physical activity: the National Health and Nutrition Examination Survey I Epidemiological Follow-Up Study. *Cancer Epidemiology Biomarkers & Prevention,* 9**,** 875-881.

COLDITZ, G. 1996. Consensus conference: smoking and prostate cancer. *Cancer Causes and Control,* 7**,** 560-562.

COOPERBERG, M. R., ODISHO, A. Y. & CARROLL, P. R. 2012. Outcomes for radical prostatectomy: is it the singer, the song, or both? *Journal of Clinical Oncology,* 30**,** 476-478.

COURNEYA, K. S. & FRIEDENREICH, C. M. Physical activity and cancer control. Seminars in oncology nursing, 2007. Elsevier, 242-252.

DAVIS, M., LANE, J., GILBERT, R., MARTIN, R., DONOVAN, J., NEAL, D. & HAMDY, F. 2008. Raised blood pressure predicts prostate cancer and elevated PSA in a large unselected population of men participating in the ProtecT study. *European Urology Supplements,* 7**,** 198.

DAVIS, M. K., RAJALA, J. L., TYLDESLEY, S., PICKLES, T. & VIRANI, S. A. 2015. The Prevalence of Cardiac Risk Factors in Men with Localized Prostate Cancer Undergoing Androgen Deprivation Therapy in British Columbia, Canada. *Journal of Oncology*.

DE MUGA, S., HERNÁNDEZ, S., AGELL, L., SALIDO, M., JUANPERE, N., LORENZO, M., LORENTE, J. A., SERRANO, S. & LLORETA, J. 2010. Molecular alterations of EGFR and PTEN in prostate cancer: association with high-grade and advanced-stage carcinomas. *Modern Pathology,* 23**,** 703-712.

DENU, J. M., ZHOU, G., GUO, Y. & DIXON, J. E. 1995. The catalytic role of aspartic acid-92 in the human dual-specific protein-tyrosine-phosphatase vaccinia h1-related. *Biochemistry,* 34**,** 3396-3403.

DISCACCIATI, A., ORSINI, N. & WOLK, A. 2012. Body mass index and incidence of localized and advanced prostate cancer—a dose–response meta-analysis of prospective studies. *Annals of Oncology***,** mdr603.

DONG, J.-T., LI, C.-L., SIPE, T. W. & FRIERSON, H. F. 2001. Mutations of PTEN/MMAC1 in primary prostate cancers from Chinese patients. *Clinical cancer research,* 7**,** 304-308.

DOSEMECI, M., HAYES, R. B., VETTER, R., HOOVER, R. N., TUCKER, M., ENGIN, K., UNSAL, M. & BLAIR, A. 1993. Occupational physical activity, socioeconomic status, and risks of 15 cancer sites in Turkey. *Cancer Causes & Control,* 4**,** 313-321.

DOWNING, S., RUSSELL, P. & JACKSON, P. 2003. Alterations of p53 are common in early stage prostate cancer. *The Canadian journal of urology,* 10**,** 1924-1933.

DUMAZ, N., MILNE, D., JARDINE, L. & MEEK, D. 2001. Critical roles for the serine 20, but not the serine 15, phosphorylation site and for the polyproline domain in regulating p53 turnover. *Biochem. J,* 359**,** 459-464.

DUMONT, P., LEU, J.-J., DELLA PIETRA, A. C., GEORGE, D. L. & MURPHY, M. 2003. The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. *Nature genetics,* 33**,** 357-365.

ECKE, T. H., SCHLECHTE, H. H., SCHIEMENZ, K., SACHS, M. D., LENK, S. V., RUDOLPH, B. D. & LOENING, S. A. 2010. TP53 gene mutations in prostate cancer progression. *Anticancer research,* 30**,** 1579-1586.

ELSHAMY, W. M. & DUHE, R. J. 2013. Overview: cellular plasticity, cancer stem cells and metastasis. *Cancer letters,* 341**,** 2-8.

ENG, C. 2003. PTEN: one gene, many syndromes. *Human mutation,* 22**,** 183-198.

ENGELAND, A., TRETLI, S. & BJØRGE, T. 2003. Height, body mass index, and prostate cancer: a follow-up of 950 000 Norwegian men. *British journal of cancer,* 89**,** 1237-1242.

EVERHART, J. & WRIGHT, D. 1995. Diabetes mellitus as a risk factor for pancreatic cancer: a meta-analysis. *Jama,* 273**,** 1605-1609.

FITZPATRICK, A. L., DALING, J. R., FURBERG, C. D., KRONMAL, R. A. & WEISSFELD, J. L. 2001. Hypertension, heart rate, use of antihypertensives, and incident prostate cancer. *Annals of epidemiology,* 11**,** 534-542.

FONG, D. Y., HO, J. W., HUI, B. P., LEE, A. M., MACFARLANE, D. J., LEUNG, S. S., CERIN, E., CHAN, W. Y., LEUNG, I. P. & LAM, S. H. 2012. Physical

activity for cancer survivors: meta-analysis of randomised controlled trials. *Bmj,* 344**,** e70.

FRIEDENREICH, C. M. & ORENSTEIN, M. R. 2002. Physical activity and cancer prevention: etiologic evidence and biological mechanisms. *The Journal of nutrition,* 132**,** 3456S-3464S.

FRIEDMAN, G. D. 1997. Blood pressure and heart rate: no evidence for a positive association with prostate cancer. *Annals of epidemiology,* 7**,** 486-489.

GADDIPATI, J. P., MCLEOD, D. G., HEIDENBERG, H. B., SESTERHENN, I. A., FINGER, M. J., MOUL, J. W. & SRIVASTAVA, S. 1994. Frequent detection of codon 877 mutation in the androgen receptor gene in advanced prostate cancers. *Cancer research,* 54**,** 2861-2864.

GANN, P. H. 2002. Risk factors for prostate cancer. *Reviews in urology,* 4**,** S3.

GEORGESCU, M.-M., KIRSCH, K. H., KALOUDIS, P., YANG, H., PAVLETICH, N. P. & HANAFUSA, H. 2000. Stabilization and productive positioning roles of the C2 domain of PTEN tumor suppressor. *Cancer research,* 60**,** 7033-7038.

GEYER, R. K., ZHONG, K. Y. & MAKI, C. G. 2000. The MDM2 RING-finger domain is required to promote p53 nuclear export. *Nature cell biology,* 2**,** 569-573.

GIOVANNUCCI, E., HARLAN, D. M., ARCHER, M. C., BERGENSTAL, R. M., GAPSTUR, S. M., HABEL, L. A., POLLAK, M., REGENSTEINER, J. G. & YEE, D. 2010. Diabetes and cancer: a consensus report. *CA: a cancer journal for clinicians,* 60**,** 207-221.

GIOVANNUCCI, E., RIMM, E. B., STAMPFER, M. J., COLDITZ, G. A. & WILLETT, W. C. 1998. Diabetes mellitus and risk of prostate cancer (United States). *Cancer Causes & Control,* 9**,** 3-9.

GLINSKY, G. V., GLINSKII, A. B., STEPHENSON, A. J., HOFFMAN, R. M. & GERALD, W. L. 2004. Gene expression profiling predicts clinical outcome of prostate cancer. *Journal of Clinical Investigation,* 113**,** 913.

GOLDZWEIG, G., ANDRITSCH, E., HUBERT, A., BRENNER, B., WALACH, N., PERRY, S. & BAIDER, L. 2009. Psychological distress among male patients and male spouses: what do oncologists need to know? *Annals of Oncology***,** mdp398.

GONG, Z., NEUHOUSER, M. L., GOODMAN, P. J., ALBANES, D., CHI, C., HSING, A. W., LIPPMAN, S. M., PLATZ, E. A., POLLAK, M. N. & THOMPSON, I. M. 2006. Obesity, diabetes, and risk of prostate cancer: results from the prostate cancer prevention trial. *Cancer Epidemiology Biomarkers & Prevention,* 15**,** 1977-1983.

GONI, J., ESTEBAN, F. J., DE MENDIZABAL, N. V., SEPULCRE, J., ARDANZA-TREVIJANO, S., AGIRREZABAL, I. & VILLOSLADA, P. 2008. A computational analysis of protein-protein interaction networks in neurodegenerative diseases. *BMC systems biology,* 2**,** 52.

GROSSMAN, E., MESSERLI, F. H., BOYKO, V. & GOLDBOURT, U. 2002. Is there an association between hypertension and cancer mortality? *The American journal of medicine,* 112**,** 479-486.

GSUR, A., PREYER, M., HAIDINGER, G., ZIDEK, T., MADERSBACHER, S., SCHATZL, G., MARBERGER, M., VUTUC, C. & MICKSCHE, M. 2002. Polymorphic CAG repeats in the androgen receptor gene, prostate-specific antigen polymorphism and prostate cancer risk. *Carcinogenesis,* 23**,** 1647-1651.

GU, M., DONG, X., ZHANG, X. & NIU, W. 2012. The CAG repeat polymorphism of androgen receptor gene and prostate cancer: a meta-analysis. *Molecular biology reports,* 39**,** 2615-2624.

GUO, Z., YANG, X., SUN, F., JIANG, R., LINN, D. E., CHEN, H., CHEN, H., KONG, X., MELAMED, J. & TEPPER, C. G. 2009. A novel androgen receptor splice variant is up-regulated during prostate cancer progression and promotes androgen depletion–resistant growth. *Cancer research,* 69**,** 2305-2313.

HAINAUT, P. & HOLLSTEIN, M. 1999. p 53 and Human Cancer: The First Ten Thousand Mutations. *Advances in cancer research,* 77**,** 81-137.

HAN, J.-D. J., BERTIN, N., HAO, T., GOLDBERG, D. S., BERRIZ, G. F., ZHANG, L. V., DUPUY, D., WALHOUT, A. J., CUSICK, M. E. & ROTH, F. P. 2004. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature,* 430**,** 88-93.

HARTMAN, T. J., ALBANES, D., RAUTALAHTI, M., TANGREA, J. A., VIRTAMO, J., STOLZENBERG, R. & TAYLOR, P. R. 1998. Physical activity and prostate cancer in the alpha-tocopherol, beta-carotene (ATBC) cancer prevention study (Finland). *Cancer Causes & Control,* 9**,** 11-18.

HASHIMOTO, Y., LOFTIS, D. W. & ADAMS, J. C. 2009. Fascin-1 promoter activity is regulated by CREB and the aryl hydrocarbon receptor in human carcinoma cells. *PLoS One,* 4**,** e5130.

HEINZER, H. & STEUBER, T. Prostate cancer in the elderly. Urologic Oncology: Seminars and Original Investigations, 2009. Elsevier, 668-672.

HICKEY, K., DO, K.-A. & GREEN, A. 2001. Smoking and prostate cancer. *Epidemiologic reviews,* 23**,** 115-125.

HOLLSTEIN, M., RICE, K., GREENBLATT, M., SOUSSI, T., FUCHS, R., SØRLIE, T., HOVIG, E., SMITH-SØRENSEN, B., MONTESANO, R. &

HARRIS, C. 1994. Database of p53 gene somatic mutations in human tumors and cell lines. *Nucleic acids research,* 22**,** 3551.

HSING, A. W., MCLAUGHLIN, J. K., ZHENG, W., GAO, Y.-T. & BLOT, W. J. 1994. Occupation, physical activity, and risk of prostate cancer in Shanghai, People's Republic of China. *Cancer Causes & Control,* 5**,** 136-140.

HSING, A. W., TSAO, L. & DEVESA, S. S. 2000. International trends and patterns of prostate cancer incidence and mortality. *International Journal of Cancer,* 85**,** 60-67.

HU, S.-Y., LIU, T., LIU, Z.-Z., LEDET, E., VELASCO-GONZALEZ, C., MANDAL, D. M. & KOOCHEKPOUR, S. 2010. Identification of a novel germline missense mutation of the androgen receptor in African American men with familial prostate cancer. *Asian journal of andrology,* 12**,** 336.

HUNCHAREK, M., HADDOCK, K. S., REID, R. & KUPELNICK, B. 2010. Smoking as a risk factor for prostate cancer: a meta-analysis of 24 prospective cohort studies. *American journal of public health,* 100**,** 693.

HUTHMACHER, C., GILLE, C. & HOLZHUTTER, H.-G. Computational analysis of protein-protein interactions in metabolic networks of Escherichia coli and yeast. Genome Inform, 2007. World Scientific, 162-172.

INVERSO, G., MAHAL, B. A., AIZER, A. A., DONOFF, R. B., CHAU, N. G. & HADDAD, R. I. 2014. Marital status and head and neck cancer outcomes. *Cancer*.

ISLAMI, F., MOREIRA, D. M., BOFFETTA, P. & FREEDLAND, S. J. 2014. A systematic review and meta-analysis of tobacco use and prostate cancer mortality and incidence in prospective cohort studies. *European urology,* 66**,** 1054-1064.

ISMAEEL, A. G. 2013. New Approach for Prediction Pre-cancer via Detecting Mutated in Tumor Protein P53. *arXiv preprint arXiv:1310.2182*.

ITO, K. 2014. Prostate cancer in Asian men. *Nature Reviews Urology*.

JAYARAMAN, A., JAMIL, K. & RAJU, S. 2011. The interaction of p53 and MDM2 genes in cancers, in silico studies and phylogenetic analysis. *Biology & Medicine,* 3.

JEMAL, A., BRAY, F., CENTER, M. M., FERLAY, J., WARD, E. & FORMAN, D. 2011. Global cancer statistics. *CA: a cancer journal for clinicians,* 61**,** 69-90.

JEMAL, A., MURRAY, T., WARD, E., SAMUELS, A., TIWARI, R. C., GHAFOOR, A., FEUER, E. J. & THUN, M. J. 2005. Cancer statistics, 2005. *CA: a cancer journal for clinicians,* 55**,** 10-30.

JEMAL, A., SIEGEL, R., XU, J. & WARD, E. 2010. Cancer statistics, 2010. *CA: a cancer journal for clinicians,* 60**,** 277-300.

JENSEN, L. J., KUHN, M., STARK, M., CHAFFRON, S., CREEVEY, C., MULLER, J., DOERKS, T., JULIEN, P., ROTH, A. & SIMONOVIC, M. 2009. STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic acids research,* 37**,** D412-D416.

JEONG, H., MASON, S. P., BARABÁSI, A.-L. & OLTVAI, Z. N. 2001. Lethality and centrality in protein networks. *Nature,* 411**,** 41-42.

JIA, Z., BARFORD, D., FLINT, A. J. & TONKS, N. K. 1995. Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. *Science,* 268**,** 1754-1758.

JIMENEZ, G. S., KHAN, S. H., STOMMEL, J. M. & WAHL, G. M. 1999. p53 regulation by post-translational modification and nuclear retention in response to diverse stresses. *Oncogene,* 18**,** 7656-7665.

KAISER, N. C., HARTOONIAN, N. & OWEN, J. E. 2010. Toward a cancer-specific model of psychological distress: population data from the 2003–2005 National Health Interview Surveys. *Journal of Cancer Survivorship,* 4**,** 291-302.

KASPER, J. S., LIU, Y. & GIOVANNUCCI, E. 2009. Diabetes mellitus and risk of prostate cancer in the health professionals follow-up study. *International Journal of Cancer,* 124**,** 1398-1403.

KATO, H., KATO, S., KUMABE, T., SONODA, Y., YOSHIMOTO, T., KATO, S., HAN, S.-Y., SUZUKI, T., SHIBATA, H. & KANAMARU, R. 2000. Functional evaluation of p53 and PTEN gene mutations in gliomas. *Clinical Cancer Research,* 6**,** 3937-3943.

KAVOUSSI, L. R. & WEIN, A. J. 2007. *Campbell-Walsh Urology,* Saunders Elsevier.

KENIRY, M. & PARSONS, R. 2008. The role of PTEN signaling perturbations in cancer and in targeted therapy. *Oncogene,* 27**,** 5477-5485.

KHAN, M. H., RASHID, H. & MIR, A. 2013. Phylogenetic analysis of human Tp53 gene using computational approach. *African Journal of Biotechnology,* 10**,** 344-349.

KHOO, K. H., ANDREEVA, A. & FERSHT, A. R. 2009. Adaptive evolution of p53 thermodynamic stability. *Journal of molecular biology,* 393**,** 161-175.

KIM, K. K. & KIM, H. B. 2009. Protein interaction network related to Helicobacter pylori infection response. *World journal of gastroenterology: WJG,* 15**,** 4518.

KIRAGA, J., MACKIEWICZ, P., MACKIEWICZ, D., KOWALCZUK, M., BIECEK, P., POLAK, N., SMOLARCZYK, K., DUDEK, M. R. & CEBRAT, S. 2007. The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. *BMC genomics,* 8**,** 163.

KO, C. & CHAUDHRY, S. 2002. The need for a multidisciplinary approach to cancer care. *Journal of Surgical Research,* 105**,** 53-57.

KOOCHEKPOUR, S., BUCKLES, E., SHOURIDEH, M., HU, S., CHANDRA, D., ZABALETA, J. & ATTWOOD, K. 2014. Androgen Receptor Mutations and Polymorphisms in African American Prostate Cancer. *International journal of biological sciences,* 10**,** 643.

KUDO, N., WOLFF, B., SEKIMOTO, T., SCHREINER, E. P., YONEDA, Y., YANAGIDA, M., HORINOUCHI, S. & YOSHIDA, M. 1998. Leptomycin B inhibition of signal-mediated nuclear export by direct binding to CRM1. *Experimental cell research,* 242**,** 540-547.

KUSSIE, P. H., GORINA, S., MARECHAL, V., ELENBAAS, B., MOREAU, J., LEVINE, A. J. & PAVLETICH, N. P. 1996. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science,* 274**,** 948-953.

LACEY, J. V., KREIMER, A. R., BUYS, S. S., MARCUS, P. M., CHANG, S.-C., LEITZMANN, M. F., HOOVER, R. N., PROROK, P. C., BERG, C. D. & HARTGE, P. 2009. Breast cancer epidemiology according to recognized breast cancer risk factors in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial Cohort. *BMC cancer,* 9**,** 84.

LAPOINTE, J., LI, C., GIACOMINI, C. P., SALARI, K., HUANG, S., WANG, P., FERRARI, M., HERNANDEZ-BOUSSARD, T., BROOKS, J. D. & POLLACK, J. R. 2007. Genomic profiling reveals alternative genetic pathways of prostate tumorigenesis. *Cancer research,* 67**,** 8504-8510.

LATULIPPE, E., SATAGOPAN, J., SMITH, A., SCHER, H., SCARDINO, P., REUTER, V. & GERALD, W. L. 2002. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer research,* 62**,** 4499-4506.

LEE, I.-M., SESSO, H. D. & PAFFENBARGER, R. S. 2001. A prospective cohort study of physical activity and body size in relation to prostate cancer risk (United States). *Cancer Causes & Control,* 12**,** 187-193.

LEE, J.-O., YANG, H., GEORGESCU, M.-M., DI CRISTOFANO, A., MAEHAMA, T., SHI, Y., DIXON, J. E., PANDOLFI, P. & PAVLETICH, N. P. 1999. Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association. *Cell,* 99**,** 323-334.

LEONGAMORNLERT, D., MAHMUD, N., TYMRAKIEWICZ, M., SAUNDERS, E., DADAEV, T., CASTRO, E., GOH, C., GOVINDASAMI, K., GUY, M. & O'BRIEN, L. 2012. Germline BRCA1 mutations increase prostate cancer risk. *British journal of cancer,* 106**,** 1697-1701.

LESLIE, N. & DOWNES, C. 2004. PTEN function: how normal cells control it and tumour cells lose it. *Biochem. J,* 382**,** 1-11.

LESLIE, N. R. & FOTI, M. 2011. Non-genomic loss of PTEN function in cancer: not in my genes. *Trends in pharmacological sciences,* 32**,** 131-140.

LI, M., ZHANG, H., WANG, J.-X. & PAN, Y. 2012. A new essential protein discovery method based on the integration of protein-protein interaction and gene expression data. *BMC systems biology,* 6**,** 15.

LIANG, S. H. & CLARKE, M. F. 2001. Regulation of p53 localization. *European journal of biochemistry,* 268**,** 2779-2783.

LIMA-MENDEZ, G. & VAN HELDEN, J. 2009. The powerful law of the power law and other myths in network biology. *Molecular BioSystems,* 5**,** 1482-1493.

LIU, S., LEE, I.-M., LINSON, P., AJANI, U., BURING, J. E. & HENNEKENS, C. H. 2000. A prospective study of physical activity and risk of prostate cancer in US physicians. *International Journal of Epidemiology,* 29**,** 29-35.

LOHRUM, M. A., WOODS, D. B., LUDWIG, R. L., BÁLINT, É. & VOUSDEN, K. H. 2001. C-terminal ubiquitination of p53 contributes to nuclear export. *Molecular and cellular biology,* 21**,** 8521-8532.

LONERGAN, P. E. & TINDALL, D. J. 2011. Androgen receptor signaling in prostate cancer development and progression. *Journal of carcinogenesis,* 10**,** 20.

LU, Y., LIU, Y., ZENG, J., HE, Y., PENG, Q., DENG, Y., WANG, J., XIE, L., LI, T. & QIN, X. 2014. Association of p53 codon 72 polymorphism with prostate cancer: an update meta-analysis. *Tumor Biology,* 35**,** 3997-4005.

MACINNIS, R. J., ENGLISH, D. R., GERTIG, D. M., HOPPER, J. L. & GILES, G. G. 2003. Body size and composition and prostate cancer risk. *Cancer Epidemiology Biomarkers & Prevention,* 12**,** 1417-1421.

MAEHAMA, T. & DIXON, J. E. 1998. The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3, 4, 5-trisphosphate. *Journal of Biological Chemistry,* 273**,** 13375-13378.

MAHAN, L. K. & ESCOTT-STUMP, S. 2004. Krause's food, nutrition, & diet therapy.

MANDAIR, D., ROSSI, R. E., PERICLEOUS, M., WHYAND, T. & CAPLIN, M. E. 2014. Prostate cancer and the influence of dietary factors and supplements: a systematic review. *Nutr. Metab.(Lond.),* 11.

MARTIN, P., LIU, Y.-N., PIERCE, R., ABOU-KHEIR, W., CASEY, O., SENG, V., CAMACHO, D., SIMPSON, R. M. & KELLY, K. 2011. Prostate epithelial Pten/TP53 loss leads to transformation of multipotential progenitors and epithelial to mesenchymal transition. *The American journal of pathology,* 179**,** 422-435.

MICHAEL, D. & OREN, M. The p53–Mdm2 module and the ubiquitin system. Seminars in cancer biology, 2003. Elsevier, 49-58.

MURPHY, M. 2006. Polymorphic variants in the p53 pathway. *Cell Death & Differentiation,* 13**,** 916-920.

NAVONE, N. M., LABATE, M. E., TRONCOSO, P., PISTERS, L. L., CONTI, C. J., VON ESCHENBACH, A. C. & LOGOTHETIS, C. J. 1999. p53 mutations in prostate cancer bone metastases suggest that selected p53 mutants in the primary site define foci with metastatic potential. *The Journal of urology,* 161**,** 304-308.

NEWMARK, J. R., HARDY, D. O., TONB, D. C., CARTER, B. S., EPSTEIN, J. I., ISAACS, W. B., BROWN, T. R. & BARRACK, E. R. 1992. Androgen receptor gene mutations in human prostate cancer. *Proceedings of the National Academy of Sciences,* 89**,** 6319-6323.

NG, P. C. & HENIKOFF, S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research,* 31**,** 3812-3814.

NGUYEN, D., JOSHI-DATAR, A., LEPINE, F., BAUERLE, E., OLAKANMI, O., BEER, K., MCKAY, G., SIEHNEL, R., SCHAFHAUSER, J. & WANG, Y. 2011. Active starvation responses mediate antibiotic tolerance in biofilms and nutrient-limited bacteria. *Science,* 334**,** 982-986.

OLIVERIA, S. A., KOHL 3RD, H., TRICHOPOULOS, D. & BLAIR, S. N. 1996. The association between cardiorespiratory fitness and prostate cancer. *Medicine and science in sports and exercise,* 28**,** 97-104.

OSBORNE, C., OSTIR, G. V., DU, X., PEEK, M. K. & GOODWIN, J. S. 2005. The influence of marital status on the stage at diagnosis, treatment, and survival of older women with breast cancer. *Breast cancer research and treatment,* 93**,** 41-47.

OSMAN, I., DROBNJAK, M., FAZZARI, M., FERRARA, J., SCHER, H. I. & CORDON-CARDO, C. 1999. Inactivation of the p53 pathway in prostate cancer: impact on tumor progression. *Clinical cancer research,* 5**,** 2082-2088.

PAFFENBARGER, R. S., HYDE, R. & WING, A. 1987. Physical activity and incidence of cancer in diverse populations: a preliminary report. *The American journal of clinical nutrition,* 45**,** 312-317.

PATTIN, K. A. & MOORE, J. H. 2009. Role for protein–protein interaction databases in human genetics. *Expert review of proteomics,* 6**,** 647-659.

PENNEFATHER, J., LAU, W., MITCHELSON, F. & VENTURA, S. 2000. The autonomic and sensory innervation of the smooth muscle of the prostate gland: a review of pharmacological and histological studies. *Journal of autonomic pharmacology,* 20**,** 193-206.

PETTERSSON, A., KASPERZYK, J. L., KENFIELD, S. A., RICHMAN, E. L., CHAN, J. M., WILLETT, W. C., STAMPFER, M. J., MUCCI, L. A. & GIOVANNUCCI, E. L. 2012. Milk and dairy consumption among men with prostate cancer and risk of metastases and prostate cancer death. *Cancer Epidemiology Biomarkers & Prevention,* 21**,** 428-436.

PEZZOLESI, M. G., ZBUK, K. M., WAITE, K. A. & ENG, C. 2007. Comparative genomic and functional analyses reveal a novel cis-acting PTEN regulatory element as a highly conserved functional E-box motif deleted in Cowden syndrome. *Human molecular genetics,* 16**,** 1058-1071.

PINTUS, S., FOMIN, E., IVANISENKO, V. & KOLCHANOV, N. 2006. Phylogenetic analysis of the p53 family. *Biophysics,* 51**,** 571-580.

PLANCHON, S. M., WAITE, K. A. & ENG, C. 2008. The nuclear affairs of PTEN. *Journal of cell science,* 121**,** 249-253.

POURMAND, G., ZIAEE, A.-A., SALEM, S., ABEDI, A. R., MEHRSAI, A., ALAVI, H. A., AHMADI, A. & SAADATI, H. R. 2009. Role of PTEN gene in progression of prostate cancer. *Urology journal,* 4**,** 95-100.

POWELL, I. J. 2011. The precise role of ethnicity and family history on aggressive prostate cancer: a review analysis. *Archivos espanoles de urologia,* 64**,** 711.

PRINS, G. S., CHANG, W. Y., WANG, Y. & VAN BREEMEN, R. B. 2002. Retinoic acid receptors and retinoids are up-regulated in the developing and adult rat prostate by neonatal estrogen exposure. *Endocrinology,* 143**,** 3628-3640.

PRINS, G. S. & PUTZ, O. 2008. Molecular signaling pathways that regulate prostate gland development. *Differentiation,* 76**,** 641-659.

PUKKALA, E., KAPRIO, J., KOSKENVUO, M., KUJALA, U. & SARNA, S. 2000. Cancer incidence among Finnish world class male athletes. *International journal of sports medicine,* 21**,** 216-220.

QIN, L.-Q., XU, J.-Y., WANG, P.-Y., KANEKO, T., HOSHI, K. & SATO, A. 2004. Milk consumption is a risk factor for prostate cancer: meta-analysis of case-control studies. *Nutrition and cancer,* 48**,** 22-27.

REISMAN, D., GREENBERG, M. & ROTTER, V. 1988. Human p53 oncogene contains one promoter upstream of exon 1 and a second, stronger promoter within intron 1. *Proceedings of the National Academy of Sciences,* 85**,** 5146-5150.

RENDALL, M. S., WEDEN, M. M., FAVREAULT, M. M. & WALDRON, H. 2011. The protective effect of marriage for survival: a review and update. *Demography,* 48**,** 481-506.

RENEHAN, A. G., TYSON, M., EGGER, M., HELLER, R. F. & ZWAHLEN, M. 2008. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *The Lancet,* 371**,** 569-578.

REVA, B., ANTIPIN, Y. & SANDER, C. 2011. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic acids research*, gkr407.

RICKS-SANTI, L., APPREY, V., MASON, T., WILSON, B., ABBAS, M., HERNANDEZ, W., HOOKER, S., DOURA, M., BONNEY, G. & DUNSTON, G. 2012. Identification of genetic risk associated with prostate cancer using ancestry informative markers. *Prostate cancer and prostatic diseases,* 15**,** 359-364.

RIES, L. A. G., REICHMAN, M. E., LEWIS, D. R., HANKEY, B. F. & EDWARDS, B. K. 2003. Cancer survival and incidence from the Surveillance, Epidemiology, and End Results (SEER) program. *The oncologist,* 8**,** 541-552.

RIZO, J. & SÜDHOF, T. C. 1998. C2-domains, structure and function of a universal Ca2+-binding domain. *Journal of Biological Chemistry,* 273**,** 15879-15882.

RODRÍGUEZ-ESCUDERO, I., OLIVER, M. D., ANDRÉS-PONS, A., MOLINA, M., CID, V. J. & PULIDO, R. 2011. A comprehensive functional analysis of PTEN mutations: implications in tumor-and autism-related syndromes. *Human molecular genetics,* 20**,** 4132-4142.

RODRIGUEZ, C., PATEL, A. V., CALLE, E. E., JACOBS, E. J., CHAO, A. & THUN, M. J. 2001. Body mass index, height, and prostate cancer mortality in two large cohorts of adult men in the United States. *Cancer Epidemiology Biomarkers & Prevention,* 10**,** 345-353.

RODRIGUEZ, C., PATEL, A. V., MONDUL, A. M., JACOBS, E. J., THUN, M. J. & CALLE, E. E. 2005. Diabetes and risk of prostate cancer in a prospective cohort of US men. *American journal of epidemiology,* 161**,** 147-152.

ROSENBLATT, K. A., WICKLUND, K. G. & STANFORD, J. L. 2001. Sexual factors and the risk of prostate cancer. *American journal of epidemiology,* 153**,** 1152-1158.

ROUDIER, M. P., TRUE, L. D., HIGANO, C. S., VESSELLE, H., ELLIS, W., LANGE, P. & VESSELLA, R. L. 2003. Phenotypic heterogeneity of end-stage prostate carcinoma metastatic to bone. *Human pathology,* 34**,** 646-653.

RUNDQVIST, H., AUGSTEN, M., STRÖMBERG, A., RULLMAN, E., MIJWEL, S., KHARAZIHA, P., PANARETAKIS, T., GUSTAFSSON, T. & ÖSTMAN, A. 2013. Effect of acute exercise on prostate cancer cell growth. *PLoS One,* 8**,** e67579.

SALMENA, L., CARRACEDO, A. & PANDOLFI, P. P. 2008. Tenets of PTEN tumor suppression. *Cell,* 133**,** 403-414.

SAMBROOK, J. & RUSSELL, D. 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, New York.

SARKAR, C., MUKHOPADHYAY, S. & SHARMA, M. C. 2002. p53 in Brain Tumors: Basic science illuminates clinical oncology. *Indian Journal of Human Genetics,* 8**,** 52.

SCHAEFFER, E. M., CARTER, H. B., KETTERMANN, A., LOEB, S., FERRUCCI, L., LANDIS, P., TROCK, B. J. & METTER, E. J. 2009. Prostate specific antigen testing among the elderly—when to stop? *The Journal of urology,* 181**,** 1606-1614.

SCHATZL, G., MARBERGER, M., REMZI, M., GRÖSSER, P., UNTERLECHNER, J., HAIDINGER, G., ZIDEK, T., PREYER, M., MICKSCHE, M. & GSUR, A. 2005. Polymorphism in ARE-I region of prostate-specific antigen gene associated with low serum testosterone level and high-grade prostate cancer. *Urology,* 65**,** 1141-1145.

SCHER, H. I. & HELLER, G. 2000. Clinical states in prostate cancer: toward a dynamic model of disease progression. *Urology,* 55**,** 323-327.

SCHLOMM, T., IWERS, L., KIRSTEIN, P., JESSEN, B., KÖLLERMANN, J., MINNER, S., PASSOW-DROLET, A., MIRLACHER, M., MILDE-LANGOSCH, K. & GRAEFEN, M. 2008. Clinical significance of p53 alterations in surgically treated prostate cancers. *Modern Pathology,* 21**,** 1371-1378.

SCHREIBER, F., PATRICIO, M., MUFFATO, M., PIGNATELLI, M. & BATEMAN, A. 2013. TreeFam v9: a new website, more species and orthology-on-the-fly. *Nucleic acids research***,** gkt1055.

SECRETAN, B., STRAIF, K., BAAN, R., GROSSE, Y., EL GHISSASSI, F., BOUVARD, V., BENBRAHIM-TALLAA, L., GUHA, N., FREEMAN, C. & GALICHET, L. 2009. A review of human carcinogens—Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. *The lancet oncology,* 10**,** 1033-1034.

SEVERSON, R. K., NOMURA, A. M., GROVE, J. S. & STEMMERMANN, G. N. 1989. A prospective analysis of physical activity and cancer. *American journal of epidemiology,* 130**,** 522-529.

SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. & IDEKER, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research,* 13**,** 2498-2504.

SHI, D. & GU, W. 2012. Dual roles of MDM2 in the regulation of p53 ubiquitination dependent and ubiquitination independent mechanisms of MDM2 repression of p53 activity. *Genes & cancer,* 3**,** 240-248.

SINGH, D., FEBBO, P. G., ROSS, K., JACKSON, D. G., MANOLA, J., LADD, C., TAMAYO, P., RENSHAW, A. A., D'AMICO, A. V. & RICHIE, J. P. 2002. Gene expression correlates of clinical prostate cancer behavior. *Cancer cell,* 1**,** 203-209.

SIRCAR, K., YOSHIMOTO, M., MONZON, F. A., KOUMAKPAYI, I. H., KATZ, R. L., KHANNA, A., ALVAREZ, K., CHEN, G., DARNEL, A. D. & APRIKIAN, A. G. 2009. PTEN genomic deletion is associated with p-Akt and AR signalling in poorer outcome, hormone refractory prostate cancer. *The Journal of pathology,* 218**,** 505-513.

SIVONOVA, M. K., VILCKOVA, M., KLIMENT, J., MAHMOOD, S., JURECEKOVA, J., DUSENKOVA, S., WACZULÍKOVA, I., SLEZAK, P. & DOBROTA, D. 2015. Association of p53 and p21 polymorphisms with prostate cancer. *Biomedical reports,* 3**,** 707-714.

SOLIMAN, S., ARONSON, W. J. & BARNARD, R. J. 2011. Analyzing serum-stimulated prostate cancer cell lines after low-fat, high-fiber diet and exercise intervention. *Evidence-Based Complementary and Alternative Medicine,* 2011.

SONG, M. S., CARRACEDO, A., SALMENA, L., SONG, S. J., EGIA, A., MALUMBRES, M. & PANDOLFI, P. P. 2011. Nuclear PTEN regulates the APC-CDH1 tumor-suppressive complex in a phosphatase-independent manner. *Cell,* 144**,** 187-199.

STAMBOLIC, V., SUZUKI, A., DE LA POMPA, J. L., BROTHERS, G. M., MIRTSOS, C., SASAKI, T., RULAND, J., PENNINGER, J. M., SIDEROVSKI, D. P. & MAK, T. W. 1998. Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell,* 95**,** 29-39.

STAMEY, T. A., YANG, N., HAY, A. R., MCNEAL, J. E., FREIHA, F. S. & REDWINE, E. 1987. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. *New England Journal of Medicine,* 317**,** 909-916.

STECK, P. A., PERSHOUSE, M. A., JASSER, S. A., YUNG, W. A., LIN, H., LIGON, A. H., LANGFORD, L. A., BAUMGARD, M. L., HATTIER, T. & DAVIS, T. 1997. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23. 3 that is mutated in multiple advanced cancers. *Nature genetics,* 15**,** 356-362.

STEENLAND, K., NOWLIN, S. & PALU, S. 1995. Cancer incidence in the National Health and Nutrition Survey I. Follow-up data: diabetes, cholesterol, pulse and physical activity. *Cancer Epidemiology Biomarkers & Prevention,* 4**,** 807-811.

STOMMEL, J. M., MARCHENKO, N. D., JIMENEZ, G. S., MOLL, U. M., HOPE, T. J. & WAHL, G. M. 1999. A leucine-rich nuclear export signal in the p53 tetramerization domain: regulation of subcellular localization and p53 activity by NES masking. *The EMBO journal,* 18**,** 1660-1672.

STRATTON, M. R., CAMPBELL, P. J. & FUTREAL, P. A. 2009. The cancer genome. *Nature,* 458**,** 719-724.

SUN, J.-H. & LEE, S.-A. 2013. Association between CAG repeat polymorphisms and the risk of prostate cancer: A meta-analysis by race, study design and the

number of (CAG) n repeat polymorphisms. *International journal of molecular medicine,* 32**,** 1195-1203.

SZKLARCZYK, D., FRANCESCHINI, A., KUHN, M., SIMONOVIC, M., ROTH, A., MINGUEZ, P., DOERKS, T., STARK, M., MULLER, J. & BORK, P. 2011. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic acids research,* 39**,** D561-D568.

TAFRIHI, M., TOOSI, S., MINAEI, T., GOHARI, A. R., NIKNAM, V. & ARAB, N. S. 2013. Anticancer properties of Teucrium persicum in PC-3 prostate cancer cells. *Asian Pacific journal of cancer prevention: APJCP,* 15**,** 785-791.

TAPLIN, M.-E., BUBLEY, G. J., SHUSTER, T. D., FRANTZ, M. E., SPOONER, A. E., OGATA, G. K., KEER, H. N. & BALK, S. P. 1995. Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer. *New England Journal of Medicine,* 332**,** 1393-1398.

TAYLOR, B. S., SCHULTZ, N., HIERONYMUS, H., GOPALAN, A., XIAO, Y., CARVER, B. S., ARORA, V. K., KAUSHIK, P., CERAMI, E. & REVA, B. 2010. Integrative genomic profiling of human prostate cancer. *Cancer cell,* 18**,** 11-22.

THOMAS, P. D., KEJARIWAL, A., GUO, N., MI, H., CAMPBELL, M. J., MURUGANUJAN, A. & LAZAREVA-ULITSKY, B. 2006. Applications for protein sequence–function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic acids research,* 34**,** W645-W650.

THOMPSON, I. M., ANKERST, D. P., CHI, C., LUCIA, M. S., GOODMAN, P. J., CROWLEY, J. J., PARNES, H. L. & COLTMAN, C. A. 2005. Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. *Jama,* 294**,** 66-70.

THUNE, I. & LUND, E. 1994. Physical activity and the risk of prostate and testicular cancer: a cohort study of 53,000 Norwegian men. *Cancer Causes & Control,* 5**,** 549-556.

THUT, C. J., CHEN, J.-L., KLEMM, R. & TJIAN, R. 1995. p53 transcriptional activation mediated by coactivators TAFII40 and TAFII60. *Science,* 267**,** 100-104.

TOLEDO, F. & WAHL, G. M. 2006. Regulating the p53 pathway: in vitro hypotheses, in vivo veritas. *Nature Reviews Cancer,* 6**,** 909-923.

VAN HAUTE, C., DE RIDDER, D. & NILIUS, B. 2010. TRP channels in human prostate. *The Scientific World Journal,* 10**,** 1597-1611.

VAN HEMELRIJCK, M., GARMO, H., HOLMBERG, L., INGELSSON, E., BRATT, O., BILL-AXELSON, A., LAMBE, M., STATTIN, P. & ADOLFSSON, J. 2010. Absolute and relative risk of cardiovascular disease in

men with prostate cancer: results from the Population-Based PCBaSe Sweden. *Journal of Clinical Oncology***,** JCO. 2010.29. 1567.

VENA, J. E., GRAHAM, S., ZIELEZNY, M., BRASURE, J. & SWANSON, M. 1987. Occupational exercise and risk of cancer. *The American journal of clinical nutrition,* 45**,** 318-327.

VICKERS, A. J., CRONIN, A. M., BJÖRK, T., MANJER, J., NILSSON, P. M., DAHLIN, A., BJARTELL, A., SCARDINO, P. T., ULMERT, D. & LILJA, H. 2010. Prostate specific antigen concentration at age 60 and death or metastasis from prostate cancer: case-control study. *Bmj,* 341**,** c4521.

VOGELSTEIN, B., LANE, D. & LEVINE, A. J. 2000. Surfing the p53 network. *Nature,* 408**,** 307-310.

VOUSDEN, K. H. & LANE, D. P. 2007. p53 in health and disease. *Nature reviews Molecular cell biology,* 8**,** 275-283.

WAHEED, R., KHAN, M. H., BANO, R. & RASHID, H. 2012. Sequence and structure based assessment of nonsynonymous SNPs in hypertrichosis universalis. *Bioinformation,* 8**,** 316.

WALTERING, K. K., HELENIUS, M. A., SAHU, B., MANNI, V., LINJA, M. J., JÄNNE, O. A. & VISAKORPI, T. 2009. Increased expression of androgen receptor sensitizes prostate cancer cells to low levels of androgens. *Cancer research,* 69**,** 8141-8149.

WANG, Y. & DAI, B. 2015. PTEN genomic deletion defines favorable prognostic biomarkers in localized prostate cancer: a systematic review and meta-analysis. *International journal of clinical and experimental medicine,* 8**,** 5430.

WARDE-FARLEY, D., DONALDSON, S. L., COMES, O., ZUBERI, K., BADRAWI, R., CHAO, P., FRANZ, M., GROUIOS, C., KAZI, F. & LOPES, C. T. 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research,* 38**,** W214-W220.

WATANABE, M., FUKUTOME, K., SHIRAISHI, T., MURATA, M., KAWAMURA, J., SHIMAZAKI, J., KOTAKE, T. & YATANI, R. 1997. Differences in the p53 gene mutational spectra of prostate cancers between Japan and Western countries. *Carcinogenesis,* 18**,** 1355-1358.

WATTERS, J. L., PARK, Y., HOLLENBECK, A., SCHATZKIN, A. & ALBANES, D. 2009. Cigarette smoking and prostate cancer in a prospective US cohort study. *Cancer Epidemiology Biomarkers & Prevention,* 18**,** 2427-2435.

WHITTEMORE, A. S., KOLONEL, L. N., WU, A. H., JOHN, E. M., GALLAGHER, R. P., HOWE, G. R., BURCH, J. D., HANKIN, J., DREON, D. M. & WEST, D. W. 1995. Prostate cancer in relation to diet, physical activity, and body size

in blacks, whites, and Asians in the United States and Canada. *Journal of the National Cancer Institute,* 87**,** 652-661.

WILLIAMS, J. L., GREER, P. A. & SQUIRE, J. A. 2014. Recurrent copy number alterations in prostate cancer: an in silico meta-analysis of publicly available genomic data. *Cancer genetics,* 207**,** 474-488.

WILT, T. J., BRAWER, M. K., JONES, K. M., BARRY, M. J., ARONSON, W. J., FOX, S., GINGRICH, J. R., WEI, J. T., GILHOOLY, P. & GROB, B. M. 2012. Radical prostatectomy versus observation for localized prostate cancer. *New England Journal of Medicine,* 367**,** 203-213.

WOLFF, B., SANGLIER, J.-J. & WANG, Y. 1997. Leptomycin B is an inhibitor of nuclear export: inhibition of nucleo-cytoplasmic translocation of the human immunodeficiency virus type 1 (HIV-1) Rev protein and Rev-dependent mRNA. *Chemistry & biology,* 4**,** 139-147.

XIAO, Y., CHIA, J. Y. C., GAJEWSKI, J. E., LIO, D. S. S., MULHERN, T. D., ZHU, H.-J., NANDURKAR, H. & CHENG, H.-C. 2007. PTEN catalysis of phospholipid dephosphorylation reaction follows a two-step mechanism in which the conserved aspartate-92 does not function as the general acid—mechanistic analysis of a familial Cowden disease-associated PTEN mutation. *Cellular signalling,* 19**,** 1434-1445.

XU, J., SUN, J. & ZHENG, S. L. 2013. Prostate cancer risk-associated genetic markers and their potential clinical utility. *Asian J Androl,* 15**,** 314-322.

YU, H., HARRIS, R. E. & WYNDER, E. L. 1988. Case-control study of prostate cancer and socioeconomic factors. *The Prostate,* 13**,** 317-325.

ZHANG, H., XU, Y., ZHANG, Z., LIU, R. & MA, B. 2012. Association between COX-2 rs2745557 polymorphism and prostate cancer risk: a systematic review and meta-analysis. *BMC immunology,* 13**,** 14.

ZHANG, Y. & XIONG, Y. 2001. Control of p53 ubiquitination and nuclear export by MDM2 and ARF. *Cell growth & differentiation: the molecular biology journal of the American Association for Cancer Research,* 12**,** 175-186.

ZHAO, C.-X., LIU, M., XU, Y., YANG, K., WEI, D., SHI, X.-H., YANG, F., ZHANG, Y.-G., WANG, X. & LIANG, S.-Y. 2013. 8q24 rs4242382 Polymorphism is a Risk Factor for Prostate Cancer among Multi-Ethnic Populations: Evidence from Clinical Detection in China and a Meta-analysis. *Asian Pacific journal of cancer prevention: APJCP,* 15**,** 8311-8317.

ZHENG, L., WANG, F., QIAN, C., NEUMANN, R. M., CHEVILLE, J. C., TINDALL, D. J. & LIU, W. 2006. Unique substitution of CHEK2 and TP53 mutations implicated in primary prostate tumors and cancer cell lines. *Human mutation,* 27**,** 1062-1063.

ZHOU, X.-P., WAITE, K. A., PILARSKI, R., HAMPEL, H., FERNANDEZ, M. J., BOS, C., DASOUKI, M., FELDMAN, G. L., GREENBERG, L. A. & IVANOVICH, J. 2003. Germline PTEN promoter mutations and deletions in Cowden/Bannayan-Riley-Ruvalcaba syndrome result in aberrant PTEN protein and dysregulation of the phosphoinositol-3-kinase/Akt pathway. *The American Journal of Human Genetics,* 73**,** 404-411.

Annexure-1

7.1 Cell lysis buffer preparation:

One liter of cell lysis buffer was prepared by weighting 8.29 gm of ammonium chloride (NH_4CL) and 2 gm of potassium carbonate ($KHCO_3$) was added to a graduated beaker. De-ionized water was added and the volume was raised to 800 ml. Total 200 µl of 0.5M ethylene diamine tetra acetic acid (EDTA) was added to the beaker and the whole solution was dissolved by using magnetic stirrer until the reagents were fully dissolved and particles were disappeared. The volume of solution was raised upto 1000 ml by adding deionized water. Its pH was recorded and was brought to pH 7.4 by the addition of acids and basis as required.Cell lysis buffer was stored in 1 liter bottle at 4°C.

7.2 STE (Saline Tris EDTA) preparation:

One liter of STE was prepared by adding 33.3ml of 3M NaCl,100 ml of 1M Tric-HCl and 2 ml of 0.5M EDTA to a beaker and were dissolved in 1000ml of dH2O. Its pH was adjusted on 8.0.The prepared solution was filtered with 0.4µl filter paper and stored at 4°C.

7.3 Equilibration of phenol:

1 kg phenol was melted with the addition of 1gm 8-hydroxy quinoline and then was extracted with equal volume of 1M Tris (1L, pH-8.0). Extraction was revised with an equal volume of 0.1M Tric until the pH of the aqueous layer was reached to 8.0.Finaly 100ml of 0.1M Tris containing 0.2% β-macaptoethanol (2ml) was added and stored at -4°C.

7.4 Preparation of 10X TBE (Tris BorateEDTA) 1 Liter:

- Tric = 108 gm
- Borate= 55 gm
- EDTA= 7.4 gm

These reagents were dissolved in 1000 ml deionized water and were stored in a labeled bottle. Preparation of 25% ammonium persulfate (Giovannucci et al.) (5 ml):1 gm ammonium per sulfate was added to 4 ml deionized water and was dissolved by shaking. Similarly 25% APS (5 ml) was prepared.

Annexure-2

8.1 Performa for Prostate Cancer Samples Collection

COMPUTATIONAL AND MOLECULAR ANALYSIS OF TP53, PTEN AND AR GENES IN PROSTATE ADENOCRCINOMA

Department of Bioinformatics and Biosciences
Capital University of Science and Technology, Islamabad, Pakistan

You are invited to take part in this research study. This form tells you why this research study is being done, what will happen in the research study, and possible risks and benefits to you. If there is anything you do not understand, please ask questions. Then you can decide if you want to join this study or not. If you have read this form and have decided to participate in this project, please understand your participation is voluntary.

Your identity and personal data will NOT be known to any personnel other than the investigators and they will not be disclosed in any published and written material resulting from the study. It is also to be noted that, you will not be paid to participate in this study.

Hospital ID: _____	Deptt. ID: _____	Sample No. _____
Status: _____	Sample Type: _____	Date: _____

1. PATIENTS DETAILS:

Sex: _____ DOB: _____ Religion: _____

Ethnicity/Caste: _____ Weight: _____ Height: _____

Marital status: _____ No. of Children:_____ Contact No: _____

Address: _____

PSA: _____ BP: _____ Smoking: _____

Physical Activity: _____ Dairy Products consumption: _____

2. DISEASE HISTORY:

Age of disease onset: _____ Symptoms: _____

Treatment details: _____

Side effects: _____

Previous diagnosis: _____

Have you ever had any type of cancer? _____ Surgery status: _____

Do you walk (or do other moderate activity) for at least 30 min on most days: _____

3. FAMILY HISTORY:

Do you have any family member who has had any type of cancer? _____

If yes, please specify: _____

Type of Cancer diagnosed in the family member: _____

4. DISEASE COMPLICATIONS:

- Need to urinate frequently, especially at night**: _____**
- Week or interrupted urine flow: _____
- Difficulty in starting urination or holding back urine: _____
- Painful or burning urination: _____
- Difficulty in having an erection: _____
- Painful ejaculation: _____
- Blood in urine or semen: _____
- Frequent pain or stiffness in the lower back, hips or upper thighs: _____
- Weight loss: _____
- Any other complication: _____

5. INFORMED CONSENT:

Since information about you and your health is personal and private, it generally cannot be used in this research study without your written authorization. If you sign this form, it will provide that authorization. Please read it carefully before signing it. I am donating my blood/tissue sample for research purposes only and not for commercial purposes.

Signature: _____

Annexure-3

9.1 Gel images for PCR conditions optimization

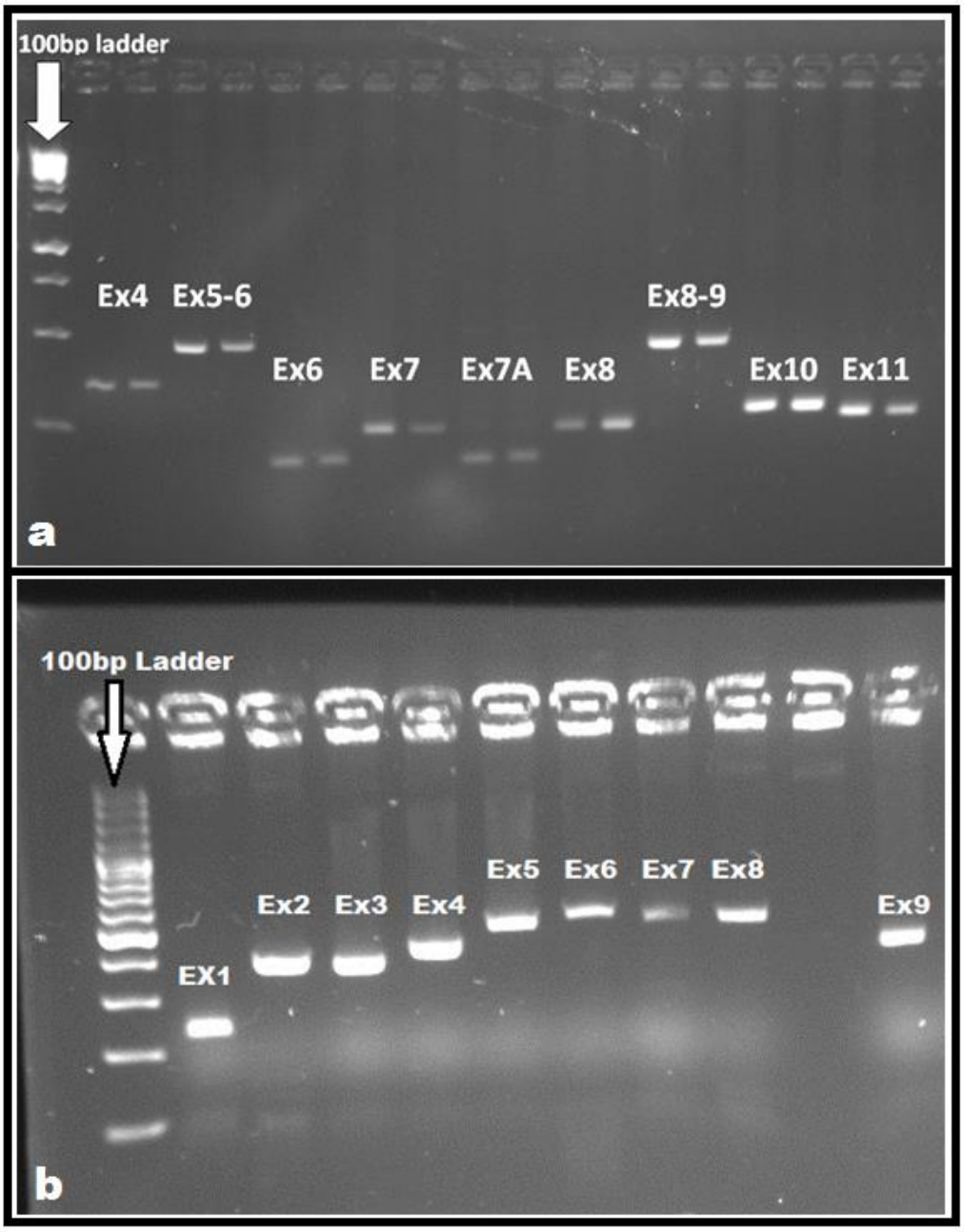

Figure 9.1: PCR optimization for TP53 and PTEN Exons.

a. Optimization of TP53 Ex4, Ex5-6, Ex6, Ex7, Ex7A, Ex8, Ex8-9, Ex10 and Ex11 with Prostate cancer samples PC-81 and PC-82.

b. Optimization of PTEN Ex1, Ex2, Ex3, Ex4, Ex5, Ex6, Ex7, Ex8 and Ex9 with Prostate cancer samples PC-81 and PC-82.