Identification of Genetic Loci Associated with Risk for Type-2 Diabetes Mellitus (T2DM) in Indians from Rajasthan

THESIS

Submitted in partial fulfillment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

URVASHI DUBE
2005PH29101H

Under the supervision of
Prof. Suman Kapur

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE
PILANI, INDIA
2012

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI

Certificate

This is to certify that this thesis entitled *"Identification of Genetic Loci Associated with Risk for Type-2 Diabetes Mellitus (T2DM) in Indians from Rajasthan"* and submitted by Ms. Urvashi Dube, ID No. 2005PH29101H for the award of Ph.D degree of the institute embodies original work done by her under my supervision.

Signature in full of Supervisor: …………………………...

Name in capital block letters: ………………………….

Designation:………………………………………………

Date:

Acknowledgements

My journey for Ph.D has been a very memorable and overwhelming experience. The whole journey of compiling a PhD thesis is almost like living a lifetime and it is difficult to summarise what has been the real learning experience during the course and who all to thank in this wonderful journey as there are so many people whom the heart remembers on reaching the pinnacle of this personal Everest climbing venture. To be honest, this intriguing journey has been instrumental in teaching me what life is all about and hence would like to express my thanks to some of the wonderful people without whom I would have probably never reached this point.

First of all, my gratitude is due to my mentor Dr Suman Kapur. It is hard to say what I have not learnt from her both professionally and personally; be it working on my Ph.D topic itself which has been the real learning experience, writing papers and proposals, giving talks, working in a group as a valuable team member, staying up in lab till someone drags you out, or to stay focussed and deliver. One thing which I have learnt from her in these years is once committed you have to deliver. To work with her has been a real pleasure for me with heaps of fun and excitement. She has listened to me as mentor and simultaneously corrected me when required. Her ability to select and to approach compelling research problems, high scientific standards, and hard work set an example. She has been a great mentor throughout my Ph.D. career; she has oriented and supported me with promptness and care, and has always been patient & encouraging. Thank you for being a part of my family, my guardian and my awesome mentor for all these years and for the years to come.

My very special thanks are due to our collaborators Dr R.P. Pareek, Dr R.P. Agarwal, Dr Khinsara, Dr S.D. Sharma, Dr Bhupendra Singh and Dr Lokesh Chaturvedi for helping me with sampling and data collection.

My heartfelt thanks are due to my DAC memebers – Dr Vidya Rajesh, Dr Pranav Kumar Narayan who have put in considerable time and effort, not only to judge the overall results but also to comment in detail on the text and the format of the thesis. They deserve mentionable special thanks for all the suggestions and support. My special thanks are due to Dr Jayati Ray Dutta,

encouraging, for believing in me, and for the many precious memories along the way. I couldn't have achieved this without your support. Thank you for everything. Furthermore, I also want to thank all other friends who put up with me through the whole Ph.D. process and helped me with personal challenges, in particular Vipin, Preeti, Prashant, Bubu, Susheel, Neetu, Devesh, Rajan and Ankit. Thank you guys for those occasional motivating calls, talks, gossips and your unconditional support!

My thanks would be incomplete without mentioning the super awesome Genomics team: Padma, Sruthi, Pavan, Mr Rao, Rupak, Madhu maam and Chinmayi for all the help, love, encouragement and support. My special thanks to my friend Anuradha for the fun and the encouraging discussions on and off the lab. I am eternally grateful to my dearest friend Shivani. We've laughed, cried and among other things had many rants together. Thank you for being a friend, sister, confidant, conscience, humorist, phone comrade and much more. In you I have found a life-long friend and a wonderful sister. Thanks to our old colleagues Sribash, Krishnaveni, Rosaline, Gaurav and Susan for making our lab a welcoming place to work and also for helping me during the cumbersome experiments. Words are not enough to express my gratitude for you all. Thank you all for your love and support!

I would also like to thank some people from early days of my Ph.D for their endless support and help. My special thanks are due to Manav sir for training me how to do research, how to tackle new problems and how to develop techniques to solve them. You really taught me a great deal about the technical groundwork and statistics. My special thanks are due to my wonderful senior's Shashwat sir and Shipra for your much needed suggestions which helped me shape my thesis.

I am also indebted to the students I had the pleasure to work with. You have been an invaluable support day in, day out, during all these years. Worth mentionable thanks to Sachin, Ipshita, Punita, Apoorva and Atul for the awesome help and company. Also a spcial mention to my cute and wonderful first degree friends especially Chhavi, Amrita, Sneha, Srishti, Kriti and Pranavi for all the amazing moments and fun.

My special thanks to my hostel in charges, Anita maam and Sunita di for their love, concern, care and support. Sunita di special thanks to you for opening the hostel gate at late night without any complaints and to you Anita maam for making us feel at home. My thanks cannot be complete without acknowledging our three technicians, Bhagwati didi, Satish and Ali bhaiya. Heartfelt thanks to Bhagwati didi and Satish for the wonderful support while packing the lab and helping us move it from Pilani. Special mention to Ali Bhaiya and Shakeel for making a lot of things happen in lab without any hassles.

Last but not the least thank you god for everything you have blessed me with in life.

At this moment, with mixed feelings of exuberance and heaviness in heart, I would like to pen down four lines by Robert Frost which kept me motivated and on the F1 race track of my PhD. I had first read them on a summer afternoon in Pilani -

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I--
I took the one less traveled by,
And that has made all the difference

(Urvashi Dube)

Abstract

The global epidemic of type 2 diabetes is increasing at an alarming rate in both developed and developing countries around the world. The emerging pandemic is driven by the combined effects of population ageing, rising levels of obesity and inactivity, and greater longevity among patients with diabetes that is attributable to improved management. The micro and macro vascular complications associated with type 2 diabetes account for the majority of the social and economic burden among patients and society more broadly. Though the genetic basis of type 2 diabetes is not understood completely, the familial nature of the disease is well recognized. Studies on hereditary nature of Type 2 Diabetes Mellitus (T2DM) have indicated both strong and suggestive linkage signals at loci 2q32-37, 3q22-29 and 10q25-26 on chromosome 2, 3 and 10 respectively. The genetic variations, such as single nucleotide polymorphisms (SNPs) reported in the genes present in these regions are associated (slightly) with increased risk for type-2 diabetes, but they only marginally improve the odds of predicting whether an individual will get type-2 diabetes based on the traditional clinical characteristics combining age, sex and weight. Additionally diabetes is known to be influenced by a number of environmental factors that affect body composition markers, which in turn are regulated by Quantitative Trait Loci (QTLs) mapped to these loci as well. We have therefore undertaken population-based genetic screening using microsatellite markers to identify the potential association of these loci with T2DM in North Indian population and to identify the most crucial risk loci (among the three) associated with the pathogenesis of the disease in North Indians.

The study recruited two hundred and ninety one (291) subjects with Type 2 Diabetes Mellitus and around four hundred and seventy five (475) unrelated healthy volunteers

as controls without any history of diabetes and related comorbidities. The subjects were examined by genotyping using PCR-SSLP technique to ascertain disease association with seven highly polymorphic microsatellite markers that were on an average 10cM apart within 2q32-37, 3q22-29 and 10q25-26 regions. Relationship of alleles/genotypes of these STR marker with anthropometric, clinical or/biochemical markers was accessed using relevant statistical tools. Online available bioinformatics tools were used to identify the effect poised by the repeats in these STR on the stability and folding of the structure formed. Since these STR repeats and their association with T2DM has yet not been studied in Indian population, it is crucial to see their direct and indirect associations with the disease in order to identify the T2DM associated marker specific for Indian population.

Several significant associations were recognized with the alleles for markers D2S1384, D2S2944 and D2S439 present within 2q32-37 region, with markers D3S2398 and D3S3609 in the 3q22-29 region and with markers D10S521 and D10S1237 in the region 10q25-26. Adult subjects possessing the risk alleles of D2S1384, D2S439, D10S1237 and D10S521 were found to have increased susceptibility to diabetes without adjustment for confounding factors. When analyzed for allele length the shorter alleles (or <12 (CTAT) repeats) of D2S1384, were found to be more prevalent in the control population. The <145bp allele constitutes 25% in adult cases and 38% in controls (OR=1.91, χ^2=14.679, 95 % CI=1.3755 to 2.6468, p=0.0001). The presence of the longer base pair allele increases the odds of having T2DM by nearly 2 times (OR= 1.82, 95 % CI=1.3461 to 2.4650, p =0.0001). When analyzed with the allele length significant differences were observed in levels of total cholesterol (<145bp: Non-significant; ≥145bp: t=1.770, p=0.0788), triglyceride level (<145bp: t=2.803, p=

0.0060; $\geq$145bp: t=3.722, p =0.0003) and VLDL (Non-significant; $\geq$145bp: t=3.721 p = 0.0003) levels among cases and controls in presence of longer allele

Further comparison of allelic data of D2S2944 in the recruited subjects it was found that the $\geq$ 124bp allele was the risk allele and was able to increase the risk for T2DM by a marginal significance (OR=1.34, χ2=4.62, 95% CI=1.0347 to 1.7497, p = 0.0268). On further analysis significant differences were observed in levels of Fasting blood glucose FBG (<124bp: t=14.768, p< 0.0001; $\geq$124bp: t=10.091, p< 0.0001), among cases and controls in presence of longer allele and significant reduction was observed in mean total cholesterol (<124bp: t=4.281, p <0.0001; $\geq$124bp: NS), triglyceride level (<124bp: t=3.673, p=0.0003; $\geq$124bp: t=2.720, p = 0.0077), HDL(<124bp: t=2.923, p = 0.0039; $\geq$124bp: NS) and VLDL(<124bp: t=3.603, p= 0.0004; $\geq$124bp: t=2.729, p = 0.0075) levels in the presence of longer base pair allele of D2S2944 among cases and controls. The marker D2S439 was evaluated for its association with T2DM and it was seen that the risk allele seems to increase the risk for T2DM by nearly two folds (χ^2=20.223, OR=1.95, 95% CI=1.4660 to 2.6196, p< 0.0001). In presence of longer base pair allele of D2S439 the mean SBP (<195bp: t=4.110, p=0.0001; $\geq$195bp: t= 1.982, p=0.04), DBP (<195bp: t=3.279, p=0.0012; $\geq$195bp: NS) BMI (<195bp: t= 6.362, p<0.0001; $\geq$195bp: t=4.65, p<0.0001) and WHR (<195bp: t= 5.208, p < 0.0001; $\geq$195bp: t=6.003, p<0.0001) also differed significantly among cases and controls. Subsequently we looked at the two markers on chromosome 3 and it was seen that the $\geq$171bp allele of the marker D3S3609 was higher in controls as compared to cases (χ^2=24.777, OR=0.4182, 95% CI=0.2968 to 0.5893, p<0.0001). Both D3S3609 and D3S2398 markers were marginally associated with T2DM and were also associated with obesity markers especially WHR. Further very good significance was observed with the longer base pair allele of D10S521. It was seen that the $\geq$179 bp allele of

D2S521 constitutes 91% in adult cases and only 64% in controls (χ^2=83.127, OR= 5.6639, 95% CI=3.8169 to 8.4047, p<0.0001). Infact the presence of longer at least one copy of longer base pair allele increased the risk two folds (OR=5.66 to 10.86). The second marker of chromosome 10 also showed highly robust association with T2DM (χ^2=48.220, OR=3.0185, 95% CI=3 2.2033 to 4.1352, p< 0.0001) with the increase in number of repeats. Studies have shown that microsatellite motifs in the UTR form structural elements (stem-loops) and contribute to mRNA regulation. Our prediction using the MFOLD program for all the markers showed a drastic change in the structure of the folded sequence of the longer base pair allele or remodeling or increase in stem loop with stabilized energy stable structure for five out of seven markers. These changing folding patterns may be associated with regulation of gene expression and needs to be evaluated in detail further.

Although the association of the STR repeats and the folding changes reported in the study still remains to be elucidated, the findings of the studies presented in this thesis is of considerable interest as such modulation of gene expression would be able to answer many unanswered questions. The study indeed proves the presence of novel diabetes causing lesions in loci studied. In the future, screening susceptible subjects by combining the use of anthropometric measurements and genotyping of these associated markers especially D10S521 and D10S1237 would probably enable early lifestyle interventions or a personalized therapy to delay the onset of T2DM by early targeted detection and intervention.

Table of Contents

List of Tables

List of Figures

List of Abbreviations/Symbols

ABCA12 - ATP-binding cassette sub-family A member 12
ADRA2A -Alpha-2-adrenergic receptor
AIF - Apoptosis-inducing factor)
ALFRED - Allele Frequency Database
AP-1 - Activating protein-1
ATF-6 - Activating transcription factor-6
AUC - Area under the curve

BiP - Immunoglobulin heavy-chain-binding protein
BMI – Body mass index
bp – base pair

CAD - Coronary Artery Disease
CAPN10 - calpain-10
CI $_{95\%}$ - Confidence Interval at 95% level
COGA - Collaborative Studies on Genetics of Alcoholism
CRP - C-reactive protein
CVD - Cardiovascular disease

DBP - Diastolic Blood Pressure

EDTA – Ethylene diamine tetra acetic acid
EGR-1 - Early growth response-1
EMBL - European Molecular Biology Laboratory
ERN-1 - ER to nucleus signalling-1
EtBr - Ethidium bromide

FPG - Fasting plasma glucose

GLA - Gamma-linolenic acid
GLP-1 - Glucagon-like peptide 1
GWAS - Genome wide association scans

HDL - High-density lipoprotein
HHEX - Hematopoietically Expressed Homeobox
HSP - Heat-shock protein
HWE - Hardy–Weinberg equilibrium
IDF - International Diabetes Federation

IGF2-Insulin-like growth factor 2
IGFBP2 - Insulin-like growth factor-binding protein 2
IGFBP5 - Insulin-like growth factor-binding protein 5
IGT- Impaired glucose tolerance
IHEC – Institutional Human Ethics Committee
IL-6 - Interleukin-6
IRE-1 - Inositol-requiring enzyme-1
IRS - Insulin resistance syndrome
IRS1 - Insulin receptor substrate 1

JNK- c-Jun N-terminal kinase

KCNJ11 - β-cell Potassium Channel Gene

LBW- Low birth weight
LDL – Low density lipoprotein
LOD – Log of Odds Ratio
MCP-1- Monocyte chemo attractant protein-1
MetS – Metabolic Syndrome
MMS – Multiple Metabolic Syndrome

NCBI – National Center for Biotechnology Information
NCDs – Non-communicable disease
NEFAs - Non-esterified fatty acids
NF-Kb- Nuclear factor Kb
NGT- Normal glucose tolerance
NIDDM1 - Noninsulin-dependent diabetic patients

OD – Optical Density
OGTT - Oral glucose tolerance test
OR – Odds Ratio
ORP150 - Oxygen-regulated protein 150
OXPHOS -Oxidative phosphorylation

PAGE -Polyacrylamide Gel electrophoresis
PBA - 4-phenyl butyric acid
Pc – Bonferenni corrected p value

PCR - Polymerase chain reaction
PERK - PKR (double-stranded-RNA-dependent protein kinase)-like ER kinase
PGC-1- Peroxisome proliferator-activated receptor-γ coactivator
PPAR- Peroxisome-proliferator-activated receptor
PPARγ - Peroxisome proliferator-activated receptor gamma

QTLs - Quantitative trait loci

SBP - Systolic Blood Pressure
SD – Standard Deviation
SLC2A2 - Solute carrier family 2 (facilitated glucose transporter), member 2
SNP - Single Nucleotide Polymorphism
SS - Sum of skin folds
SSLP – Short Sequence Length Polymorphism

STR – Short Tandem Repeats
STS – Sequence Tag Site

T2DM - Type II Diabetes Mellitus
TC – Total Cholesterol
TCF7L2 - Transcription factor 7-like 2
TNF-α - Tumour necrosis factor-α
TPMD - Taiwan Polymorphic Marker Database
TUDCA- Taurine-conjugated ursodeoxycholic acid

VLDL - Very low-density lipoprotein

WHO - World Health Organization
WHR – Waist to Hip Ratio

XBP-1- X-box-binding protein-1
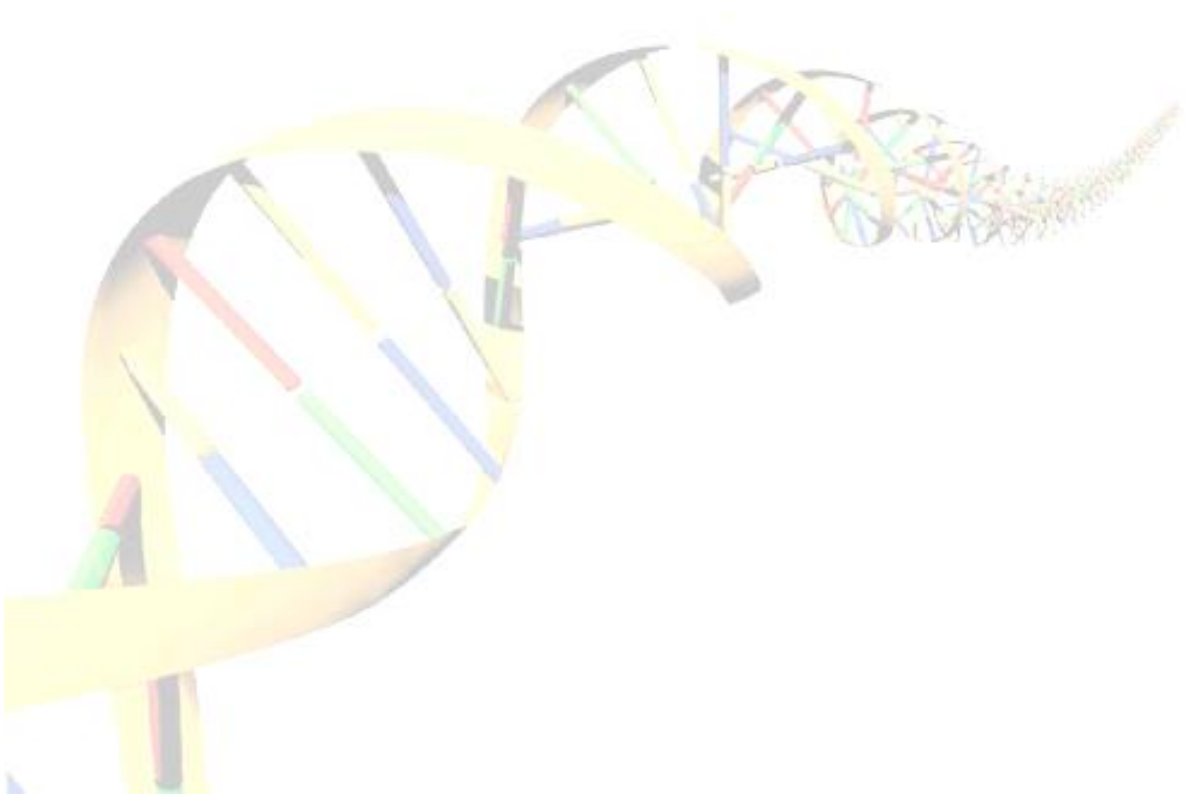
Introduction

Chapter 1

Chapter 1

Introduction

1.1 Type II Diabetes: The Silent Killer

Type II diabetes mellitus (T2DM) belongs to the heterogeneous group of metabolic disorders characterized by elevated glucose levels or impaired insulin secretion/action. It is a multifactorial disorder (controlled by both genetic and environmental players) involving dysfunctions in multiple organs/tissues. These defects might be in insulin-mediated glucose uptake in muscle, a dysfunction of the pancreatic β-cells, a disruption of secretory function of adipocytes, and an impaired insulin action in liver . T2DM ranks as the fourth most common cause of mortality with coronary artery disease (CAD) posing high global burden. During the last fifteen years the prevalence of T2DM has been increased by five folds worldwide (Roup *et al.,* 2009). International Diabetes Federation (IDF) reports that the disease has already reached to a global epidemic level with 285 million reported diabetics in 2010, an increase of 39 million people as compared to the reported number in 2007 (Whiting *et al.,* 2011). The recent projection by IDF for year 2030 is 439 million diabetics across the world, an equivalent of 65% increase since 2010 (Whiting *et al.,* 2011; Leonard and Egede, 2010). These non-communicable diseases like diabetes, CVD, depression, obesity etc. are emerging as major health challenges especially in South East Asian countries. About one-fifth of the adult diabetics, across the world, live in the South-East Asian region. Current estimates indicate that 8.3% of the adult South East Asian population, or 71.4 million people, have diabetes in 2011, 61.3 million of whom are in India (Whiting *et al.,* 2011). India, currently is second in lead after China with maximum number of diabetics with in an age group of 20-79 years with

an estimated 51 million diabetics in 2010, a figure projected to leap to 87 million (58% increase) in 2030 (Leonard and Egede, 2010). The impact of the disease is tremendous with an associated increase in mortality/morbidity; there by having huge effect on the quality of life of an individual per-se. The economic burden associated with the disease also poses heavy toll on the individual as well as the family and the society.

1.2 Gap in Existing Research

To date, genome-wide association studies (GWAS) have identified variants in and around 42 genes as associated with susceptibility to T2DM (Ramachandran and Snehalatha, 2009). However, these studies have been predominantly performed in populations of European ancestry. Initial approaches focused on the comparison of single nucleotide polymorphism (SNP) and structural polymorphism frequencies between diabetes patients and controls in a small number of candidate genes such as insulin, the insulin receptor, the GLUT4 glucose transporter, the insulin receptor substrate 1 (IRS1), and control of glucose-stimulated insulin secretion (glucokinase). These early studies, which focused almost entirely on coding variations, were largely viewed as non-reproducible and relatively uninformative for common forms of T2DM. Nonetheless, candidate gene studies identified two genes, now considered widely replicated: PPARγ (Peroxisome proliferator-activated receptor gamma) and the β-cell potassium channel gene, KCNJ11 (Kooner *et al.,* 2011). Since 2007, the T2DM genetic field has come full circle, returning to the association approach along with linkage analysis. These studies have provided a logarithmic growth in the number of polymorphisms contributing to human disease. With the availability of hypothesis-free methods such as linkage studies in families and affected sibling pairs using simple tandem repeat markers, T2DM genetics entered a new

Chapter 1

era that promised the identification of new pathways not identified by studies of pathophysiology. A plethora of linkage studies across populations identified some reproducible signals, as noted later, but few signals reached levels of genome-wide significance observed in monogenic diseases, and no linkage peaks could be explained by single major genetic risk factors. Whole genome studies conducted in the western world have revealed the role of specific chromosomal loci/genes in the etiology of T2DM but to date there is still a paucity of data even on T2DM susceptibility genes (candidate gene approach) or associated chromosomal locations in the Asian Indian population. Given the likely extensive role of intronic and intergenic DNA in determining phenotype, a major role of sequence variants in non-coding regions in T2DM pathogenesis should be anticipated.

Our increased understanding of such phenomena will open new doors to understanding that how common variants can alter T2DM susceptibility and will be essential in understanding the physiologic importance of the genetic associations that are uncovered. Moreover there are several key questions that need to addressed with reference to the interpretation of all the data available related to the genetic studies across the globe. We need to look at the various key behavioral and environmental factors along with these genetic markers to tailor the various existing detection/prevention approaches. These detection and prevention strategies need to be cost effective at the societal levels and accessible to one and all to ensure that the outcome is available for use by the masses. The outcomes of these genetic studies need to be translated in diverse settings and populations to prevent T2DM among youths and adults.

Chapter 1

1.3 Objective of the Proposed Research

The aim of this genetic epidemiological thesis is to study chromosomal regions/candidate genes that may play a role in the etiology and pathogenesis of T2DM. The broader objective of this work is to obtain new insights into the biological pathways that may be involved in T2DM and identifying other disorders that have common etiological pathways and hence may share the same underlying genes as T2DM. In brief the two main objectives of this thesis are as follows-

1. Ascertaining the role of hitherto identified loci on chromosomes 2, 3 and 10 in the etiology of T2DM in Indians from Rajasthan.

2. To screen for most critical sequence variations that could lead to diabetes.

1.4 Approach Taken

Advances in human genetic map, genetic linkage and association studies in complex inheritable traits have led to an increased understanding of the role of genetics in several chronic multi-factorial diseases. Several approaches to locate the genes responsible for T2DM have been adopted by different groups worldwide. Overall, there is limited or no high extent of success from the candidate-gene approach based studies in identifying the genetic determinants involved in T2DM. The other commonly used approach, linkage and linkage disequilibrium analysis, relies on the fact that genes with similar chromosome positions will only rarely be separated during recombination so the susceptibility to causative genes can be localized by searching for genetic markers that co-segregate with disease. Linkage studies, conducted predominantly in the developed western world, have observed a strong racial and sex bias in the genetic constitution of T2DM patients. Till date only a single gene TCF7L2, identified first during a search for

Chapter 1

392935 SNP's by Sladek *et al.* (2007) in a French case–control cohort could be replicated across various populations without any loss of associated significance. These associations explain a substantial portion of disease risk and constitute proof of principle for linkage based studies to elucidate of complex genetic traits. To date a similar approach has not been undertaken to study T2DM in subjects of Indian origin. Infact, there is a scarcity of data even on T2DM susceptibility genes (candidate gene approach) in the Asian Indian population.

Thus, a population based association study was undertaken which considered genomic regions- directly implicated in T2DM or regions indirectly associated with systemic disorders where T2DM is co-morbid, and/or body composition markers related with T2DM. Such an approach will help impart unraveling the complex pathogenesis of T2DM. This involved reviewing the regions (by literature search) in the genome depending on LOD scores (A LOD score of 3 or more was considered significant, a LOD score between 2.2 and 3 was considered suggestive and a LOD score between 1 and 2.2 was considered nominal). Only those loci were included that were significant or suggestive in at least one study and at least nominal in two or more studies. Finally the critical regions having the attribute of all i.e. meeting the inclusion criteria laid down in present study and are overlapping chromosomal regions showing high probability of presence of susceptible genes for T2DM were studied.

1.5 Region chosen for present study

Three such regions on chromosome 2q32-37, 3q22–q29 and 10q25-26 are studied in detail in this thesis for their genetic contribution in causing T2DM.

Chapter 1

1.6 Significance of the study

Several research approaches suggest that both genetic factors and environmental factors play critical role in T2DM. The extra molecular players in this disease are still being undiscovered. If this is understood better it will shed light on genetic architecture of T2DM, which could serve as a platform for early diagnosis of individuals at risk from the high-susceptibility families/population and could allow for the design of strategies for early and timely intervention of the disease. The study is particularly relevant in the Indian context, as ours is a developing nation with a large lower income group and inadequate medical facilities. Till date, there is large epidemiological data available which shows that T2DM is the leading cause of death in India but still not much of genetic data on the Indian populations in reference to T2DM and its related genes and phenotypes is available.

1.7 Outline of thesis

The present thesis has focused on the role of genetic factors in T2DM and has been divided into 6 chapters. The Chapter 2 summarizes the prevalence, incidence, disease burden and etiology of T2DM. Chapter 3 focuses on the first aim and provides a detailed review of all genetic studies on T2DM and the meta-analyses of region on chromosome 2 (2q33-37) 3(3q22–q29) and 10 (10q25-26) that have been investigated in any of the global studies. The association of these loci with T2DM was investigated using STR markers in chapter 5 following the methodology described in chapter 4. These microsatellite markers analyzed in chapter 5 have either shown strong linkage signals or alleles of these markers were found to be a part of disease associated haplotype constantly in several familial linkage or association studies on T2DM in the past. The

region showing evidence of robust association with T2DM is further analyzed using bioinformatics to narrow down to candidate genes. Additionally the relationship of shared genetic and environmental factors in the occurrence of symptoms of T2DM has also been investigated. Chapter 6 discusses the results described in chapter 5 and identify the biological pathways that may be involved in the etiology of T2DM and the role of common genetic factors in the incidence of T2DM. The relationship of genetic and other lifestyle related disease risk factors such as BMI, WHR, blood pressure, and lipid parameters in the occurrence of T2DM symptoms are also discussed in chapter 6. Finally, the last section of the thesis provides conclusion of this work along with specific contribution and future scope of the study.
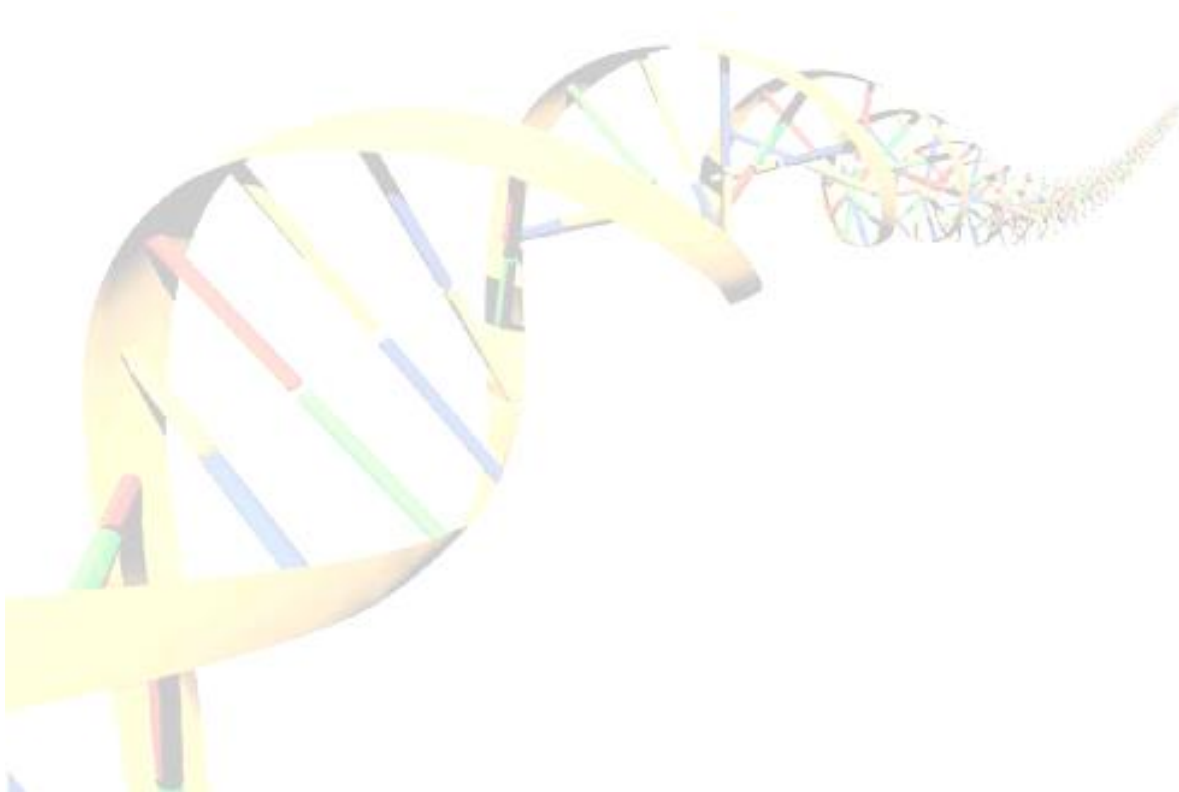
Literature Review

Chapter 2

Literature Review

2.1 Diabetes Mellitus

Diabetes mellitus, or simply diabetes, is a group of metabolic diseases characterised by elevated blood sugar (hyperglycemia), due to inability of pancreas to produce sufficient insulin, or inability of cells to respond to the insulin that is produced. This multifactorial disorder is often accompanied by polyuria (frequent urination), polydipsia (increased thirst) and polyphagia (increased hunger).

2.2 Etiologic Classifications of Diabetes Mellitus

The previous classification of diabetes was given in 1979 when the National Diabetes Data Group produced a consensus document standardizing the nomenclature and definitions for diabetes mellitus (National Diabetes Data Group, 1995). This document was endorsed one year later by WHO (WHO, 1980; WHO, 1985; Mayfield, 1998). The two major types of diabetes mellitus were given names descriptive of their clinical presentation: "Insulin-Dependent Diabetes Mellitus" (IDDM) and "Non–Insulin-Dependent Diabetes Mellitus" (NIDDM). However, as treatment recommendations evolved, correct classification of the type of diabetes mellitus became confusing. This confusion led to the incorrect classification of a large number of patients with diabetes mellitus, complicating epidemiologic evaluation and clinical management. The discovery of other types of diabetes with specific pathophysiology that did not fit into this classification system further complicated the situation. These difficulties, along with new insights into the mechanisms of diabetes mellitus, provided a major impetus for the development of a new classification system. The

Chapter 2

new classification system identifies four types of diabetes mellitus: Type 1, Type 2, "other specific types" and gestational diabetes.

2.2.1 Type 2 diabetes mellitus

Type 2 diabetes mellitus (formerly called NIDDM, type II or adult-onset) is characterized by insulin resistance in peripheral tissue and an insulin secretory defect of the beta cell (ADA, 2012). This is the most common form of diabetes mellitus and is highly associated with a family history of diabetes, older age, obesity and lack of exercise. It is more common in women, especially women with a history of gestational diabetes, in individuals with hypertension or dyslipidaemia and in Blacks, Hispanics and Native Americans. Insulin resistance and hyper-insulinemia eventually lead to impaired glucose tolerance. Defective beta cells become exhausted, further fuelling the cycle of glucose intolerance and hyperglycaemia. It is a leading public health problem; accounting for nearly 90% of diabetic cases reported worldwide, the remaining 10% comprises of Type 1 Diabetes Mellitus (T1DM), gestational Diabetes and other specific types (Sordi *et al.,* 2008; Defronzo *et al.*, 2009). It is a complex metabolic disorder characterised by elevated glucose levels due to impaired insulin secretion/action. It is a polygenic, multifactorial, and multi-organ disorder which comprises of a defect in insulin-mediated glucose uptake in muscle, a dysfunction of the pancreatic β-cells, a disruption of secretory function of adipocytes, or an impaired insulin action in liver. At least initially, and often throughout their lifetime, these individuals do not need insulin treatment to survive. Most patients with this form of diabetes are obese, and obesity itself causes some degree of insulin resistance. Patients who are not obese by traditional weight criteria may have an increased percentage of body fat distributed predominantly in the abdominal region. This form

of diabetes frequently goes undiagnosed for many years because the hyperglycaemia develops gradually and the classic symptoms of diabetes are not severe enough for patient to notice at earlier stages. Nevertheless, such patients are at increased risk of developing macrovascular and microvascular complications. The patients with this form of diabetes may have insulin levels that appear normal or elevated; the higher blood glucose levels in these diabetic patients would be expected to result in even higher insulin values had their β-cell function been normal. Thus, insulin secretion is defective in these patients and insufficient to compensate for insulin resistance. Insulin resistance may improve with weight reduction and/or pharmacological treatment of hyperglycaemia but is seldom restored to normal. It is often associated with a strong genetic predisposition, more so than is the autoimmune form of Type 1 Diabetes. However, the genetics of this form of diabetes is complex and not clearly defined (ADA, 2012).

2.3 Normal Glucose Metabolism

For the normal functioning of our body energy is required. There are various sources of energy in our body. Many tissues can use fat or protein as an energy source but others, such as the brain and red blood cells, can only use glucose. Infact one of the major sources of energy in the body is carbohydrate metabolism. Glucose is one such carbohydrate which acts as source of energy and is also an important precursor to a lot of biosynthetic pathways. Thus glucose metabolism is important for the normal physiological functioning of our body (Fig 2.1). Glucose metabolism is the process of conversion of glucose into utilisable energy (ATP) for the cells and this process is referred as glycolysis. Normally the body derives glucose from the dietary carbohydrate rich food. Glucose metabolism is highly dependent on hormones

secreted by the islets of Langerhans especially insulin. As the level of glucose in the body increases there is an increase in the level of insulin secretion by the pancreas, which is responsible for transporting this glucose to cells for its conversion to energy using glycolysis in the cytoplasm of the cell (WHO, 1994).
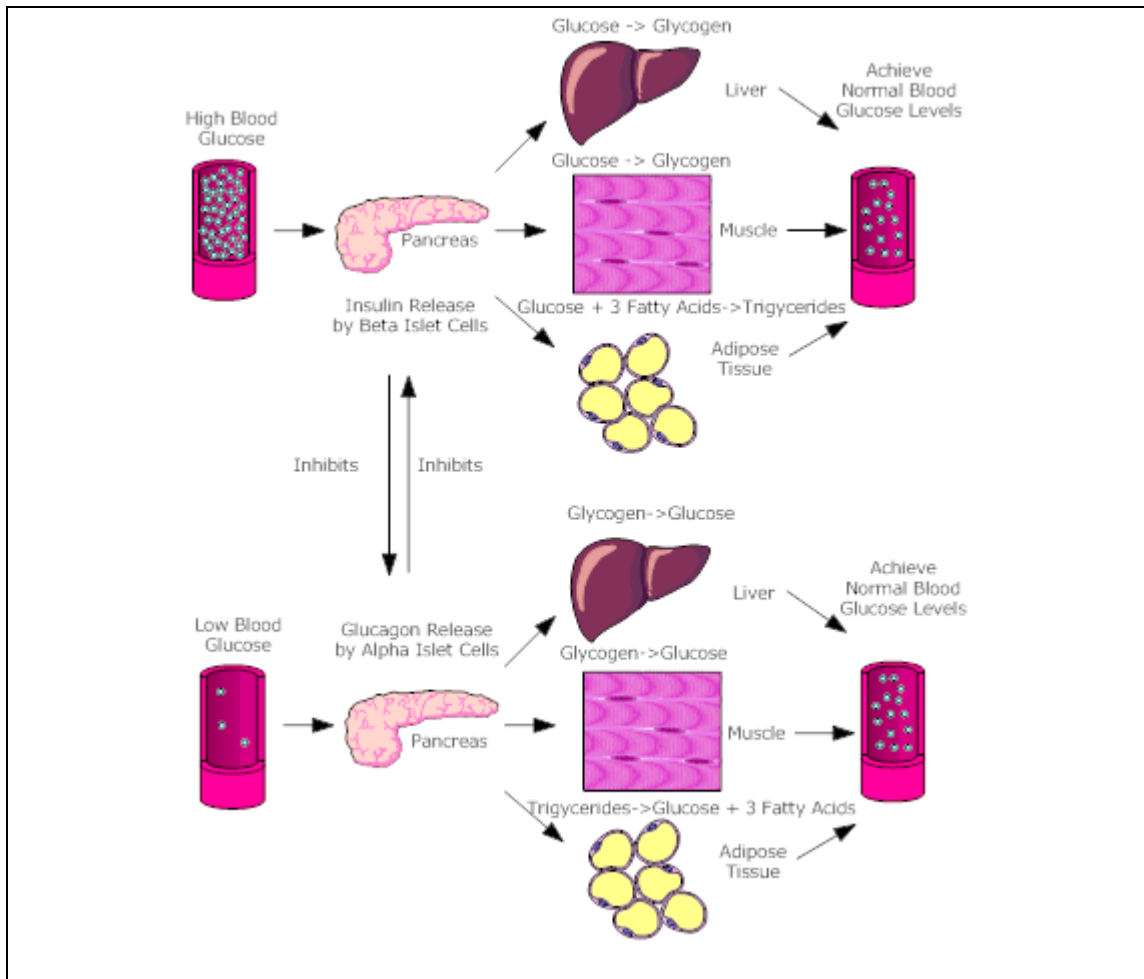
Fig 2.1: Glucose metabolism under normal physiological condition in human body. In humans the glucose metabolism is tightly regulated and blood glucose level is maintained in a very narrow range via regulated interaction between two pancreatic endocrine hormones insulin and glucagon. The picture above shows the role of pancreas and its secreted hormones in maintaining the glucose homeostasis in human body.
Adapted from http://sahely.com/medicine/normal_glucose_regulation.htm

Under normal physiologic conditions, glucose concentrations remain within a narrow range in the fasting as well as in the fed state. This tight glucose regulation is

Chapter 2

maintained by a delicate balance between insulin secretion and insulin sensitivity (Kahn, 2003). A hyperbolic relationship governs this balance (Fig 2.2), such as the product of insulin sensitivity and insulin secretion is a constant at a given glucose tolerance in any individual (Arslanian, 2005).
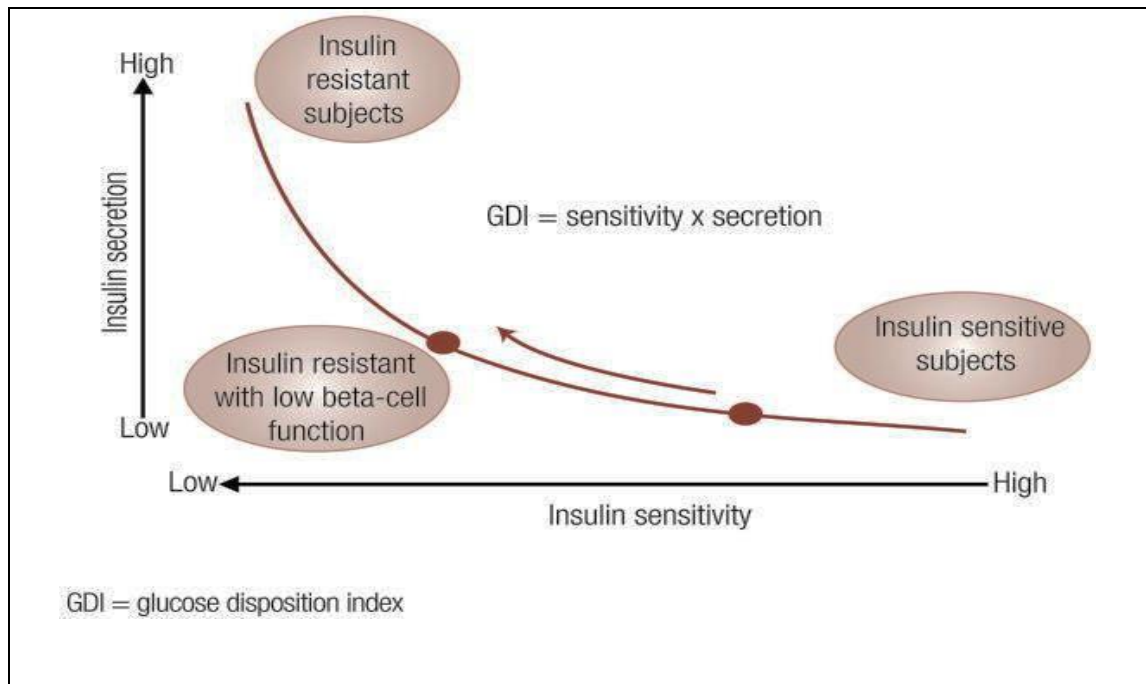
Fig 2.2: Hyperbolic relationship between insulin sensitivity and insulin secretion
A hyperbolic relationship governs the balance between insulin sensitivity and secretion at a given glucose tolerance in any individual. The right-hand side of the curve is representing high insulin sensitivity, so very minimal insulin resistance. In individuals with normal glucose tolerance (NGT), a decrease in insulin sensitivity of the peripheral tissues results in a compensatory increase in insulin secretion, and normoglycaemia is maintained. Failure of this compensatory response can lead to glucose intolerance and diabetes.
Adapted from (Arslanian, 2005; WHO, 1985)

2.4 Impaired Glucose Metabolism in T2DM

The chronic hyperglycaemia, an outcome of T2DM, is associated with long term morbidity and multiple organ failure. Insulin resistance in muscle, liver and β-cell failure represent the core pathophysiologic defects in type 2 diabetes. Infact the β-cell failure occurs much earlier and is more severe. Subjects in the upper tertile of

impaired glucose tolerance (IGT) are maximally/near-maximally insulin resistant and have lost over 80% of their β-cell function. In addition to the muscle, liver, and β-cell (triumvirate), the fat cell (accelerated lipolysis), gastrointestinal tract (incretin deficiency/resistance), α-cell (hyperglucagonemia), kidney (increased glucose reabsorption), and brain (insulin resistance) all play important roles in the development of glucose intolerance in T2DM individuals (Defronzo *et al.*, 2009). Fig 2.3 summarises the regulation and the metabolism of the glucose in diabetics and non-diabetics. Thus, the potentially important mechanisms associated with T2DM and insulin resistance include: insufficient insulin production from β-cell in the setting of insulin resistance, increased breakdown of lipids within fat cells, resistance to and lack of incretin, high glucagon levels in the blood, increased retention of salt and water by the kidneys, and inappropriate regulation of metabolism by the central nervous system. Fig 2.3 summarise the regulation and metabolism of glucose in both diabetics and non diabetics. Although both IGT and IFG are associated with resistance to insulin and increased insulin secretion, they do not identify the identical patient populations and are not equivalent in predicting development of T2DM or cardiovascular events. Both IFG and IGT have been associated with other features of insulin resistance, including dyslipidaemia, hypertension, abdominal obesity, microalbuminuria, endothelial dysfunction, and markers of inflammation and hypercoagulability, traits collectively referred to as the metabolic syndrome. Analyses of combinations of these components have also been associated with progression to T2DM, cardiovascular disease and increased mortality (Peterson., 2005)
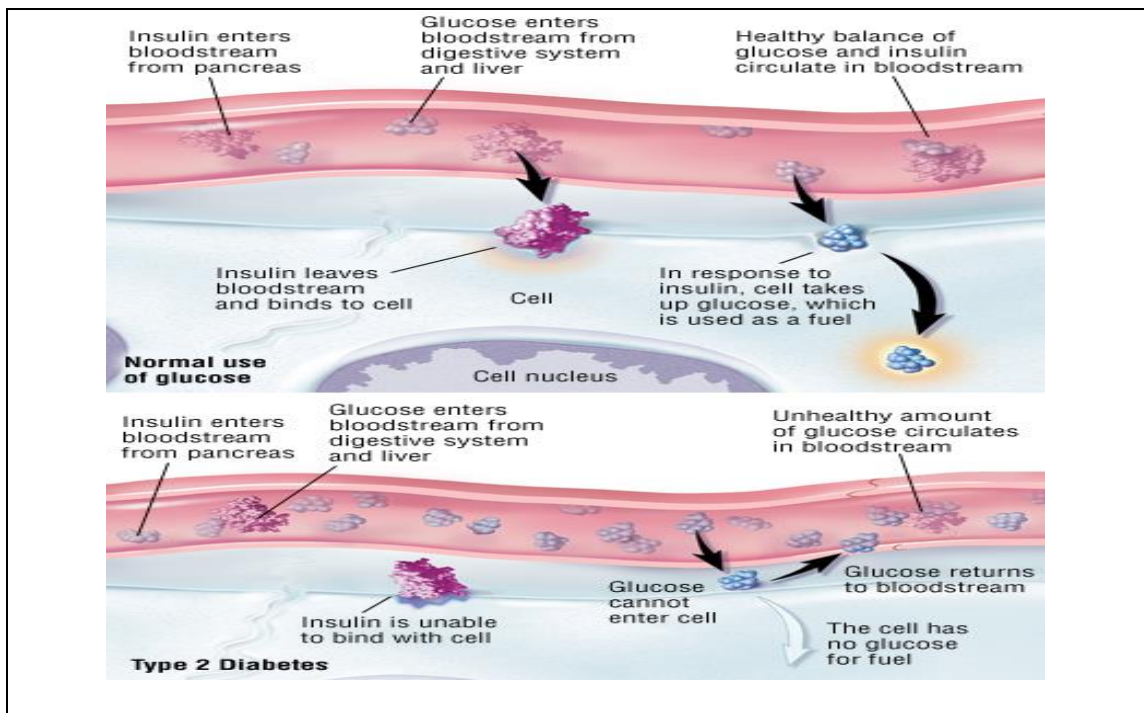
Fig 2.3: Normal vs. Impaired Glucose Metabolism
With the onset of diabetes, the body loses its ability to adequately utilize sugar. When this occurs, blood sugar levels increase due to the body's inability to transport sugar into the cells and out of the blood stream. The figure above projects the difference in blood sugar transport in diabetics as compared to normal individuals.
Adapted from http://www.drugs.com/health-guide/type-2-diabetes-mellitus.html

2.5 Risk Factors for Diabetes and Progression to T2DM

Risk factors for the development and/or progression of T2DM broadly include genetics/family history and/or both pre- and post-natal environmental factors, including suboptimal intrauterine environment, LBW (low birth weight), obesity, inactivity, gestational diabetes and advancing age. Each of these risk factors can, via largely undefined mechanisms, lead to skeletal muscle, adipose and hepatic insulin resistance, and/or β-cell dysfunction. Ultimately, insulin resistance accompanied by inadequate insulin secretory responses results in postprandial and fasting hyperglycaemia. In turn, diabetes-related hyperglycaemia and associated metabolic abnormalities can further alter signal transduction and gene expression (glucolipotoxicity), thus contributing to a vicious cycle (Jin & Patti, 2009).

Chapter 2

Both environmental and genetic factors play a crucial role in the pathogenesis of T2DM. There is almost 50% contribution by the lifestyle factors and 50% contribution by the genetics of an individual which is responsible for the disease. The lifestyle factors include obesity, lack of physical activity, stress, improper diet. The genetic factors include a wide range of polymorphism cumulatively acting for causing the disease (Fig 2.4).
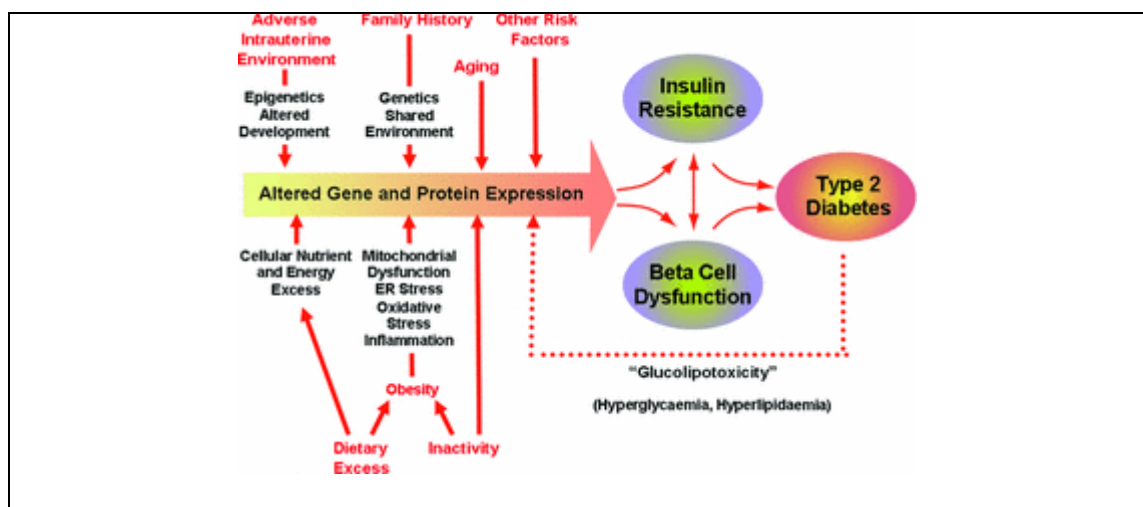
Fig 2.4: Risk Factors for T2DM
The figure above shows the multiple risk factors involved in the pathogenesis of T2DM, including classical genetic risk (family history) as well as a prominent contribution from multiple environmental risk factors
Adapted from Jin and Patti, 2009

2.5.1 Family History

The importance of genetic risk factors is exemplified by the high concordance of T2DM in identical twins (Poulsen *et al.,* 1999), the strong influence of family history and ethnicity on risk, and the identification of DNA sequence alterations in both rare and common forms of T2DM (Jin & Patti, 2009). Low SI (insulin sensitivity) and low SG (insulin-independent glucose effectiveness), both measures of glucose disposal, predict T2DM development several decades later in family-history-positive

individuals, but not in family-history-negative individuals, indicating that family history may contribute to specific alterations in systemic physiology distinct from other risk factors (Martin *et al.,* 1992; Goldfine *et al.,* 2003).

2.5.2 Intrauterine Environment

The nutritional environment during both pre- and post-natal life is increasingly recognized as a significant contributor to the pathogenesis of T2DM. Both epidemiological and animal studies have linked a suboptimal nutritional environment during pregnancy and LBW with increased susceptibility to diabetes and obesity during adult life (Chillaron *et al.,* 2005; Hovi *et al.,* 2007; Harder *et al.,* 2007). Such findings have led to the formulation of the developmental programming hypothesis, which proposes that insults or stimuli acting during critical windows of development, including fetal and/or early postnatal periods, can produce permanent alterations in cell/tissue structure and function (Jin & Patti, 2009; Chillaron *et al.,* 2007).

2.5.3 Obesity

Although increased intake of calorie-dense foods and decreased energy expenditure are clear contributors to the development of obesity, the mechanisms responsible for either obesity or its effects on systemic metabolism have not been completely elucidated (Jin & Patti, 2009). Adipose tissue is now recognized as an important endocrine tissue, which modulates systemic metabolism by releasing NEFAs (non-esterified fatty acids; 'free fatty acids') and glycerol, hormones (Graham *et al.,* 2006) and pro-inflammatory cytokines such as TNF-α (tumour necrosis factor-α), IL-6 (interleukin-6), and MCP-1 (monocyte chemoattractant protein-1) (Wellen *et al.,* 2005; Shoelson *et al.,* 2006; Fried *et al.,* 1998). With obesity, patterns of adipose

tissue metabolism are altered, with impaired production of adiponectin and increased release of NEFAs and pro-inflammatory cytokines, a pattern which may contribute to systemic insulin resistance. These relationships are particularly prominent for intra-abdominal adipose tissue, contributing to the heightened risk for T2DM associated with abdominal pattern obesity (Jin & Patti, 2009; Gesta *et al.,* 2007).

2.5.4 Aging

Life expectancy from birth has increased from approx. 45 years in the early 1900s to approx. 75 years for men and 80 years for women today (Oeppen & Vaupel, 2002). These trends have certainly contributed to the increasing prevalence of age-related diseases, including obesity and T2DM (Davidson, 1979), perhaps related to aging-associated loss of lean body mass and increased adipose tissue (Enzi *et al.,* 1986 ; Borkan *et al.,* 1983). In addition, pre-adipocyte capacity to replicate, differentiate and store lipids declines with age (Kirkland & Dobson, 1997). This too, may contribute to increased accumulation of ectopic lipids in muscle and liver, and induction of insulin resistance.

2.5.5 Inactivity

Decreased physical activity also contributes to T2DM risk (Manson *et al.,* 1991; Levine *et al.,* 2005). Exercise is critical for maintenance of muscle oxidative function and systemic insulin sensitivity. Both acute exercise and chronic exercise training increase mitochondrial gene expression and oxidative capacity, Via pathways activating AMP kinase, calcium/calmodulin dependent protein kinase and the expression of the PGC-1 (PPAR (peroxisome-proliferator-activated receptor) γ co-

activator-1) family of genes controlling oxidative mitochondrial function (Jin & Patti, 2009; Kuhl *et al.,* 2006; Russell AP *et al.,* 200).

2.6 Systemic Metabolic Defects in T2DM

2.6.1 Inflammation

Although it is clear that tissue-specific defects contribute to T2DM pathophysiology, studies have emphasized common molecular events mediating defects in multiple tissues. Adipose tissue is probably the central player and site of activation of systemic inflammation. More than 10 years ago, in1993 Hotamisligil *et al*. demonstrated that adipocytes produced TNF-α and could contribute to insulin resistance. Subsequent studies from multiple investigators have shown that high-fat feeding and obesity can induce inflammation within adipose tissue (Xu *et al.,* 2003 and Weisberg *et al.,* 2003), with production of IL-6, MCP-1 and other inflammatory mediators. At a molecular level, this may be mediated through increased adipocyte death and remodelling (Strissel *et al.,* 2007) leading to recruitment of pro-inflammatory effector cells, increased oxidative stress and increased activity of transcription factors, including NF-κB, AP-1 (activating protein-1) and EGR-1 (early growth response-1) (Aljada *et al.,* 2004).

2.6.2 Endoplasmic Reticulum (ER) Stress

Newly synthesized proteins are folded to attain their proper three-dimensional structures and undergo additional post-translational modification inside the ER, a network of intracellular membranes. Obesity, viral infections, toxins and other environmental stressors can all trigger ER stress (Gregor *et al.,* 2007). Unfortunately, the resultant cascades also disrupt insulin signalling, a feature potentially exacerbated

by the reductions in heat-shock chaperone proteins in T2DM (Kurucz *et al.,* 2002). ER stress may also contribute to impaired β-cell function, as demonstrated during the progression of both T1DM (Type 1 diabetes mellitus), T2DM, and Wolfram syndrome (Lipson *et al.,* 2006; Fonseca *et al.,* 2005, Jin & Patti, 2009).

2.6.3 Mitochondrial Dysfunction

Over the past few years, multiple lines of experimental evidence have converged to provide support for the hypothesis that mitochondrial dysfunction is intimately linked with the complex pathophysiology of T2DM. Genomic analysis of skeletal muscle biopsy samples from diverse populations has demonstrated that a dominant pattern associated with diabetes is reduced nuclear-encoded mitochondrial gene expression (Patti *et al.,* 2003; Mootha *et al.,* 2003; Sreekumar *et al.,* 2002). These findings have been extended by enzymatic and NMR spectroscopy studies further linking T2DM with decreases in mitochondrial oxidative activity (Kelley *et al.,* 2002), impaired basal and insulin-stimulated ATP synthesis (Petersen *et al.,* 2004) and reduced numbers of subsarcolemmal mitochondria (Kelley *et al.,* 2002). Importantly, similar patterns are also observed in some populations of insulin-resistant, but completely normoglycaemic, individuals (Patti *et al.,* 2003 and Morino *et al.,* 2005). Thus mitochondrial oxidative dysfunction is a key feature of T2DM, potentially contributing to metabolic inflexibility, reduced lipid oxidation and increased accumulation of intramyocellular lipid (Boushel *et al.,* 2007).

2.7 Symptoms of T2DM

Chapter 2

The various symptoms of marked hyperglycaemia include polyuria, polydipsia, weight loss, sometimes with polyphagia, and blurred vision (ADA, 2009) (Table 2.1).

Table 2.1: Symptoms of T2DM and their association with the elevated blood glucose levels

Symptoms	Cause
Polydypsia and Polyurea	Excessive level of unutilised Glucose is spilled into urine by kidney.
Extreme hunger, fatigue and weight loss	During T2DM the body is unable to utilise the excess glucose and hence results in the feeling of hunger and fatigue, despite of consuming ample food. The resultant lack of cell nutrition is a cause of thus observed weight loss.
Headaches, dizziness.	Shortage of energy (due to inability to utilise glucose by the body) supply to brain causes headaches and tiredness.
Dry itchy skin	Lack of gamma-linolenic acid (GLA), a resultant of metabolic process during diabetes, causes dry itchy skin seen in diabetics.
Blurred vision	High blood glucose levels damage the blood vessels in different organs of the body. This usually starts from retina of the eye and kidney. Either these blood vessels are lost, or are leaky which causes blood and fat to seep out of damaged vessels, leading to swelled retina and blurry vision. This is also called as Diabetic Nephropathy.
Tingling pain in the feet	The excessive glucose level in the body causes the degeneration of peripheral nerve fibers throughout the body leading to a lack of sensation feeling in the feet followed by legs, hands and muscles. The auto healing of these nervous fibers causes the needle like pain and tingling in the various body parts.

Impairment of growth and susceptibility to certain infections may also accompany chronic hyperglycaemia. Acute, life-threatening consequences of uncontrolled diabetes are hyperglycaemia with ketoacidosis or the nonketotic hyperosmolar

syndrome. The various long-term comorbidities associated with diabetes primarily include obesity and dyslipidaemia, retinopathy with potential loss of vision; nephropathy leading to renal failure; peripheral neuropathy with risk of foot ulcers, amputations, and Charcot joints; autonomic neuropathy causing gastrointestinal, genitourinary, and cardiovascular symptoms and sexual dysfunction (ADA, 2009).

2.8 Diagnostic criteria for T2DM

The established glucose criteria for the diagnosis of diabetes include the FPG and 2-h Post Prandial Glucose (Table 2.2). Additionally, patients with severe hyperglycaemia such as those who present with severe classic hyperglycaemic symptoms can continue to be diagnosed when a random (or casual) plasma glucose of >200 mg/dl (11.1 mmol/l) is found. It is likely that in such cases the health care professional would also measure an HbA1c test as part of the initial assessment of the severity of the diabetes and that it would (in most cases) be above the diagnostic cut point for diabetes. The following tests are used for diagnosis of T2DM (WHO, 2006; Vijan *et al.,* 2010).

- A **fasting plasma glucose (FPG)** test measures blood glucose in a person who has not eaten anything for at least 8 hours. This test is used to detect diabetes and pre-diabetes.

- An **oral glucose tolerance test (OGTT)** measures blood glucose after a person fasts at least 8 hours and 2 hours after the person drinks a glucose-containing beverage. This test can be used to diagnose diabetes and pre-diabetes.

- A **random plasma glucose test**, also called a casual plasma glucose test, measures blood glucose without regard to when the person being tested last

ate. This test, along with an assessment of symptoms, is used to diagnose diabetes but not pre-diabetes.

Table 2.2: Diagnostic criteria for Diabetes

Condition	2 hour glucose	Fasting glucose	Random glucose	HbA$_{1c}$
	mmol/l(mg/dl)	mmol/l(mg/dl)	mmol/l(mg/dl)	%
Normal	<7.8 (<140)	<6.1 (<110)	<11 (<200)	<6.0
IFT	<7.8 (<140)	≥ 6.1(≥110) & <7.0(<126)	-	6.0–6.4
IGT	≥7.8 (≥140)	<7.0 (<126)	-	6.0–6.4
DM	≥11.1 (≥200)	≥7.0 (≥126)	≥11 (≥200)	≥6.5

IFT: Impaired fasting glycaemia; IGT: Impaired glucose tolerance; DM: Diabetes Mellitus; PG: Post parandial glucose; HbA1c: Glycosylated Haemoglobin

2.9 Epidemiology of T2DM

2.9.1 International Status

Non-communicable diseases (including diabetes) account for 60% of all deaths worldwide (Arslanian, 2005). In the past three decades, the global prevalence of diabetes (8.3%) has doubled, as a result of population ageing, urbanization and associated lifestyle changes making it one of the most important public health challenges across the globe. In addition to the early onset of T2DM in young adults, an increasing trend of T2DM and pre-diabetes is noticeable among children and adolescents. This epidemic is due to complex group of genetic and epigenetic systems interacting within an equally complex societal framework that determines behaviour and environmental influences. In 2010, an estimated 285 million people worldwide had diabetes mellitus (Aronoffd, 2004), 90% of whom had type 2 diabetes mellitus (T2DM). The number of people globally with diabetes mellitus is projected to rise to 552 million by 2030, which represents 9.9% of the total adult population of the world aged 20–79 years (IDF, 2009). Some 4.6 million people 20-79 years of age died from diabetes in 2011, accounting for 8.2% of global all-cause mortality of people in this

age group (IDF, 2011). The number of deaths attributable to diabetes in 2011 shows a 13.3% increase over the estimates for the year 2010 which is actually very high leading to the raised alarm among the health fraternity globally (IDF, 2011).

2.9.2 National Status

Few decades back T2DM was relatively rare in developing countries, however, higher rates observed in Asian Indians and Asian Indian immigrants in Western countries strongly predicted the potential epidemic of T2DM that has now emerged in mainland China and India. The major burden of diabetes mellitus is now taking place in developing rather than in developed countries. 80% of cases of diabetes mellitus worldwide live in less developed countries and areas (Shaw *et al.,* 2010). Asia has emerged as the 'diabetes epicentre' in the world, as a result of rapid economic development, urbanization and nutrition transition over a relatively short period of time. Among the 10 countries with the largest numbers of people predicted to have diabetes mellitus in 2030 (Fig 2.5), five are in Asia (China, India, Pakistan, Indonesia and Bangladesh) (Shaw *et al.,* 2010). Compared with developed countries, the proportion of young to middle-aged individuals with T2DM is higher in developing countries (Shaw *et al.,* 2010). Furthermore, T2DM is not necessarily less prevalent in rural than in urban areas of developing countries, as is generally believed. The rural–urban difference in prevalence is predicted to narrow owing to urbanization, rural to urban migration and its associated lifestyle changes. A study from India showed a significant increase in diabetes mellitus prevalence in both urban (from 13.9% in 2000 to 18.2% in 2006) and rural areas (from 6.4% in 2000 to 9.2% in 2006) (Ramachandran, *et al.,* 2008). Similar findings have been reported from other Asian

Chapter 2

countries (Chan *et al.,* 2009). In 2011, India had 61.3 million people with type 2 diabetes, compared with 50.8 million the previous year, according to the International Diabetes Federation (IDF) (2011) and the Madras Diabetes Research Foundation, a figure which is projected to rise to 101.2 million by 2030 (IDF, 2011). As the economy started growing, so did the incidence of diabetes. The nationwide prevalence of diabetes in India now tops 9%, and is as high as 20% in the relatively prosperous southern cities. Health experts are alarmed because, although the onset of type 2 diabetes tends to affect people in the West in their 40s and 50s, the disease strikes Indians much younger. Indians as young as 25 are being diagnosed with the disease, a trend that threatens to seriously hamper the country's economic development (Shetty, 2012). Although the exact reasons why Asian Indians are more prone to type 2 diabetes at a younger age and premature cardiovascular disease (CVD) remain speculative, there is a growing body of evidence to support the concept of the "Asian Indian Phenotype'' (Mohan *et al.,* 2005). This term refers to the peculiar metabolic features of Asian Indians characterized by a propensity to excess visceral adiposity, dyslipidaemia with low HDL cholesterol, elevated serum triglycerides and increased small, dense LDL cholesterol, and an increased ethnic (possibly genetic) susceptibility to diabetes and premature coronary artery disease (Mohan *et al.,* 2005; Deepa *et al.,* 2006; Joshi, 2003)
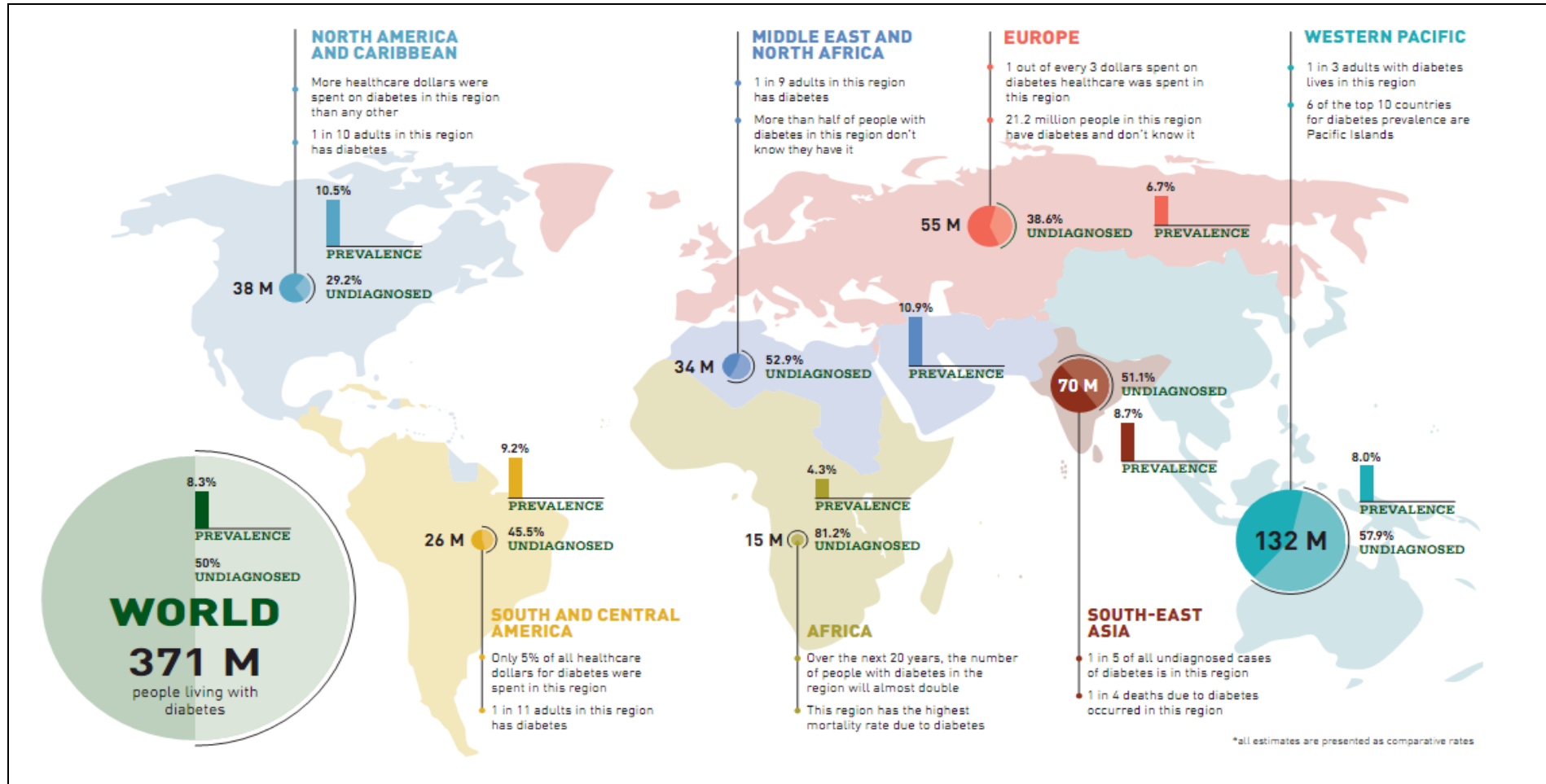
Fig 2.5: Global prevalence of T2DM. The figure presents the Global prevalence of T2DM along with the list of top 10 countries for numbers of people with diabetes in millions. All but two of these countries are middle-income countries and rapidly developing. Combined, these countries make up 75% of the total prevalence of diabetes in the world. *Adapted from IDF Diabetes Atlas 2009 and IDF Diabetes Atlas 2012*

Chapter 2

2.10 Economic Toll Due to T2DM

The financial burden borne by people with diabetes and their families as a result of their disease depends on their economic status and the social insurance policies of their countries. In the poorest countries, people with diabetes and their families bear almost the whole cost of the medical care they can afford (IDF, 2009).

The global expenditures on diabetes $418 billion in 2010, $465 in 2011 and at least $595 billion estimated for the year 2030. An estimated average of $878 per person was estimated to be spent on diabetes in 2010 globally. Besides excess healthcare expenditure, diabetes also imposes large economic burdens in the form of lost productivity and foregone economic growth. The largest economic burden is associated with the long term morbidity and disability associated with the disease along with the actual death toll due to the disease itself and associated complications (Fig 2.6). The World Health Organization (WHO) predicted net losses in national income from diabetes and cardiovascular disease totalling to ~$2000 with $336.6 billion in India, between 2005 and 2015 (IDF, 2009). The global healthcare expenditures on diabetes was 11% of total healthcare expenditures in adults (20-79 years) worldwide and is expected to rise to ID 654 billion in 2030 (IDF, 2011).

Only 20% of global healthcare expenditures due to diabetes were made in low- and middle-income countries. On average, the estimated healthcare spending due to diabetes was USD 5,063 per person with diabetes in high-income countries compared to USD 271 in low- and middle-income countries (IDF, 2011). Close to one-fifth of all adults with diabetes in the world live in the South-East Asia Region (Fig 2.7).
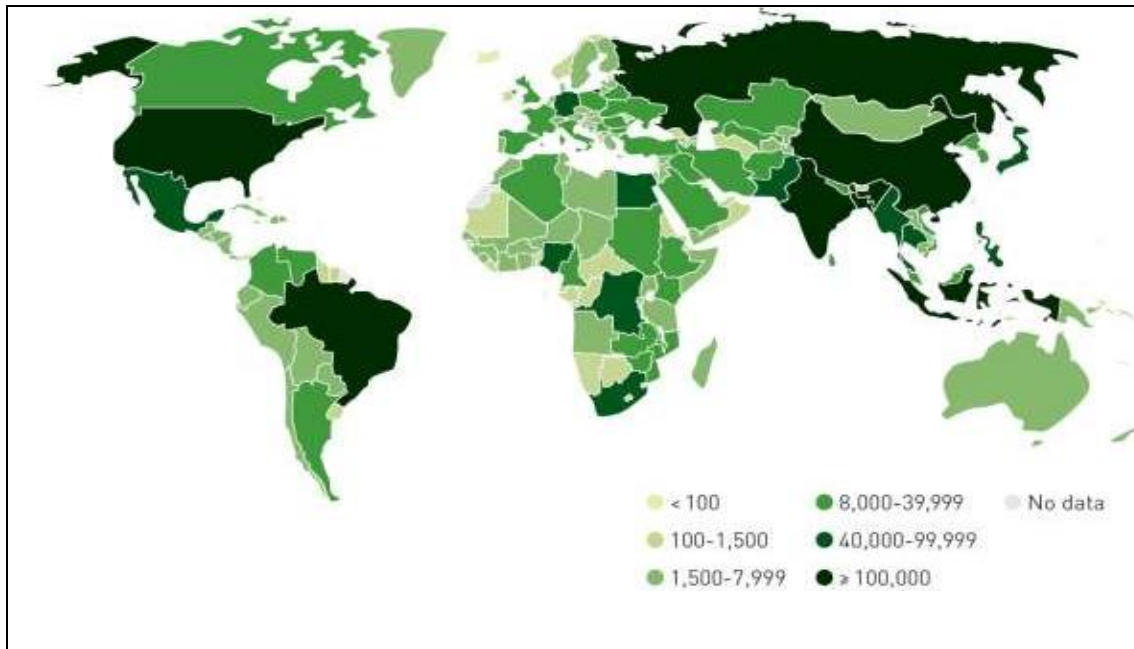
Fig 2.6: Deaths attributable to Diabetes (20-79 years) in year 2011
Adapted from IDF, 2011
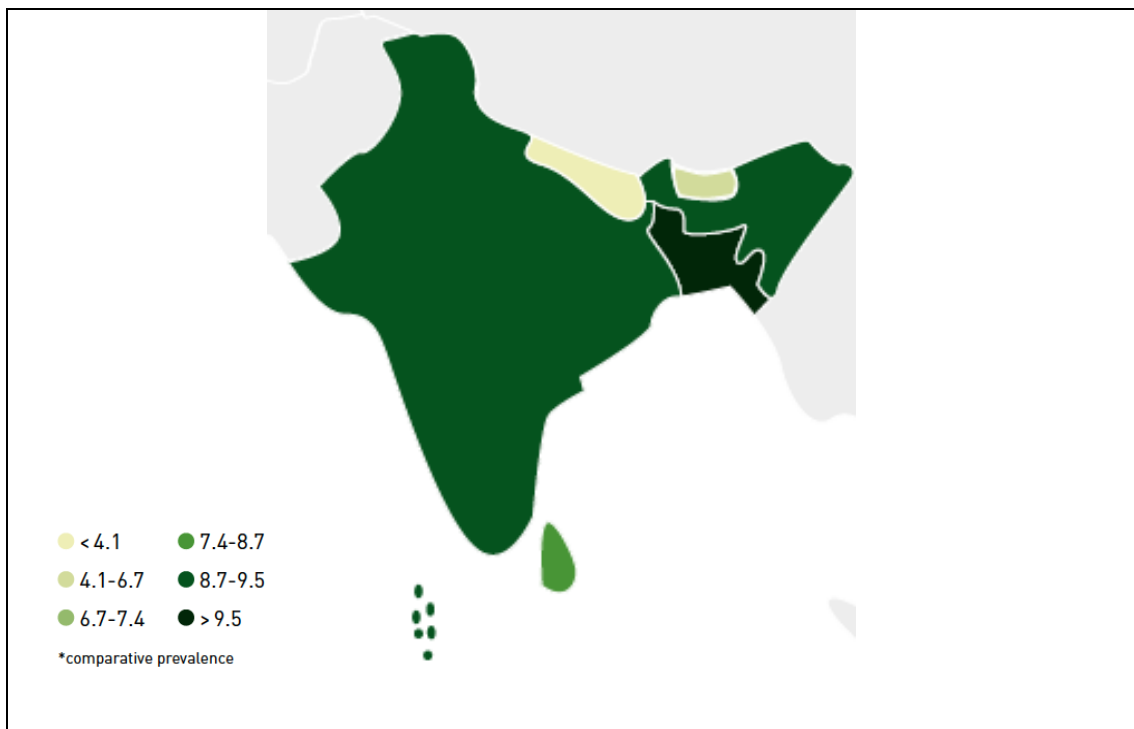
Fig 2.7: Prevalence estimates of T2DM in South East Asia
Adapted from IDF, 2011
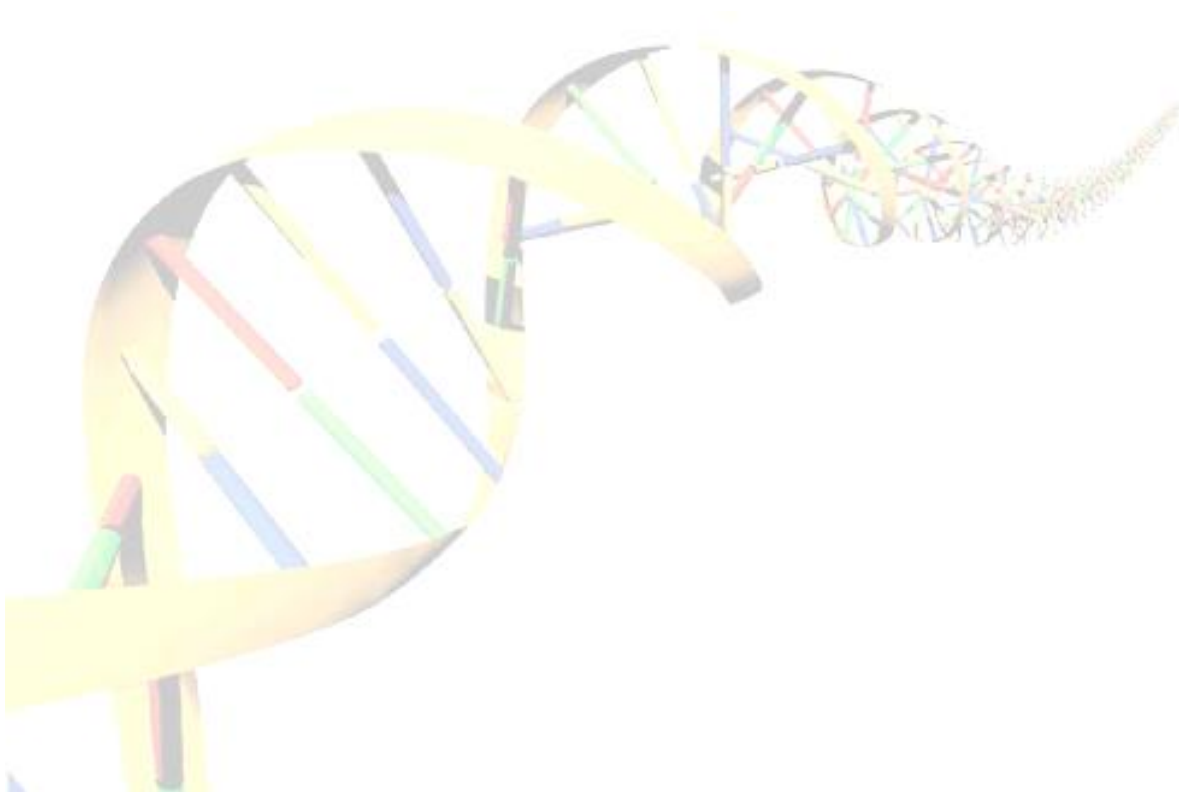
Chromosome 2, 3 and 10: Genetic studies on T2DM till date

Chapter 3

Chromosome 2, 3 and 10: Genetic studies on T2DM till date

The evidence for a role of genetic factors in susceptibility to common, multifactorial type 2 diabetes (T2DM) has been known for decades. A lot of familial studies (Relative risk in the families of the affected individuals) have shown the role of these genetic factors in T2DM (Mlynarski, 2012). An extensive effort and money exchequer has gone in last few years to catalogue, analyse and correlate the human genetic variations with the observed phenotypic differences in T2DM. Especially with the advent of genome wide association scans (GWAS) most of the polymorphisms have been assessed for their significant association with the complex phenotypic traits and multifactorial diseases like T2DM. Despite the amount of information generated by such huge genome wide scans and multicentre powerful studies, only a limited amount of results are reproducible. Although these studies have provided new biological insights, only a limited amount of the heritable component of any complex trait has been identified and it remains a challenge to elucidate the functional link between associated variants and phenotypic traits (Frazer *et al.,* 2009). Complex diseases such as T2DM are likely to be influenced by incomplete penetrance, genotype-by-environment interactions and multilocus effects, in addition to genetic heterogeneity making identification of genes by linkage analysis less successful (Stambolian *et al.*, 2004, Pezzolesi *et al.,* 2004). Efforts to identify the genetic factors underlying susceptibility to T2DM have used both candidate gene approaches and linkage studies (Stern, 2002). More than a dozen genome wide scans have been completed from which various chromosomal regions with suggestive or significant evidence of linkage with type 2 diabetes have emerged. However, despite intense

Chapter 3

efforts to clone the susceptibility genes, the identification of these genes has been unsuccessful, with the single exception of calpain-10 (CAPN10) or NIDDM1 (Horikawa *et al.,* 2000). So there is need to explore more about genetic basis of T2DM. Factors contributing to the failure to replicate linkage results include the presence of environmentally determined phenocopies of T2DM as well as the complex nature of the inheritance of type 2 diabetes. The latter includes genetic and allelic heterogeneity, epistasis (gene-to-gene interaction), and gene-to-environment interactions (Pezzolesi *et al.,* 2004). For this, strategies based on analysing first the specific region of chromosome implicated in disorder where T2DM is co-morbid, as well as region with QTLs that regulate body composition parameters associated with T2DM should be considered to hunt the locus or causative gene/s.

In the following review we will analyse three such loci namely 2q32-37, 3q22-29 and 10q25-26 for the possibility of locating region harbouring the causative regions for T2DM.

3.1 Chromosome 2 (region 2q32-37) involvement in T2DM

The search for the genetic basis of T2DM is on-going, generating an abundance of new data. These data consist of a large number of candidate genes, association of previously known and novel candidate genes with various facets of diabetes, detection of new quantitative trait loci and identification of genes that mediate susceptibility to diabetes. In the present study, for a better understanding of the pathogenesis of type 2 diabetes mellitus, an approach has been taken where the number of screened markers can be narrowed down by literature review and the selected panel of STRs can then be studied for their association with T2DM. One such region identified during literature

Chapter 3

review was the region 2q32-37. Chromosome 2 is the second largest human chromosome, spanning more than 243 million base pairs. The genetic distance for this region, 2q32-37 is roughly 60M bp (183M-243M bp) and containing nearly 628 genes according to NCBI (build 37.1). However, the number of genes involved, their chromosomal location and the magnitude of their effect on NIDDM susceptibility are still unknown.

3.1.1 Genetic studies on association of 2q32-37 region with T2DM

Early in this decade, studies across various ethnicities have reported moderate (suggestive) association (Odds 1.25 – 2.3) of few STRs located in the region 2q32-37 with T2DM (Vionnet *et al.,* 2000; Ehm *et al,* 2000*;* Wiltshire *et al*, 2001; Mori *et al.,* 2002; Larkin *et al.,* 2004). Although the ethnicities mainly studied were the European, Caucasians, Japanese, African Americans, Whites, Finnish and Mexicans, very few studies tried to study the Asian population and almost none looked at the Indian population (specifically the north Indians). Evidence suggest that the genetic basis of several diseases in Indians might be different from that of Europeans which could be due to differences in the risk allele frequency and pattern of linkage disequilibrium. Infact the report from the Indian Genome Variation Consortium also suggested that most of the populations in the Indian subcontinent are distinct from HapMap populations (Chauhan *et al.,* 2012).

Apart from the reported associations of the STR repeats in this region in last five years few candidate genes have also been shown to be strongly associated with T2DM and associated risks. The five major ones include ABCA12 (Li *et al.*, 2008; Li *et al.*, 2008), IGFBP5 (Kallio *et al.,* 2008; Narayanan *et al.*, 2011), IGFBP2 (Narayanan *et al.*, 2012), IRS1 (Rung *et al.,* 2009; Teslovich *et al.,* 2010; Billings *et al.*, 2010;

Chapter 3

Yiannakouris *et al.*, 2011; Sharma *et al.*, 2011) and CAPN10 (Ezzidi *et al.,*2010). Although apart from the gene CAPN10 most of these associations still need to be confirmed in other ethnicities. In case of the calpain-10 gene (CAPN10) also many groups have tried to replicate the association of CAPN10 with T2DM but could establish only no (Baier *et al.* 2000; Evans *et al.* 2001; Hegele *et al.* 2001; Tsai *et al.* 2001; Xiang *et al.* 2001; Daimon *et al.* 2002; Fingerlin *et al.* 2002; Malecki *et al.* 2002; Rasmussen *et al.* 2002; Sun *et al.* 2002; Horikawa *et al.* 2003; Iwasaki *et al.* 2003) to moderate (Schwarz *et al.* 2001; Cassell *et al.* 2002; Garant *et al.* 2002; Orho-Melander *et al.* 2002) association of this SNP with T2DM.

Table 3.1 summarizes all the genetic studies done on T2DM in this region with the physical position of marker/s forming the disease haplotype and the marker with peak LOD scores are indicated. Reports show that that mutations reported in the potential candidate genes for diabetes do not account for significant proportion of cases studied. This suggested the presence of novel genes or sequence elements that are yet to be discovered at this chromosomal region. Table 3.1 summarises the list of studies (both candidate and linkage) which have tried to study this region for its association with T2DM and which have reported a LOD of >1.2 (moderate association) and a significant *p* value.

Table 3.1: Genetic studies on Type 2 Diabetes Mellitus (T2DM) with LOD score ≥ 1.2 (*p*<0.01) within 2q32-37

Chr. loci	Gene	SNP/variation	Population/Phenotype	*p*-value (OR)	References
2q34	-	D2S325	Japanese / T2DM risk	0.00366 (1.79)	Mori *et al.,* 2002
2q35	-	D2S126	Chinese/T2DM	<0.05 (2.1)	Luo *et al.,* 2001
2q36.3	-	D2S396	British/ T2DM	0.007 (1.26)	Wiltshire *et al.,* 2001
2q36.3	-	D2S396	Japaneses/ T2DM risk, young onset	0.004 (1.48)	Mori *et al.,* 2002
2q36.3	IRS 1	rs2943641	European/T2DM	0.056 (0.89)*	Yiannakouris *et al.,* 2011
2q36.3	IRS1	rs1801278	Mexico/ T2DM	2.43	Gómez *et al.,* 2011
2q36.3	IRS 1	rs6725556	European/T2DM	0.015 (0.82)*	Yiannakouris *et al.,* 2011
2q37.3	-	D2S125	Indo-Mauritius/T2DM	<0.05 (3.0)	Francke *et al.,* 2001
2q37.3	CAPN10	UCSNP-43	German populations/ T2DM	<0.05 (4.98)	Schwarz *et al.,* 2001
2q37.3	CAPN10	UCSNP-43	Czech populations/ T2DM	<0.05 (2.80)	Schwarz *et al.,* 2001
2q37.3	CAPN10	UCSNP-43	Mixed /T2DM	<0.05 (6.52)	Cassell *et al.,* 2002

* Represents the odds and the significance calculated are as per the minor or the protective allele (G allele).
The shaded area in the table represents the studies assessing the association of STR repeats in this region of interest.
Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio

Chapter 3

3.1.2 Anthropometric or Metabolic markers associated with T2DM in region 2q32-37

Searching for quantitative trait loci (QTLs) that explain the variation in the "intermediate" phenotypes of T2DM along with evaluating the direct association with T2DM has been considered as a more powerful approach to dissect the genetic factors involved in different pathogenetic pathways that lead to T2DM. Epidemiologic studies have showed an increased risk of T2DM with higher body weight, BMI (measure of overall obesity), and waist-to-hip ratio (WHR-- measure of abdominal adiposity) (Hennis *et al.,* 2002; Jacques *et al.,* 2003 & Cheung *et al.,* 2007). Likewise, the relationship of blood pressure with T2DM was observed to be moderate and subject to confounding by other risk factors (Schaumberg *et al.*, 2001). Number of metabolic syndrome associated phenotypes like LDL (Bossé *et al.,* 2003; Deng & Shen, 2007), apoB (Bossé *et al,* 2004) total cholesterol (Deng & Shen, 2007; Dong *et al*. in 2008) FBG (Li *et al,* 2004) have been mapped to 2q33.3 and 2q37.2-q37.3. Associations of this region with BMI (Guo *et al.,* 2006), percentage fat mass (Guo *et al.,* 2006) and essential hypertension (Yatsu *et al.,* 2007) have also been reported. Infact human homologue to mouse QTLs for HDL-C levels were also mapped between 2q23–31 and 2q32 (Wang & Paigen, 2005). This region has also been reported to be associated with the important cytokines like adiponectin (Kallio *et al.,* 2008; Dastani *et al.*, 2012), leptin (Larkin *et al.*,2004). A specific mutation of IRS1 gene has also been shown to be associated with hyperinsulinemia and insulin resistance (Rung *et al.,* 2009) while a nearby marker D2S434 was shown to have a moderate association with fasting plasma glucose (Li *et al.*, 2004). Table 3.2 and 3.3 summarises the studies associating various anthropometric and demographic QTLs in this region reported to be associated with T2DM.

Table 3.2: Summary of genome scans on anthropometric QTLs with LOD score $\geq$ 1.2 (*p*<0.05) in the region 2q32-37

Chr. loci	Gene	SNP/variation	Population/Phenotype	*p*-value (OR)	References
2q33.3	-	D2S1384-D2S2944	White/BMI	<0.05 (1.19)	Wu *et al.*, 2002
2q33.3	-	D2S1353-D2S2944	Japanese/BMI	.008 (1.28) .005 (1.45)	Iwasaki *et al.*, 2003
2q34	ABCA12	rs4673937	BMI BMI adj TC %fat Waist	$7.6 \times 10^{-7\#}$ $0.00007^{\#}$ $0.001^{\#}$ $0.003^{\#}$	Dong *et al.*, 2008
2q35	IGFBP2	Circulating IGFBP2	Caucasian /DBP	0.037(1.14)	Narayanan *et al.*, 2012

\# Represents the linkage P value obtained by multi/univariate analysis.
  The shaded area in the table represents the studies assessing the association of STR repeats in this region of interest.
  Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio

Table 3.3: Summary of genome scans on various QTLs with LOD score $\geq$ 1.2 (*p*<0.05) in the region 2q32-37

Chr. loci	Gene	SNP/variation	Population/Phenotype	*p*-value (OR)	References
2q33.3	-	D2S1384	French-Canadian families/ LDL	<0.05 (2.3)	Bossé *et al.,* 2003
2q33.3	-	D2S1384	French-Canadian families/ Total apo B Lipoprotein	$0.0044^{\#}$	Bossé *et al.,* 2004
2q33.3	-	D2S1384 - D2S2944	African Americans/ Obesity, Leptin levels	.00170 ($\geq$3)	Larkin *et al.*,2004

\# Represents the linkage P value obtained by multi/univariate analysis
  The shaded area in the table represents the studies assessing the association of STR repeats in this region of interest.
  Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio

Table 3.3: Summary of genome scans on various QTLs with LOD score $\geq$ 1.2 ($p<0.05$) in the region 2q32-37 (Contd.)

Chr. loci	Gene	SNP/variation	Population/Phenotype	p-value (OR)	References
2q34	ABCA12	rs1980846 rs4673937	Cholesterol FBG	0.0001# 0.00009#	Li *et al.*, 2007
2q34	-	D2S2944	TC LDL	<0.00001 (4.36) 0.003 (1.65)	Deng & Shen, 2007
2q35	-	D2S434	Mixed (obese)/ FPG	<0.05 (1.58)	Li *et al.*, 2004
2q35	IGFBP5	rs9341234	Finnish/ Adiponectin level	<0.05	Kallio *et al.*, 2008
2q35	IGFBP5	rs1978346	Caucasian / Circulating IGFBP2	0.005#	Narayanan *et al.*, 2011
2q36.3	IRS 1	rs2943641	Multiple populations/ HDL TG	$3 \times 10^{-6\#}$ $7 \times 10^{-6\#}$	Teslovich *et al.,* 2010
2q36.3	IRS 1	rs2943641	Whites (Americans)/ HDL TG	0.0045 0.018	Sharma *et al.*, 2011
2q36.3	IRS1	rs935735	Multi ethinicities/Adiponectin levels	$1.88 \times 10^{-8}(\beta=.02)$	Dastani *et al.*, 2012
2q37.1	-	D2S172-D2S427	Mexican American/ HDL	(1.30)	Arya *et al.* 2002

Represents the linkage P value obtained by multi/univariate analysis

The shaded area in the table represents the studies assessing the association of STR repeats in this region of interest.

Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio

Chapter 3

3.1.3 Association of 2q32-37 with other systemic disorders related to T2DM

Type 2 diabetes mellitus and obesity are also considered to be closely related. The region 2q33.3 was mapped and shown to be associated with obesity and altered leptin levels with almost a 3 fold increased risk. Several population based studies have associated metabolic syndrome, its components with an increased risk of T2DM. Type 2 diabetes is now widely considered to be one component within a group of disorders called the metabolic syndrome. Tang *et al.*, 2003 mapped significant genetic and phenotypic correlation underlying the clustering of traits involved in the metabolic syndrome on between the interval 223 - 246cM on chromosome 2q. The clustering of risk factors for diabetes and cardiovascular disease (CVD) has been known for many decades. The marker D2S162 in this region was reported to be associated with CHD (Francke *et al.,* 2001). Regions on chromosome 2q33 and 34 are suggested to be in linkage with early onset Alzheimer's disease (onset between 50 and 60 years) and rheumatoid arthritis as well (Scott *et al.*, 2003; Amos *et al.*, 2006; Remmers, *et al.*, 2007). Other disorders which are associated with adulthood, like lifetime alcoholism (Davis *et al.,* 1998), depression (Hmuely-Dulitzki *et al.,* 1997), Alzheimer's (Melov *et al.*, 2005) and rheumatoid arthritus (Mazzantini *et al.*, 2010) showed increased prevalence of T2DM. The risk increased with increasing alcohol consumption. Major depressive disorder (MDD) co-morbid with alcoholism was also reported to have a LOD of 2.2 between, 237–248 cM at chromosome 2 (Nurnberger *et al.*, 2001; Zubenko *et al.*, 2003; Kapoor *et al.* 2009). Table 3.4 summarizes linkage signals at markers in studies done on systemic disorders where T2DM is co-morbid.

Table 3.4: Summary of genome scans on systemic disorders with T2DM as co-morbidity and LOD score $\geq$ 1.2 (p<0.05) between the   region 2q32-37.

Chr. loci	Gene	SNP/variation	Population/Phenotype	p-value (OR)	References
2q33.3	-	D2S1384- D2S2944	African Americans/ Obesity	.00170 ($\geq$3)	Larkin *et al.*,2004
2q34	-	D2S2944	Indians/ Depression	0.02(1.43)	Kapoor *et al.*, 2009
2q37.1	-	D2S427-D2S1279	Marshfield & Utah / MMS	.0004 (3.34)	Tang *et al.*, 2003

The shaded area in the table represents the studies assessing the association of STR repeats in this region of interest.

Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio

Fig 3.1 Summary of the reported direct and indirect associations of the region 2q32-37 with T2DM
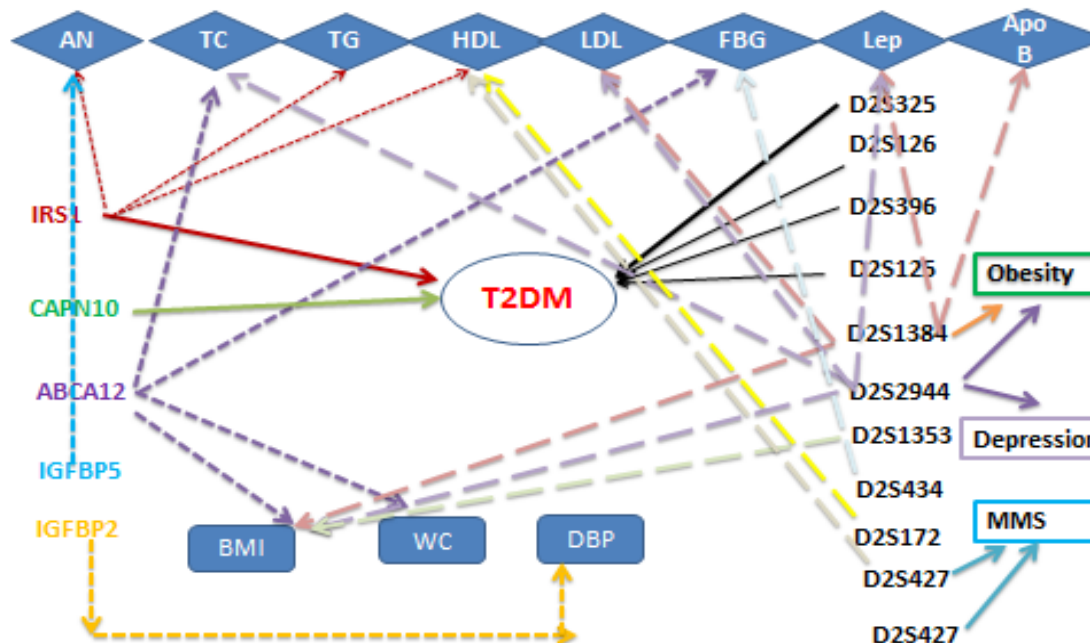
AN: Adiponectin; TC: Total Cholesterol; TG: Triglyceride; HDL: High Density Lipoprotein; LDL: Low Density Lipoprotein; FBG: Fasting Blood glucose; Lep: Plasma Leptin; ApoB: Apolipoprotein B; BMI: Body Mass Index; WC: Waist Circumference; DBP: Diastolic Blood Pressure; MMS: Multiple metabolic Syndrome; IRS1: Insulin Resistance Substrate 1; CAPN10: Calpain10; ABCA12: ATP-binding cassette sub-family A member 12 ; IGFBP(2/5): insulin-like growth factor binding protein (2/5); T2DM: Type 2 Diabetes Mellitus.

Solid lines represent direct association while dotted lines indicate indirect association to T2DM

Chapter 3

3.2 Chromosome 3 (region 3q22-29) involvement in T2DM

Another promising region which was identified during literature survey was the region 3q22-29. This region 3q22-29 is roughly 68M bp (130M-198M bp) in genetic distance which contains 692 genes according to NCBI (build 37.1).

3.2.1 Genetic studies on association of 3q22-29 region with T2DM

In last one decade most of the studies which have reported direct association of this region with T2DM have used specific candidate genes and SNPs (Table 3.4). These mainly include SLC2A2 (Wagner *et al.,* 2011), ADIPOQ1 (Pollin *et al.,* 2005, Bostrom *et al.,* 2009), IGF2BP2 (Omori *et al.,* 2008; Hertel *et al.*, 2008; Duesing *et al.*, 2008; Huang *et al.,* 2010; Nemr *et al.*, 2012; Horikawa *et al.*, 2008; Mtiraoui *et al.*, 2012) and ST6GAL1 (Kooner *et al.,* 2011). Very few studies have used individual marker panels to study the linkage of this region with T2DM. Moderate linkages of STRs D2S1292 (Francke *et al.,* 2001), D3S1565 (Mori *et al.,* 2002) to strong linkages of the STRs D3S1262 (Vionnet *et al.,* 2000) and D3S1580 (Vionnet *et al.,* 2000) with an odds ranging from 1.4 - 4.7 have been reported to be associated directly with T2DM (Table 3.4). Very recently, efforts to identify novel diabetes susceptibility genes in the region have been also reported from the developing world countries evaluating Lebanese Arabs (Nemr *et al.*, 2012), Japanese population (Horikawa *et al.*, 2011) and South East Asian populations (Kooner *et al.,* 2011). However, not much of this region has been evaluated and replicated in multiple ethnicities using STR markers. Pertaining to the 3q22-29 region, the question of whether the linkage and association signals result solely from variation in the candidate gene or include other repeat elements is still undecided. Moreover the Indian population still remains unexplored in terms of association of this region with T2DM.

Table 3.5 Genetic studies on Type 2 Diabetes Mellitus (T2DM) with LOD score > 1.2 ($p<0.01$) within 3q22-29

Chr Loc	Gene	SNP/variation	Population/Phenotype	p-value (OR)	References
3q22	-	D3S1292	Pondicherry/T2DM	.0019 (2.16)	Francke *et al.*, 2001
3q26.2	SLC2A2	rs11920090	T2DM	<.05 (2.98)	Wagner *et al.*, 2011
3q26.31	-	D3S1565	Japanese/ T2DM	0.00623 (1.57)	Mori *et al.*, 2002
3q27	ADIPOQ	rs2275737	Amish/ T2DM	.007 (1.65)	Coleen *et al.*, 2005
3q27	ADIPOQ	rs1342387	Amish/T2DM	.014 (1.61)	Coleen *et al.*, 2005
3q27	ADIPOQ	rs182052	African Americans/T2DM	0.0054 (1.46)	Bostrom *et al.*, 2009
3q27.2	IGF2BP2	rs9826022	French Caucasian/T2DM	0.0002 (1.53)	Duesing *et al.*, 2008
3q27.2	IGF2BP2	rs4402960	Chinese/ T2DM	<0.001	Huang *et al.*, 2010
3q27.2	IGF2BP2	rs4402960	Lebanese Arabs/ T2DM	6.5×10^{-6}	Nemr *et al.*, 2012
3q27.2	IGF2BP2	rs4402960	Japanese/ T2DM	8.1×10^{-5} (1.23)	Horikawa *et al.*, 2008
3q27.2	IGF2BP2	rs4402960	Japanese/ T2DM	0.00009	Omori *et al.*, 2008
3q27.2	IGF2BP2	rs4402960	Norwegian/ T2DM	<0.05 (1.10)	Hertel *et al.*, 2008

The shaded area in the table represents the studies assessing the association of STR repeats in this region of interest.
Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio

Table 3.5 Genetic studies on Type 2 Diabetes Mellitus (T2DM) with LOD score > 1.2 (p<0.01) within 3q22-29 (Contd.)

Chr Loc	Gene	SNP/variation	Population/Phenotype	*p* -value (OR)	References
3q27.2	IGF2BP2	rs1470579	Chinese/ T2DM	<0.05	Huang *et al.*, 2010
3q27.2	IGF2BP2	rs1470579	Lebanese Arabs/ T2DM	1.3×10^{-5} (1.66)	Mtiraoui *et al.*, 2012
3q27.2	IGF2BP2	rs9826022	Caucasian/T2DM	0.0002 (1.53)	Duesing K *et al.*, 2008
3q27.3	ST6GAL1	rs16861329	South east Asians/T2DM	1.12	Kooner *et al.*, 2011
3q27-qter	-	D3S1262	French /T2DM	.00002 (3.91)	Vionnet *et al.*, 2000
3q27-qter	-	D3S1580	French /T2DM	00011 (2.97)	Vionnet *et al.*, 2000
3q28	-	D3S1580	French Caucasians/ T2DM	<.05 (4.7)	Vionnet *et al.*, 2000
3q28	-	D3S1580	Pima Indians/ T2DM	<.05 (1.8)	Weyer *et al.*, 2000
3q28	-	D3S1580	Australian aborigines/ T2DM	<.05 (1.8)	Busfield *et al.*, 2002
3q28	-	D3S1580	Japanese / T2DM	<.05 (1.4)	Mori *et al.*, 2002

The shaded area in the table represents the studies assessing the association of STR repeats in this region of interest.
SLC2A2: Solute carrier family 2 (facilitated glucose transporter); ADIPOQ: Adiponectin Q1; IGF2BP2: Insulin-like growth factor 2 mRNA-binding protein 2 ST6GAL1:Glycosyltransferase, MRAS: Muscle RAS oncogene homolog
Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio

Chapter 3

3.2.2 QTLs for Glucose-lipid metabolism and obesity in 3q22-29

Recent research has emphasized the importance of the central obesity indicators (BMI) and weight as a strong predictor of the T2DM-related morbidities and premature mortality. Univariate linkage analyses have demonstrated evidence of linkage (defined by univariate $p<0.05$) for BMI with SNPs of adiponectin gene on chromosome 3q22-27 (Nakatani *et al.*, 2005, Richardson *et al.,* 2006). Similar univariate linkages of ADIPOQ gene SNPs located on this region has been shown to be associated with FBG and fasting insulin (Richardson *et al.*, 2006). Another SNP on 3q26.2 rs11920090 (SLC2A2) has been shown recently to be associated with FPG with a significant univariate p value (Liana *et al.,* 2010; Dupuis *et al.,* 2010). Also an STR repeat D3S2398 located at 209.41 cM position on the locus 2q22-27 has been shown to be associated with BMI (Kissebah *et al.,* 2000), hip circumference (Kissebah *et al.*, 2000), body weight change (Kissebah *et al.,* 2000; Golla *et al.,* 2003) and insulin with the odds for risk for T2DM ranging from 1.8 to 3.54. The association of this region with markers for central as well as abdominal obesity seems robust but still a lot of exploration needs to be done in terms of its replicability in various populations and in terms of extensively evaluating this locus with pleiotropic effects on correlated traits like BMI, hip circumference and insulin. Till date much of the study of the association of this region with T2DM or associated QTLs has been done mainly in Caucasians, Mexican Americans, Japanese with a only a few reports evaluating the South East Asian populations. Table 3.6 and 3.7 summarised the reported associations of the various quantitative traits (reported to be associated with T2DM) with the region 3q22-29 and the fig 3.2 summarises the network of associations (either candidate or STR based) as compiled from the various studies in tables 3.5 -3.8

Chapter 3

3.2.3 Association of 3q22-29 with other systemic disorders related to T2DM

Patients with T2DM have an increased incidence of disease in several internal organs and tissues. Chronic microvascular and macrovascular diseases have greater influence on the long-term prognosis of patients with T2DM than acute complications. The co-occurrence of various comorbidities and type 2 diabetes mellitus (T2DM) has been well documented. The hypothesis that the co-occurrence of these comorbidities and T2DM may be, at least in part, driven by shared genetic factors can be a strong basis of looking at the association of common lesion of T2DM and associated comorbidities to narrow down to a cause. The region 3q22-29 has been shown to be moderately associated with multiple metabolic syndromes (Francke *et al.,* 2001). The STR marker located at 142.6cM position on this locus was found to be moderately associated with MMS in the Mauritius population while another marker set D2S1571-D3S3686 was shown to be associated with Coronary heart disease and Myocardial infraction (Francke *et al.,* 2001) among the same population with and odds of 2.37. The remaining association of this region with diabetic comorbidities like CVD (Erdmann *et al.,* 2009; Ellis *et al.,* 2011; Ong *et al.,* 2010), IGT (Laukkanen *et al.,* 2005), Insulin sensitivity (Nakatani *et al.,* 2005) have been candidate gene specific (MRAS, SLC2A2 and ADIPOQ) with an odds ranging from 1.2 to 3.04 (Table 3.8). Patients with T2DM often have irregular diet patterns, which deleteriously influences glucose control, lipid metabolism, and micronutrient intake. In addition, T2DM is progressive and generally incurable, precluding several complications related to poor glucose regulation. Thus it is very important to study the association of the STR repeats in this region with T2DM as well as the indirect association obtained as an effect to the various important T2DM associated QTLs and comorbidities to obtain a clearer picture.

Table 3.6 Summary of genome scans on anthropometric QTLs with LOD score $\geq$ 1.2 ($p<0.05$) in region 3q22-29

Chr Loc	Gene	SNP/variation	Population/Phenotype	P-value (OR)	References
3q27	ADIPOQ	SNP45T<G	Japanese/BMI	0.043	Nakatani *et al.,* 2005
3q27	ADIPOQ	rs4632532	Mexican American / BMI	0.029	Richardson *et al.,* 2006
3q27	ADIPOQ	rs182052	Mexican American / BMI	0.032	Richardson *et al.,* 2006
3q28	-	D3S2398	Caucasians/ BMI	.016 (3.3)	Kissebah *et al.,* 2000
3q28	-	D3S2398	Caucasians/ Weight	.022 (3.17)	Kissebah *et al.,* 2000
3q28	-	D3S2398	Body weight change	<.05 (1.8)	Golla *et al.,* 2003

The shaded area in the table represents the studies assessing the association of STR repeats in this region of interest.
ADIPOQ: Adiponectin Q1; Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds: OR: Odds Ratio

Table 3.7 Summary of genome scans on various QTLs with LOD score $\geq$ 1.2 ($p<0.05$) in loci 3q22-29

Chr Loc	Gene	SNP/variation	Population/Phenotype	P-value (OR)	References
3q26.2	SLC2A2	rs11920090	FPG	8×10^{-11}	Liana *et al.,* 2010
3q26.2	SLC2A2	rs11920090	FPG	3.3×10^{-6}	Dupuis *et al.,* 2010
3q27	ADIPOQ	rs10848569	Mexican American /FBG	0.00029	Richardson *et al.,* 200
3q27	ADIPOQ	rs4632532	Mexican American / fasting	0.023	Richardson *et al.,* 2006
3q27	ADIPOQ	rs182052	Mexican American / fasting	0.026	Richardson *et al.,* 2006
3q27	ADIPOQ	rs929434	Mexican American / FPG	0.00016	Richardson *et al.,* 2006

SLC2A2: Solute carrier family 2 (facilitated glucose transporter); ADIPOQ: Adiponectin Q1
Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio

Table 3.7 Summary of genome scans on various QTLs with LOD score ≥ 1.2 (p<0.05) in loci 3q22-29 (Contd.)

Chr Loc	Gene	SNP/variation	Population/Phenotype	P-value/OR	References
3q27	ADIPOQ	rs3809266	Mexican American /FPG	0.00027	Richardson *et al.*, 2006
3q27	ADIPOQ	rs12342	Mexican American /FPG	0.00021	Richardson *et al.*, 2006
3q28	-	D3S2398	Caucasians/ Hip	.009 (3.54)	Kissebah *et al.*, 2000
3q28	-	D3S2398	Caucasians/ Insulin	.032 (3.01)	Kissebah *et al.*, 2000

The shaded area in the table represents the studies assessing the association of STR repeats in this region of interest.
ADIPOQ: Adiponectin Q1;Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio; FPG: Fasting Plasma Glucose

Table 3.8 Summary of genome scans on systemic disorders with T2DM as co-morbidity and LOD score ≥ 1.2 (*p*<0.05) between the region 3q22-29

Chr Loc	Gene	SNP/variation	Population/Phenotype	P-value (OR)	References
3q22	-	D3S1292	Mauritius/ MMS	.0106 (1.36)	Francke *et al.*, 2001
3q22.3	MRAS	rs9818870	New Zealand/ CVD	.045	Ellis *et al.*, 2011
3q26.2	SLC2A2	rs5393	Finnish/IGT to T2DM	0.008(3.04)	Laukkanen *et al.*, 2005
3q26.2	SLC2A2	rs5394	Finnish/IGT to T2DM	0.026 (2.54)	Laukkanen *et al.*, 2005
3q26.2	SLC2A2	rs5400	Finnish/IGT to T2DM	0.009 (2.60)	Laukkanen *et al.*, 2005
3q26.2	SLC2A2	rs5404	Finnish/IGT to T2DM	0.025 (2.57)	Laukkanen *et al.*, 2005
3q26.31	-	D3S1571-D3S3686	Mauritius/CHD_MI	.0009 (2.37)	Francke *et al.*, 2001
3q27	ADIPOQ	SNP45T<G	Japanese/ Insulin Sensitivity	0.046	Nakatani *et al.*, 2005
3q27	ADIPOQ	rs266729	Han Chinese / CVD	0.0044 (1.49)	Ong *et al.*, 2009

The shaded area in the table represents the studies assessing the association of STR repeats in this region of interest.
ADIPOQ: Adiponectin Q1; SLC2A2: Solute carrier family 2; MRAS: Muscle RAS oncogene homolog; Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio;

Fig 3.2 Summary of the reported direct and indirect associations of the region 3q22-29 with T2DM
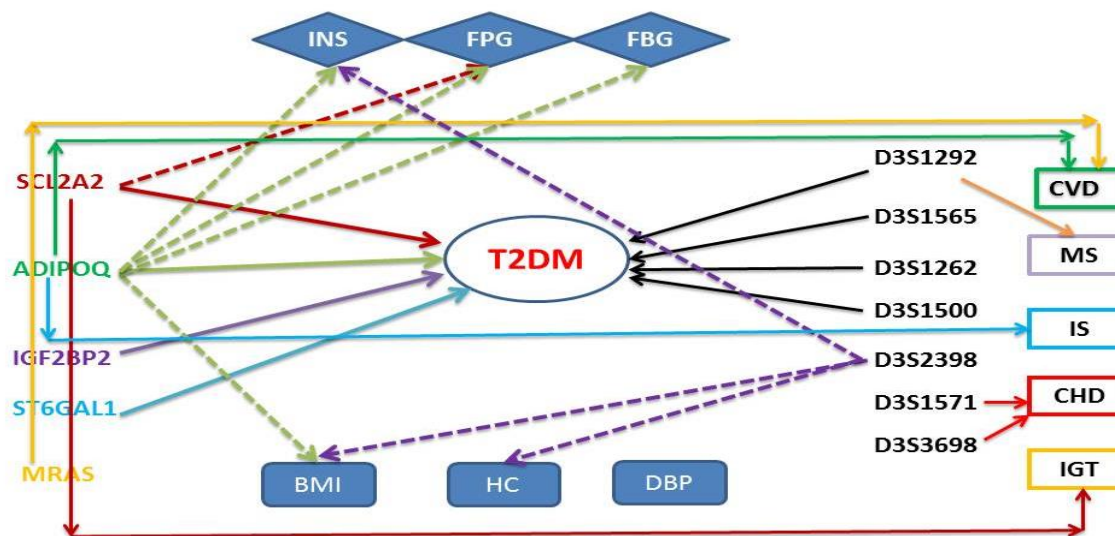
The figure here summarises the associations of the key studies discussed above and potray an overview of the various reported genetic players on the loci 3q22-29 acting directly or indirectly in the pathogenesis of T2DM

INS: Insulin, FPG: Fasting Plasma Glucose, FBG: Fasting Blood Glucose, BMI : Body Mass Index, HC: Hip Circumference, DBP: Diastolic Blood Pressure, CVD: Cardio Vascular Disease, MS: Metabolic Syndrome, IS: Insulin Sensitivity, CHD: Coronary Heart Disease, IGT: Impaired Glucose Tolerance.AD1POQ: Adiponectin Q1, IGF2BP2: Insulin-like growth factor 2 mRNA-binding protein 2 ST6GAL1: Glycosyltransferase, MRAS: muscle RAS oncogene homolog.SLC2A2: solute carrier family 2 (facilitated glucose transporter)

Solid lines represent direct association while dotted lines indicate indirect association to T2DM

Chapter 3

3.3 Chromosome 10 (region 10q25-26) involvement in T2DM

The next promising region which was identified during literature review was the region 10q25-26. This region 10q25-26 is roughly 48M bp (71M-119M bp) in genetic distance which contains 581 genes according to NCBI (build 37.1).

3.3.1 Genetic studies on association of 10q25-26 region with T2DM

In the previously reported studies this region also has mainly been evaluated using candidate genes for its association with T2DM. Reports have shown that this region harbours many important genes like ADRA2A (Alpha-2-adrenergic receptor), TCF7L2 (Transcription factor 7-like 2), HTRA1 (High temperature requirement A serine peptidase), ARMS2 (Age related Maculopathy succeptibility gene 2), HHEX1 (Hematopoietically Expressed Homeobox), IDE (Insulin Degrading Enzyme) etc. These genes have shown moderate association (ADRA2A: Talmud *et al.,* 2011; TCF7L2: Grant *et al.,* 2006, Sladek *et al.,* 2007, Salpea *et al.,* 2009, Thorsby *et al.,* 2009, Takeuchi *et al.,* 2009, Lin *et al.,* 2010a, Gupta *et al.,* 2010, Dupuis *et al.,* 2010, Dabelea *et al.,* 2011, Silbernagel *et al.,* 2011, Gómez *et al.,* 2011, Lin *et al.,* 2010b; HHEX1: Horikawa *et al.,* 2008, Furukawa *et al.*, 2008, Zhou *et al.,* 2010, Horikawa *et al.*, 2008, Furukawa *et al.,* 2008; IDE: Kwak *et al.,* 2008, Rudovich *et al.,* 2009, Zhou *et al.,* 2010, Bartl *et al.,* 2011) with T2DM when evaluated across multiple ethnicities. However, the association studies have been consistent across various groups and ethinicities.

These three genes TCF7L2, HHEX and IDE on chromosome 10q25-26 reside within the linkage region for type 2 diabetes (T2DM). Since the initial report in early 2006, several single nucleotide polymorphisms (SNPs) in the TCF7L2 gene located on chromosome 10q25 have been shown to be associated with T2DM in multiple

Chapter 3

populations. For the first time, a gene is consistently involved in T2D susceptibility in all major ethnic groups. At the individual level, carrying the TCF7L2 risk allele increases T2D risk 50%. However, at the population level, the attributable risk is lower than 25% and varies with the allele frequency. The presence of the TCF7L2 rs7903146 risk allele increases TCF7L2 gene expression in β cells, possibly impairing glucagon-like peptide-1-induced insulin secretion and/or the production of new mature β cells (Cauchi, 2008). This makes TCF7L2 variants the strongest known genetic risk factors for type 2 diabetes. Apart from these candidate genes two STR repeats D10S1655 and D10S212 have also been shown to have an association with T2DM with the odds ratio ranging from 1.22-1.59. In Europeans, genome-wide association scans showed that TCF7L2 has been the most important locus predisposing to T2DM so far.

Although the specific mechanism driving the development of type 2 diabetes remains unclear, there is sufficient evidence to demonstrate that TCF7L2 variants strongly predict the development of type 2diabetes and/or the progression to diabetes from impaired glucose tolerance. The discovery of TCF7L2 as a diabetes gene illustrates that novel true diabetes genes can be found, their association with type 2 diabetes replicated and their effect incorporated into risk prediction models (Florez, 2007). Table 3.9 summarises the various reports which have directly associated the region 10q25-26 with T2DM.

Table 3.9 Genetic studies on Type 2 Diabetes Mellitus (T2DM) with LOD score $\geq$ 1.2 ($p<0.01$) within 10q25-26

Chr Loc	Gene	SNP/ variation	Population/Phenotype	*p*-value (OR)	References
10q25.2	ADRA2A	rs553668	British/T2DM risk	0.007 (1.38)	Talmud *et al.,* 2011
10q25.3	TCF7L2	rs7903146	British/T2DM	<0.05 (1.37)	Grant *et al.,* 2006
10q25.3	TCF7L2	rs4506565	British/T2DM	$5.68 \times 10^{-13\$}$	Wellcome Trust 2007
10q25.3	TCF7L2	rs7903146	Japanese/T2DM	<0.0001(2.07)	Salpea *et al.,* 2009
10q25.3	TCF7L2	rs7903146	Norway/T2DM	<0.05 (1.48)	Thorsby *et al.,* 2009
10q25.3	TCF7L2	rs7903146	Japanese/T2DM	7.6×10^{-12} (1.54)	Takeuchi *et al.,* 2009
10q25.3	TCF7L2	rs7903146	Han chinese/T2DM	0.001 (1.58)	Lin *et al.,* 2010
10q25.3	TCF7L2	rs7903146	Indian/ T2DM	0.00179 (2.0)	Gupta *et al.,* 2010
10q25.3	TCF7L2	rs7903146	European/T2DM	2.2×10^{-51} (1.40)	Dupuis *et al.,* 2010
10q25.3	TCF7L2	rs7903146	Multi ethnic youth/ T2DM	< 0.0001(1.97)	Dabelea *et al.,* 2011
10q25.3	TCF7L2	rs7901346	Caucasians/T2DM	<.001	Silbernagel *et al.,* 2011

TCF7L2: Transcription factor 7-like 2; T2DM: Type 2 Diabetes Mellitus; Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio;

Table 3.9 Genetic studies on Type 2 Diabetes Mellitus (T2DM) with LOD score $\geq$ 1.2 ($p<0.01$) within 10q25-26 (Contd.)

Chr Loc	Gene	SNP/ variation	Population/Phenotype	p-value (OR)	References
10q25.3	TCF7L2	rs7903146	Austria/ T2DM	0.001(1.91)	Muendlein *et al.*, 2011
10q25.3	TCF7L2	rs1225537	Caucasians/T2DM	<.001	Silbernagel *et al.*, 2011
10q25.3	TCF7L2	rs1225537	Caucasians/HbA1c	<.001	Silbernagel *et al.*, 2011
10q25.3	TCF7L2	rs12255372	Mexico/T2DM	<.05 (1.78)	Gómez *et al.*, 2011
10q25.3	TCF7L2	rs6585205	Han chinese/T2DM	4.0×10^{-4} (1.31)	Lin *et al.*, 2010
10q25.3	TCF7L2	rs4506565	Indian/ T2DM	0.00191 (2.1)	Gupta *et al.*, 2010
10q26	-	D10S1655	French /T2DM	.003 (1.59)	Vionnet *et al.*, 2000
10q26.3	-	D10S212	French/T2DM	.008 (1.22)	Vionnet *et al.*, 2000
10q26.3	-	D10S212	French/T2DM	.005 (1.44)	Vionnet *et al.*, 2000
10q23.3	HHEX1	rs1111875	Japanese/T2DM	1.4×10^{-5} (1.27)	Horikawa *et al.*, 2008
10q23.3	HHEX1	rs1111875	Japanese/T2DM	0.0024 (1.42)	Furukawa *et al.*, 2008

The shaded area in the table represents the studies assessing the association of STR repeats in this region of interest.
TCF7L2: Transcription factor 7-like 2; HHEX1: Hematopoietically-expressed homeobox protein;
Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio

Table 3.9 Genetic studies on Type 2 Diabetes Mellitus (T2DM) with LOD score ≥ 1.2 (*p*<0.01) within 10q25-26 (Contd.)

Chr Loc	Gene	SNP/ variation	Population/Phenotype	*p*-value (OR)	References
10q23.3	HHEX1	rs1111875	Japanese/ T2DM	6.7×10^{-12} (1.21)	Takeuchi *et al.*, 2009
10q23.3	HHEX1	rs1111875	Han Chinese/ T2DM	0.009 (1.15)	Zhou *et al.*, 2010
10q23.3	HHEX1	rs7923837	Japanese/ T2DM	1.0×10^{-4} (1.27)	Horikawa *et al.*, 2008
10q23.3	HHEX1	rs7923837	Japanese/ T2DM	0.00014 (1.66)	Furukawa *et al.*, 2008
10q23-25	IDE	g.-179T>C	Korean/ T2DM	0.04 (1.73)	Kwak *et al.*, 2008
10q23-25	IDE	g.IVS18+99G>A	Korean/ T2DM	0.02 (1.23)	Kwak *et al.*, 2008
10q23-25	IDE	g.IVS18+99G>A	Mixed/ T2DM	.005 (1.18)	Kwak *et al.*, 2008
10q23-25	IDE	rs1887922	German/ T2DM	0.003, (1.26^)	Rudovich *et al.*, 2009
10q23-25	IDE	rs2149632	German/ T2DM	<.0001, (1.33^)	Rudovich *et al.*, 2009
10q23-25	IDE	11187007	Han Chinese/ T2DM	0.009 (1.15)	Zhou *et al.*, 2010
10q23-25	IDE	rs2251101	Austria/ T2DM	0.009 (2.4)	Bartl *et al.*, 2011

HHEX1: Hematopoietically-expressed homeobox protein; IDE: Insulin Degrading Enzyme
Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio

Chapter 3

3.3.2 Anthropometric markers in this region associated with T2DM

Two SNPs of ADRA2A located at 10q25.6 have also been shown to be associated with SBP and BMI in the British population (Talmud *et al.,* 2011). This group reported that the Variants of ADRA2A are associated with fasting glucose, blood pressure, body mass index and type 2 diabetes risks by conducting a meta-analysis of four prospective studies. Interesting not many studies have been able to associate this region to any of the diabetes related QTLs.

Other reported QTLs in this region associated with T2DM

Recently few reports have associated SNPs in the gene ADRA2A with FBG (Dupuis *et al.,* 2010, Talmud *et al.,* 2011, Bo *et al.,* 2012) and SBP (Talmud *et al.,* 2011). The gene TCF7L2 has also been shown to be associated with FBG (Stolerman *et al.,* 2009; Dupuis *et al.,* 2010), Elevated HbA1c (Silbernagel *et al.,* 2011). This gene has also been reported to be associated with the pro-insulin to insulin ratio in the GWAS (Stolerman *et al.,* 2009)**.** Infact these reports have been relatively new only.

3.3.3 Association of 10q25-26 with systemic disorders comorbid with T2DM

Two SNPs rs11200638 and rs10490924 in the PRASS11 gene and ARMS2 gene have been recently shown to be associated with diabetic retinopathy in south Indian population (Balasubbu *et al.,* 2010). Apart from this the SNP rs290487 has also been shown to be associated with NODM. Majority of the association of this region is direct with T2DM. Table 3.10-3.12 summarises the reported association of these regions with QTLs for T2DM and associated comorbidities. Also Fig 3.3 summarises the network of the associations drawn from the studies discussed in this section.

Table 3.10 Summary of genome scans on systemic disorders with T2DM as co-morbidity and LOD score $\geq$ 1.2 ($p<0.05$) between the region 10q25-26

Chr Loc	Gene	SNP/ variation	Population/Phenotype	p-value (OR)	References
10q25.3	TCF7L2	rs290487	Han Chinese/ NODM	.015 (2.17)	Ling *et al.,* 2012
10q26.3	HTRA1 (PRSS11)	rs11200638	South Indian / Diabetic retinopathy	0.05 (0.68)	Balasubbu *et al.,* 2010
10q26.3	ARMS2	rs10490924	South Indian /Diabetic retinopathy	0.07 (0.77)	Balasubbu *et al.,* 2010

TCF7L2: Transcription factor 7-like 2; HTRA1: High Temperature Requirement Factor A1; ARMS2: Age-related maculopathy susceptibility protein 2; Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio

Table 3.11 Summary of genome scans on various QTLs with LOD score $\geq$ 1.2 ($p<0.05$) in the region 10q25-26 and are associated with T2DM

Chr Loc	Gene	SNP/ variation	Population/Phenotype	p-value (OR)	References
10q25.2	ADRA2A	rs10885122	European/ FBG	9.7×10^{-8}	Dupuis *et al.,* 2010
10q25.2	ADRA2A	rs553668	British/FBG	0.011	Talmud *et al.,* 2011
10q25.2	ADRA2A	rs17128356	British/FBG	.008	Talmud *et al.,* 2011
10q25.2	ADRA2A	rs553668	Caucasian/ FBG	$\beta = 0.48$	Bo *et al.,* 2012
10q25.3	TCF7L2	rs7903146	FHS / FBG	0.01	Stolerman *et al.,* 2009
10q25.3	TCF7L2	rs7903146	European/ FBG	2.8×10^{-8}	Dupuis *et al.,* 2010
10q25.3	TCF7L2	rs7901346	Caucasians/ HbA1c	<.001	Silbernagel *et al.,* 2011
10q25.3	TCF7L2	rs1225537	Caucasians/ HbA1c	<.001	Silbernagel *et al.,* 2011

TCF7L2: Transcription factor 7-like 2; ADRA2A: Alpha-2-adrenergic receptors; FHS: Framingham Heart Study
Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio

Table 3.12 Summary of genome scans on anthropometric QTLs with LOD score ≥ 1.2 (*p*<0.05) in the loci 10q25-26

Chr Loc	Gene	SNP/ variation	Population/Phenotype	*p*-value	References
10q25.2	ADRA2A	rs491589	British/SBP	.002	Talmud *et al.,* 2011
10q25.2	ADRA2A	rs36022820	British/BMI	.009	Talmud *et al.,* 2011

TCF7L2: Transcription factor 7-like 2; ADRA2A: Alpha-2-adrenergic receptors;
Chr Loc: Chromosomal Location; SNP: Single Nucleotide polymorphism; LOD: Log of odds; OR: Odds Ratio

Fig 3.3 Summary of the reported direct and indirect associations of the region 10q25-26 with T2DM
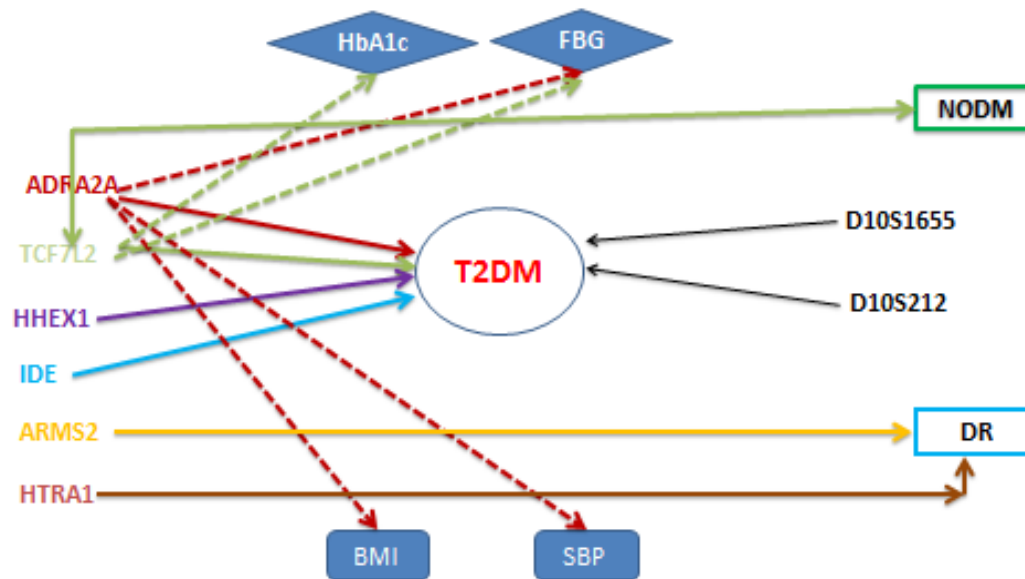
The figure here summarises the associations of the key studies discussed above and portrays an overview of the various genetic players on the loci 10q25-26 acting directly or indirectly in the pathogenesis of T2DM.

FBG: Fasting Blood Glucose, HbA1c: Glycated Haemoglobin, NODM: New Onset of Diabetes Mellitus, DR: Diabetic Retinopathy, ADRA2A: Alpha-2-adrenergic receptors, TCF7L2: Transcription factor 7-like 2, HHEX1: Hematopoietically-expressed homeobox protein, IDE: Insulin-degrading enzyme, ARMS2: Age-related maculopathy susceptibility protein 2.

Solid lines represent direct association while dotted lines indicate indirect association to T2DM
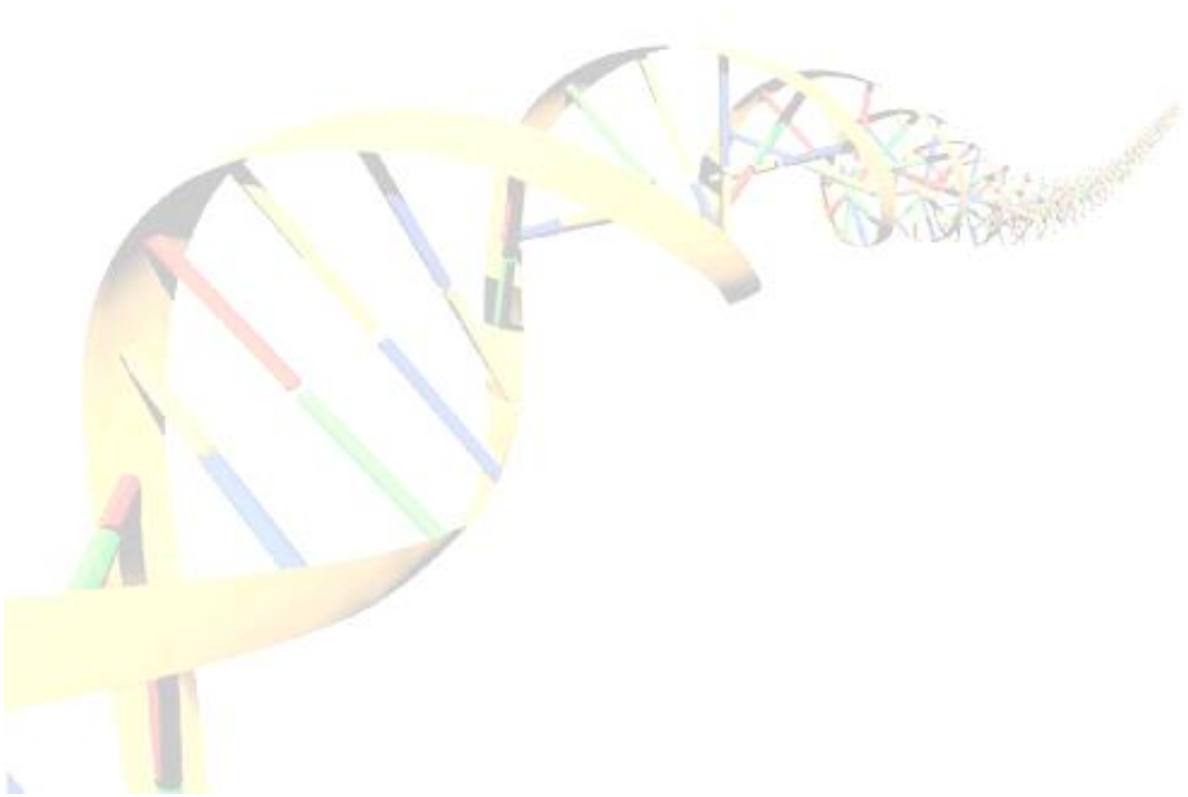
Materials and Methods

Chapter 4

Materials and Methods

4.1 Recruitment of Subjects

This case control association study was conducted according to the norms of declaration of Helsinki for human experimentation and with prior approval (IHEC- 24/09-10) by the Institutional Human Ethics Committee (IHEC) at Birla Institute of Technology and Science (BITS), Pilani and the collaborative institutes. The study involved the participation of individuals comprising two hundred ninty one (291) diabetics recruited from the four collaborating centers in Rajasthan namely Sardar Patel Medical College (Bikaner), Siddhi Nursing Home (Jodhpur), Diabetes Thyroid And Endocrine Centre (Jaipur) and Medical center (Pilani) and four hundred seventy five (475) ethnically matched controlled group from the same geographical region without any history of diabetes or associated co morbidities like obesity, dyslipidemia etc. The patients who agreed to participate and signed the consent (Appendices I) form were only included in the study and were explained in detail the information regarding the study procedure. The subjects were interviewed by participating clinicians using a questionnaire to capture information on demographic details (Height, weight, SBP, DBP, Waist circumference), diabetes related characteristics and comorbidities.

4.1.1 Inclusion criteria
  ❖ **Diabetic Cases**

The inclusion criteria and rationale for the study is presented in the table given below. The age range for the selection of cases and control was chosen to cover a broad range namely (20-60 years) adult group. Along with diabetes symptoms (i.e. polyuria,

polydipsia and unexplained weight loss) the cases recruited need to have at-least one of the conditions listed below.

Table 4.1: Values for diagnosis of diabetes mellitus and other categories of hyperglycemia

	Glucose concentration, mmol l-1 (mg dl-1) (Venous plasma)
Diabetes mellitus:	
Fasting and/or	=> 7.0 (=> 126)
2-h post glucose load	=> 11.1 (=> 200)
Impaired glucose tolerance (IGT)	< 7.0 (< 126)
Fasting (if measured)	=>7.8 (=>140)
and 2-h post glucose load	and < 11.1 (< 200)
Impaired fasting glycaemia or impaired fasting glucose (IFG)	
Fasting	=> 6.1 (=> 110) and <7.0 (<126)
and (if measured)	< 7.8 (< 140
2-h post glucose load	

* Corresponding values for capillary plasma are: for Diabetes Mellitus, fasting 7.0 (126), 2-h 12.2 (220); for Impaired Glucose Tolerance, fasting < 7.0 (< 126) and 2-h 8.9 (160) and < 12.2 (< 220); and for Impaired Fasting Glycaemia 6.1 (110) and < 7.0 (< 126) and if measured, 2-h < 8.9 (< 160).For epidemiological or population screening purposes, the fasting or 2-h value after 75 g oral glucose may be used alone. Glucose concentrations should not be determined on serum unless red cells are immediately removed; otherwise glycolysis will result in an unpredictable under-estimation of the true concentrations. It should be stressed that glucose preservatives do not totally prevent glycolysis. If whole blood is used, the sample should be kept at 0-4 °C or centrifuged immediately, or assayed immediately.
*adapted from WHO report 1999

Also the recruited cases must have a history of diabetes for as minimum of 6 months to reduce the effect of confounding by subjects that were still in initial phase of medication adjustment or initial emotional adjustment to the disease.

❖ **Control Group**

A random selection of healthy controls from the same geographical regions as a representation to the population in Rajasthan was carried out (Population based control group). The selected healthy control group was independent to the presence of diabetes related symptoms and risk factors/co-morbidities in order to be included in the study

group. The unrelated population controls had normal glucose tolerance (NGT) based on ADA (2003) definitions (fasting plasma glucose <6.1mM and 2 hour postload glucose during an OGTT <7.8 mM) at last clinical visit and no first degree family history of type 2 diabetes.

4.1.2 Exclusion criteria

1. Patients with FBG levels <126mg dl^{-1}

2. Pregnant women

3. Lactating mothers were excluded from the study.

4. Individuals <20 years of age

4.2 Sample Collection

2-5 ml of blood was collected in 0.5 M ethylenediamine tetra acetic acid (EDTA) coated 15 ml sterile tubes (EDTA from Sigma Aldrich) from anti-cubital vein under aseptic conditions.

4.2.1 Separation of plasma from the blood

The blood samples were centrifuged at a speed of 3500 rpm for ten minutes and the plasma was separated from the blood using 1ml micropipette into fresh 1.5ml microcentrifuge tube.

4.2.2 Storage and further processing of samples

Both the blood and plasma samples were stored at -70^{0}C till further analysis. It was ensured that the PCV samples were properly stored so as to obtain maximum yield of DNA

Chapter 4

4.3 Demographic details

After a comprehensive survey of literature, a suitable questionnaire (Appendices II) to obtain the required information about demographic variables was designed. During data collection personal interview was held with each subject. Information about their, personal history and other life style characteristics was collected..

4.3.1 Anthropometric measurements

All the subjects under study were extensively characterized for different anthropometric and quantitative metabolic traits. Anthropometric measurements, including height, weight, and waist and hip circumferences were done per standardized protocols, and BMI and waist-to-hip ratio were calculated according to standard techniques (WHO classification).

4.3.1.1 Body Mass Index (BMI)

Body mass Index or Quetelet index, is a heuristic proxy for human body fat based on an individual's weight and height and was calculated using formula:

BMI= Weight (in Kg) / Height (in meters) ^{2}

4.3.1.2 Waist Hip Ratio (WHR)

Waist to hip ratio was calculated by using standard formula:

WHR= Waist Circumference/ Hip Circumference

4.3.1.3 Blood Pressure

Blood pressure was measured using calibrated mercury sphygmomanometer following the Joint National Committee (JNC VII) guidelines. Two separate measures of systolic and diastolic blood pressure were taken after 10 minutes rest. Average of the two measured of systolic and diastolic blood pressures were used for the analysis.

4.3.1.4 Duration of illness

Duration of illness in years was recorded for each patient. The information was either self-reported by the subject or was provided by the consulting diabetologist/clinician.

4.3.1.5 Family history

Subjects were asked about their history of diabetes, metabolic syndrome, obesity in the family and disease phenotype was recorded for the affected family members of the subjects.

4.3.2 Clinical tests or Blood parameters estimation

Biochemical measurements including fasting plasma glucose (FPG), total cholesterol, HDL, LDL and VLDL cholesterol, and triglycerides were performed using commercially available kits.

4.3.2.1 Fasting plasma glucose: Plasma glucose level was measured using glucose oxidase-peroxidase method (Glucose Eco-Pak - GOD – POD; Accurex Biomedical Pvt. Ltd.). The estimation of glucose was done within 1 hour of sample collection.

4.3.2.2 Plasma cholesterol levels

Plasma cholesterol levels were measured using commercially available kit (AutoZyme New Cholesterol, Accruex Biomedical Pvt. Ltd). This kit contains a reagent set for determination of total cholesterol based on enzymatic method using Cholesterol esterase, Cholesterol oxidase and Peroxidase. Tests were performed in duplicate as per the manufacturer's instructions using a semi-auto biochemistry analyzer.

4.3.2.3 Plasma triglyceride levels

Plasma triglyceride levels were measured using commercially available kit (AutoZyme New Triglycerides, Accruex Biomedical Pvt. Ltd). This kit contains a reagent set for

determination of triglycerides, based on enzymatic method using Lipoprotein lipase, Glycerol kinase, Glycerol phosphate oxidase and Peroxidase. Tests were performed in duplicate as per the manufacturer's instructions using a semi-auto biochemistry analyzer.

4.3.2.4 Plasma HDL levels

Plasma HDL levels were measured using commercially available kit (Spinreact). This kit contains a reagent set for determination of cholestrol HDL. The very low density (VLDL) and low density (LDL) lipoproteins from serum or plasma are precipitated by phosphotungstate in the presence of magnesium ions. After removed by centrifugation the clear supernatant containing high density lipoproteins (HDL) is used for the determination of HDL cholesterol.

4.3.2.5 Plasma LDL levels

Plasma LDL levels were calculated according to the standard Friedewald Formula: LDL Cholesterol=Total cholesterol-(Triglyceride/5+HDL)

4.3.2.6 Plasma VLDL levels

There is no direct measurement method for VLDL. It is usually inferred from other serum cholesterol numbers.

The plasma VLDL levels were calculated using the standard formula:

VLDL= (Triglyceride)/5

4.3.2.7 HbA1c: The HbA1c was determined using commercially available spinreact kit. It is a quantitative turbidometric test for the measurement of glycated hemoglobin in human whole blood.

Chapter 4

4.4 Genomic DNA Extraction

Genomic DNA was extracted using standard procedure described by Hammond *et al.*, (1996).

4.4.1 Materials Used

Ammonium Chloride (NH_4Cl); Potassium Bicarbonate ($KHCO_3$); EDTA; Tris-HCl (2-amino-2 hydroxy methyl 1-3 propandiol); Guanidium Thiocynate (GTC); Sodium Dodecyl Sulphate (SDS); Tris saturated Phenol; Isoamyl alcohol; Chloroform; 3M Sodium Acetate (CH_3COONa); Ethanol 70% and 100% and 50 ml, 15 ml, 1.5 ml sterile tarson tubes.

4.4.2 Reagents and working solutions

1. Lysis buffer I (RBC lysis buffer): 155 mM Ammonium Chloride; 10 mM Potassium Bicarbonate; 0.1 mM EDTA pH 8.0

2. Lysis buffer II (WBC lysis buffer or extraction buffer): 5 M Guanidium Thiocynate; 10 mM Tris HCl; pH 8.0, 10 mM EDTA; pH 8.0, 0.5% SDS.

3. Phenol Chloroform Isoamylalcohol (PCA): Tris Saturated Phenol: Chloroform: Isoamyl Alcohol in 25:24:1 ratio

4. Chloroform Isoamyl Alcohol (CA): Chloroform: Isoamyl Alcohol in 24:1 ratio

4.4.3 Procedure for DNA isolation from peripheral human blood

Cell Lysis

1. To the packed cell volume, three volume of chilled RBC Lysis Buffer I was added and mixed well by wrist shaking. It was then placed on ice for 20 minutes.

2. After 20 minutes, tubes were shaken vigorously and centrifuged at 3500 rpm for 20 minutes.

3. Supernatant was discarded and the above steps (2 and 3) were repeated till the white pellet of WBC was obtained.

4. This pellet was stored at -20ºC for until further processing for DNA extraction.

Extraction of DNA from WBC

1. To the white pellet 500 μl of Lysis Buffer II was added and the tube was placed on ice (4ºC) for 10 minutes. The tubes were then vortexed for 2 minutes.

2. To this equal volume of PCA mix was added, mixed gently by inverting the tube, centrifuged at 3500 rpm for 10 minutes.

3. The aqueous phase was collected in a new eppendorf and equal volume of CA was added, mixed gently and centrifuged at 3500 rpm 10 minutes.

4. The aqueous phase was aspirated in a fresh autoclaved eppendorf, to which ½ volume of chilled Isopropanol was added subsequently.

5. The tube was gently swirled to precipitate the DNA till a thread like structure appeared, and was then kept at $-20^{O}C$ for overnight precipitation.

6. The precipitated DNA was centrifuged at 10,000 rpm for 20 minutes and the supernatant was discarded.

7. The DNA pellet was washed with 70% chilled Ethanol and centrifuged at 10,000 rpm for 20 minutes and again the supernatant was discarded carefully.

8. The pellet was air dried at room temperature and DNA was dissolved in 500 μl of sterile Milli Q grade water;

9. Samples were stored at $-20^{O}C$ for analysis & at $-70^{O}C$ for long term storage.

4.5 Quantitation and Purity of Nucleic Acids

The concentration of purified DNA was determined by measuring the absorbance at 260

nm and 280 nm in a spectrophotometer. The reading at 260 nm allows for the calculation of the nucleic acids concentration in the sample. . The ratio of the absorbance readings at 260 nm and 280 nm (OD_{260}/OD_{280}) provides an estimate of the purity of the nucleic acid. Pure preparations of DNA have OD_{260}/OD_{280} of 1.8 and 2.0 respectively. A ratio of the readings at 260 nm and 230 nm (OD_{260}/OD_{230}) can aid in evaluating the level of salt carryover in the purified nucleic acid. As a guideline, the OD_{260}/OD_{230} is best if it is greater than 1.5.

4.5.1 Materials Used

1.5 ml microcentrifuge tubes, Tips and Milli Q water for dissolution, Kim wipes

Equipment: Nanodrop -8000 (UV spectroscopy).

4.5.2 Procedure

1. 2 µl of purified genomic DNA sample was diluted with 8 µl of water (1/5 dilution).

2. Each receptor (8) of Nano drop was first cleaned and wiped with Kim wipes.

3. The system was then set blank with autoclaved Milli Q water by loading 2µl of water on to each receptor.

4. 2µl of each of the diluted genomic DNA sample was loaded with the help of 2-20µl multichannel micropipette on to the receptor platform of Nano drop.

5. The absorbance of the diluted sample was then measured at 260, 280, 230 nm

4.5.3 Precautions

All the tips and tubes used in extraction and quantification were sterilized by autoclaving at 15 lbs pressure for 15 minutes. The water used in quantification was of Milli Q grade.

Chapter 4

4.6 Molecular genetic analysis of chromosomes 2q32-q37, 3q22-29 and 10q25-26

Microsatellites were used as molecular markers in genetics, for determining kinship, population and other studies. The repeated sequence of two, three or four nucleotides were amplified for identification by the polymerase chain reaction (PCR) process and visualized on polyacrylamide gels.

4.6.1 Repeat polymorphism in chromosome 2q32-q37, 3q22-29 and 10q25-26

The microsatellite were selected such that either they are mentioned in literature matching the criteria laid down in the study or are simple sequence length polymorphism (SSLP) markers chosen from UniSTS (Database of sequence tagged sites) and ALFRED (Allele frequency database). The markers had a heterozygosity of between 0.5-0.8. Table 4.2 list the marker used for the study along with their genetic position on chromosome in centi Morgan on the following page (cM).

Table 4.2: List of selected microsatellite markers in the study

S. No	Chromosomal Location	Marshfield Marker	Position (In cM)	PCR amplicon (in bp)
1	2q32-q37	D2S1384	200.43	141-161
		D2S2944	210.43	96-124
		D2S439	231.27	165-193
2	3q22-29	D3S3609	195.60	163-185
		D3S2398	209.41	266-298
3	10q25-26	D10S521	127.11	155-189
		D10S1237	134.70	376-432

Primers for repeat polymorphism

Primer sequences for tetra dinucleotide repeat polymorphism were obtained from databases such as UniSTS (http://www.ncbi.nlm.nih.gov/sites/entrez?db=unists) and

Human Genome Database (http://www.gdb.org). Primers sequences procured from Ocimum Biosolution, Hyderabad, India are listed in Table 4.3:

Table 4.3: Primers used for microsatellite amplification along with the codes

Markers	cM	Primer sequence-5'-3'
D2S1384	200.43	2GF2: AATAGAGGGCCCTTGCTTAA
		2GR2: TTTGGGATAAAAGGTATTTTGC
D2S2944	210.43	4F: TCTGTCTTAGATGGATGAATGG
		5R: GGGATGGATGGAAAAGATTC
D2S439	231.27	2GF4: AGGAACAAAGTCTCATTCTCTTG
		2GR4:ACAGATATCATTTACAACACATGTG
D3S3609	195.60	3GW6F: AGCTGGGGACCAGTCT
		3GW6R: CGAGAGTAACTTGTACGGTG
D3S2398	209.41	3GW7F: AGCCTGAGCAAAACAGTGAA
		3GW7R: GAAGACCTACGGATTGGGTC
D10S521	127.11	10GW1F: CTCCAGAGAAAACAGACCAA
		10GW1R: CCTACCATCAATCAACTGAG
D10S1237	134.70	10GW2F: CTTGTCCTGCCTTTGGACTA
		10GW2R: CTCTGTCCTAGTCAGGTTCTCC

4.6.2 Amplification by polymerase chain reaction (PCR)

Amplification was carried out on "Veriti"- ABI Biosolutions thermal cycler in duplex for two markers together in a final volume a 10 μl reaction. Reaction cocktail for each sample contained a total of 2.5 μl of 4X cocktail mix (4X-CT) of the marker , 2.0 μl (0.3-0.5μg/ul) of Genomic DNA, 0.15 μl (5U/ μl) of Bioron Taq DNA Polymerase and the final volume was made up to 10 μl using ultra-pure water. The basic composition of the 4X-CT for each marker is listed in table 4.4. PCR cycling conditions are amplification of markers in duplex is summarized in table 4.5. The PCR tubes were maintained at 4ºC during dispensing of the master mix.

Chapter 4

The final concentration of salts or reagents in the 10 µl reaction mixture was 5mM ammonium sulfate, 10 mM KCl, 3.2mM MgCl$_2$, 200µM of dNTP's mix, 1 picomol/ µl each of forward and reverse primer.

Table 4.4: Composition of basic 4X cocktail for each marker for 100 reactions

Components	4X CT
10X Complete Buffer (Bioron)	31.3µl
0.5M KCl (Bioron)	20 µl
10mM dNTPS	20 µl
100mM MgCl$_2$ (Bioron)	25 µl
Forward Primer (marker)	1 µl
Reverse Primer (marker)	1 µl
Ultrapure Water	152 µl
Total	250 µl

Table 4.5: PCR amplification conditions for SSLP markers

Steps	Markers D2S2944 & D2S439	D2S1384	D3S3609*	D3S2398	D10S521	D10S1237
Denaturation	95ºC-5'	95ºC-5'	95ºC-5'	95ºC-5'	95ºC-5'	95ºC-5'
Denaturation	95ºC-1'	95ºC-45''	95ºC-30''	95ºC-30''	95ºC-30"	95ºC-30"
Re-annealing	58ºC-1'	58ºC+Δ0.1ºC-30''	47ºC+Δ0.1ºC-75"	51ºC+Δ0.1ºC-75"	54ºC-1'15"	55ºC-1'15"
Extension	72ºC-1'	72ºC-30''	72ºC-1''	72ºC-1''	72ºC-1'	72ºC-1'
Cycles	40 cycles	40 cycles	35cycles	35cycles	40 cycles	40 cycles
Final extension	72ºC-6'	72ºC-6'	72ºC-6'	72ºC-6'	72ºC-7'	72ºC-7'
*In case of D3S3609 Triton was added as additive to obtain specific priming since the primer was only 16mer and was showing problems by probably forming strong secondary structures.						

4.6.3 Gel Electrophoresis

Non-denaturing Polyacrylamide gel electrophoresis (PAGE)

PCR amplified products of microsatellite markers having simple sequence length polymorphism (SSLP) were resolved on native medium length 16cm PAGE.

4.6.3.1 Materials Used

50 and 15 ml microcentrifuge tubes, Tips, Milli Q water, Acrylamide, N,N-Bisacrylamide, Tris base, Boric acid, Ethylenediaminetetraceticacid (0.5M, pH=8),Ammonium persulfate, N,N,N',N'-Tetramethylethylenediamine, Ethidium bromide.

4.6.3.2 Reagents and working solutions

1. 30% Acrylamide: 29:1- Acrylamaide: Bisacrylamide made 100ml with Milli Q water.

2. 5X TBE: 0.09M tris-borate, 0.002M EDTA.

3. 50X TAE: TAE; 0.04M tris-acetate, 0.001M EDTA.

4. 10% APS: 0.1mg APS in 1ml Milli Q water.

5. 6X loading dye: 0.25% bromophenol blue, 0.25% xylene cyanol, 30% v/v glycerol.

6. Ethidium Bromide (EtBr)- 10mg/ml.

4.6.3.3 Polyacrylamide gels for SSLP markers

PCR products of microsatellite markers were resolved on midi-polyacrylamide native gels. Required percentage of polyacrylamide gels was prepared from the 30% acrylamide stock solution with 1X TBE buffer. Final volume was adjusted with autoclaved deionized water to 20 ml. Polymerization was done with the addition of 200 µl of 10%APS and 20 µl of TEMED. The mix was then poured between glass plates (20 X 16 cm) with 1.5 mm spacers. After polymerization, the sandwich was set onto to vertical gel electrophoresis unit (Bangalore Genie, India) and given a pre-run for 2 hrs in 1X TBE running buffer for equilibration. PCR product samples were prepared with 6X loading dye such that the final concentration is 1X in solution and loaded into the well leaving the corner most wells along with DNA size standard (MBI Fermentas, Lithuania). Electrophoresis was

performed at a voltage of 80-100 V for 16 hrs (2hrs at 60V followed by 80-100V for 12hrs and final 2 hrs at 30V). .

4.6.4 Genotyping for repeat polymorphism

Amplified PCR products were resolved on a 12% gradient non-denaturing polyacrylamide gel run on manual sequencer (Banglore Genie) for 16 hrs. The gel was stained with Ethidium Bromide (EtBr) at the concentration 0.5µg/ml and visualized on UV-trans-illuminator (Alphaimager, Alpha Innotech Corporation, USA). Fragment size was determined by comparison to pBR322 *Hae*III digest (MBI Fermentas, Lithuania) and 100bp ladder (MBI Fermentas, Lithuania) according to the fragment resolution required. Details of markers i.e. their PCR product size and the gel run conditions are summarized in Table 4.6.

Table 4.6: Electrophoresis conditions for SSLP markers

Marker	Pdt size	Repeats	% Resolving Gel	Running conditions
D2S1384	133-161bp	4bp (CTAT)		
D2S439	167-203bp	4bp (CTAT)		
D2S2944	92-124bp	4bp (GATA)		
D3S3609	163-185bp	2bp (CA)	12%	16hrs, 80V for 16cm for native gel
D3S2398	266-298bp	4bp (GATA)		
D10S521	155-189	4bp (GATA)		
D10S1237	376-432	4bp (GATA)		

Genotypes of individual subjects were independently read with inbuilt software of Alpha-Imager by two independent observers. The independent readings were compared for consistency before genotypes were entered into the laboratory database. Accurate genotype assignment after repeats and without errors or ambiguity for 2 bp resolution was completed for 406 individuals for D3S3609 polymorphism (205 cases and 201 controls). Similarly for 4bp repeats, accurate assignment of alleles could be allocated for 407

individuals for D2S1384 polymorphism (193 Diabetics and 214 healthy controls), 398 individuals (233 diabetics and 165 healthy controls) for D2S439, 436 individuals for D2S2944 polymorphism (231 cases and 205 controls), 426 individuals (224 cases and 202 controls) for D3S2398 polymorphism , 403 individuals (216 cases and 187 controls) for D10S521 and 413 individuals (223 cases and 190 controls) for D10S1237 polymorphism.

4.7 Statistical Analysis

The association of the above-mentioned polymorphism in the three chromosomes with type 2 diabetes and related traits was examined in the population under study. The distributions of genotype for all the polymorphism were analyzed for deviation from Hardy-Weinberg equilibrium (HWE) using χ2 analysis (Guo *et. al.,* 1992). Descriptive statistics were performed using MedCalc (MedCalc Software, Mariakerke, Belgium) software. Continuous measures were compared using the t-statistic. Categorical variables were compared using χ^2 test and Odds Ratios (OR) with 95% Confidence interval (CI). Frequencies of multiple alleles or genotypes of markers were compared between groups and multivariate associations were calculated using multivariate regression analysis for the dichotomous variables and linear regression analysis for the continuous variables. *p*<0.05 was considered to be statistically significant.

4.8 Bioinformatics approach for evaluation of Secondary Structure of derived transcript/s

The study of RNA structure calls for a distinct set of computational tools designed expressly for RNA applications. For predicting and evaluating the secondary structures online available software tools like mfold and RNAFold (Michael Zuker, Rensselaer

Chapter 4

Polytechnic Institute, U.S.A.) were used. The UNAFold and mfold use the Zuker algorithms to compute the minimum free-energy structure for a given sequence and systematically sample structures within a percentage of free-energy range to create a set of diverse suboptimal structures. The predicted Insilco folding of these markers were associated with the results obtained from the case control analysis to predict the functional importance of the marker. Further, online available software tools like PROMO were used to study the transcription factor binding pattern and tried to establish any change due to increase in the number of repeat expansions further correlating it with the data obtained from the case control study.
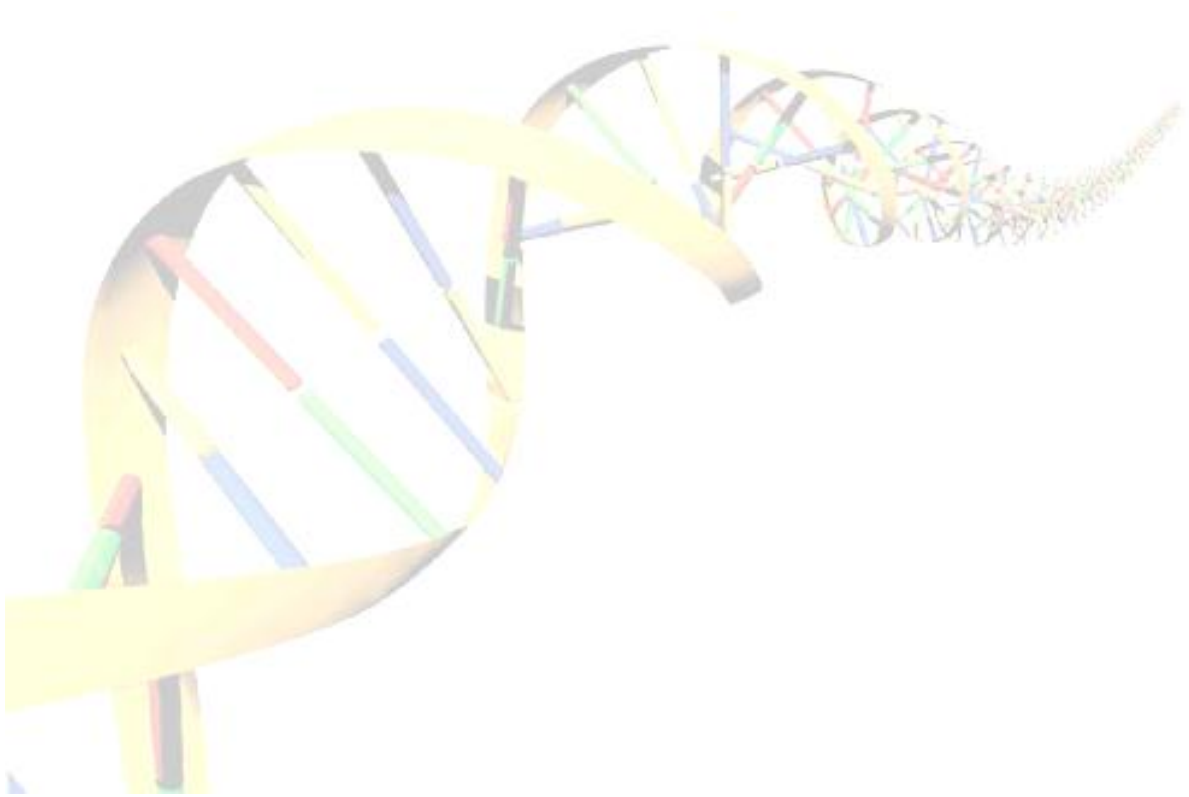
Results

Chapter 5

Chapter 5

Results

5.1 Cohort of Cases and Controls

Two hundred and ninety one (291) subjects with Type 2 Diabetes Mellitus and around four hundred and seventy five (475) unrelated healthy volunteers as controls without any history of diabetes and related comorbidities were recruited for the study initially. Fig 5.1 describes the number of adult diabetic patients and healthy controls for which various clinical, anthropometric markers were determined prior to genetic study. For few samples, only PCV was available which prevented the determination of clinical parameters
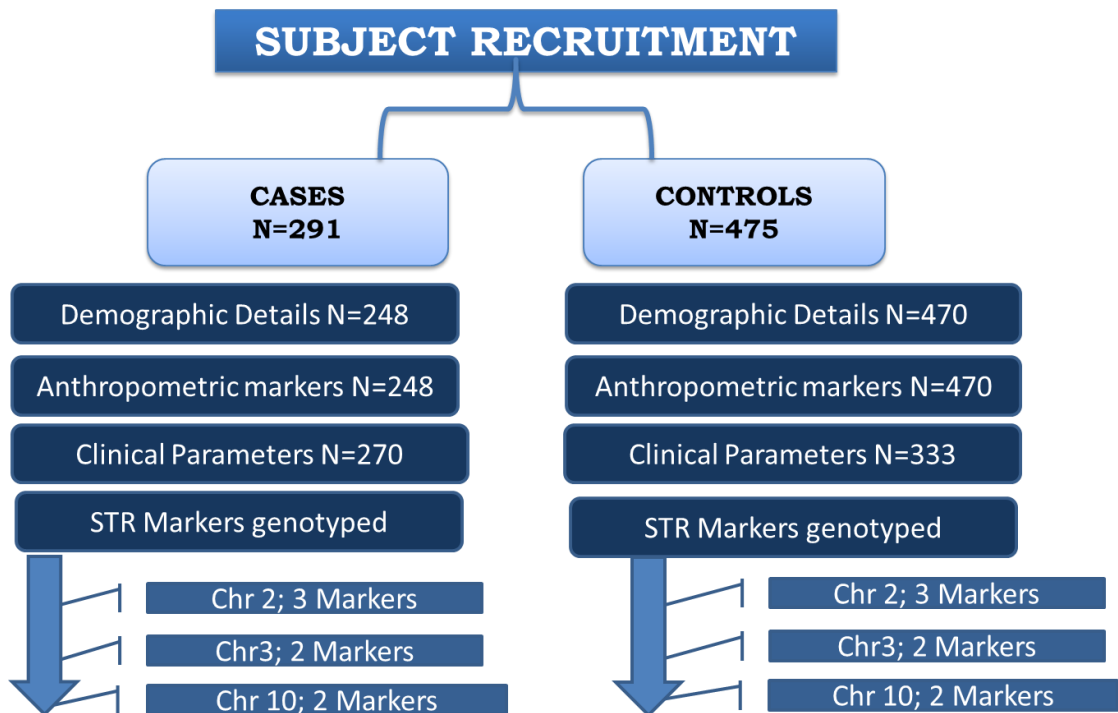
Fig 5.1: Summary of subject recruitment and data collection

Chapter 5

A case control study was carried out in 766 subjects (291 cases and 475 controls). Cases were (mean age 52.78$\pm$11.92 years) who visited our collaborative centers and were examined by senior physicians. Ethnicity matched control subjects (15 to 90 years age) were healthy adults of mean age 49.88$\pm$15.13 with no history of diabetes and associated co morbidities. The control cohort matching to the cases was also recruited from the same region from where the cases were recruited to remove any kind of regional bias. The hyperglycemia (FBG) was taken as the marker for segregating cases from controls. The adult controls included 171 healthy women and 304 healthy males of mean age 53.14$\pm$14.40 and 48.03$\pm$15.25 while the cases included 115 women and 175 men of mean age 48.87$\pm$11.80 and 52.01$\pm$8.7 respectively.

5.1.1 Demographic characteristics of the studied population

The cases and controls differ significantly for the studied anthropometric indices i.e. BMI (Cases vs Controls: 95% CI= 2.781 to 4.299, t=9.153, $p < 0.0001$), WHR (Cases vs Controls: 95% CI= 0.029 to 0.151, t = 2.88, p=0.0040). (Table 5.1) The significance was not lost even when segregated for gender; BMI (Males: t = 6.71, p <0.0001; Females: t= 6.71, p <0.0001) and WHR (Males: t = 5.79, p <0.0001; Females: t = 4.45, p <0.0001) (Table 5.2). The cases showed higher mean abdominal obesity; as reflected by altered WHR (marker for abdominal obesity) as compared to controls (Table 5.1& 5.2). Cases had higher mean values of SBP (95% CI= 3.92 to 9.56, t = 4.693, $p <$ 0.0001), although the difference in mean DBP was found to be non-significant (NS) when compared to the controls. This significance was not lost even on segregation for gender (Males t = 1.87, p = 0.05; Females: t = 5.82, p <0.0001) and DBP (Males: NS; Females: t = 2.75, p = 0.006) than controls (Table 5.2).

Table 5.1: Demographic and anthropometric profile of cases and controls

Variables	Cases		Controls		Cases vs Controls
	N	Mean±SD	N	Mean±SD	
Age (in yr)	248	52.78±11.92$^{\$}$	473	49.88±15.13$^{\$}$	$^{\$}$ 95% CI=0.728 to 5.072, t = 2.622, p = 0.0089
BMI (Kg/m^2)	217	25.74±4.86$^{\wedge}$	473	22.20±4.65$^{\wedge}$	$^{\wedge}$95% CI= 2.781 to 4.299, t=9.153, $p < 0.0001$
WHR	196	1.00±0.60*	406	0.91±0.13*	*95% CI= 0.029 to 0.151, t = 2.88, p=0.0040
SBP (mm/Hg)	248	135.38±15.60$^{\#}$	471	128.64±19.58$^{\#}$	$^{\#}$95% CI= 3.92 to 9.56, t = 4.693, $p < 0.0001$
DBP (mm/Hg)	248	85.02±9.41$^{\%}$	471	84.16±11.51$^{\%}$	$^{\%}$NS

Yr: year, BMI: Body Mass Index, WHR: Waist Hip Ratio, SBP: Systolic Blood pressure, DBP: Diastolic Blood Pressure, SD: Standard Deviation, NS: Non Significant; Values are presented as means ± standard deviation, *p* values differed significantly among cases and controls (*p*<0.05)

Table 5.2: Gender segregation for demographic and anthropometric profile of cases and controls

Variables	Cases, N=248		Controls, N=473	
	Females, N=97	Females, N=97	Females, N=170	Males, N=303
	Mean±SD	Mean±SD	Mean±SD	Mean±SD
Age (in yr)	48.76±11.80$^{\$}$	52.007±11.88$^{\$\$}$	53.14±14.40$^{\$}$	48.03±15.25$^{\$\$}$
BMI (Kg/m^2)	26.74±5.39$^{\wedge}$	25.09±4.37$^{\wedge\wedge}$	22.22±5.23$^{\wedge}$	22.20±4.30$^{\wedge\wedge}$
WHR	1.02±0.96*	1.00±0.07**	0.87±0.08*	0.93±0.14**
SBP (mm/Hg)	136.16±14.58$^{\#}$	134.89±16.26$^{\#\#}$	123.87±17.64$^{\#}$	131.36±20.13$^{\#\#}$
DBP (mm/Hg)	84.4±9.10$^{\%}$	85.46±9.60$^{\%\%}$	80.71±10.61$^{\%}$	86.12±11.55$^{\%\%}$
$^{\$\$}$95% CI=1.18 to 6.75, t= 2.80, p =0.005			$^{\$}$95% CI=0.994 to 7.766, t = 2.547, p = 0.01	
$^{\wedge\wedge}$95% CI=2.04 to 3.73, t = 6.71, p <0.0001			$^{\wedge}$95% CI=5.845 to 3.195, t= 6.71, p <0.0001	
**95% CI=0.04 to 0.09, t = 5.79, p <0.0001			*95% CI=0.034 to 0.086, t = 4.45, p <0.0001	
$^{\#\#}$95% CI=0.177 to 7.23, t = 1.87, p = 0.05			$^{\#}$95% CI=8.13 to 16.448, t = 5.82, p <0.0001	
$^{\%\%}$ NS			$^{\%}$95% CI=1.00 to 6.05, t = 2.75, p = 0.006	

Yr: year, BMI: Body Mass Index, WHR: Waist Hip Ratio, SBP: Systolic Blood Pressure, DBP: Diastolic Blood Pressure, SD: Standard Deviation, NS: Non Significant; Values are presented as means ± standard deviation, *p* values differed significantly among male and female cases (*p*<0.05)

Fig 5.2 describes the distribution of BMI and WHR among cases and controls using a Box-whisker plot. It was found that both overall obesity (marked by BMI) and central

Chapter 5

obesity (marked by WHR) were found significantly elevated in cases as compared to controls. This observation also supports the significant difference observed in the mean BMI and WHR values represented in Table 5.2. This distribution of the obesity parameters was clearly and significantly varied in cases as compared to controls, which is very clearly evident from the Fig 5.3.

Fig 5.3 describes the distribution of SBP and DBP among cases and controls. It is clearly evident that the SBP was elevated in cases as compared to the controls while the DBP seems to be almost comparable in cases and controls. Again the observation is supported by the significant difference observed in mean systolic blood pressure and the non-significant difference observed in mean diastolic blood pressure among cases and controls (Table 5.2). The significance in terms of DBP also seems to be lost due to the presence of a set of outlier values in control population which in turn changes the mean value of the parameter. Though few extreme values of SBP can also be seen in controls, however they are unable to impart much effect on the mean of the group since the cases also had a similar percentage outliers which negates the observed effect. Also as mentioned above the Table 5.2 clearly shows that the DBP was significantly raised in females cases as compared to female controls, which was not there in case of males.

It is evident that there was significant difference in the distribution of the anthropometric indices among the recruited cases and controls. This difference in most of the parameters was not lost even on gender segregation.

Distribution of BMI among cases and controls

Distribution of WHR among cases and controls

Fig 5.2: Distribution of BMI and WHR among cases and Controls

Distribution of SBP among cases and controls
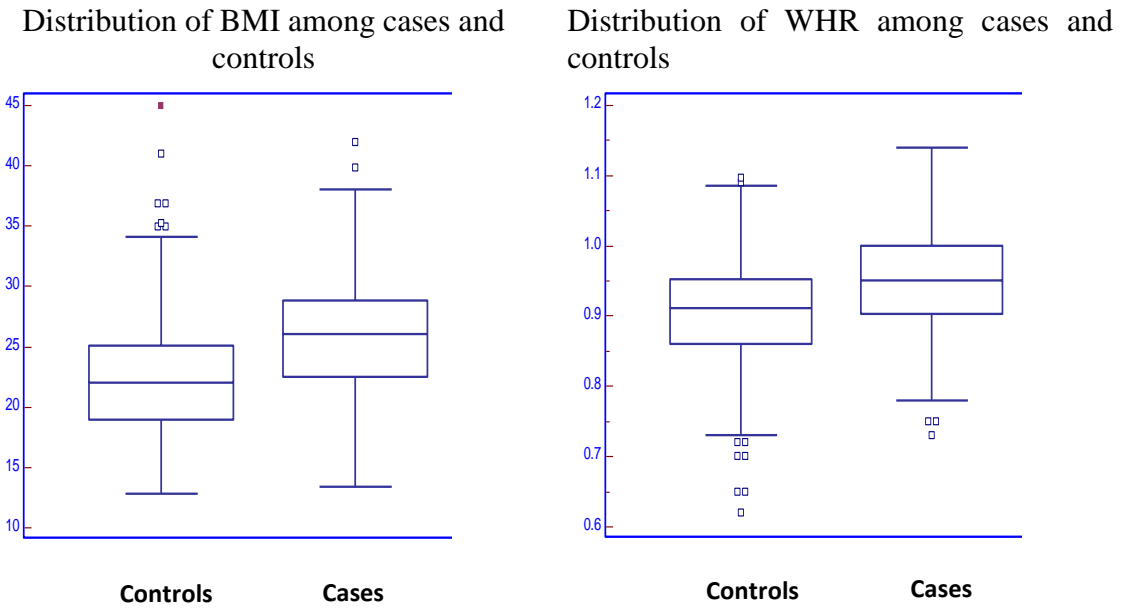
Distribution of DBP among cases and controls
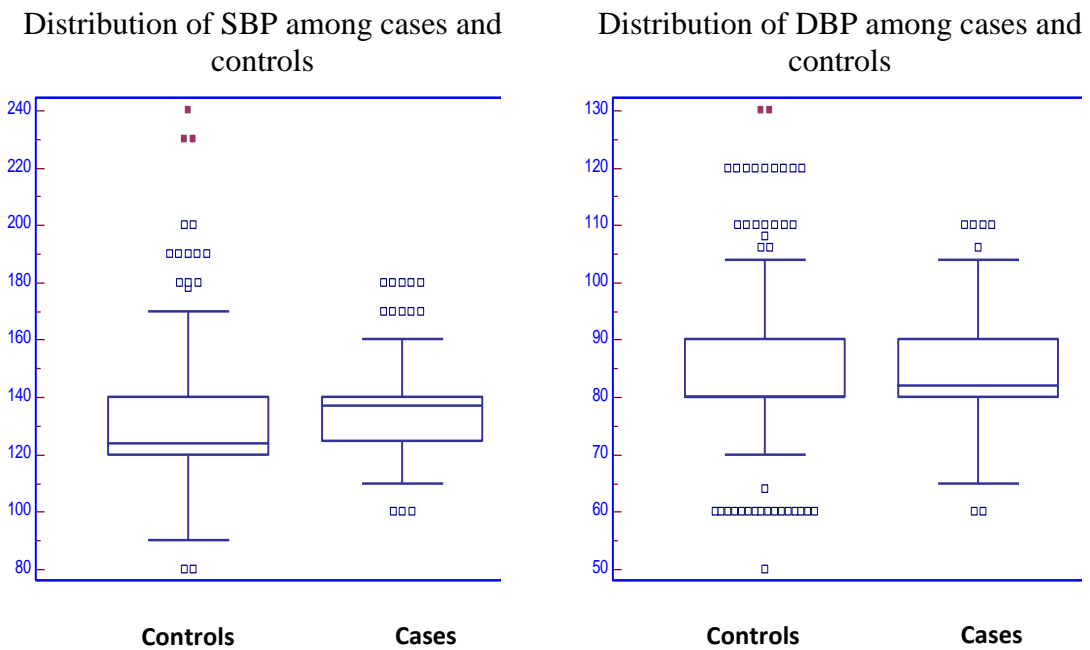
Fig 5.3: Distribution of SBP and DBP among cases and Controls

Chapter 5

5.1.2 Biochemical and clinical characteristics among Cases and Controls

The cases and controls significantly differed in clinical and biochemical indicators. Taking FBG as the indicator for hyperglycemia, the mean FBG among cases and controls was found to be significantly different (95% CI=81.232 to 92.648, t = 29.904, $p < 0.0001$) and this significance again was not lost with gender segregation (Males: NS; Females: t = 2.75, $p = 0.006$).

Table 5.3 Biochemical and clinical characteristics among Cases and Controls

	Cases		Controls		Cases vs controls
	N	Mean+SD	N	Mean+SD	
FBG (mg/dl)	250	175.31+60.1	474	88.36+14.5	95% CI=81.232 to 92.648, t = 29.904, $p < 0.0001$
Cholesterol (mg/dl)	270	196.93+51.0	333	173.31+42.7	95% CI=16.125 to 31.115, t = 6.190, $p < 0.0001$
Triglyceride (mg/dl)	270	171.59+100.9	333	148.38+78.499	95% CI=8.926 to 37.494, t = 3.191, $p = 0.0015$
HDL (mg/dl)	270	45.92+21.5	333	36.52+18.9	95% CI=6.135 to 12.665, t = 5.655, $p < 0.0001$
LDL (mg/dl)	270	117.11+49.7	333	107.25+41.02	95% CI=2.604 to 17.116, t = 2.669, $p = 0.0078$
VLDL (mg/dl)	270	34.06+19.9	333	29.66+15.7	95% CI=1.553 to 7.247, t = 3.035, $p = 0.0025$

FBG: Fasting Blood Glucose, HDL: High Density lipoprotein, LDL: Low density lipoprotein, VLDL: Very Low Density Lipoprotein; SD: Standard Deviation; Values are presented as means $\pm$ standard deviation, p values differed significantly among cases and controls (p<0.05)

The cases also had significantly higher mean for Lipid markers; Total Cholesterol (Cases vs Controls: 95% CI=16.125 to 31.115, t = 6.190, $p < 0.0001$), Triglyceride (Cases vs Controls: 95% CI=2.569 to 36.969, t = 1.70, $p = 0.08$); HDL (Cases vs Controls: 95% CI=6.135 to 12.665, t = 5.655, $p < 0.0001$); LDL (95% CI=2.604 to 17.116, t = 2.669, $p = 0.0078$) and VLDL (Cases vs Controls: 95% CI=1.553 to 7.247, t = 3.035, $p = 0.0025$). In fact significant differences were even observed after

Chapter 5

segregating for gender (Table 5.4). In control group females had significantly raised level of mean FBG (females vs males: t = 18.362; *p*< 0.0001), TC (females vs males: t = 4.769; *p*< 0.0001), TG (females vs males: t = 5.712; *p*< 0.0001), HDL (females vs males t = 5.712; *p*< 0.0001) and VLDL (females vs males: t = 5.513; *p*< 0.0001) as compared to males. Whereas in cases significant differences were only seen in case of FBG (females vs males: t = 22.054; *p*< 0.0001), TC (females vs males: t=3.337; *p*=0.0009) and HDL (females vs males: t = 4.022; *p*=0.0001). Fig 5.4 and 5.5 are a visual representation of how the various biochemical parameters are spread across the studied cohort there by highlight the differences in the individual values among cases and controls. As evident from the tables and the graphs below all the lipid markers were significantly altered in cases as compared to controls.

The graph below (Fig 5.4) describes the distribution of fasting blood glucose among cases and controls using a Box whisker plot. Since FBG or hyperglycemia was used as the marker to segregate cases and controls it was expected that the FBG levels among cases would be significantly higher than that of controls (Fig 5.4).
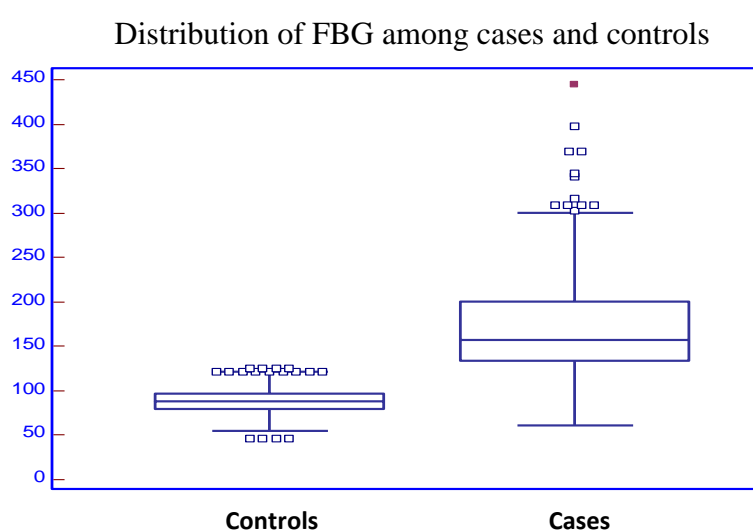
Fig 5.4: Distribution of Fasting blood glucose among cases and Controls

Chapter 5

Table 5.4: Gender segregation for biochemical and clinical characteristics among Cases and Controls

Variable	Controls				Cases			
	Females		Males		Females		Males	
	N	Mean±SD	N	Mean±SD	N	Mean±SD	N	Mean±SD
FBG(mg/dl)	97	$171.61\pm56.0^{\#}$	153	$177.66\pm62.7^{\#\#}$	97	$171.61\pm56.0^{\#}$	153	$177.66\pm62.7^{\#\#}$
Cholesterol (mg/dl)	94	$215.63\pm59.7^{\wedge}$	146	$186.13\pm41.3^{\wedge\wedge}$	94	$215.63\pm59.7^{\wedge}$	146	$186.13\pm41.3^{\wedge\wedge}$
Triglyceride (mg/dl)	94	$180.81\pm98.5^{\$}$	146	$170.48\pm99.1^{\$\$}$	94	$180.81\pm98.5^{\$}$	146	$170.48\pm99.1^{\$\$}$
HDL(mg/dl)	94	$46.69\pm23.6^{*}$	146	$43.80\pm19.8^{**}$	94	$46.69\pm23.6^{*}$	146	$43.80\pm19.8^{**}$
LDL(mg/dl)	94	$129.47\pm58.6^{\%}$	146	$109.48\pm41.3^{\%\%}$	94	$129.47\pm58.6^{\%}$	146	$109.48\pm41.3^{\%\%}$
VLDL(mg/dl)	94	$35.86\pm20.1^{\beta}$	146	$33.81\pm19.2^{\beta\beta}$	94	$35.86\pm20.1^{\beta}$	146	$33.81\pm19.2^{\beta\beta}$
$^{\#}$95% CI= 73.984 to 91.756;t = 18.362;$p< 0.0001$				$^{\#\#}$95% CI= 76.078 to 90.962;t = 22.054;$p< 0.0001$				
$^{\wedge}$95% CI= 22.846 to 55.074;t = 4.769;$p< 0.0001$				$^{\wedge\wedge\wedge}$95% CI= 5.841 to 22.599;t=3.337;$p=0.0009$				
$^{\$}$95% CI= 45.722 to 93.958;t = 5.712;$p< 0.0001$				$^{\$\$}$NS				
*95% CI= 2.372 to 15.328;t = 5.712;$p< 0.0001$				**95% CI= 4.002 to 11.658;t = 4.022;p=0.0001				
$^{\%}$NS				$^{\%\%}$NS				
$^{\beta}$95% CI= 8.779 to 18.561;t = 5.513;$p< 0.0001$				$^{\beta\beta}$NS				

FBG: Fasting Blood Glucose, HDL: High Density lipoprotein, LDL: Low density lipoprotein, VLDL: Very Low Density Lipoprotein; SD: Standard Deviation; Values are presented as means ± standard deviation, *p* values differed significantly among cases and controls (p<0.05)

Chapter 5

Fig 5.5 A & B: Distribution of Lipid Markers (TC&TG) among cases and controls
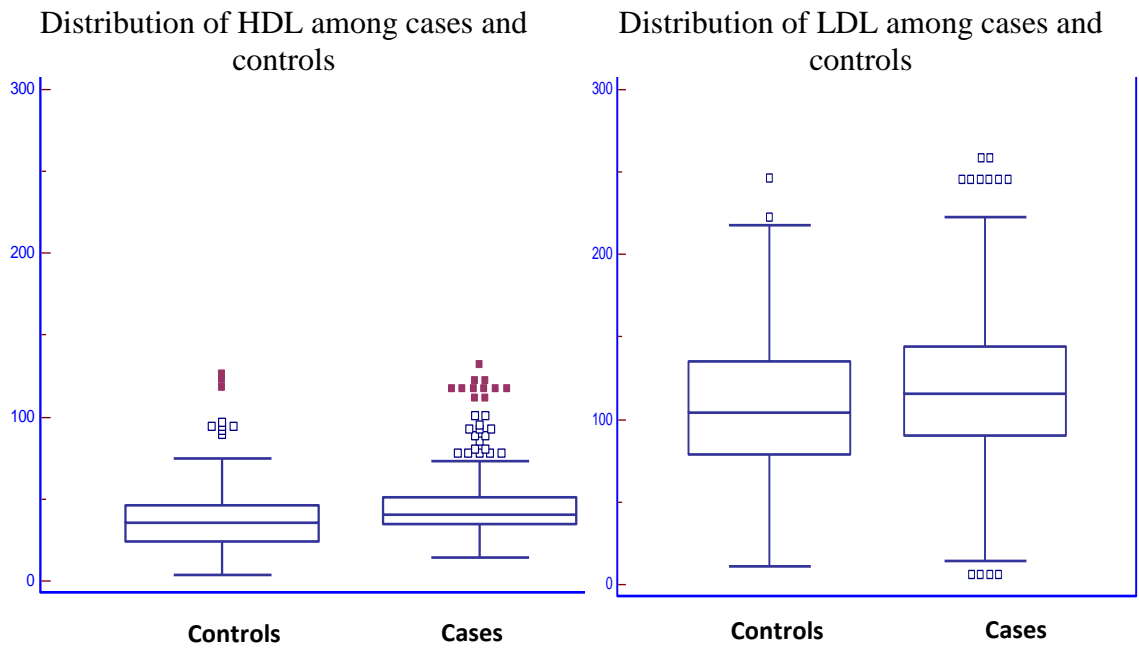
Fig 5.5 C & D: Distribution of Lipid Markers (HDL & LDL) among cases and Controls

Chapter 5

5.1.3 Prevalence of Hypertension among Cases and Controls

The cases and controls were distributed according to JNC VII (Joint National committee on prevention, detection, evaluation, and treatment of High Blood Pressure) criteria for the hypertension. The distribution among cases and controls is represented in Table 5.5. Among the total recruited diabetics only 7% cases were normotensives as compared to the 16% controls (Controls vs Cases: 95% CI = 3.9 to 13.6; χ^2 = 10.209; p= 0.0014); 37% cases were pre-hypertensive in comparison to 51% controls (Controls vs Cases: 95% CI = 6.1 to 21.6; χ^2 = 11.435; p= 0.0007). While 35% controls and 55% cases were clearly hypertensive (Controls vs Cases: 95% CI = 12.1 to 27.6; χ^2 = 24.149; p< 0.0001). As evident the cases had a significantly higher prevalence of hypertension as compared to the healthy control population, though it seems from the data that probably the control population itself had a higher prevalence (~35%) of hypertension.

Table 5.5: Prevalence of hypertension among cases and controls

SBP	DBP Group	<80 N (%)	80-89 N (%)	>90 N (%)	Cases vs controls : Hypertension
<120	Controls	70 (16)	25 (6)	5 (1)	Normotensives: Controls vs Cases: 16% vs 7%: 95% CI = 3.9 to 13.6; χ^2 = 10.209; p= 0.0014
	Cases	17 (7)	9 (4)	1 (0)	
120-139	Controls	13 (3)	184 (42)	14 (3)	Pre Hypertensive Controls vs Cases:51% vs 37%: 95% CI = 6.1 to 21.6; χ^2 = 11.435; p= 0.0007
	Cases	10 (4)	72 (29)	18 (7)	
140-159	Controls	2 (0)	17 (4)	104 (24)	Hypertensive : Controls vs Cases: 35% vs 55%: 95% CI = 12.1 to 27.6; χ^2 = 24.149 ; p< 0.0001
	Cases	3 (1)	38 (15)	80 (32)	

SBP: Systolic Blood Pressure; DBP: Diastolic Blood pressure; The Values are presented as percentages; p values differed significantly among cases and controls (p<0.05)

Chapter 5

5.1.4 Prevalence of overall obesity among cases and controls according to their gender

Table 5.6 & 5.7 presents the overall prevalence of overall obesity and abdominal obesity among cases & controls. The data was also evaluated after segregating for gender. All cut offs are according to cutoffs prescribes for Asian Indians by Snehalatha *et al*., 2003 and Misra *et al*., 2009. There was statistically significant difference in the distribution of BMI among the cases as compared the controls (Controls vs Cases: 95% CI=22.56 to 37.85, χ^2=58.620, $p < 0.0001$). This significant difference in BMI among cases and controls was still observed on gender segregation, Males (cases vs Controls): 95% CI=13.97 to 33.65, χ^2=21.31, $p < 0.0001$; Females (cases vs Controls): 95% CI=28.11 to 51.70, χ^2=39.25, $p <0.0001$). Nearly 62.7% females and 51.9% males had BMI above 25 as compared to the control population (22% females and 27.9% males)

Table 5.6: Prevalence of Overall obesity among cases & controls according to gender

BMI	Controls (N=469)			Cases (N=217)		
	Total N (%)	Male N (%)	Female N (%)	Total N (%)	Male N (%)	Female N (%)
Below 18	90 (19.3)	57 (19.1)	33 (19.6)	11 (5.06)	8 (6.1)	3 (3.4)
18 – 22.9	188 (40.3)	114 (38.2)	74 (44.0)	53 (24.4)	33 (25.1)	20 (23.3)
23 – 24.9	68 (14.6)	44 (14.8)	24 (14.3)	31 (14.2)	22 (16.8)	9 (10.4)
Above 25	120 (25.8)[*]	83 (27.9)[#]	37 (22.0)[$]	122(56)[*]	68 (51.9)[#]	54 (62.7)[$]

[*]BMI>25 Cases vs Controls : 95% CI=22.56 to 37.85, χ^2=58.620, $p < 0.0001$;
[#]BMI>25 Males (cases vs Controls): 95% CI=13.97 to 33.65, χ^2=21.31, $p < 0.0001$
[$]BMI>25 Females (cases vs Controls): 95% CI=28.11 to 51.70, χ^2=39.25, $p <0.0001$

The Values are presented as percentages; *p* values differed significantly among cases and controls (p<0.05); BMI: Body Mass Index; N: Number of Individuals; %: Frequency

Chapter 5

5.1.5 Prevalence of Abdominal obesity among cases and controls according to their gender

Table 5.7 summarizes Waist Hip Ratio (WHR) cutoffs in cases and controls. All cut offs are according to cutoffs prescribes for Asian Indians by Snehalatha *et al.,* 2003 and Misra *et al.,* 2009. It can also be seen from the table 5.7 that cases had higher prevalence of abdominal obesity as compared to the control group (Males vs females: $\chi^2 = 17.766$, *p*< 0.0001 vs $\chi^2 = 11.095$, *p*=0.0009). Nearly 93% cases (males) had altered Waist Hip Ratio (WHR $\geq$0.89), as compared to 73.7% controls (males). Infact in case of females also nearly 97.4% females has altered waist hip ratio as compared to 80.3 controls (females). The data may reflect the probable association of central adiposity marker (WHR) and overall obesity marker (BMI) with T2DM independent of any other risk factor. It is well evident that the Indians are predisposed to the abdominal obesity and it is well evident from the demographics seen in the tables below that the prevalence of central adiposity was also higher in the control cohort (general population) both in males (73.7%) and females (80.3%), which probably is not a very good indicator in terms of the health parameters.

Table 5.7: Prevalence of abdominal obesity among cases and controls according to their gender

WHR	Males		WHR	Females	
	Controls N (%)	Cases N (%)		Controls N (%)	Cases N (%)
<0.89	72 (26.2)	8 (6.8)	<0.81	26 (19.7)	02 (2.5)
$\geq$0.89	202 (73.7)*	109 (93.1)*	$\geq$0.81	106 (80.3)[#]	77 (97.4)[#]

*Male cases vs male controls: 95% CI= 11.61 to 25.85, χ^2= 17.766, *p*< 0.0001; [#] Female cases vs female controls: 95% CI= 8.50 to 24.93, χ^2= 11.095, *p*=0.0009; WHR cut offs for Asian Indians;

WHR :Waist Hip Ratio; N:Number of Individuals; %:Percentage

Chapter 5

5.2 Association of Chromosome 2 loci with T2DM

The fig 5.6 shows the region studied on chromosome 2 and the three markers D2S1384, D2S2944 and D2S439 studied in the present study are encircled. The details of the region as per UniSTS along with the hetrozygosity score of each marker is:

Region: 2q32-27

Bp size: 182700000-242951000

Marker position: 183M-243M
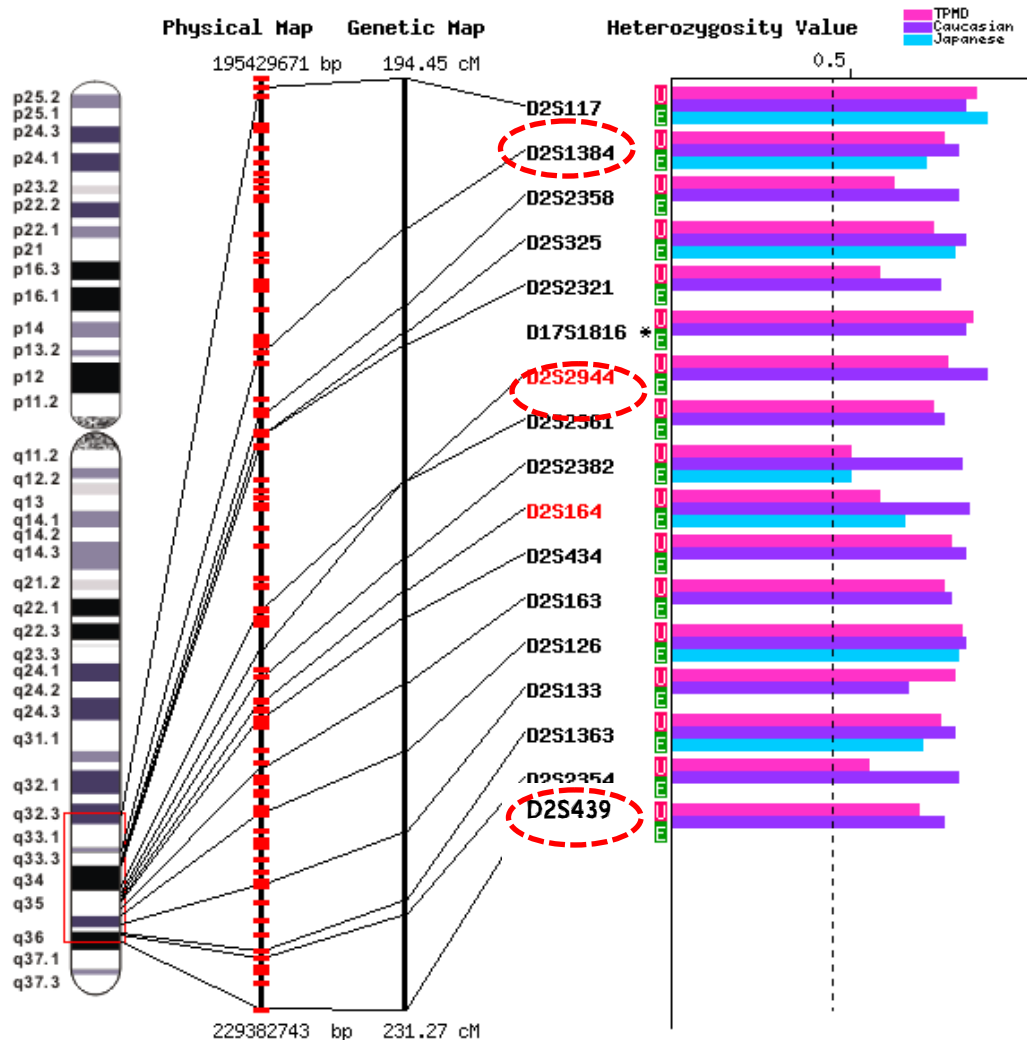
Fig 5.6: Position and Hetrozygosity score of STR markers analyzed on chromosome 2. The red square box represents the region chosen for the study. The encircled regions represent the markers studied in the current study.

5.2.1   D2S1384 polymorphism & its association with T2DM

To study the role of genetic variants of Short Tandem Repeat (STR) marker D2S1384 in diabetes we examined alleles of the, D2S1384. The alleles ranged in size from 133 to 171 bp in PCR product length corresponding to 9-19(CTAT) n tetra repeats (Gel Image in Appendices III: Fig A1). The control population was in HWE and the Exact test (Guo and Thompson 1992) gave a *p* value of >0.05. Population data showed that this is a highly polymorphic STR with a heterozygosity and polymorphic information content of more than 0.80.

5.2.1.1 Comparison with world frequencies

The distribution of major alleles in our control population was compared with the reported world frequencies obtained from ALFRED. The significant interethnic variations for the marker frequency can be seen across the groups. The frequency of the most common alleles observed in our population matches with those reported for Africans, Europeans and East Asians (Fig 5.7)
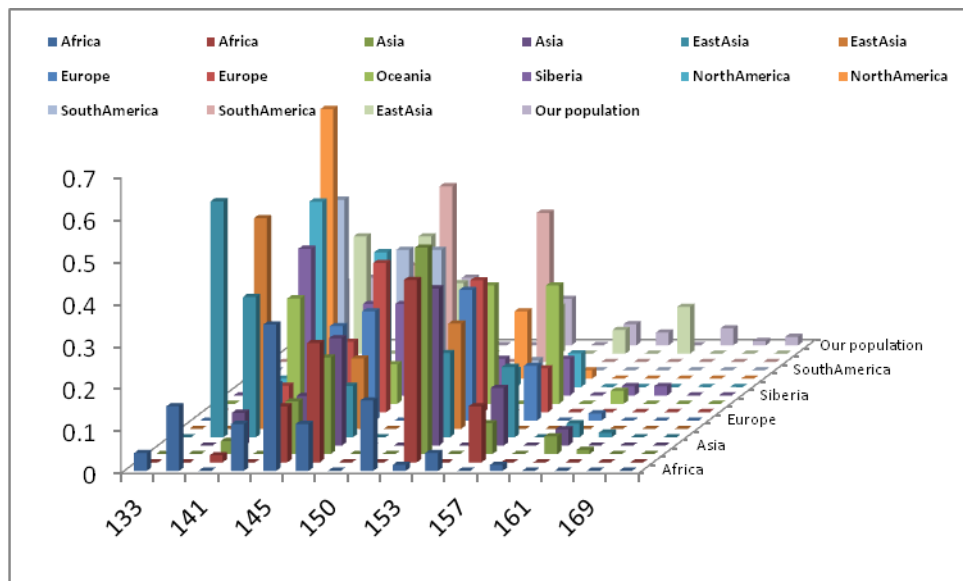
Fig 5.7: Distribution of D2S1384 allele frequencies across different regions of the world

Chapter 5

5.2.1.2 Allelic Distribution of D2S1384 marker among cases and controls

The various alleles of D2S1384 as observed in the studied cohort are compiled in table 5.8 and fig 5.8

Table 5.8: Allelic Distribution of D2S1384 marker among cases and controls

PCR Pdt Size (bp)	(CTAT)n repeats	Allele	Cases (N=191)	Controls (N=214)
			N (%)	N (%)
133	9	1	34 (0.09)	29 (0.07)
137	10	2	43 (0.11)	63 (0.15)
141	11	3	19 (0.05)[1]	67 (0.16)[1]
145	12	4	87 (0.23)	91 (0.21)
149	13	5	106 (0.28)[2]	70 (0.16)[2]
153	14	6	53 (0.14)	47 (0.11)
157	15	7	15 (0.04)	21 (0.05)
161	16	8	08 (0.02)	13 (0.03)
165	17	9	09 (0.02)	16 (0.04)
169	18	10	08 (0.02)	06 (0.01)
173	19	11	00 (0.00)	08 (0.02)

[1]=95% CI = 0.1829 to 0.5333; χ^2= 18.7, $p<.0001$; OR=0.3123
[2]=95% CI 1.4096 to 2.7830; χ^2= 15.139, $p=0.0001$, OR=1.9806

PCR Pdt: PCR product; bp: Base pair; N: Number; %: Frequency
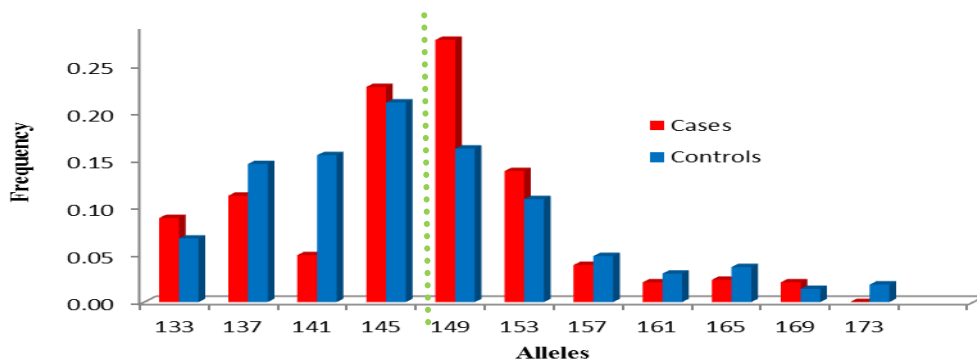
Fig 5.8: Distribution of major allele frequencies of D2S1384 STR in cases vs controls. Dotted lines represent the cut off made according to natural tendency for bimodal distribution of alleles. Red color represents cases; Blue represents controls.

Chapter 5

5.2.1.3 Bimodal Distribution of D2S1384 Alleles

The cohort was further analyzed using a case-control design on the genotype data of D2S1384 from 193 diabetics and 214 controls. It is evident from table 5.8 that possession of 141bp allele (95% CI = 0.1829 to 0.5333; $\chi 2$= 18.7, p<.0001; OR=0.3123) of D2S1384 marker protects an individual from T2DM while presence of the longer 149 bp allele seems to make individuals more susceptible to T2DM (95% CI 1.4096 to 2.7830; $\chi 2$= 15.139, p=0.0001, OR=1.9806). Others alleles did not exhibit any significant difference between cases and controls groups. The cohort was further divided into approximately equal groups, at 145 bp or $(CTAT)_{12}$ repeats, according to their natural tendency for bimodal distribution. Hence the resulting genotypes were <145 bp/<145 bp, <145 bp//$\geq$145 bp and >145 bp/$\geq$145 bp. The shorter alleles (or <12 (CTAT) repeats) of D2S1384, were found to be more prevalent in the control population. The <145bp allele constitutes 25% in adult cases and 38% in controls (OR=1.91, χ^2=14.679, 95 % CI= 1.3755 to 2.6468, p = 0.0001)

Table 5.9: Allelic and genotypic frequencies of D2S1384 (Bimodal distribution) Microsatellite Repeat in cases and controls

Group (N)	Genotype Frequency				Allele frequency	
	<145	<145/$\geq$145	$\geq$145		<145*	$\geq$145*
	N (%)	N (%)	N (%)	2N	N (%)	N (%)
Cases (193)	14 (0.07)	68 (0.35)	111 (51.8)	386	96 (0.25)	290 (0.75)
Controls (214)	47 (0.21)	67 (0.31)	100 (0.47)	428	161 (0.38)	267 (0.62)
Difference among genotypes: χ^2=14.679 , p=0.0001; *Longer allele ($\geq$145bp) as risk allele: OR = 1.82, 95 % CI= 1.3461 to 2.4650, p = 0.0001 (<145 & <145/$\geq$145) as protective genotype: OR = 0.29, 95 % CI = 0.1479 to 0.5825, p = 0.0005; (<145 vs $\geq$145bp): OR = 3.73, 95 % CI = 1.9355 to 7.1747, p=0.0001						

N: Number; %: Frequency

Chapter 5

Fig 5.9 A : Genotypic distribution of D2S1384 STR Repeat after Bimodal distribution. Red color represents cases; Blue represents controls.

Fig 5.9 B: Allelic distribution of D2S1384 STR Repeat after Bimodal distribution. Red color represents cases; Blue represents controls.

5.2.1.4 Distribution of different variables according to D2S1384 alleles

The distribution of anthropometric in adult cohort alleles of D2S1384 is presented in Table 5.10. Significant difference was observed in mean values of systolic blood pressure, SBP (<145bp: t = 2.455, p = 0.01; ≥145bp: t = 2.595, p = 0.0101) however no significant difference was observed in diastolic blood pressure, DBP (<145bp: Non-Significant; ≥145bp: Non-Significant) among cases and controls with either long or short allele of D2S1384. However Body mass Index, BMI and Waist Hip Ratio, WHR levels differ significantly in cases and controls in presence of longer base pair allele (<145bp: Non-Significant; ≥145bp: t = 6.086, p < 0.0001) and (<145bp: NS; ≥145bp: t = 4.250 , p < 0.0001) respectively. Probably individuals with shorter base pair allele tend to have lower BMI and WHR as compared to the individuals with longer base pair allele. Both of these indicators are markers for obesity. It is well known that diabetes & obesity are interrelated and share similar causes and indicators. Thus the longer base pair allele, seems to be associated with higher BMI and WHR, although this

Chapter 5

observation will only validate if this significance is maintained even after controlling for the obesity markers (BMI and WHR)

Table 5.10: Distribution of different anthropometric variables according to D2S1384 genotype among cases and controls with gender segregation

		<145		<145/ ≥145		≥145	
	N	Mean±SD	N	Mean±SD	N	Mean±SD	
SBP	Controls 47	125.75±15.55	Controls 67	130.1±19.1	Controls 100	128.21±21.9	
	Cases 14	137.92±18.65	Cases 68	137.2±14.8	Cases 109	135.04±15.9	
	t = 2.455, p = 0.01		t=2.416, p = 0.0170		t = 2.595, p = 0.0101		
DBP	Controls 47	80.68±10.41	Controls 67	84.1±10.70	Controls 100	82.40±10.72	
	Cases 14	85.28±10.62	Cases 68	84.6±8.62	Cases 109	84.15±9.70	
	NS		NS		NS		
BMI	Controls 47	21.47±6.06	Controls 67	20.7±4.19	Controls 99	21.14±4.69	
	14	23.87±5.84	Cases 64	26.3±4.73	Cases 97	25.34±4.97	
	NS		t = 7.278, p < 0.0001		t = 6.086, p < 0.0001		
WHR	Controls 40	0.934±0.085	Controls 51	0.91±0.072	Controls 67	0.89±0.095	
	Cases 11	0.931±0.0411	Cases 57	0.96±0.082	Cases 85	0.95±0.0798	
	NS		t = 3.350, p = 0.0011		t = 4.250 , p < 0.0001		

BMI: Body Mass Index, WHR: Waist Hip Ratio, SBP: Systolic Blood pressure, DBP: Diastolic Blood Pressure, SD: Standard Deviation, NS: Non-Significant. Values are presented as means $\pm$ standard deviation, p values differed significantly among cases and controls (p<0.05)

The distribution of biochemical parameters in adult cohort alleles of D2S1384 is presented in Table 5.11. No significant difference was observed in mean values of Fasting blood glucose, FBG (<145bp: t = 12.407, p < 0.0001; ≥145bp: t = 13.553, p < 0.0001) and Low density lipoproteins, LDL (<145bp: Non-Significant; ≥145bp: Non-Significant) among cases and controls with either long or short allele of D2S1384. However, significant differences were observed in levels of Total cholesterol, TC

(<145bp: Non-significant; ≥145bp: t = 1.770, *p* = 0.0788), Triglyceride level, TG (<145bp: t = 2.803, *p* = 0.0060; ≥145bp: t = 3.722, *p* =0.0003) and VLDL (Non-significant; ≥145bp: t = 3.721 *p* = 0.0003) cases and controls in presence of longer allele.

Fig 5.11: Distribution of different biochemical variables according to D2S1384 genotype among cases and controls with gender segregation

		<145		<145/ ≥145		≥145	
	N	Mean±SD	N	Mean±SD	N	Mean±SD	
FBG	Controls 47	87.5±13.11	Controls 67	92.3±12.05	Controls 99	90.50±16.08	
	Cases 14	172.92±41.38	Cases 68	168.3±53.9	Cases 111	179.3±63.46	
	t = 12.407, *p* < 0.0001		t = 11.267, *p* < 0.0001		t = 13.553, *p* < 0.0001		
CHL	Controls 30	162.36±42.9	Controls 43	166.1±38.6	Controls 43	178.55±49.7	
	Cases 11	188.00±45.1	Cases 64	195.3±42.8	Cases 109	194.9±51.96	
	NS		t = 3.597, *p* = 0.0005		t = 1.770, *p* = 0.0788		
TG	Controls 30	112.28±58.0	Controls 43	101.9±44.7	Controls 43	112.9±65.68	
	Cases 11	145.99±62.9	Cases 64	161.0±67.9	Cases 109	176.4±103.9	
	t = 2.803, *p* = 0.0060		t = 5.020, *p* < 0.0001		t = 3.722, *p* =0.0003		
HDL	Controls 30	31.37±19.51	Controls 43	30.6±18.38	Controls 43	37.94±23.82	
	Cases 11	47.79±27.76	Cases 64	41.0±16.58	Cases 109	45.7±21.84	
	t=2.125, *p* = 0.0400		t = 3.045 , *p* = 0.0029		t = 1.923, *p* = 0.0564		
LDL	Controls 30	108.53±42.2	Controls 43	115.1±37.9	Controls 43	116.4±51.19	
	Cases 11	108.32±55.6	Cases 64	122.4±45.1	Cases 109	114.5±48.74	
	NS		NS		NS		
VLDL	Controls 30	22.45±11.61	Controls 43	20.3±8.95	Controls 43	22.4±13.30	
	Cases 11	28.71±13.54	Cases 64	32.1±13.72	Cases 109	34.7±19.98	
	NS		t = 4.970, *p* < 0.0001		t = 3.721 *p* = 0.0003		

FBG: Fasting Blood Glucose, HDL: High Density lipoprotein, LDL: Low density lipoprotein, VLDL: Very Low Density Lipoprotein; CHL: Total cholesterol. Values are presented as means ± standard deviation, *p* values differed significantly among cases and controls (*p*<0.05)
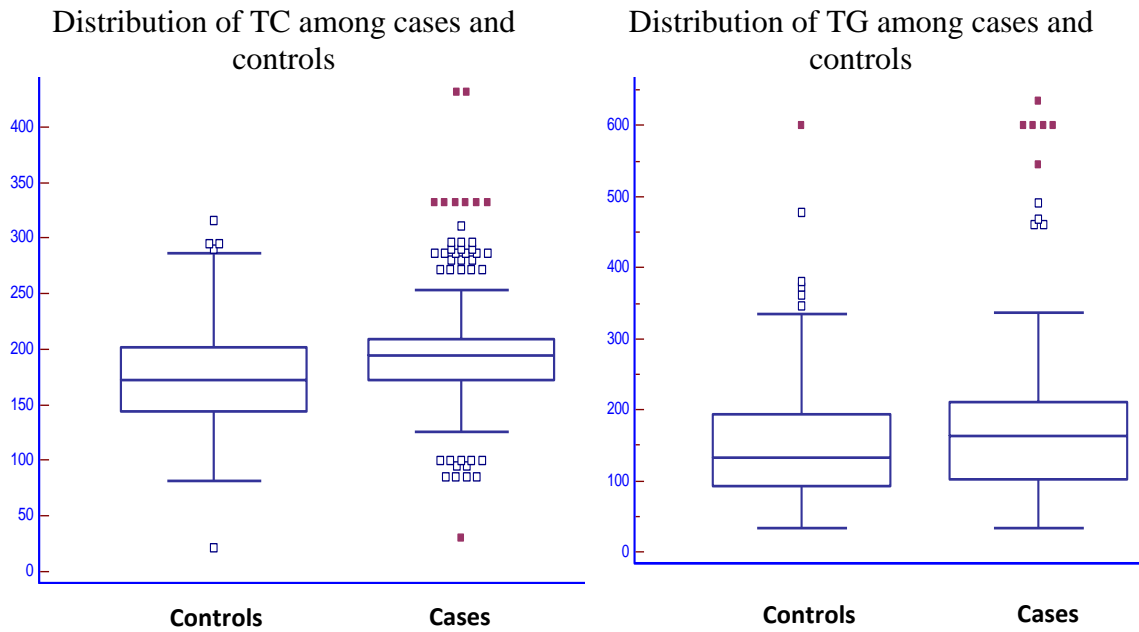
Chapter 5

Table 5.12 summarizes the distribution of D2S1384 genotype and allele after an adjustment for BMI (BMI<22.9kg/m^2 and >23kg/m^2) among cases and controls. The difference in frequency of longer allele is lost on adjustment for BMI. Neither at genotypic level nor at allelic level after adjustment for BMI there was any difference observed in the presence of longer base pair allele. Thus the probable difference in the mean BMI as seen in table 5.11 was probably due to the significant difference in BMI levels already existing among cases and controls than due to the presence of longer base pair risk allele.

Table 5.12: **Genotype distribution and allele frequencies of the D2S1384 polymorphism in cases and controls after adjusting for BMI**

Group (n)	Genotype Frequency			Allele frequency		
	<145/<145	<145/≥145	>145/≥145	2N	<145*	>145*
	N (%)	N (%)	N (%)		N (%)	N (%)
Adjusted for BMI ≤ 22.9 kg/m^2						
Cases (69)	15 (0.22)	19 (0.28)	35 (0.50)	138	49 (0.36)	89 (0.65)
Controls (143)	32 (0.22)	47 (0.33)	64 (0.45)	286	111 (0.38)	175 (0.61)
Long allele (≥145bp) as risk allele: NS						
Adjusted for BMI >23 kg/m^2						
Cases (121)	6 (0.04)	50 (0.41)	65 (0.54)	242	62 (0.25)	180 (0.74)
Controls (54)	8 (0.15)	14 (0.26)	32 (0.59)	108	30 (0.28)	78 (0.72)
Long allele (≥145bp) as risk allele: NS						

BMI : Body Mass Index; NS: Non Significant; N:Number; %: Frequency

Table 5.13 summarizes the distribution of D2S1384 genotype and allele after an adjustment for WHR (WHR<0.81 and WHR≥0.81) among female cases and controls. The difference in frequency of longer allele was non-significant in non-obese cases. Overweight/obese individual showed significant difference in the frequency of longer base pair allele (risk allele) among cases and controls (χ^2 = 7.543, p =0.0060,

OR=1.6330, $CI_{95\%}$ 0.96 to 2.77) but the presence of at least one single longer allele significantly increase the odds for T2DM ~ 2 folds ($\chi^2 = 4.003$, $p = 0.03$, OR=3.1869 , $CI_{95\%}$ 1.1082 to 9.1650) in females.

Table 5.13: Genotype distribution and allele frequencies of the D2S1384 polymorphism in cases and controls after adjusting for WHR in females

Group (n)	Genotype Frequency			Allele frequency		
	<145/<145	<145/$\geq$145	>145/$\geq$145	2N	<145*	>145*
	N (%)	N (%)	N (%)		N (%)	N (%)
	Adjusted for WHR<0.81					
Cases (2)	0 (0)	1 (0.5)	1 (0.5)	4	1 (0.25)	2 (0.75)
Controls (13)	1 (0.08)	1(0.08)	11 (0.84)	26	3 (0.12)	23 (0.88)
Long allele ($\geq$145bp) as risk allele: NS						
Adjusted for WHR$\geq$0.81						
Cases (59)	5 (0.08)	21 (0.36)	33 (0.56)	118	31 (0.26)	87 (0.74)
Controls (79)	18 (0.22)	32 (0.40)	29 (0.36)	158	68 (0.43)	90 (0.57)
Long allele ($\geq$145bp) as risk allele: $\chi^2 = 7.543$, $p = 0.004$, OR=2.12, $CI_{95\%=}$ 1.26 to 3.55; Presence of atleast one $\geq$145 bp allele: $\chi^2 = 4.003$, $p = 0.03$, OR=3.1869 , $CI_{95\%=}$ 1.1082 to 9.1650						

WHR: Waist Hip Ratio; NS: Non Significant; N: Number; %: Frequency

Table 5.14 summarizes the distribution of D2S1384 genotype and allele after an adjustment for WHR (WHR<0.81 and WHR$\geq$0.81) among male cases and controls. The difference in frequency of longer allele remained significant in non-obese cases ($\chi^2 = 7.61$, $p = 0.005$, OR= 7.8, $CI_{95\%}$ 1.26 to 3.55). Overweight/obese individual also showed significant difference in the frequency of longer base pair allele (risk allele) among cases and controls ($\chi^2 = 14.50$, $p = 0.0001$, OR= 2.8, $CI_{95\%}$ 1.66 to 4.72) but the presence of atleast one single longer allele significantly increase the odds for T2DM ~three folds (OR=2.8 vs OR 8.33: $\chi^2 = 18.56$, $p < 0.0001$, OR= 8.33, $CI_{95\%}$ 3.0386 to 22.8844) in males subjects.

Chapter 5

Table 5.14: Genotype distribution and allele frequencies of the D2S1384 polymorphism in cases and controls after adjusting for WHR in males

Group (N)	Genotype Frequency			Allele frequency		
	<145/<145	<145/≥145	>145/≥145	2N	<145*	>145*
Adjusted for WHR<0.89						
	N (%)	N (%)	N (%)		N (%)	N (%)
Cases (7)	0 (0)	1 (0.14)	6 (0.86)	14	1 (0.07)	13 (0.93)
Controls (8)	2 (0.25)	6 (0.75)	0 (0.0)	16	10 (0.63)	6 (0.38)
Long allele (≥145bp) as risk allele: $\chi^2 = 7.61$, $p = 0.005$, OR= 7.8, $CI_{95\%}$ 1.26 to 3.55						
Adjusted for WHR≥0.89						
Cases (85)	6 (0.07)	34 (0.40)	45 (0.53)	170	46 (0.27)	124 (0.73)
Controls (49)	19 (0.38)	12 (0.24)	18 (0.37)	98	50 (0.51)	48 (0.49)
Long allele (≥145bp) as risk allele: $\chi^2 = 14.50$, $p = 0.0001$, OR= 2.8, $CI_{95\%}$ 1.66 to 4.72; Presence of at least single longer base pair allele: $\chi^2 = 18.56$, $p < 0.0001$, OR= 8.33, $CI_{95\%}$ 3.0386 to 22.8844						

WHR: Waist Hip Ratio; N:Number; %:Frequency

5.2.1.4 Logistic regression to find other factors influencing the relation between longer allele of D2S1384 marker and T2DM in obese individuals.

In present study logistic regression model was used to assess the risk of T2DM in obese individuals in presence of at least one copy of longer allele (x/≥145 or ≥145/≥145 genotype) of D2S1384 marker after controlling for different variables. Variables such as age, sex, TC, TG, HDL, LDL, SBP and DBP were added to the models. It was observed that presence of risk allele was nearly significantly associated with risk for T2DM in model 2 after adjusting for age and the significance for each variable improves after added adjustment for gender (model 3)

Table 5.15 shows the odds ratios, confidence intervals and level of significance (p values) explaining the effect of different variables on risk allele (X/<145 or <145/<145) in three regression models studied. There was not much change in the Odds ratio on

adjustment for age or age and sex together, although in case of each of the variable the significance was drastically improved after age adjustment, which was further strengthened after age and sex adjustment together.

Table 5.15: Multivariate logistic regression analysis of factors associated with longer allele of D2S1384 among T2DM patients

	Model 1	Model 2	Model 3
	p OR (CI$_{95\%}$)	p OR (CI$_{95\%}$)	p OR (CI$_{95\%}$)
	unadjusted	Adjusted for age	Adjusted for age & sex
SBP	$p = 0.0484$ OR= 1.01 CI$_{95\%}$ 0.99 to 1.038	$p = 0.0001$ OR= 1.01 CI$_{95\%}$ 0.99 to 1.03	$p = 0.0001$ OR= 1.02 CI$_{95\%}$ 1.00 to 1.04
DBP	$p = 0.0766$ OR= 1.02 CI$_{95\%}$ 0.99 to 1.06	$p = 0.0001$ OR= 1.02 CI$_{95\%}$ 0.99 to 1.05	$p = 0.0002$ OR= 1.03 CI$_{95\%}$ 0.99 to 1.06
FBG	$p = 0.004923$ OR= 1.01 CI$_{95\%}$ 1.00 to 1.01	$p < 0.0001$ OR= 1.01 CI$_{95\%}$ 1.00 to 1.01	$p < 0.0001$ OR= 1.01 CI$_{95\%}$ 1.00 to 1.02
TC	$p = 0.0711$ OR= 1.01 CI$_{95\%}$ 0.99 to 1.01	$p = 0.0010$ OR=1.00 CI$_{95\%}$ 0.99 to 1.01	$p = 0.0029$ OR= 1.00 CI$_{95\%}$ 0.99 to 1.01
TG	$p = 0.1729$ OR= 1.00 CI$_{95\%}$ 0.99 to 1.00	$p = 0.0015$ OR= 1.01 CI$_{95\%}$ 0.99 to 1.007	$p = 0.0039$ OR= 1.01 CI$_{95\%}$ 0.99 to 1.008
HDL	$p = 0.7589$ OR= 1.00 CI$_{95\%}$ 0.9841 to 1.0222	$p = 0.0018$ OR= 0.99 CI$_{95\%}$ 0.97 to 1.017	$p = 0.0047$ OR= 0.99 CI$_{95\%}$ 0.979 to 1.01
LDL	$p = 0.1622$ OR= 1.00 CI$_{95\%}$ 0.9975 to 1.0147	$p = 0.0010$ OR= 1.01 CI$_{95\%}$ 0.99 to 1.01	$p = 0.0028$ OR= 1.00 CI$_{95\%}$ 0.9958 to 1.0139

SBP: Systolic Blood Pressure; DBP: Diastolic Blood pressure ; FBG: Fasting Blood Glucose, HDL: High Density lipoprotein, LDL: Low density lipoprotein, VLDL: Very Low Density Lipoprotein; SD: Standard Deviation; Values are presented as means $\pm$ standard deviation, p values differed significantly among cases and controls (p<0.05)

Chapter 5

5.2.2 D2S2944 polymorphism & its association with T2DM

D2S2944 is an intragenic polymorphic STR marker located in Sperm Associated Antigen 16 gene, *SPAG16* (NT_005403.17) at 2q34. Population data (ALFRFED and TPMD database) showed 11 alleles ranging from 92-132bp PCR product size, corresponding to 5-15 (GATA)n tetra repeats in general (Gel Image in Appendices III: Fig A2).

5.2.2.1 Comparison with World Frequencies

In present study a total of eleven different allelic groups namely allele 1-11 with 4-15 "GATA" repeats were found in 458 subjects with unrelated healthy (227) and T2DM (231) subjects from India. The most frequent allele was allele 9 (30%) followed by allele 8 (25%) and allele 7 (23%), in controls. The system was in HWE and the Exact test (Guo and Thompson 1992) gave a p value of >0.05. The heterozygosity coefficient and PIC, amounting to 0.81 and 0.78 respectively, suggest considerable variability in the marker. A comparison of observed allele frequencies is depicted in fig 5.10 along with the world frequencies
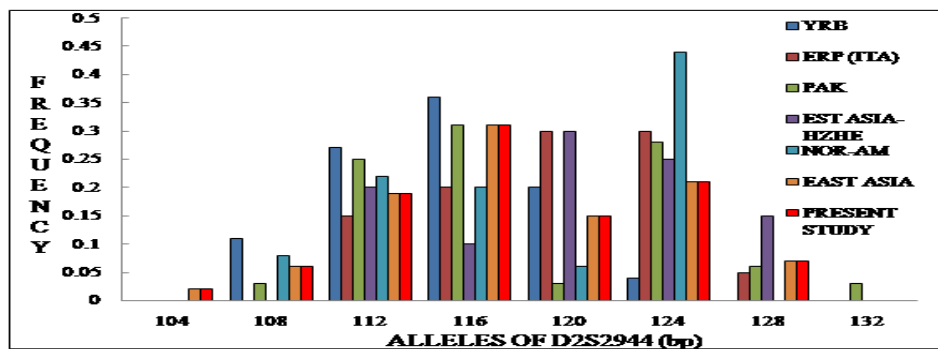
Fig 5.10 Allele frequencies of D2S2944 across different population of the world

YRB: Yourba Africa; ERP (ITA): Italian; PAK: Pakistan; EST ASIA-HZHE: Hezhen from north east China; NOR-AM:Pima Mexico; EAST ASIA: Uigur from Turpan county China; Present study : North Indians (Controls only)

5.2.2.2 Allelic Distribution of D2S2944 marker among cases and controls

Table 5.16: Allelic Distribution of D2S2944 marker among cases and controls

PCR Pdt Size (bp)	(GATA)n repeats	Allele	Cases (N=231)	Controls (N=220)
			N (%)	N (%)
92	05	1	2 (0.00)	9 (0.00)
96	06	2	0 (0.00)	3 (0.00)
100	07	3	6 (0.01)	9 (0.00)
104	08	4	19 (0.04)	12 (0.02)
108	09	5	19 (0.03)	7 (0.0.3)
112	10	6	61 (0.13)	38 (0.09)
116	11	7	84 (0.18)	55 (0.23)
120	12	8	64 (0.14)	120 (0.25)[1]
124	13	9	185 (0.40)	84 (0.30)[2]
128	14	10	19 (0.04)	89 (0.07)
132	15	11	02 (0.00)	14 (0.00)

[1] =95% CI= 0.3061 to 0.6008, χ^2= , p<0.0001, OR=0.42
[2] =95% CI= 2.09-3.82, χ^2= 4.600, p<0.0001, OR=2.8

PCR Pdt: PCR product; bp: Base pair; N: Number; %: Frequency

Fig 5.11: Distribution of major allele frequencies of D2S2944 STR in cases vs controls. Dotted lines represent the cut off made according to natural tendency for bimodal distribution of alleles. Red color represents cases; Blue represents controls.

Chapter 5

5.2.2.3 Distribution of D2S2944 Alleles

The cohort was further analyzed using a case-control design on the genotype data of D2S2944 from 231 diabetics and 227 controls. It is evident from table 5.16 that possession of 120bp allele (95% CI= 0.3061 to 0.6008, χ^2= , p<0.0001, OR=0.4288) of D2S2944 marker protects an individual from T2DM while presence of the longer 124bp allele seems to make individuals more susceptible to T2DM (95% CI= 2.09-3.82, χ2= 4.600, p<0.0001, OR=2.8). Other alleles did not exhibit any significant difference between cases and controls groups.

The cohort was further divided into approximately equal groups, at 124 bp or $(CTAT)_9$ repeats, according to their natural tendency for bimodal distribution. Hence the resulting genotypes were <124 bp/<124 bp, <124 bp//$\geq$124 bp and >124 bp/$\geq$124 bp. The shorter alleles (< $(CTAT)_{13}$ repeats) of D2S2944, were found to be more prevalent in the control population (62.5 vs 55.3). The <124bp allele constitutes 56% in adult cases and 63% in controls (OR = 1.34, χ^2=4.62, 95 % CI= 1.0347 to 1.7497, p = 0.0268)

> **Table 5.17: Allelic and genotypic frequencies of D2S2944 (Bimodal distribution) Microsatellite Repeat in cases and controls**

Group (N)	Genotype Frequency				Allele frequency	
	<124	<124/$\geq$124	$\geq$124	2N	<124*	$\geq$124*
	N (%)	N (%)	N (%)		N (%)	N (%)
Cases (232)	100 (43)	57 (24)	75(32)	464	257 (55.3)	207 (44.6)
Controls (231)	119 (52)	51 (22)	61 (27)	462	289 (62.5)	173 (37.4)

Difference among genotypes: χ^2=4.622 , p=0.03; *Longer allele ($\geq$124bp) as risk allele: OR = 1.34, 95 % CI= 1.0347 to 1.7497, p = 0.0268; (<124 vs $\geq$124bp): OR = 0.6835, 95 % CI = 0.4446 to 1.0507, p=0.082
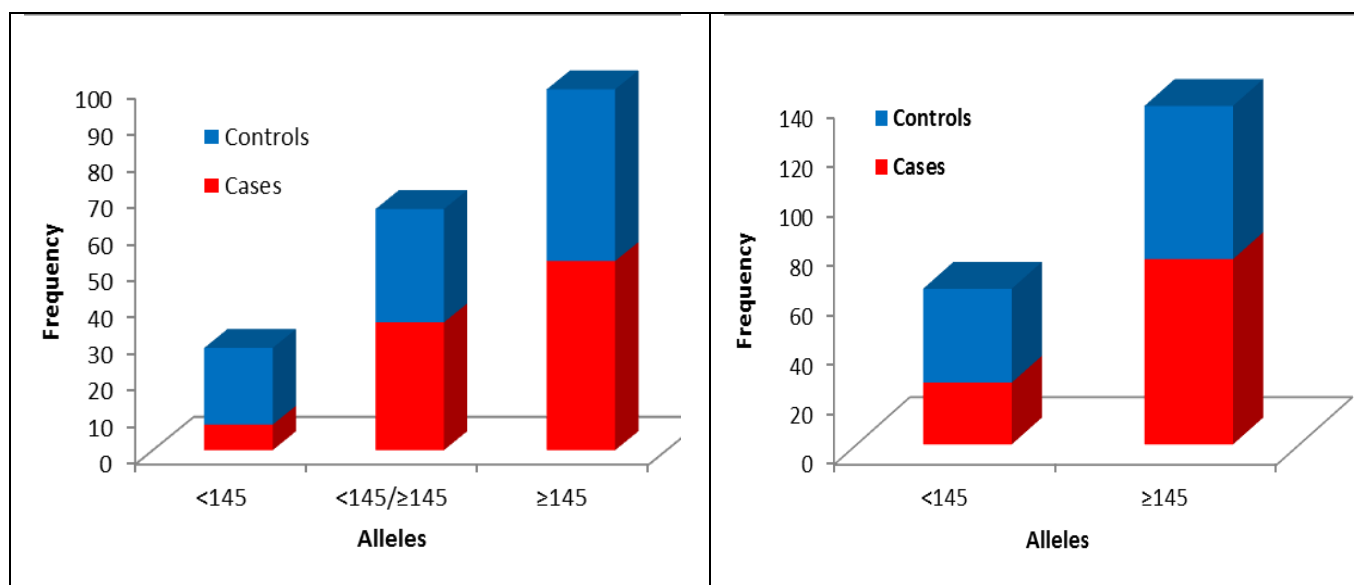
N: Number; %:Frequency
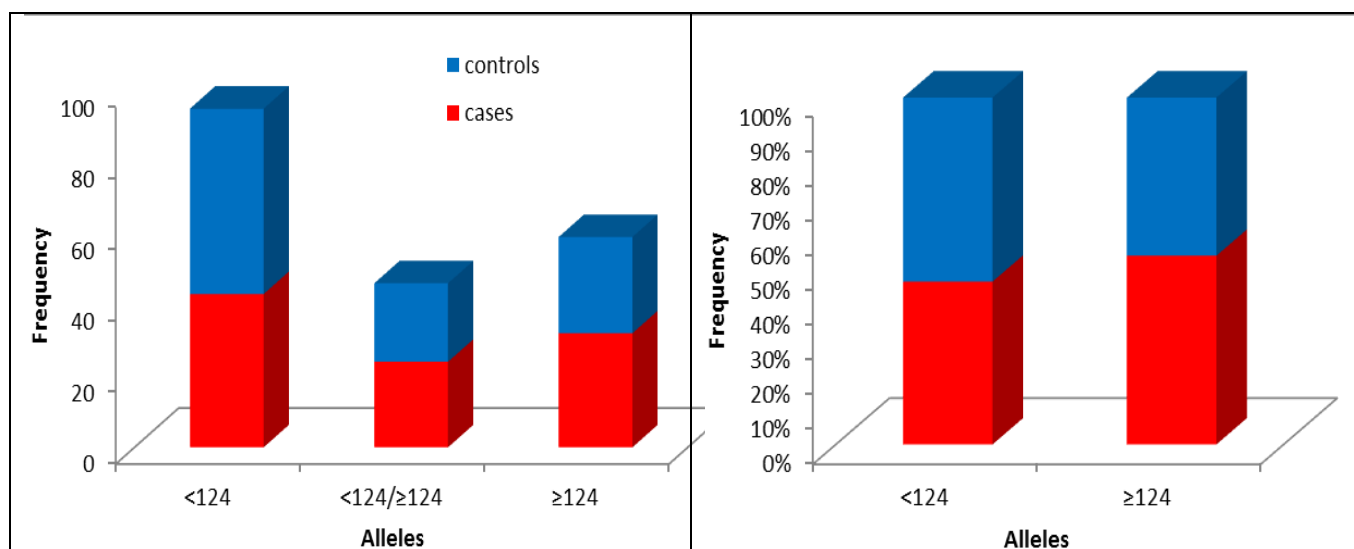
Fig5.12A : Genotypic distribution of D2S2944 STR Repeat after Bimodal distribution. Red color represents cases; Blue represents controls.

Fig 5.12 B: Allelic distribution of D2S2944 STR Repeat after Bimodal distribution. Red color represents cases; Blue represents controls.

5.2.2.4 Distribution of different variables according to D2S2944 alleles

The distribution of anthropometric in adult cohort alleles of D2S2944 is presented in Table 5.18. There was significant difference observed in the mean SBP, DBP, Body BMI and WHR levels in cases and controls when segregated on the basis of long and short alleles. In presence of longer base pair allele the mean SBP (<124bp: t = 2.08, p = 0.0379; ≥124bp: t = 4.588, p =0.0001), DBP (<124bp: NS; ≥124bp: t = 3.132, p =0.0022) and WHR (<124bp: t = 3.25 p < 0.0013; ≥124bp: t = 5.302, p < 0.0001) differed significantly with a nearly two fold increase in t value among cases and controls. While the presence of shorter base pair allele significantly reduced the mean BMI (<124bp: t = 6.59, p < 0.0001; ≥124bp: t = 4.61, p < 0.0001) value among cases and controls. Probably individuals with longer base pair allele tend to have higher BMI as compared to the individuals with shorter base pair allele. Also the individuals with longer base pair allele tend to have higher WHR, SBP and DBP.

Chapter 5

Table 5.18: Distribution of different anthropometric variables according to D2S2944 genotype among cases and controls

		<124		<124/ ≥124		≥124	
		N	Mean±SD	N	Mean±SD	N	Mean±SD
SBP	Controls 119		129.235±24.5	Controls 51	133.2±20.2	Controls 61	121.18±17.1
	Cases 100		135.13±15.34	Cases 53	138.2±16.9	Cases 62	133.95±13.6
	t = 2.08, *p* = 0.0379			NS		t = 4.588, *p* =0.0001	
DBP	Controls 118		83.38±12.72	Controls 51	84.82±10.4	Controls 61	78.54±11.41
	Cases 100		85.15±8.94	Cases 53	84.24±9.5	Cases 62	84.40±9.26
	NS			NS		t = 3.132, *p* =0.0022	
BMI	Controls 118		21.51±5.29	Controls 51	21.60±5.20	Controls 61	21.33±5.83
	85		26.44±5.21	Cases 46	25.64±4.89	Cases 56	25.90±4.78
	t = 6.59, *p* < 0.0001			t = 3.93, *p* = 0.0002		t = 4.61, *p* < 0.0001	
WHR	Controls 98		0.911±0.08	Controls 43	0.91±0.085	Controls 51	0.87±0.095
	Cases 82		0.95±0.08	Cases 41	0.95±0.086	Cases 47	0.96±0.07
	t = 3.25 *p* < 0.0013			t = 2.144, *p* = 0.0350		t = 5.302 , *p* < 0.0001	

BMI: Body Mass Index, WHR: Waist Hip Ratio, SBP: Systolic Blood pressure, DBP: Diastolic Blood Pressure, SD: Standard Deviation, NS: Non Significant; Values are presented as means ± standard deviation, *p* values differed significantly among cases and controls (*p*<0.05)

The distribution of biochemical parameters in adult cohort alleles of D2S2944 is presented in Table 5.19. No significant difference was observed in mean values of low density lipoproteins, LDL (<124bp: Non-Significant; ≥124bp: Non-Significant) among cases and controls with either long or short allele of D2S2944. Significant reduction was observed in mean Total cholesterol, TC(<124bp: t = 4.281, *p* <0.0001; ≥124bp: NS), Triglyceride level, TG(<124bp: t = 3.673, *p* = 0.0003 ; ≥124bp: t = 2.720, *p* =0.0077), HDL(<124bp: t=2.923, *p* = 0.0039; ≥124bp: NS) and VLDL(<124bp: t=3.603, *p*= 0.0004; ≥124bp: t =2.729, *p* = 0.0075) in the presence of longer base pair allele among cases and controls.

Chapter 5

Table 5.19: Distribution of different biochemical variables according to D2S2944 genotype among cases and controls

		<124		<124/ ≥124		≥124	
	N	Mean±SD	N	Mean±SD	N	Mean±SD
FBG	Controls 119	91.12± 13.81	Controls 51	89.64±16.9	Controls 61	92.213±14.4
	Cases 100	173.65 ±59.1	Cases 54	176.6±60.7	Cases 62	165.01±54.5
	t = 14.768, $p < 0.0001$		t = 9.874, $p < 0.0001$		t = 10.091, $p < 0.0001$	
CHL	Controls 76	172.1 ± 46.97	Controls 33	162.4±40.8	Controls 31	184.5± 46.2
	Cases 97	203.2±53.70	Cases 52	191.7±42.6	Cases 73	197.5± 53.5
	t = 4.281, $p <0.0001$		t = 3.141, $p =0.0023$		NS	
TG	Controls 76	122.9± 77.2	Controls 33	116.0 ±77.4	Controls 31	125.4± 53.1
	Cases 97	179.5 ±115.6	Cases 52	153.2± 73.3	Cases 73	170.8± 86.1
	t = 3.673, $p = 0.0003$		t = 2.231, $p =0.0284$		t = 2.720, $p =0.0077$	
HDL	Controls 76	37.80±25.56	Controls 33	37.3± 13.95	Controls 31	36.9± 22.5
	Cases 97	48.7± 23.4 4	Cases 52	44.9± 22.3	Cases 73	43.5± 19.9
	t=2.923, $p = 0.0039$		NS		NS	
LDL	Controls 76	109.7±46.87	Controls 33	101.8± 41.4	Controls 31	122.5± 44.7
	Cases 97	118.8± 52.35	Cases 52	113.4±46.4	Cases 73	121.3± 49.6
	NS		NS		NS	
VLDL	Controls 76	24.5± 15.4	Controls 33	23.2± 15.49	Controls 31	25.08± 10.6
	Cases 97	35.3± 22.28	Cases 52	30.1± 15.43	Cases 73	34.18± 17.2
	t = 3.603, $p= 0.0004$		t = 2.013, $p=0.0473$		t = 2.729, $p = 0.0075$	

FBG: Fasting Blood Glucose, HDL: High Density lipoprotein, LDL: Low density lipoprotein, VLDL: Very Low Density Lipoprotein, CHL: Total cholesterol; SD: Standard Deviation, NS: Non Significant; Values are presented as means ± standard deviation, p values differed significantly among cases and controls ($p<0.05$)

Table 5.20 and 5.21 summarizes the distribution of D2S2944 genotype and allele after an adjustment for BMI (BMI<22.9kg/m2 and >23kg/m2) among cases and controls. The difference in frequency of longer allele is lost on adjustment for BMI. Neither at

genotypic level nor at allelic level after adjustment for BMI there was any difference observed in the presence of longer base pair allele. Thus the probable difference in the mean BMI as seen in table 5.20 was probably due to the significant difference in BMI levels already existing among cases and controls than due to the presence of longer base pair risk allele.

Table 5.20: Genotype distribution and allele frequencies of the D2S2944 polymorphism in cases and controls after adjusting for BMI

Group (N)	Genotype Frequency			Allele frequency		
	<124/<124	<124/≥124	>124/≥124	2N	<124*	>124*
	N (%)	N (%)	N (%)		N (%)	N (%)
Adjusted for BMI ≤ 22.9 kg/m^2						
Cases (53)	24 (0.45)	14 (0.27)	15 (0.28)	106	82 (0.77)	44 (0.42)
Controls (159)	81 (0.50)	37 (0.23)	41 (0.26)	318	199 (0.62)	119 (0.38)
Long allele (≥124bp) as risk allele: NS						
Adjusted for BMI >23 kg/m^2						
Cases (134)	61 (0.45)	32 (0.24)	41 (0.31)	268	154 (0.57)	114 (0.43)
Controls (70)	37 (0.53)	13 (0.19)	20 (0.29)	140	87 (0.62)	53 (0.38)
Long allele (≥124bp) as risk allele: NS						

BMI: Body Mass Index; NS: Non significant ; N:Number; %:Frequency

Table 5.21 summarizes the distribution of D2S2944 genotype and allele after an adjustment for WHR (WHR<0.81 and WHR≥0.81) among female cases and controls. The difference in frequency of longer allele was non-significant in non-obese cases as well as non-obese subjects. The loss in the significance could probably mean that the difference in the mean WHR was probably due to the significant difference in WHR levels already existing among cases and controls than due to the presence of longer base pair risk allele.

Chapter 5

Table 5.21: Genotype distribution and allele frequencies of the D2S2944 polymorphism in cases and controls after adjusting for WHR in females

Group (N)	Genotype Frequency			Allele frequency		
	<124/<124	<124/≥124	>124/≥124	2N	<124*	>124*
	N (%)	N (%)	N (%)		N (%)	N (%)
Adjusted for WHR<0.81						
Cases (2)	0 (0)	1 (0.5)	1 (0.5)	4	1 (0.25)	2 (0.75)
Controls (18)	8 (0.44)	4 (0.22)	06 (0.33)	36	20 (0.56)	16 (0.44)
Long allele (≥124bp) as risk allele: NS						
Adjusted for WHR≥0.81						
Cases (76)	33 (0.43)	21 (0.28)	22 (0.29)	152	87 (0.57)	65 (0..43)
Controls (81)	38 (0.47)	21 (0.26)	22 (0.27)	162	97 (0.60)	65 (0.40)
Long allele (≥124bp) as risk allele: NS						

WHR: Waist Hip Ratio; NS: Non significant; N:Number; %:Frequency; Cutoffs for WHR were taken separately for males and females.

Table 5.22 summarizes the distribution of D2S2944 genotype and allele after an adjustment for WHR among male cases and controls. No significant difference seen.

Table 5.22: Genotype distribution and allele frequencies of the D2S2944 polymorphism in cases and controls after adjusting for WHR in males

Group (N)	Genotype Frequency			Allele frequency		
	<124/<124	<124/≥124	>124/≥124	2N	<124*	>124*
Adjusted for WHR<0.89						
	N (%)	N (%)	N (%)		N (%)	N (%)
Cases (7)	4 (0.5)	1 (0.14)	2 (0.28)	14	9 (0.64)	5 (0.35)
Controls (26)	10 (0.38)	5 (0.19)	11 (0.42)	52	25 (0.48)	27 (0.52)
Long allele (≥124bp) as risk allele: NS						
Adjusted for WHR≥0.89						
Cases (85)	45 (0.52)	18 (0.21)	22 (0.26)	170	108 (0.63)	62 (0.36)

Controls (67)	42 (0.63)	13 (0.19)	12 (0.18)	134	97 (0.72)	37 (0.28)
Long allele ($\geq$124bp) as risk allele: NS						

WHR: Waist Hip Ratio; NS: Non significant ; N:Number; %:Frequency

Table 5.23 summarizes the distribution of D2S2944 genotype and allele after an adjustment for Total cholesterol (TC<200 and TC$\geq$200) among cases and controls. Marginally significant difference was observed in the frequency of longer base pair allele among individuals with TC<200, (χ^2=7.638, OR = 1.7239, 95 % CI= 1.1848 to 2.5081, p = 0.0044) though this significance is lost in individuals with TC$\geq$200.

Table 5.23: Genotype distribution and allele frequencies of the D2S2944 polymorphism in cases and controls after adjusting for Total cholesterol

Group (N)	Genotype Frequency			Allele frequency		
	<124/<124	<124/$\geq$124	>124/$\geq$124	2N	<124*	>124*
	N (%)	N (%)	N (%)		N (%)	N (%)
Adjusted for TC<200						
Cases (142)	57 (0.40)	38 (0.27)	47 (0.33)	284	152 (0.54)	132 (0.46)
Controls (100)	55 (0.55)	23 (0.23)	22 (0.22)	200	133 (0.66)	67 (0.34)
Long allele ($\geq$124bp) as risk allele: χ^2=7.638, OR = 1.7239, 95 % CI= 1.1848 to 2.5081, p = 0.0044; Presence of at least single longer base pair allele:χ^2 = 4.631,p = 0.0314, OR= 0.5487 , CI$_{95\%}$ =0.3270 to 0.9206						
Adjusted for TC $\geq$200						
Cases (80)	40 (0.50)	14 (0.18)	26 (0.33)	160	94 (0.59)	66 (0.41)
Controls (40)	21 (0.53)	10 (0.25)	9 (0.23)	80	52 (0.65)	28 (0.35)
Long allele ($\geq$124bp) as risk allele: NS						

TC: Total Cholesterol; N:Number; %:Frequency

Chapter 5

5.2.3 D2S439 Polymorphism & its association with T2DM

The association of variation in tetra repeat microsatellite marker – D2S439 lying at 231.27cM at 2q37 nearly 0.13Mb away from SPHKAP (sphingosine kinase type 1 interacting protein AKAP domain containing) gene was analyzed in the present study.

5.2.3.1 Comparison with World Frequencies

The allele frequencies for D2S439 STR marker differ from previously known CEPH population (Utah Residents with Northern and Western European Ancestry) (CEU). Fig 5.13 (the CEPH Genotype database: http:// www.cephb.fr/cephdb and the CHLC Genetic Mapping data- base: http://lpgws.nci.nih.gov/html-chlc/ChlcMarkers.html). The studied Indian population, D2S439 microsatellite has 11 types of alleles with 05-15 (CTAT) n tetra repeat nucleotides which yield PCR product of length ranging from 163-203 base pair (Gel Image in Appendices III: Fig A2). Population data showed that this is a highly polymorphic STR with a heterozygosity and polymorphic information content (PIC Value) of more than 0.80.
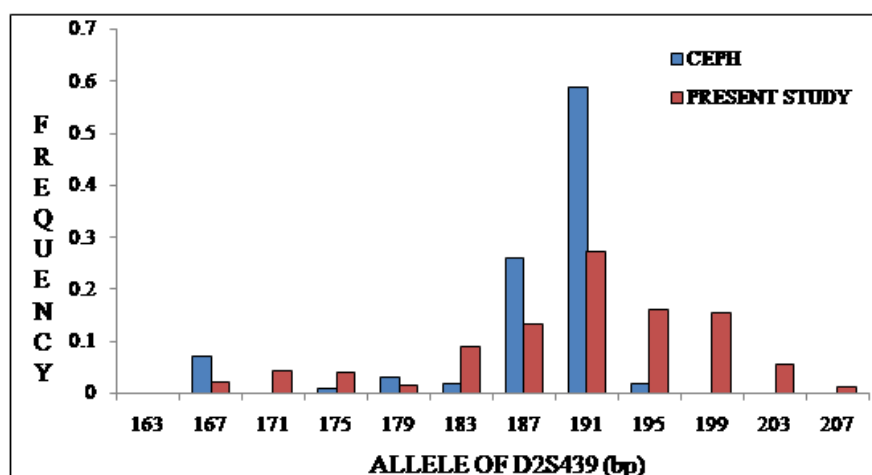
Fig 5.13: Distribution of major allele frequencies of D2S439 STR in CEPH vs Indian population (controls). The blue bars represents CEPH population frequency and the red ones represent the frequencies reported in the present study

Chapter 5

5.2.3.2 Allelic Distribution of D2S439 marker among cases and controls

The frequency distribution among cases and controls was analyzed. The most frequency allele in our population was found to be 191bp allele.

Table 5.24: Allelic Distribution of D2S439 marker among cases and controls

PCR Pdt Size (bp)	(CTAT)n Repeats	Allele	Cases (N=233)	Controls (N=165)
			N (%)	N (%)
163	05	1	6 (0.01)	10(0.03)
167	06	2	3 (0.01)	6 (0.02)
171	07	3	12 (0.03)	14 (0.04)
175	08	4	5 (0.01)	11 (0.03)
179	09	5	19 (0.04)	9 (0.03)
183	10	6	23 (0.05)	39 (0.12)[1]
187	11	7	67 (0.14)	55 (0.17)
191	12	8	105 (0.23)	86 (0.26)
195	13	9	70 (0.15)	46 (0.14)
199	14	10	73 (0.16)	35 (0.11)
203	15	11	83 (0.28)	18 (0.05)[2]

[1]=95% CI= 0.24-0.72, χ^2= 8.3, *p*= 0.003, OR=0.41
[2]=95% CI= 2.2-6.38, χ^2= 25.5, *p< 0.0001,* OR= 3.7

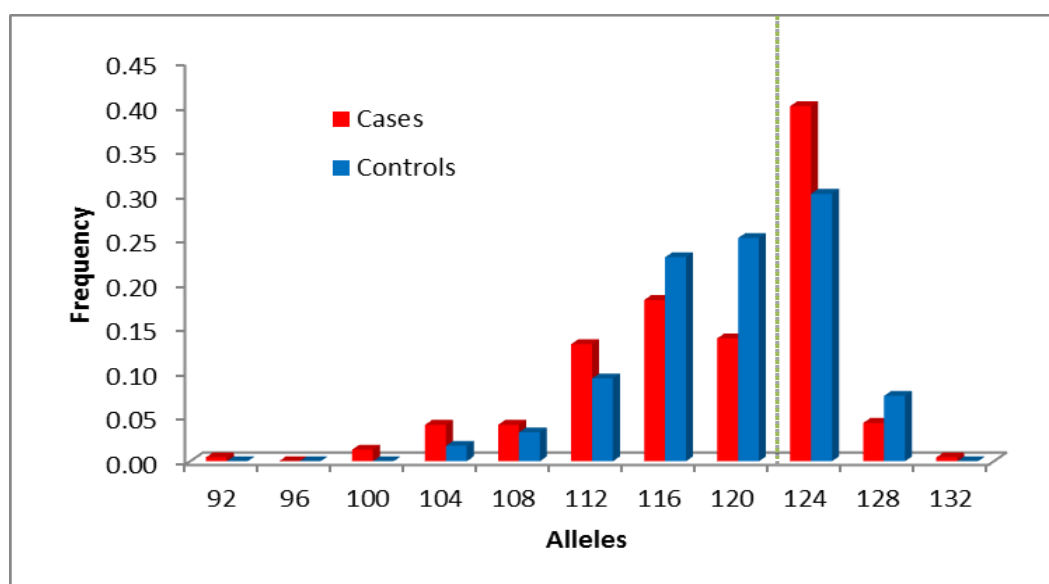PCR Pdt: PCR product; bp: Base pair; N: Number; %: Frequency
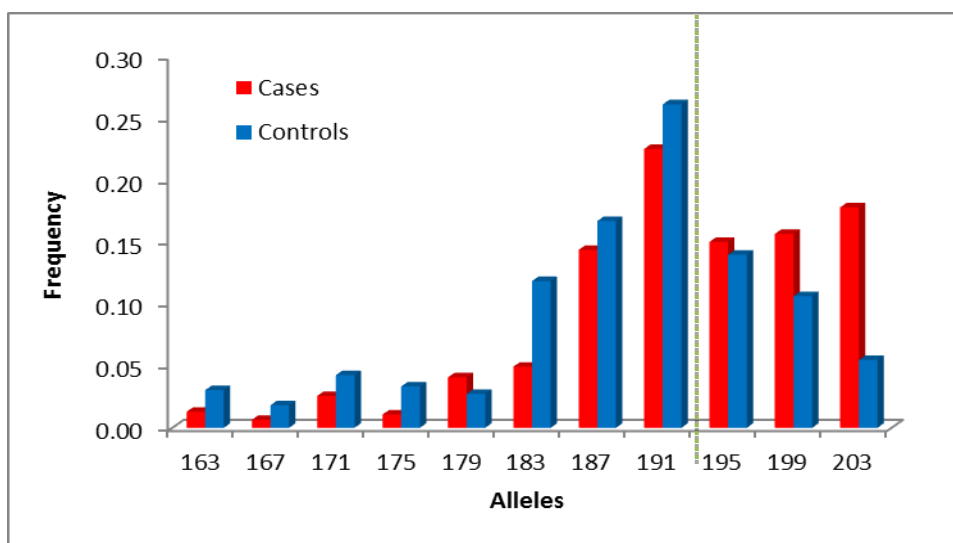
Fig 5.14: Distribution of major allele frequencies of D2S439 STR in cases vs controls. Dotted lines represent the cut off made according to natural tendency for bimodal distribution of alleles. Red color represents cases; Blue represents controls.

Chapter 5

5.2.3.3 Bimodal Distribution of D2S439 Alleles

The cohort was further analyzed using a case-control design on the genotype data of D2S439 from 233 diabetics and 171 controls. It is evident from table 5.24 that possession of 183bp allele (95% CI= 0.24-0.72, $\chi2$= 8.3, p= 0.003, OR=0.41) of D2S439 marker seems to have a protective effect against T2DM while presence of the longer 203 bp allele seems to increase risk for T2DM (95% CI= 2.2-6.38, $\chi2$= 25.5, p< 0.0001, OR= 3.7). Others alleles did not exhibit any significant difference between cases and controls groups. The cohort was further divided into approximately equal groups, at 195 bp or $(CTAT)_9$ repeats, according to their natural tendency for bimodal distribution. Hence the resulting genotypes were <195 bp/<195 bp, <195 bp//≥195 bp and >195 bp/≥195 bp. The shorter alleles (or <26 (CTAT) repeats) of D2S439, were found to be more prevalent in the control population (68 vs 52). The <195bp allele constitutes 52% in adult cases and 68% in controls (χ^2=20.223, OR = 1.95, 95 % CI= 1.4660 to 2.6196, $p < 0.0001$)

Table 5.25: Allelic and genotypic frequencies of D2S439 (Bimodal distribution) Microsatellite Repeat in cases and controls

| Group (N)	Genotype Frequency			2N	Allele frequency	
	<195	<195/≥195	≥195		<195*	≥195*
	N (%)	N (%)	N (%)		N (%)	N (%)
Cases (233)	112 (48)	16 (07)	105 (45)	464	240 (52)	226 (48)
Controls (171)	103 (60)	25 (15)	43 (25)	342	231 (68)	111 (32)
Difference among genotypes: χ^2=20.223 , $p < 0.0001$; *Longer allele (≥195bp) as risk allele: OR = 1.95, 95 % CI= 1.4660 to 2.6196, $p < 0.0001$; (<195 vs ≥195bp): OR = 1.6364 , 95 % CI 1.0971 to 2.440, $p = 0.015$, $\chi^{2=}$ 5.3						

N: Number; %:Frequency
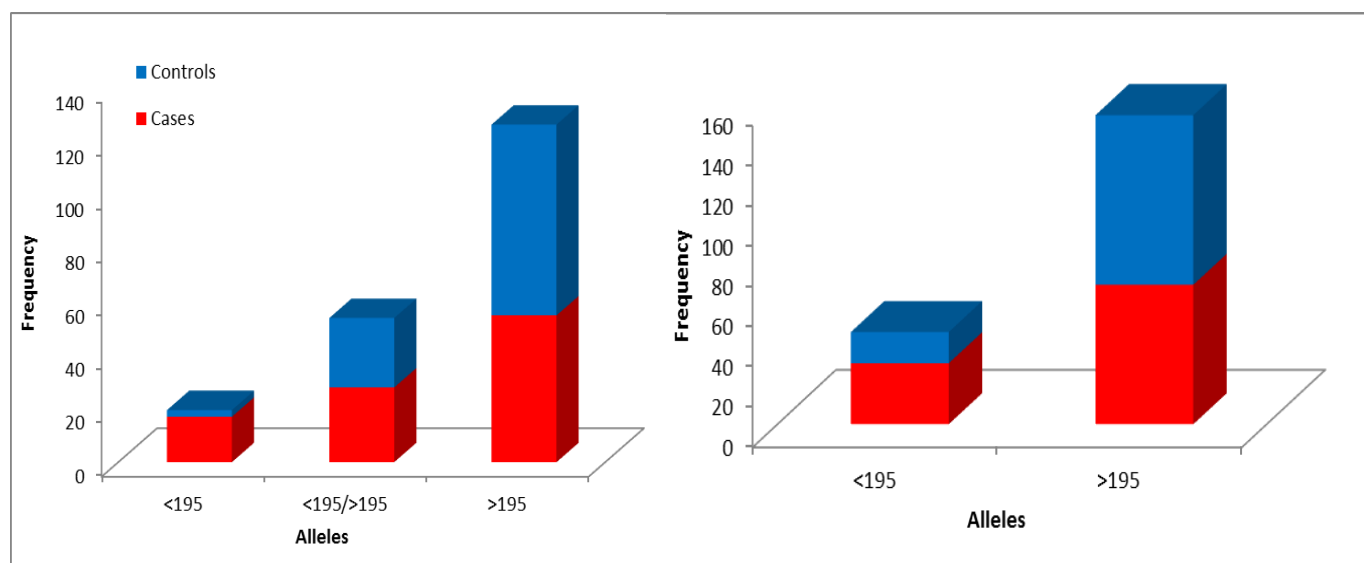
Fig 5.15 A : Genotypic distribution of D2S439 STR Repeat after Bimodal distribution. Red color represents cases; Blue represents controls.

Fig 5.15 B: Allelic distribution of D2S439 STR Repeat after Bimodal distribution. Red color represents cases; Blue represents controls.

5.2.3.4 Distribution of different variables according to D2S439 alleles

The distribution of anthropometric in adult cohort alleles of D2S439 is presented in Table 5.26. There was significant difference observed in the mean SBP, DBP, BMI and WHR levels in cases and controls when segregated on the basis of long and short alleles of D2S439. In presence of longer base pair allele the mean SBP ($<$195bp: t = 4.110, p = 0.0001; $\geq$195bp: t = 1.982, p =0.04), DBP ($<$195bp: 3.279, p = 0.0012; $\geq$195bp: NS) BMI ($<$195bp: t = 6.362, p < 0.0001; $\geq$195bp: t = 4.65, p < 0.0001) and WHR ($<$195bp: t = 5.208, p < 0.0001; $\geq$195bp: t = 6.003, p < 0.0001) differed significantly among cases and controls. In the presence of the longer base pair allele the mean SBP, DBP and BMI were significantly lower as compared to the shorter allele. Though, the mean WHR seems to be higher in case of longer base pair allele of D2S439 marker when compared among cases and controls.

Chapter 5

Table 5.26: Distribution of different anthropometric variables according to D2S439 genotype among cases and controls

		<195		<195/ ≥195		≥195	
	N	Mean±SD	N	Mean±SD	N	Mean±SD	
SBP	Controls 103	126.30±19.62	Controls 25	133.5±28.4	Controls 43	127.6 ± 25.7	
	Cases 100	136.3± 14.6	Cases 16	132.3±13.3	Cases 89	134.9± 16.3	
	t = 4.110, *p* = 0.0001		NS		t = 1.982, *p* =0.04		
DBP	Controls 103	80.9 ±10.87	Controls 25	86.9 ±11.4	Controls 42	81.3 ±13.72	
	Cases 111	85.2± 8.30	Cases 16	84.43±10.4	Cases 89	84.0± 9.94	
	t =3.279, *p* = 0.0012		NS		NS		
BMI	Controls 103	21.1±5.70	Controls 25	21.67 ±4.51	Controls 42	21.2± 5.82	
	Cases 92	26.1± 5.22	Cases 16	26.0± 3.70	Cases 80	25.9± 5.00	
	t = 6.362, *p* < 0.0001		t = 3.212, *p* = 0.002		t = 4.65, *p* < 0.0001		
WHR	Controls 76	0.88± 0.09	Controls 20	0.89± 0.07	Controls 36	0.86± 0.08	
	Cases 84	0.95± 0.08	Cases 14	0.96 ±0.06	Cases 72	0.95± 0.07	
	t = 5.208, *p* < 0.0001		t = 3.038, *p* = 0.0047		t = 6.003, *p* < 0.0001		

SBP: Systolic Blood Pressure, DBP: Diastolic Blood pressure: BMI: Body Mass Index, WHR: Waist to hip ratio, NS: Non Significant; Values are presented as means ± standard deviation, *p* values differed significantly among cases and controls (*p*<0.05)

The distribution of biochemical parameters in adult cohort alleles of D2S439 is presented in Table 5.27. Significant differences were observed in mean levels of Fasting blood glucose FBG (<195bp: t = 14.768, *p* < 0.0001; ≥195bp: t = 10.091, *p* < 0.0001), HDL (<195bp: t=2.717, *p* = 0.0073; ≥195bp: t = 3.206, *p* = 0.0018) among cases and controls in presence of longer allele. Significant reduction was observed in mean Total cholesterol, TC(<195bp: 3.004, *p*=0.0031; ≥195bp: NS), Triglyceride level, TG(<195bp: t = 2.667, *p* = 0.008; ≥195bp: t = 2.165, *p* =0.03), HDL(<195bp: t=2.923, *p* = 0.0039; ≥195bp: NS) and VLDL(<195bp: t = 2.555, *p*= 0.0116 ; ≥195bp t = 2.191, *p* = 0.03) in the presence of longer base pair allele among cases and controls. No

significant association of the marker was seen with the mean LDL levels (<195bp: NS; ≥195bp: NS) both among cases and controls and also after stratification for long and short form of D2S439

Table 5.27: Distribution of different biochemical variables according to D2S439 genotype among cases and controls

		<195		<195/ ≥195		≥195	
	N	Mean±SD	N	Mean±SD	N	Mean±SD	
FBG	Controls 103	91.6±13.8	Controls 25	91.7± 16.6	Controls 43	93.11± 14.9	
	Cases 112	167.08± 55.2	Cases 16	173.8± 57.9	Cases 89	177.03± 61.7	
	t = 13.493, p < 0.0001		t = 6.714, p < 0.0001		t = 8.780, p < 0.0001		
CHL	Controls 49	176.1± 49.8	Controls 14	167.1± 41.0	Controls 13	168.8 ± 47.8	
	Cases 108	203.9± 55.4	Cases 14	239.8± 67.3	Cases 101	186.8± 39.2	
	t = 3.004, p=0.0031		t = 3.452, p =0.0019		NS		
TG	Controls 49	123.1± 85.2	Controls 14	103.2± 45.4	Controls 13	112.8± 57.1	
	Cases 108	165.3± 94.7	Cases 14	173.4± 99.3	Cases 101	175.01± 101.3	
	t = 2.667, p = 0.008		t = 2.406, p =0.02		t = 2.165, p =0.03		
HDL	Controls 49	36.3± 22.2	Controls 14	39.3± 24.1	Controls 13	25.9± 10.6	
	Cases 108	47.08± 23.4	Cases 14	50.7± 22.9	Cases 101	44.4 ± 20.4	
	t=2.717, p = 0.0073		NS		t = 3.206, p = 0.0018		
LDL	Controls 49	115.1± 49.8	Controls 14	107.1± 39.9	Controls 13	120.3 ±42.4	
	Cases 108	122.09± 55.3	Cases 14	156.3± 54.8	Cases 101	108.9± 39.4	
	NS		t= 2.716, p = 0.0116		NS		
VLDL	Controls 49	24.6±17.04	Controls 14	20.6± 9.08	Controls 13	22.5± 11.4	
	Cases 108	32.8± 19.3	Cases 14	34.4 ± 20.4	Cases 101	34.4± 19.1	
	t = 2.555, p= 0.0116		t = 2.312, p=0.028		t = 2.191, p = 0.03		

FBG: Fasting Blood Glucose, HDL: High Density lipoprotein, LDL: Low density lipoprotein, VLDL: Very Low Density Lipoprotein, CHL: Cholesterol, TG: Triglycerides; Values are presented as means ± standard deviation, p values differed significantly among cases and controls (p<0.05)

Chapter 5

5.3   Chromosome 3 and T2DM

The fig bellow lists the region studied on chromosome 3 and the two markers D3S3609

and D3S2398 studied in the present study are encircled. The details of the region as per

UniSTS are given below.

Region: 3q22-29

Bp size: 130021000-198022000

Marker position: 130M-198M
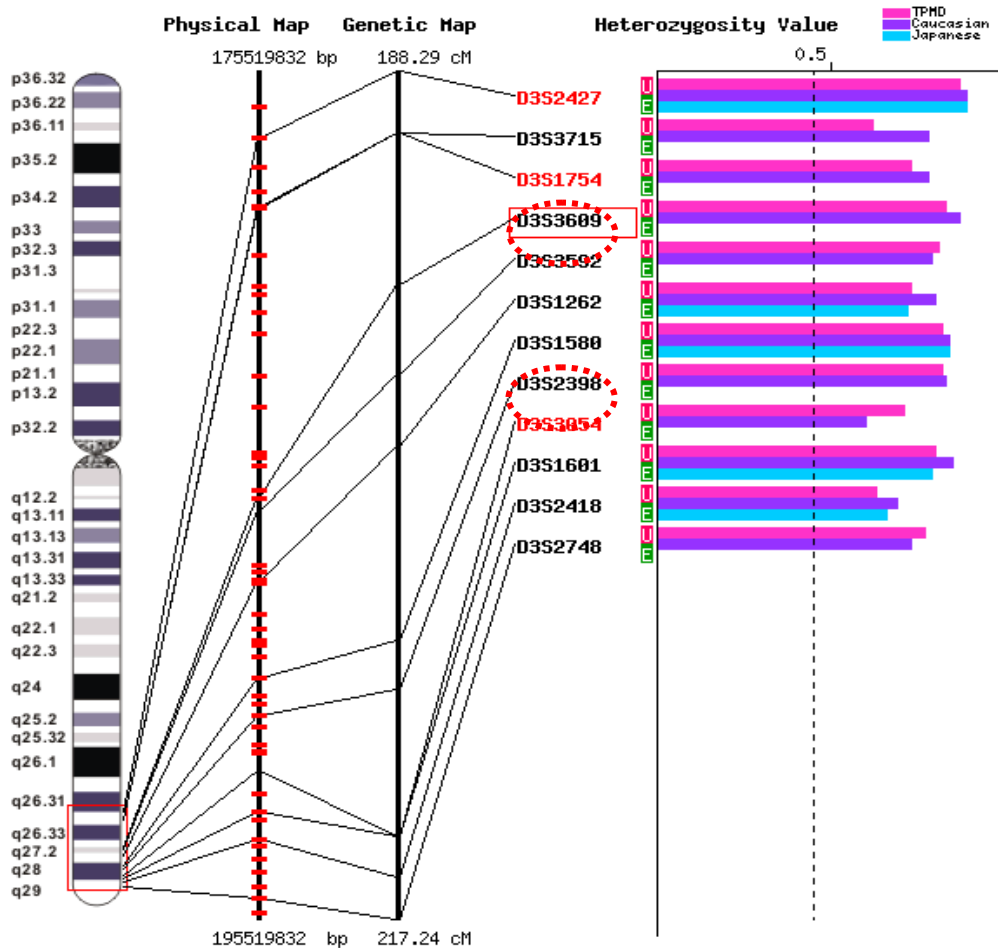
Fig 5.16: Position and Hetrozygosity score of STR markers analyzed on chromosome 3. The red square box on the ideogram represents the region chosen for the study. The encircled regions represent the markers studied in the current study

Chapter 5

5.3.1 D3S3609 polymorphism & its association with T2DM

The association of variation in di nucleotide repeat microsatellite marker D3S3609 lying at gene eukaryotic translation initiation factor 4 gamma, at 195.60 cM at 3q27-29 was analyzed in the present study.

5.3.1.1 Comparison with World Frequencies

The allele frequencies for D3S3609 STR marker differ from previously known CEPH population (Utah Residents with Northern and Western European Ancestry)(CEU): Fig 5.17 (the CEPH Genotype database: http:// www.cephb.fr/cephdb and the CHLC Genetic Mapping data- base: http://lpgws.nci.nih.gov/html-chlc/ChlcMarkers.html). The studied Indian population, D3S3609 microsatellite has 13 types of alleles with 18-30 $(CA)_n$ di nucleotide repeat nucleotides: which yield DNA fragment of length ranging from 163-187 base pair. Population data showed that this is a highly polymorphic STR with a heterozygosity and polymorphic information content (PIC Value) of more than 0.80
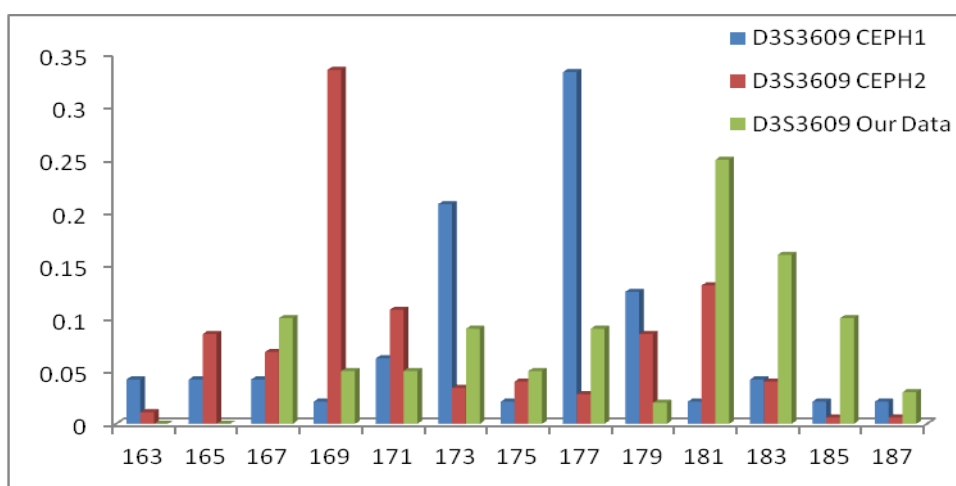
Fig 5.17: Distribution of major allele frequencies of D3S3609 STR in CEPH vs Indian population.

5.3.1.2 Allelic Distribution of D3S3609 marker among cases and controls

Table 5.28: Allelic Distribution of D3S3609 marker among cases and controls

PCR Pdt Size (bp)	(CA)n repeats	Allele	Cases (N=205)	Controls (N=201)
			N (%)	N (%)
163	18	1	51 (0.13)	02 (0.00)[1]
165	19	2	29 (0.07)	00 (0.00)[2]
167	20	3	17 (0.04)	40 (0.10)[3]
169	21	4	09 (0.02)	20 (0.05)
171	22	5	20 (0.05)	22 (0.05)
173	23	6	15 (0.04)	37 (0.09)
175	24	7	32 (0.08)	19 (0.05)
177	25	8	43 (0.11)	38 (0.09)
179	26	9	23 (0.06)	09 (0.02)
181	27	10	46 (0.11)	103 (0.25)[4]
183	28	11	69 (0.17)	67 (0.16)
185	29	12	39 (0.10)	40 (0.10)
187	30	13	09 (0.02)	12 (0.03)

[1]=95% CI= 7.2 to 122.03, χ^2= 47.2, *p<0.0001,* OR=29.4;[2] χ^2= 28.5, *p<0.0001;* [3]=95% CI 0.226 to 0.731 , χ^2= 8.73, *p*=0.003, OR=0.407 [4]=95% CI= 0.261 to 0.559, χ^2= 24.78, *p<0.0001*, OR=0.38

PCR Pdt: PCR product; bp:Base pair; N: Number
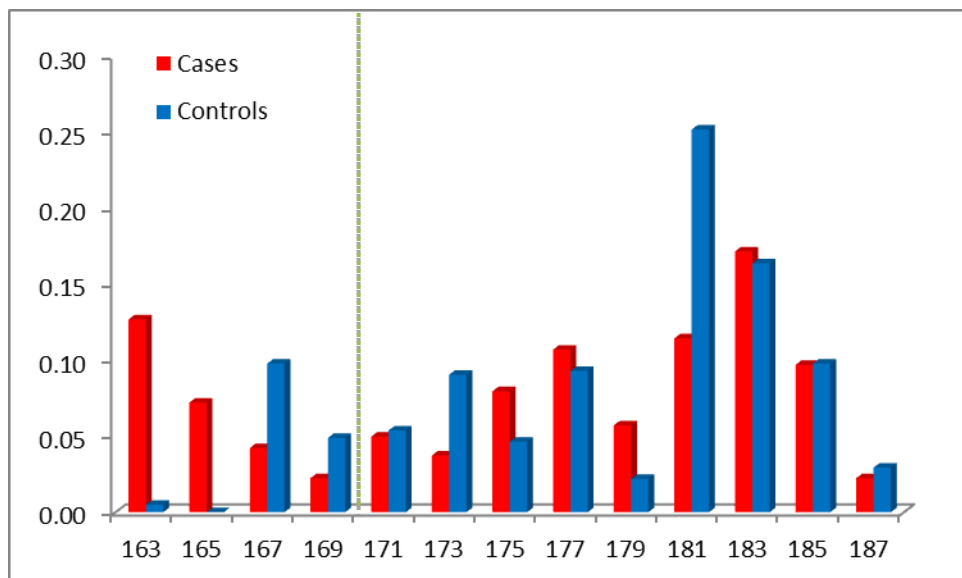
Fig 5.18: Distribution of major allele frequencies of D3S3609 STR in cases vs controls. Dotted lines represent the cut off made according to natural tendency for bimodal distribution of alleles. Red color represents cases; Blue represents controls.

Chapter 5

5.3.1.3 Bimodal Distribution of D3S3609 Alleles

The cohort was further analyzed using a case-control design on the genotype data of D3S3609 from 201 diabetics and 205 controls. It is evident from table 5.28 that possession of 167bp (95% CI= 0.226 to 0.731, $\chi 2$= 8.73, p=0.003, OR=0.407) of D3S3609 marker seems to have a protective effect against T2DM while presence of the shorter 163bp (95% CI= 7.2 to 122.03, $\chi 2$= 47.2, p<0.0001, OR=29.4) allele seems to make individuals more susceptible to T2DM. Others alleles did not exhibit any significant difference between cases and controls groups.

The cohort was further divided into two groups, at 171bp, according to their natural tendency for bimodal distribution at 171bp. Hence the resulting genotypes were <171 bp/<171 bp, <171 bp/≥171 bp and >171 bp/≥171 bp. The shorter alleles of D3S3609 were found to be more prevalent in the cases (31% vs 16%). The ≥71bp allele constitutes 70% in adult cases and 85% in controls (χ^2=24.777, OR = OR =0.4182, 95 % CI= 0.2968 to 0.5893, $p < 0.0001$). Infact the presence of a single long allele increases the protection from 0.42 times to 0.13 times.

Table 5.29: Allelic and genotypic frequencies of D3S3609 (Bimodal distribution) Microsatellite Repeat in cases and controls

Group (N)	Genotype Frequency			2N	Allele frequency	
	<171	<171/≥171	≥171		<171*	≥171*
	N (%)	N (%)	N (%)		N (%)	N (%)
Cases (205)	34 (17)	57 (28)	114 (55)	410	125 (30.4)	285 (69.5)
Controls (200)	05 (2.5)	52 (26)	143 (71.5)	400	62 (15.5)	338 (84.5)

Difference among genotypes: χ^2=24.777 , $p < 0.0001$; *Longer allele (≥171bp) as protective allele: OR =0.4182, 95 % CI= 0.2968 to 0.5893, $p < 0.0001$; (<171 vs ≥171bp): OR 0.1290, 95 % CI 0.0493 to 0.3371, $p < 0.0001$, χ^2= 21.489

N:Number; %:Frequency
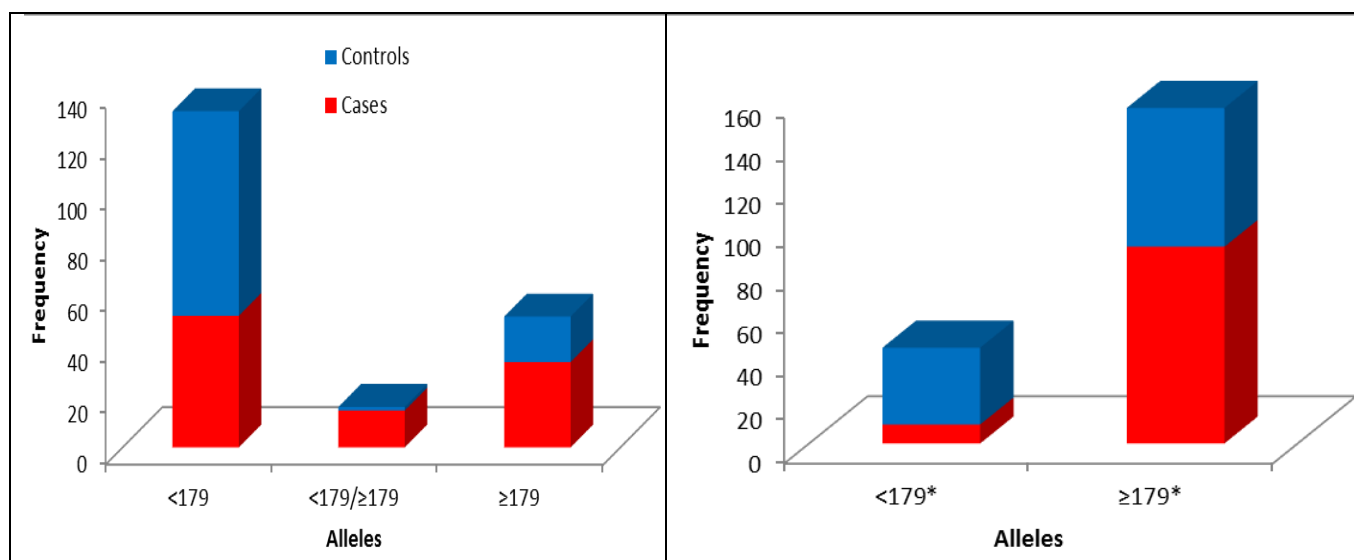
Fig5.19A : Genotypic distribution of D3S3609 STR Repeat after Bimodal distribution. Red color represents cases; Blue represents controls.

Fig 5.19 B: Allelic distribution of D3S3609 STR Repeat after Bimodal distribution. Red color represents cases; Blue represents controls.

5.3.1.4 Distribution of different variables according to D3S3609 alleles

The distribution of anthropometric in adult cohort alleles of D2S3609 is presented in Table 5.30. There was significant difference observed in the mean SBP (<171bp: NS; ≥171bp: t = 2.646, p =0.008), BMI (<171bp: NS; ≥171bp: t = 2.668, p =0.0082) and WHR (171bp: NS; ≥171bp: t = 5.525 p < 0.0001) levels in cases and controls when segregated on the basis of long and short alleles of D3S3609. As is evident from the table below there was no difference in the mean SBP and BMI among cases and controls in presence of shorter base pair allele, while in presence of even a single copy of longer base pair alleles, the mean SBP and BMI levels were significantly raised. However no significant differences in mean DBP (<171bp: NS; ≥171bp: NS) levels were observed when segregated for long and short allele. In both cases (presence or absence of longer bp allele) the mean levels remained unchanged among cases and controls.

Chapter 5

Table 5.30: Distribution of different anthropometric variables according to D3S3609 genotype among cases and controls

		<171		<171/ ≥171		≥171	
	N	Mean±SD	N	Mean±SD	N	Mean±SD	
SBP	Controls 5	126.0±15.1	Controls 50	130.3±16.4	Controls 141	130.9± 18.3	
	Cases 33	132.09± 14.9	Cases 50	137.1±15.7	Cases 96	137.1± 16.8	
	NS		t = 2.118, *p* =0.03		t = 2.646, *p* =0.008		
DBP	Controls 5	84.2 ±11.5	Controls 50	84.7 ±11.3	Controls 141	87.07±10.1	
	Cases 33	84.5± 9.0	Cases 50	84.9±10.09	Cases 96	84.5± 9.5	
	NS		NS		NS		
BMI	Controls 5	22.8±3.6	Controls 52	24.01±4.2	Controls 142	23.8± 4.1	
	Cases 30	25.6± 4.2	Cases 46	26.6± 5.9	Cases 85	25.3± 4.1	
	NS		t = 2.525, *p* = 0.013		t = 2.668, *p* =0.0082		
WHR	Controls 5	0.92± 0.06	Controls 46	0.91± 0.05	Controls 137	0.9± 0.05	
	Cases 28	0.96± 0.07	Cases 44	0.9±0.07	Cases 78	0.9± 0.07	
	NS		NS		t = 5.525 *p* < 0.0001		

SBP: Systolic Blood Pressure, DBP: Diastolic Blood pressure: BMI: Body Mass Index, WHR: Waist to hip ratio, NS: Non Significant; Values are presented as means ± standard deviation, *p* values differed significantly among cases and controls (*p*<0.05)

Further, the group was evaluated for the mean levels of various biochemical parameters according to the presence of long and short allele of D2S3609 marker (Table 5.31).

The cases and controls differed significantly in mean FBG levels (<171bp: t = 3.419, *p* 0.0016; ≥171bp: t = 16.630, *p* < 0.0001). It seems that the cases having longer base pair allele had significantly higher mean FBG level as compared to the cases with shorter base pair allele. Although there was no significant difference observed among cases and controls on stratification for long and short allele for mean levels of any of the lipid parameters except HDL wherein the longer bp allele carrying subjects had lower HDL

levels as compared to those bearing shorter allele (<171bp: t=2.061, $p = 0.04$; ≥171bp: t = t= 1.997, $p = 0.046$).

Table 5.31: Distribution of different biochemical variables according to D3S3609 genotype among cases and controls

		<171		<171/ ≥171		≥171	
		N	Mean±SD	N	Mean±SD	N	Mean±SD
FBG	Controls 5		85.4±9.6	Controls 52	90.2± 17.7	Controls 143	85.5± 14.3
	Cases 33		175.4± 58.08	Cases 52	175.09± 59.4	Cases 96	182.1± 67.3
		t = 3.419, p 0.0016		t = 9.876, $p < 0.0001$		t = 16.630, $p < 0.0001$	
CHL	Controls 5		155.9± 30.9	Controls 38	193.7± 52.6	Controls 130	179.6± 42.9
	Cases 34		188.3± 53.5	Cases 51	194.9± 54.9	Cases 110	188.3± 40.5
		NS		NS		NS	
TG	Controls 5		165.08± 83.3	Controls 38	177.2± 69.6	Controls 129	178.6± 84.4
	Cases 34		168.04± 114.3	Cases 51	172.4± 80.6	Cases 110	175.9± 101.8
		NS		NS		NS	
HDL	Controls 5		28.8± 11.9	Controls 38	39.8± 16.3	Controls 129	38.8± 16.7
	Cases 34		52.7± 25.3	Cases 51	43.2± 16.8	Cases 110	43.7± 21.2
		t=2.061, $p = 0.04$		NS		t= 1.997, $p = 0.046$	
LDL	Controls 5		94.3± 19.0	Controls 38	116.1± 50.07	Controls 129	106.1±40.2
	Cases 34		103.2± 55.8	Cases 51	117.03± 47.6	Cases 110	110.7± 42.4
		NS		NS		NS	
VLDL	Controls 5		32.9±16.7	Controls 38	35.3± 14.2	Controls 129	35.7± 16.8
	Cases 33		33.5± 23.5	Cases 51	34.3± 16.3	Cases 110	34.7± 19.4
		NS		NS		NS	

FBG: Fasting Blood Glucose, HDL: High Density lipoprotein, LDL: Low density lipoprotein, VLDL: Very Low Density Lipoprotein, TG: Triglycerides; Values are presented as means ± standard deviation, p values differed significantly among cases and controls ($p<0.05$)

Chapter 5

5.3.2 D3S2398 polymorphism in the studied case control cohort

The association of variation in tetra nucleotide repeat microsatellite marker D3S2398 lying at Abdominal obesity-metabolic syndrome QTL1, at 209.41 cM at 3q27-qter was analyzed in the present study.

5.3.2.1 Comparison with World Frequencies

The allele frequencies for D3S2398 STR marker observed in our control group differ from previously known CEPH population (Utah Residents with Northern and Western European Ancestry) (CEU): Fig 5.20 (the CEPH Genotype database: http://www.cephb.fr/cephdb and the CHLC Genetic Mapping data-base: http://lpgws.nci.nih.gov/html-chlc/ChlcMarkers.html). Our study group was closer to Asians and east Asian reported frequencies as expected. The studied Indian population, D3S2398 microsatellite has 10 types of alleles with 6-15 (GATA) $_n$ tetra nucleotide repeat nucleotides: which yield DNA fragment of length ranging from 262-298 base pair (Gel Image in Appendices V: Fig A2). Population data showed that this is a highly polymorphic STR with a heterozygosity and polymorphic information content (PIC Value) of more than 0.80
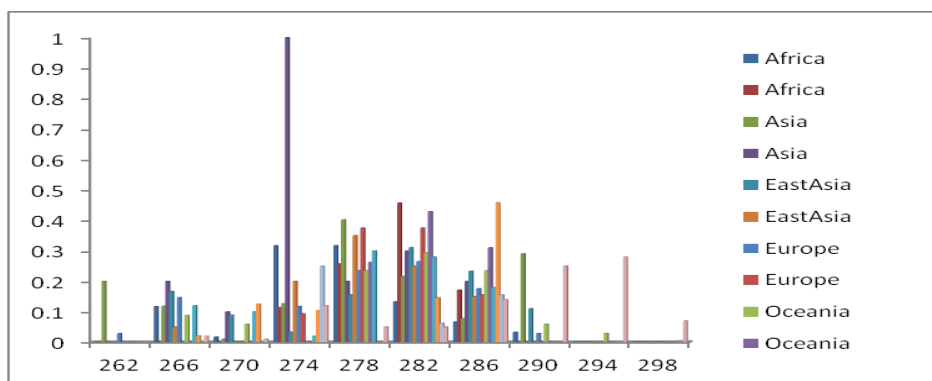
Fig 5.20: Distribution of major allele frequencies of D3S2398 STR in frequencies reported on ALFRED of various populations across the globe

Chapter 5

5.3.2.2 Allelic Distribution of D3S2398 marker among cases and controls

Table 5.32: Allelic Distribution of D3S2398 marker among cases and controls

PCR Pdt Size (bp)	(GATA)n repeats	Allele	Cases (N=224)	Controls (N=202)
			N (%)	N (%)
262	6	1	01 (0.00)	00 (0.00)
266	7	2	12 (0.03)	08 (0.02)
270	8	3	24 (0.05)	04 (0.01)[1]
274	9	4	34 (0.08)	48 (0.12)
278	10	5	44 (0.10)	21 (0.05)
282	11	6	72 (0.16)	22 (0.05)[2]
286	12	7	51 (0.11)	55 (0.14)
290	13	8	81 (0.18)	102 (0.25)
294	14	9	124 (0.28)	115 (0.28)
298	15	10	05 (0.01)	20 (0.07)[3]

[1]=95% CI= 1.90 to 16.13, χ^2= 11.06, *p= 0.0009;* OR=5.54, [2]=95% CI= 1.97 to 5.35, χ^2= 22.43, *p<0.0001* ,OR=3.35, [3]=95% CI= 0.07 to 0.57, χ^2= 9.9, *p= 0.0016,* OR=0.21

PCR Pdt: PCR product; bp: Base pair; N: Number; %: Frequency
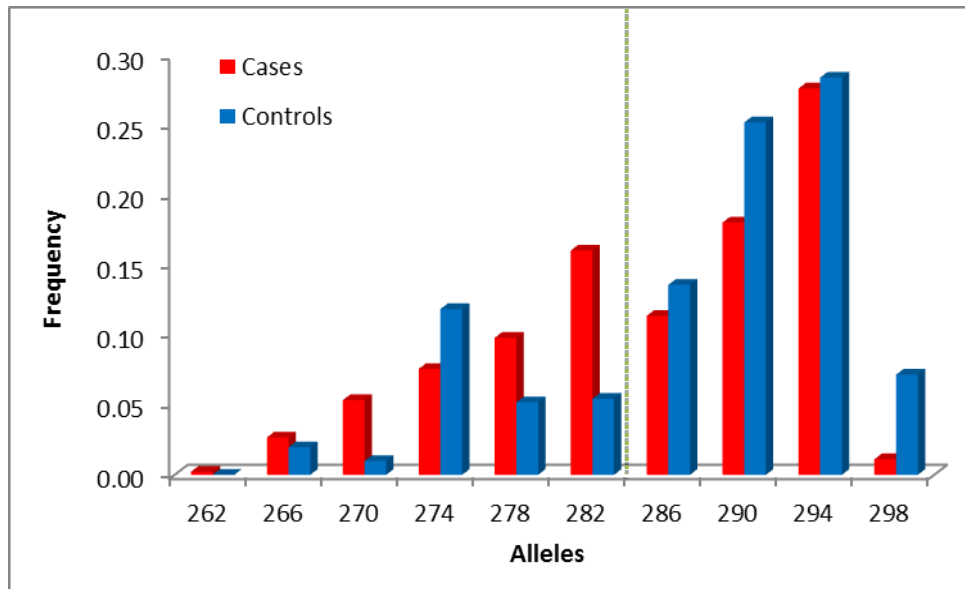
Fig 5.21: Distribution of major allele frequencies of D3S2398 STR in cases vs controls. Dotted lines represent the cut off made according to natural tendency for bimodal distribution of alleles. Red color represents cases; Blue represents controls

Chapter 5

5.3.2.3 Bimodal Distribution of D3S2398Alleles

The cohort was further analyzed using a case-control design on the genotype data of D3S2398 from 226 diabetics and 202 controls. It is evident from Table 5.32 that possession of 298 bp (95% CI= 0.07 to 0.57, $\chi 2$= 9.9, p= 0.0016, OR=0.21) of D3S2398 marker seems to have a protective effect against T2DM while presence of the shorter 270 bp (95% CI= 1.90 to 16.13, $\chi 2$= 11.06, p= 0.0009; OR=5.54) and 282 bp (95% CI= 1.97 to 5.35, $\chi 2$= 22.43, p<0.0001 ,OR=3.35) allele seems to make individuals more susceptible to T2DM. Others alleles did not exhibit any significant difference between cases and controls groups.

The cohort was further divided into approximately two groups, at 286 bp, according to their natural tendency for bimodal distribution at 286 bp. Hence the resulting genotypes were <286 bp/<286 bp, <286 bp/$\geq$286 bp and >286 bp/$\geq$286 bp. The shorter alleles of D3S2398 were found to be more prevalent in the cases (31% vs 16%). The $\geq$286 bp allele constitutes 58% in adult cases and 75% in controls (χ^2=26.51, OR = 0.4608, 95 % CI= 0.3436 to 0.6179, $p < 0.0001$). Although, the significance was lost at the genotypic level.

Table 5.33: Allelic and genotypic frequencies of D3S2398 (Bimodal distribution) Microsatellite Repeat in cases and controls

Group (n)	Genotype Frequency				Allele frequency	
	<286	<286/$\geq$286	$\geq$286	2N	<286*	$\geq$286*
	N (%)	N (%)	N (%)		N (%)	N (%)
Cases (226)	57 (25.2)	75 (33.12)	94 (41.5)	452	189 (41.5)	263 (58.1)
Controls (202)	40 (19.8)	20 (09)	141 (69.8)	404	100 (24.7)	302 (74.8)

Difference among genotypes: χ^2=26.51, p < 0.0001; *Longer allele ($\geq$286bp) as protective allele:
OR = 0.4608, 95 % CI= 0.3436 to 0.6179, p < 0.0001; (<286 vs $\geq$286bp): NS
N: Number; %: Frequency
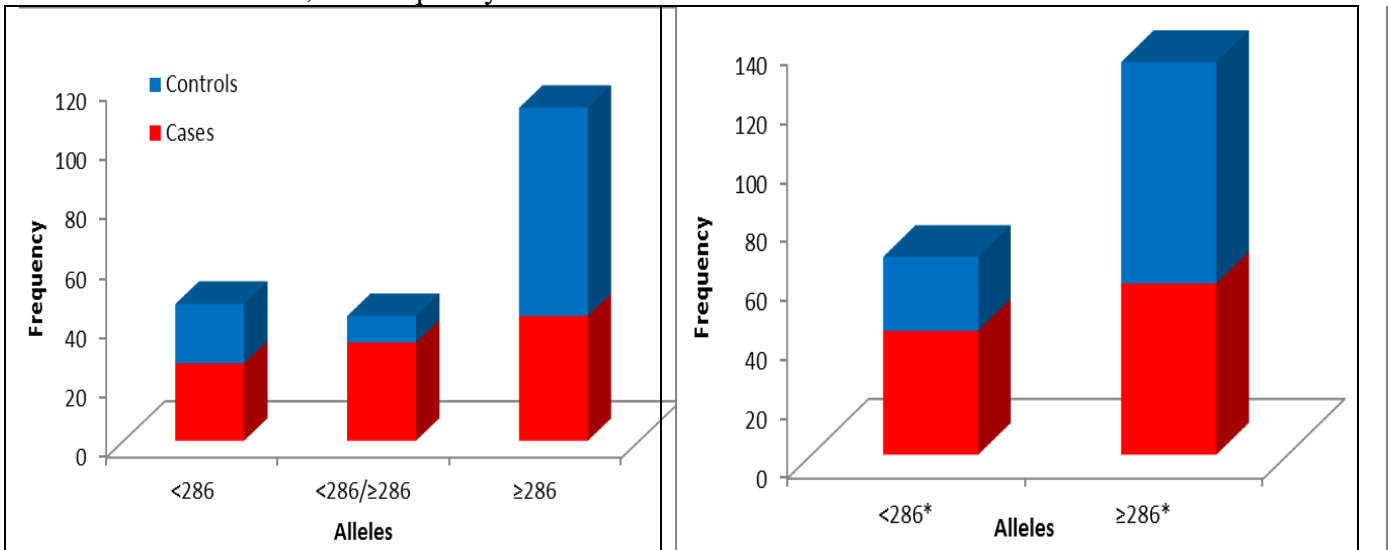
Fig5.22A : Genotypic distribution of D3S2398 STR Repeat after Bimodal distribution. . Red color represents cases; Blue represents controls.

Fig5.22B: Allelic distribution of D3S2398 STR Repeat after Bimodal distribution. . Red color represents cases; Blue represents controls.

5.2.3.4 Distribution of different variables according to D3S2398 alleles

The distribution of anthropometric markers in adult cohort alleles of D2S2398 is presented in Table 5.34. Significant difference was observed in mean WHR (<286 bp: t = 2.865 p= 0.0053; $\geq$286bp: t = 5.970 p < 0.0001). There was marginal difference observed in the mean SBP (<286bp: t = 2.583, p = 0.011; $\geq$286bp: NS) levels in cases and controls in presence of shorter base pair allele. As is evident from the table below no significant differences in mean DBP (<286bp: NS; $\geq$286bp: NS), BMI(<286bp: NS; $\geq$286bp: NS) and levels were observed when segregated for long and short allele. In both cases (presence or absence of longer bp allele) the mean levels remained unchanged among cases and controls. Although not much of the association between the studied anthropometric variable and the STR marker could be established as evident from the table below.

Chapter 5

Table 5.34: Distribution of different anthropometric variables according to D3S2398 genotype among cases and controls

		<286		<286/ ≥286		≥286	
		N	Mean±SD	N	Mean±SD	N	Mean±SD
SBP	Controls 40		128.0±13.4	Controls 18	135.2±21.5	Controls 140	130.8± 18.2
	Cases 50		136.7± 17.6	Cases 70	137.1±15.8	Cases 76	135.3± 14.8
		t = 2.583, *p* = 0.011		NS		NS	
DBP	Controls 40		83.5±10.2	Controls 18	88.9±11.06	Controls 140	87.09±10.5
	Cases 50		85.8± 11.06	Cases 70	85.6±8.6	Cases 76	84.1± 8.9
		NS		NS		NS	
BMI	Controls 40		23.5±4.04	Controls 20	25.09±4.4	Controls 141	23.9± 4.2
	Cases 46		26.1± 5.7	Cases 60	26.2± 4.4	Cases 66	25.03± 3.7
		t = 2.406 *p*= 0.0183		NS		NS	
WHR	Controls 40		0.90± 0.04	Controls 18	0.9± 0.05	Controls 130	0.9± 0.05
	Cases 46		0.94± 0.08	Cases 50	0.9±0.06	Cases 60	0.9± 0.07
		t = 2.865 *p* = 0.0053		t = 3.159 *p*= 0.0024		t = 5.970 *p* < 0.0001	

FBG: Fasting Blood Glucose, HDL: High Density lipoprotein, LDL: Low density lipoprotein, VLDL: Very Low Density Lipoprotein, CHL: Cholesterol. Values are presented as means ± standard deviation, *p* values differed significantly among cases and controls (*p*<0.05)

The distribution of the various biochemical parameters in adult cohort alleles of D3S2398 is presented in Table 5.35.

The mean FBG levels in cases and controls differed significantly in presence of longer base pair allele (<286bp: t = 9.882, *p* < 0.0001; ≥286bp: t = 15.134, *p* < 0.0001). Although there was no significant difference observed among cases and controls on stratification for long and short allele for mean levels of any of the lipid parameters

except HDL wherein the longer bp allele carrying subjects had lower HDL levels as compared to those bearing shorter allele.

Table 5.35: Distribution of different biochemical variables according to D3S2398 genotype among cases and controls

		<286	<286	<286/ ≥286	<286/ ≥286	≥286	≥286
		N	Mean±SD	N	Mean±SD	N	Mean±SD
FBG	Controls	41	86.3±17.8	20	88.4± 14.1	141	86.6± 14.1
	Cases	50	192.6± 66.9	70	169.2± 49.5	78	176.3± 67.9
		t = 9.882, $p < 0.0001$		t = 7.191, $p < 0.0001$		t = 15.134, $p < 0.0001$	
CHL	Controls	36	186.9± 50.4	19	180.6± 48.6	117	181.7± 43.3
	Cases	56	188.4± 49.7	72	194.6± 40.1	84	185.4± 45.8
		NS		NS		NS	
TG	Controls	35	188.7± 100.1	19	167.09± 59.5	117	180.0± 78.7
	Cases	56	167.7± 99.5	72	171.9± 92.6	84	176.7± 101.7
		NS		NS		NS	
HDL	Controls	35	41.1± 24.3	19	35.6± 16.1	117	38.9± 13.8
	Cases	56	46.1± 18.8	72	40.8± 18.6	84	49.8± 24.8
		NS		NS		NS	
LDL	Controls	35	106.1± 46.4	19	111.5± 43.5	117	107.8±40.8
	Cases	56	107.7± 46.06	72	119.4± 40.2	84	103.2± 45.8
		NS		NS		NS	
VLDL	Controls	35	37.7±20.01	19	33.4± 11.9	117	35.9± 15.8
	Cases	56	32.4± 17.9	72	34.3± 18.7	83	35.4± 20.6
		NS		NS		NS	

FBG: Fasting Blood Glucose, HDL: High Density lipoprotein, LDL: Low density lipoprotein, VLDL: Very Low Density Lipoprotein, TG: Triglycerides; Values are presented as means ± standard deviation, p values differed significantly among cases and controls ($p<0.05$)

Chapter 5

5.4 Chromosome 10 and T2DM

The fig bellow lists the region studied on chromosome 10 and the two markers D10S1237 and D10S521 studied in the present study are encircled. The details of the region as per UniSTS are given below.

Region: 10q25-26

Bp size: 130021000-198022000
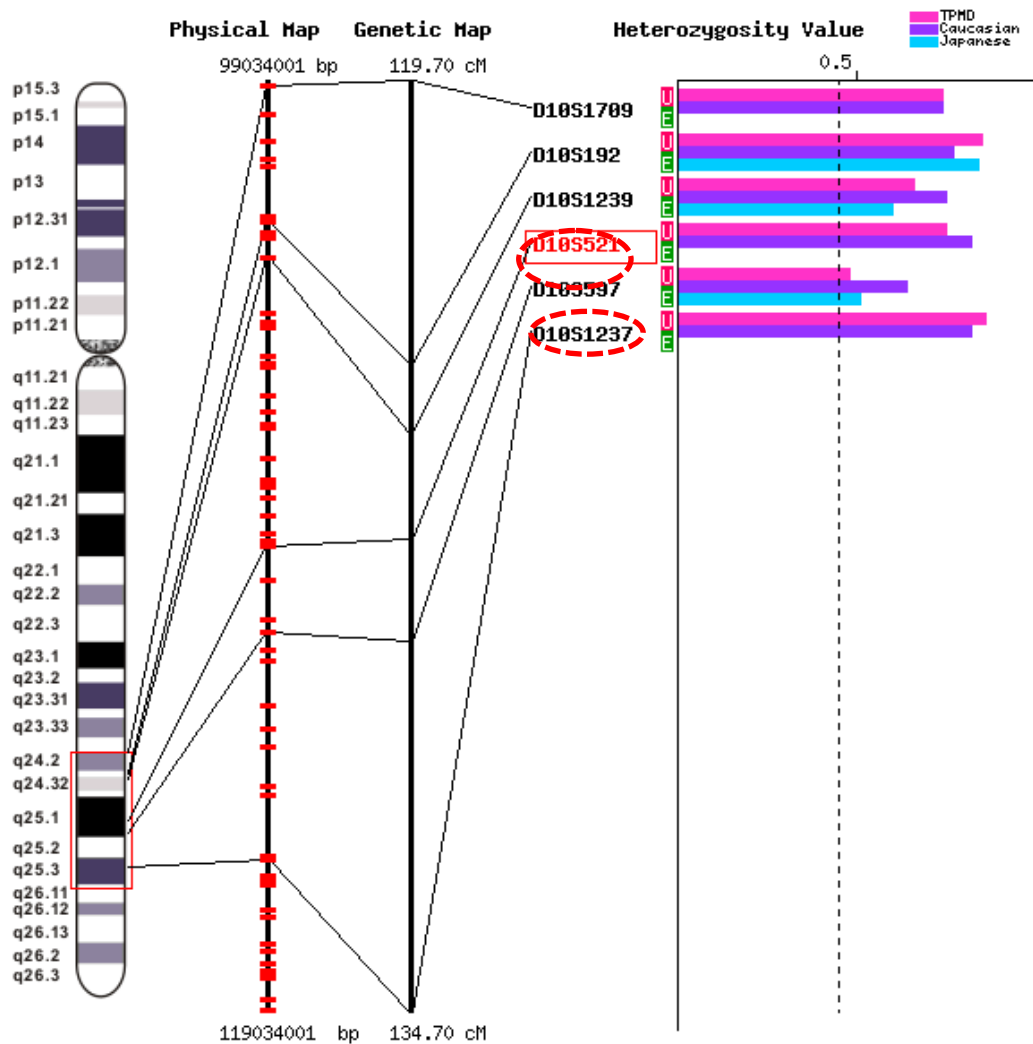
Fig 5.23: Position and Hetrozygosity score of STR markers analyzed on chromosome 10. The red square box on the ideogram represents the region chosen for the study. The encircled regions represent the markers studied in the current study

Chapter 5

5.4.1 D10S521 polymorphism & its association with T2DM

The association of variation in tetranucleotide repeat microsatellite marker D10S521 lying at 125.85 at 10q was analyzed in the present study.

5.4.1.1. Comparison with World Frequencies

The allele frequencies for D10S521 STR marker differ from previously known CEPH population (Utah Residents with Northern and Western European Ancestry)(CEU): Fig 5.24 (the CEPH Genotype database: http:// www.cephb.fr/cephdb and the CHLC Genetic Mapping data- base: http://lpgws.nci.nih.gov/html-chlc/ChlcMarkers.html). The studied Indian population, D10S521 microsatellite has 10 types of alleles with 15-24 $(GATA)_n$ tetra nucleotide repeat nucleotides: which yield DNA fragment of length ranging from 155-191 base pair. Population data showed that this is a highly polymorphic STR with a heterozygosity and polymorphic information content (PIC Value) of more than 0.80
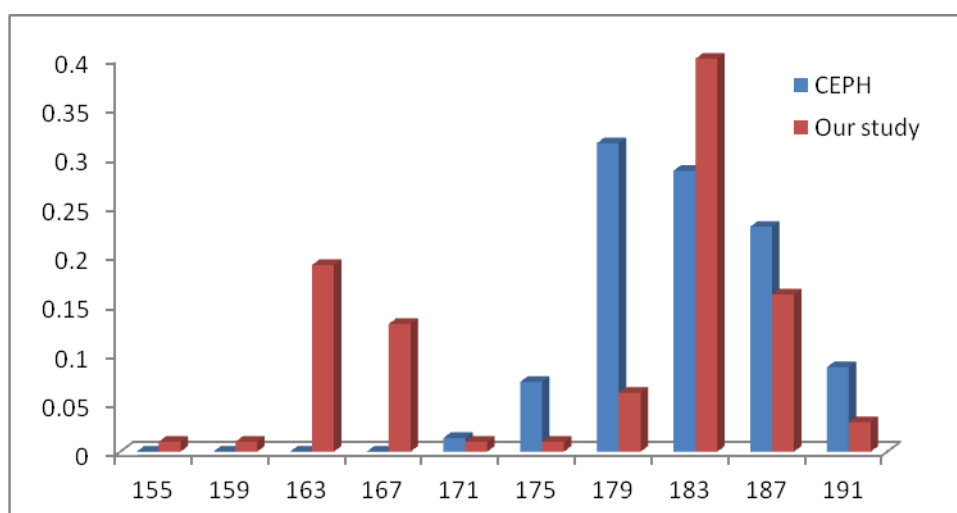
Fig 5.24: Distribution of major allele frequencies of D10S521 STR in CEPH vs Indian Population. The red color represents frequencies reported in the current study

Chapter 5

5.4.1.2 Allelic Distribution of D10S521 marker among cases and controls

Table 5.36: Allelic Distribution of D10S521 marker among cases and controls

PCR Pdt Size (bp)	(GATA)n repeats	Allele	Cases (N=216)	Controls (N=187)
			N (%)	N (%)
155	15	1	01 (0.00)	04 (0.01)
159	16	2	03 (0.01)	03 (0.01)
163	17	3	21 (0.05)	72 (0.19)[1]
167	18	4	03 (0.01)	48 (0.13)[2]
171	19	5	07 (0.02)	2 (0.01)
175	20	6	05 (0.01)	04 (0.01)
179	21	7	28 (0.06)	24 (0.06)
183	22	8	225 (0.52)	149 (0.40)[3]
187	23	9	135 (0.31)	58 (0.16)[4]
191	24	10	03 (0.01)	10 (0.03)
[1]=95% CI= 0.128 to 0.356, χ^2= 39.27, $p < 0.0001$, OR=0.21				
[2]=95% CI=0.014 to 0.15, χ^2= 47.81, $p < 0.0001$, OR=0.04				
[3]=95% CI= 1.2 to 2.17, χ^2= 11.59, $p= 0.0005$; OR=1.64				
[4]=95% CI= 1.75 to 3.5, χ^2= 26.41, $p<0.0001$, OR=2.47				

PCR Pdt: PCR product; bp: Base pair; N: Number; %:Frequency
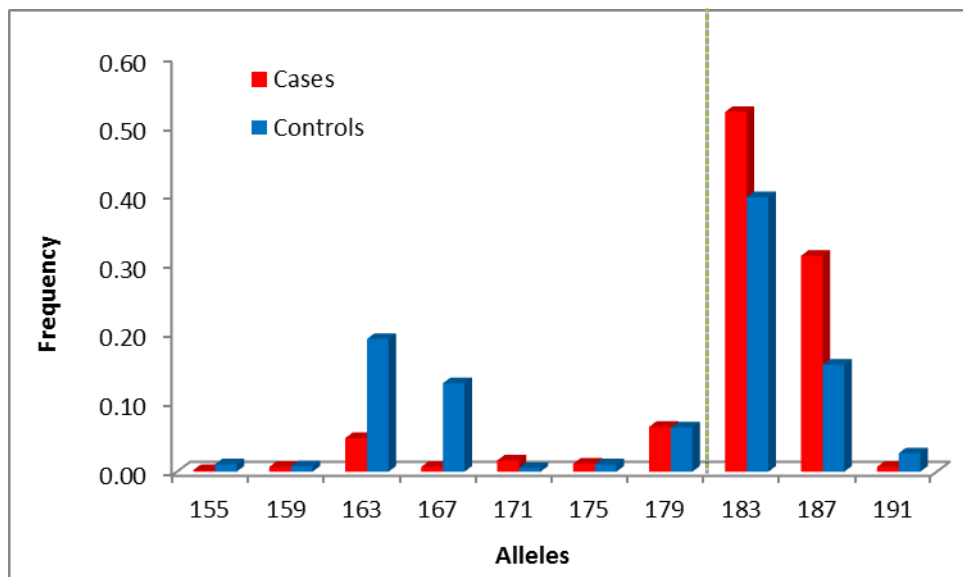
Fig 5.25: Distribution of major allele frequencies of D10S521 STR in cases vs controls. Dotted lines represent the cut off made according to natural tendency for bimodal distribution of alleles. Red color represents cases; Blue represents controls.

Chapter 5

5.4.1.3 Bimodal Distribution of D10S521 Alleles

The cohort was further analyzed using a case-control design on the genotype data of D10S521 from 216 diabetics and 187 controls. It is evident from Table 5.36 that possession of 163 bp (95% CI= 0.128 to 0.356, $\chi 2$= 39.27, $p < 0.0001$, OR=0.21) and 167bp (95% CI=0.014 to 0.15, $\chi 2$= 47.81, $p < 0.0001$,OR=0.04) of D10S521 marker seems to have a protective effect against T2DM while presence of the longer allele 183bp (95% CI= 1.2 to 2.17, $\chi 2$= 11.59, p= 0.0005;OR=1.64) and 187 bp (95% CI= 1.75 to 3.5, $\chi 2$= 26.41, p<0.0001, OR=2.47) allele seems to make individuals more susceptible to T2DM. The cohort was further divided into approximately two groups, at 179 bp, according to their natural tendency for bimodal distribution at 179 bp. Hence the resulting genotypes were <179 bp/<179 bp, <179 bp/$\geq$179 bp and >179 bp/$\geq$179 bp. The shorter alleles of D10S521 were found to be more prevalent in the cases (31% vs 16%). The $\geq$179 bp allele constitutes 91% in adult cases and 64% in controls (χ^2=83.127, OR = 5.6639, 95 % CI=3.8169 to 8.4047, $p < 0.0001$). Infact the presence of at least one copy of longer base pair allele increased the risk further by two folds. (OR=5.66 to OR=10.86)

Table 5.37: Allelic and genotypic frequencies of D10S521 (Bimodal distribution) Microsatellite Repeat in cases and controls

Group (n)	Genotype Frequency				Allele frequency	
	<179	<179/$\geq$179	$\geq$179	2N	<179*	$\geq$179*
Cases (214)	10 (4.6)	18 (8.4)	186 (86.9)	428	38 (8.8)	390 (91.2)
Controls (187)	65 (34.7)	03 (1.6)	119 (63.6)	374	133 (35.6)	241 (64.4)
Difference among genotypes: χ^2=83.127, $p < 0.0001$; *Longer allele ($\geq$179bp) as risk allele: OR = 5.6639, 95 % CI=3.8169 to 8.4047, $p < 0.0001$; (<179 vs $\geq$179bp): χ^2= 57.448 OR = 10.8689, 95 % CI= 5.3835 to 21.9431, $p < 0.0001$						
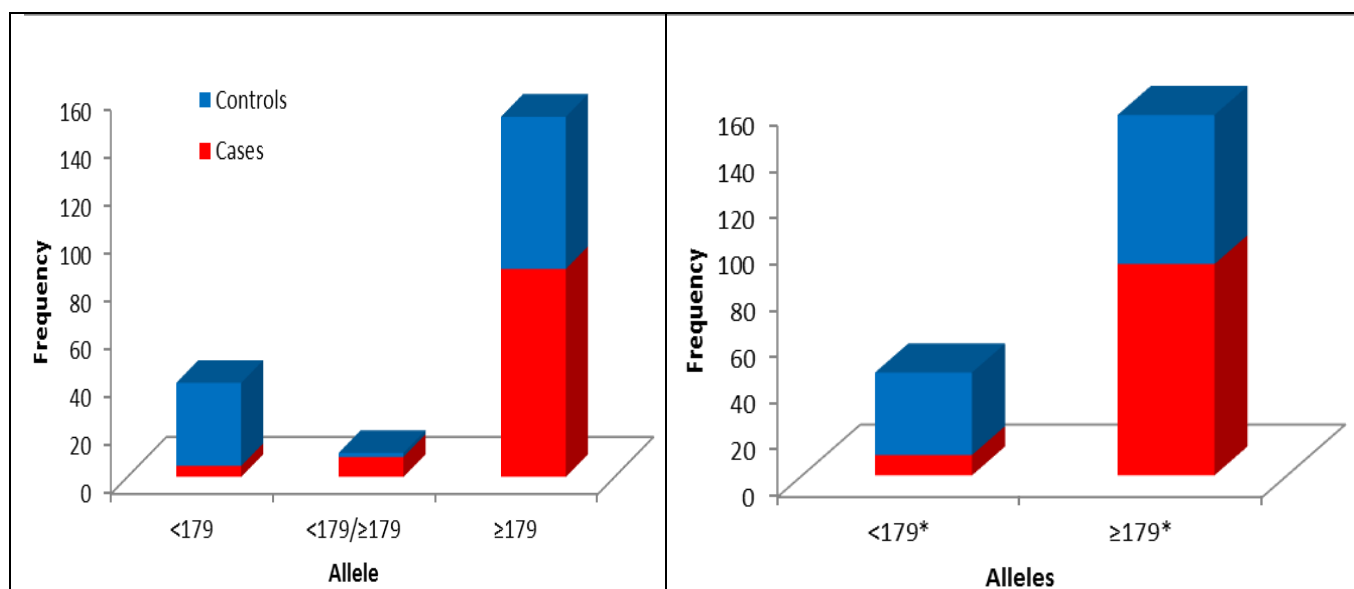
Fig5.26A : Genotypic distribution of D10S521 STR Repeat after Bimodal distribution. Red color represents cases; Blue represents controls.

Fig 5.26B: Allelic distribution of D10S521 STR Repeat after Bimodal distribution. Red color represents cases; Blue represents controls.

5.4.1.4 Distribution of different variables according to D10S521 alleles

The distribution of anthropometric in adult cohort alleles of D10S521 is presented in Table 5.38. There was significant difference observed in the mean SBP (<179bp: t = 2.327, p = 0.02 ≥179bp: NS), DBP (<179bp: NS ≥179bp: t = 2.149, p = 0.03), BMI (<179bp: t = 2.271, p =0.02; ≥179bp t = 3.278 , p =0.0012) and WHR(<179bp: t = 4.424, p < 0.0001 ≥179bp: NS),   levels in cases and controls when segregated on the basis of long and short alleles of D10521. Though the significances observed in the case of SBP, DBP and BMI were marginal only; the difference between mean WHR between cases and controls in presence of shorter base pair allele or even a single copy of shorter base pair allele was significant. (<179: t = 4.424, p < 0.0001; <179/≥179: t=2.405 p = 0.03). In case of DBP the significance was seen only in case of longer base pair allele and no significant difference in the mean DBP was observed among cases and controls in presence of shorter base pair allele or heterozygous state. In case of

SBP the significance was seen only in case of shorter base pair allele and no significant difference in the mean SBP was observed among cases and controls in presence of longer base pair allele or heterozygous state.

Table 5.38: Distribution of different anthropometric variables according to D10S521 genotype among cases and controls

		<179		<179/ ≥179		≥179	
		N	Mean±SD	N	Mean±SD	N	Mean±SD
SBP	Controls 62		130.1±19.3	Controls 3	120.0±17.3	Controls 118	131.7± 17.2
	Cases 9		145.5± 11.3	Cases 13	136.5±22.1	Cases 170	134.8± 14.8
	t = 2.327, p = 0.02			NS		NS	
DBP	Controls 62		86.3±11.7	Controls 3	80.0±10.0	Controls 118	87.3±10.4
	Cases 9		90.8± 11.3	Cases 13	79.6±9.1	Cases 170	84.8± 9.2
	NS			NS		t = 2.149, p = 0.03	
BMI	Controls 64		23.6±4.6	Controls 3	27.6±5.2	Controls 119	24.2± 3.9
	Cases 8		27.6± 5.5	Cases 13	27.2± 5.7	Cases 154	25.9± 4.5
	t = 2.271, p =0.02			NS		t = 3.278 , p =0.0012	
WHR	Controls 59		0.8± 0.06	Controls 3	0.8±0.01	Controls 114	0.9± 0.05
	Cases 8		0.9± 0.06	Cases 13	0.9±0.07	Cases 136	0.9± 0.07
	t = 4.424, p < 0.0001			t=2.405 p = 0.03		NS	

SBP: Systolic Blood Pressure, DBP: Diastolic Blood pressure: BMI: Body Mass Index, WHR : Waist to hip ratio, NS: Non Significant; Values are presented as means $\pm$ standard deviation, p values differed significantly among cases and controls ($p<0.05$)

The cohort was further analyzed on the mean levels of FBG and lipid parameters after segregating for the long and short allele. The mean FBG level among cases and controls significantly differed when segregated on the presence of long and short allele. The longer bp allele carriers had higher difference in cases and controls as compared to those carrying shorter base pair allele. (<179 t = 12.596, $p < 0.0001$; >179: t = 16.054, $p < 0.0001$).

Chapter 5

Table 5.39: Distribution of different biochemical variables according to D10S521 genotype among cases and controls

		<179		<179/ ≥179		≥179	
	N	Mean±SD	N	Mean±SD	N	Mean±SD	
FBG	Controls 65	86.5±17.2	Controls 3	97.0± 6.08	Controls 119	87.1± 13.9	
	Cases 9	204.3± 62.1	Cases 13	201.07± 82.5	Cases 172	176.2± 59.4	
	t = 12.596, *p* < 0.0001		t = 2.126, *p* =0.05		t = 16.054, *p* < 0.0001		
CHL	Controls 57	182.7± 50.06	Controls 3	259.7± 56.8	Controls 101	178.8± 39.4	
	Cases 10	176.7± 44.9	Cases 18	195.4± 53.5	Cases 178	193.2± 50.1	
	NS		t = 1.915, *p* =0.07		t = 2.485, *p* =0.013		
TG	Controls 57	180.9± 85.2	Controls 3	161.06± 76.8	Controls 101	178.8± 81.1	
	Cases 10	124.4± 50.5	Cases 18	148.2± 73.6	Cases 178	171.4± 95.0	
	t = 2.027, *p* = 0.046		NS		NS		
HDL	Controls 57	40.1± 17.5	Controls 3	46.8± 9.5	Controls 101	39.1± 16.8	
	Cases 10	54.4± 23.5	Cases 18	46.9± 18.9	Cases 178	46.4± 21.2	
	t= 2.261, *p* = 0.0271		NS		t= 2.971, *p* = 0.03		
LDL	Controls 57	109.4± 43.09	Controls 3	180.7± 56.4	Controls 101	103.2±37.5	
	Cases 10	97.4± 47.4	Cases 18	115.6± 55.9	Cases 178	113.7± 49.7	
	NS		NS		NS		
VLDL	Controls 57	36.2±17.03	Controls 3	32.2± 32.2	Controls 101	35.7± 16.3	
	Cases 10	24.8± 10.1	Cases 18	29.6± 15.3	Cases 177	33.9± 18.5	
	t =2.047, *p*= 0.04		NS		NS		

FBG: Fasting Blood Glucose, HDL: High Density lipoprotein, LDL: Low density lipoprotein, VLDL: Very Low Density Lipoprotein, TG: Triglycerides; Values are presented as means ± standard deviation, *p* values differed significantly among cases and controls (*p*<0.05)

It was observed that the TC levels were significantly higher in longer base pair bearing group as compared to those with shorter base pair allele. . (<179: NS; >179: t = 2.485, *p* =0.013). Even the presence of a single longer base pair allele showed nominal difference in the mean TC levels. (<179/≥179: t = 1.915, *p* =0.07). Also nominal

Chapter 5

significance only was seen in case of TG (<179 t = 2.027, p = 0.046; >179: NS). HDL (<179: t= 2.261, p = 0.0271; >179: NS) and VLDL<179 t =2.047, p= 0.04; >179: NS). Though no significance was seen in case of LDL (<179: NS; >179: NS).

5.4.2 D10S1237 polymorphism & its association with T2DM

The association of variation in tetra nucleotide repeat microsatellite marker D10S1237 lying at gene actin filament associated protein 1-like 2at position: 10q25.3 (134.70cM at 10q was analyzed in the present study.

5.4.2.1Comparison with World Frequencies

The allele frequencies for D10S1237 STR marker was not much different from previously known CEPH population (Utah Residents with Northern and Western European Ancestry)(CEU): Fig 5.27 (the CEPH Genotype database: http:// www.cephb.fr/cephdb and the CHLC Genetic Mapping data- base: http://lpgws.nci.nih.gov/html-chlc/ChlcMarkers.html). The studied Indian population D10S1237 microsatellite has 13 types of alleles with 18-30 (GATA) $_n$ tetra nucleotide repeat nucleotides: which yield DNA fragment of length ranging from 376-432 base pair. Population data showed that this is a highly polymorphic STR with a heterozygosity and polymorphic information content (PIC Value) of more than 0.80
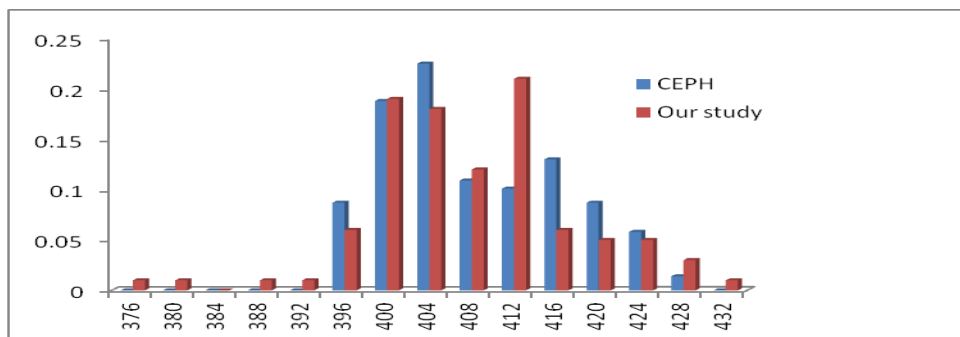
Fig 5.27: Distribution of major allele frequencies of D10S1237 STR in CEPH vs Indian Population

5.4.2.2 Allelic Distribution of D10S1237 marker among cases and controls

Table 5.40: Allelic Distribution of D10S1237 marker among cases and controls

PCR Pdt Size (bp)	(GATA)n Repeats	Allele	Cases (N=223)	Controls (N=190)
			N (%)	N (%)
376	12	1	06 (0.01)	02 (0.01)
380	13	2	10 (0.02)	02 (0.01)
384	14	3	05 (0.01)	01 (0.00)
388	15	4	04 (0.01)	4(0.01)
392	16	5	06 (0.01)	2 (0.01)
396	17	6	24 (0.05)	23 (0.06)
400	18	7	20 (0.04) [1]	73 (0.19) [1]
404	19	8	60 (0.13)	69 (0.18)
408	20	9	60 (0.13)	47 (0.12)
412	21	10	65 (0.15)	81 (0.21)
416	22	11	21 (0.05)	21 (0.06)
420	23	12	30 (0.07)	19 (0.05)
424	24	13	66 (0.15) [2]	18 (0.05) [2]
428	25	14	35 (0.08) [3]	13(0.03) [3]
432	26	15	33 (0.07) [4]	04 (0.01) [4]

Cases vs controls: [1]95% CI= 0.118 to 0.331;χ^2= 42.92 *p < 0.0001*, OR=0.197; [2]95% CI= 2.03 to (,χ^2= 21.74, *p<0.0001*, OR=3.5; [3] 95% CI= 1.2 to 4.6; χ^2= 6.5 ; *p =0.0102*, OR= 2.4; [4]95% CI= 2.6 to 21.45, χ^2= 17.9, *p <0.0001*, OR=6.9

PCR Pdt: PCR product; bp: Base pair; N: Number; %:Frequency
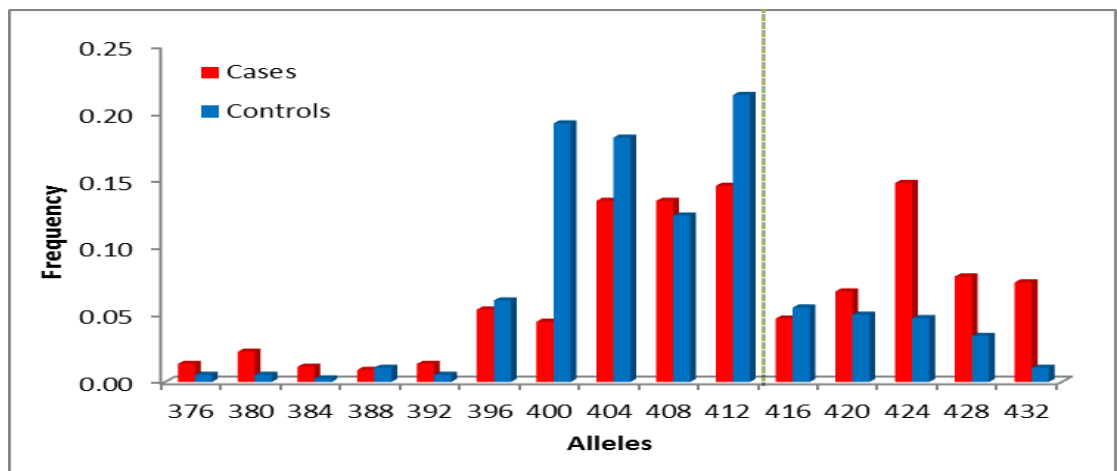
Fig 5.28: Distribution of major allele frequencies of D10S521 STR in cases vs controls. Dotted lines represent the cut off made according to natural tendency for bimodal distribution of alleles. Red color represents cases; Blue represents controls.

Chapter 5

5.4.2.3 Bimodal Distribution of D10S1237 Alleles

The cohort was further analyzed using a case-control design on the genotype data of D10S1237 from 223 diabetics and 190 controls. It is evident from table 5.40 that possession of 400 bp (195% CI= 0.118 to 0.331; χ 2= 42.92 p < 0.0001, OR=0.197) allele of D10S1237 marker seems to have a protective effect against T2DM while presence of the longer base pair allele 424bp (295% CI= 2.03 to 6.01, χ 2= 21.74, p<0.0001, OR=3.5), 428 bp (95% CI= 1.2 to 4.6; χ 2= 6.5 ; p =0.0102, OR= 2.4) and 432 (95% CI= 2.6 to 21.45, χ 2= 17.9, p <0.0001, OR=6.9) allele seems to make individuals more susceptible to T2DM. Others alleles did not exhibit any significant difference between cases and controls groups. The cohort was further divided into approximately two groups, at 416 bp, according to their natural tendency for bimodal distribution at 416 bp. Hence the resulting genotypes were <416 bp/<416 bp, <416 bp/≥416 bp and >416 bp/≥416 bp. The shorter alleles of D10S1237 were found to be more prevalent in the control cohort as compared to cases (81% vs 59%). The ≥416 bp allele constitutes 41% in adult cases and 19% in controls (χ^2=48.220, OR = 3.01, 95 % CI=3 2.2033 to 4.1352, $p < 0.0001$). The significance was not lost after genotypic segregation (χ^2= 37.263, OR = 3.8623, 95 % CI=2.4916 to 5.9872, $p < 0.0001$)

Table 5.41: Allelic and genotypic frequencies of D10S1237 (Bimodal distribution) Microsatellite Repeat in cases and controls

Group (N)	Genotype Frequency			2N	Allele frequency	
	<416	<416/≥416	≥416		<416*	≥416*
	N (%)	N (%)	N (%)		N (%)	N (%)
Cases (220)	114 (51.8)	32 (14.5)	74 (33.6)	440	260 (59.1)	180 (40.9)
Controls (201)	162 (80.5)	03 (1.4)	36 (17.9)	402	327 (81.3)	75 (18.7)
Difference among genotypes: χ^2=48.220, $p < 0.0001$; *Longer allele (≥416bp) as risk allele: OR = 3.0185 , 95 % CI=3 2.2033 to 4.1352, $p < 0.0001$; (<416 vs ≥416bp): χ^2=37.263, OR = 3.8623, 95 % CI2.4916 to 5.9872 , $p < 0.0001$						

N: Number; %: Frequency
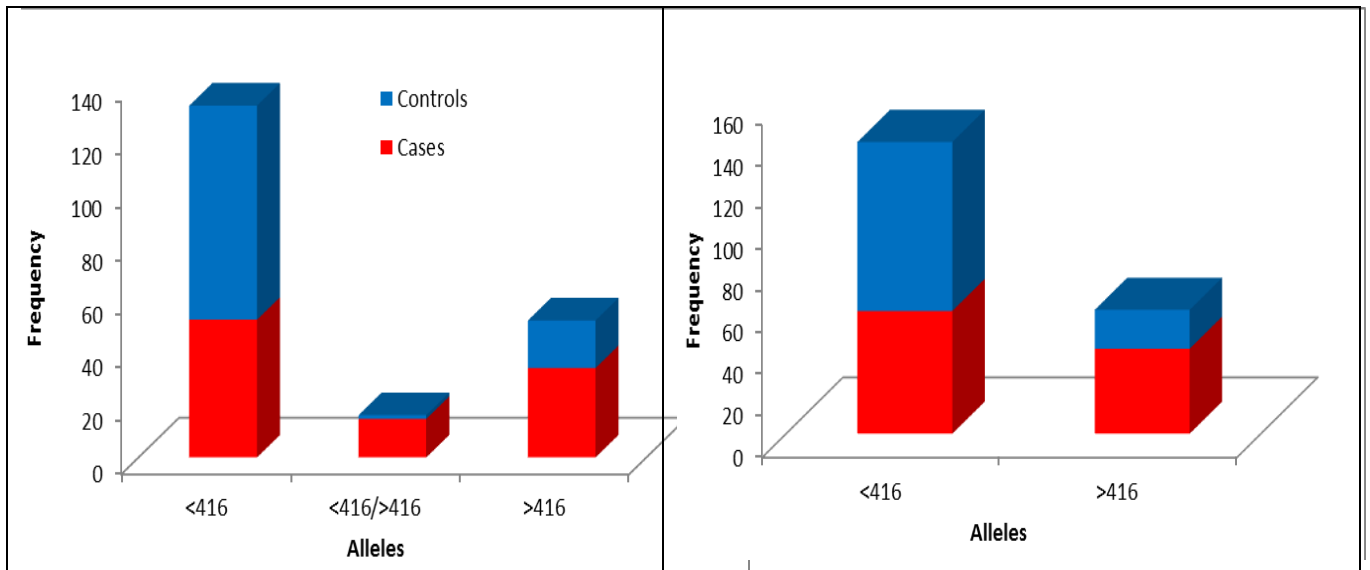
Fig5.29A:Genotypic distribution of D10S1237 STR Repeat after Bimodal distribution. Red color represents cases; Blue represents controls.

Fig 5.29B: Allelic distribution of D10S1237 STR Repeat after Bimodal distribution. Red color represents cases; Blue represents controls.

5.4.2.4 Distribution of different variables according to D10S1237 alleles

The distribution of anthropometric in adult cohort alleles of D10S1237 is presented in Table 5.42. There was significant difference observed in the mean WHR (<416bp: NS ≥416bp: t = 6.984, $p < 0.0001$) levels in cases and controls when segregated on the basis of long and short alleles of D101237. Though the significances observed in the case of SBP, DBP and BMI were marginal only; the difference between mean WHR between cases and controls in presence of longer bp allele was significant. (<416: NS; <416/≥416: t = 6.984, $p < 0.0001$). In case of shorter bp allele the significances were observed in BMI, SBP and DBP which were lost when analyzed in case of longer base pair allele.

Chapter 5

Table 5.42: Distribution of different anthropometric variables according to D10S1237 genotype among cases and controls

		<416		<416/ ≥416		≥416
	N	Mean±SD	N	Mean±SD	N	Mean±SD
SBP	Controls 159	130.6±17.5	Controls 3	121.3±10.2	Controls 35	129.6± 12.3
	Cases 104	135.9± 14.1	Cases 28	134.1±18.5	Cases 68	135.3± 15.8
	t = 2.587, *p* = 0.01		NS		NS	
DBP	Controls 159	87.03±11.04	Controls 3	81.6±10.4	Controls 35	85.8±8.9
	Cases 104	84.3± 9.08	Cases 28	82.7±9.8	Cases 68	86.07± 9.5
	t = 2.099, *p* = 0.03		NS		NS	
BMI	Controls 162	24.3±4.1	Controls 3	30.5±3.8	Controls 35	22.9± 3.3
	Cases 91	25.5± 3.8	Cases 26	27.5± 6.2	Cases 64	26.1± 4.9
	t = 2.293, *p* =0.02		NS		t = 3.455,*p* =0.0008	
WHR	Controls 152	0.9± 0.05	Controls 3	0.9± 0.06	Controls 36	0.8± 0.04
	Cases 77	0.9± 0.06	Cases 25	0.9±0.07	Cases 60	0.9± 0.08
	NS		NS		t = 6.984, *p* < 0.0001	

SBP: Systolic Blood Pressure, DBP: Diastolic Blood Pressure, BMI: Body Mass Index, WHR: Waist Hip ratio, NS: Non significant; Values are presented as means ± standard deviation, *p* values differed significantly among cases and controls (*p*<0.05)

The distribution of the various biochemical parameters in adult cohort alleles of D10S1237 is given in Table 5.43. Significant differences were observed in mean levels of FBG (<416bp: 18.171, *p* < 0.0001; ≥416bp: t = 9.869, *p* < 0.0001) among the cases and controls on segregating for the long and short alleles along with mean TC (<416bp: NS; ≥416bp: t = 3.361, *p* =0.0011) and HDL (<416bp: t=2.177, *p* = 0.03; ≥416bp: t= 3.699, *p* = 0.004) levels on grouping on the basis of long and short allele.

Table 5.43: Distribution of different biochemical variables according to D101237 genotype among cases and controls

		<416		<416/ ≥416		≥416	
	N	Mean±SD	N	Mean±SD	N	Mean±SD	
FBG	Controls 162	86.8±14.08	Controls 3	113.6± 11.8	Controls 36	82.8± 17.1	
	Cases 104	183.8± 65.7	Cases 28	165.8± 63.1	Cases 69	178.6± 56.8	
	t = 18.171, p < 0.0001		t = 9.874, p < 0.0001		t = 9.869, p < 0.0001		
CHL	Controls 140	187.2± 42.3	Controls 3	254.9± 41.3	Controls 36	160.5± 46.9	
	Cases 114	186.8± 47.5	Cases 31	224.5± 63.0	Cases 67	190.2± 40.4	
	NS		NS		t = 3.361, p =0.0011		
TG	Controls 139	174.3± 76.1	Controls 3	291.7± 270.4	Controls 36	179.3± 65.5	
	Cases 114	161.9± 86.9	Cases 31	177.6± 98.1	Cases 67	168.1± 98.1	
	NS		NS		NS		
HDL	Controls 139	40.2± 15.6	Controls 3	67.6± 44.4	Controls 36	32.2± 14.1	
	Cases 114	45.3± 21.6	Cases 31	47.9± 20.9	Cases 67	46.2± 20.2	
	t=2.177, p = 0.03		NS		t= 3.699, p = 0.004		
LDL	Controls 139	111.1± 42.0	Controls 3	128.9± 92.3	Controls 36	97.7±36.1	
	Cases 114	110.8± 48.05	Cases 31	137.3± 59.4	Cases 67	111.2± 42.6	
	NS		NS		NS		
VLDL	Controls 139	34.8±15.2	Controls 3	58.3± 54.09	Controls 36	35.8± 13.1	
	Cases 113	32.4± 17.6	Cases 31	35.1± 20.2	Cases 67	32.7± 17.8	
	NS		NS		NS		

FBG: Fasting Blood Glucose, HDL: High Density lipoprotein, LDL: Low density lipoprotein, VLDL: Very Low Density Lipoprotein, TG: Triglycerides. Values are presented as means ± standard deviation, p values differed significantly among cases and controls (p<0.05)

Chapter 5

5.5 *In-silico* analysis of transcripts containing different number of repeats of the studied STR markers

To evaluate the functional importance of the identified and studied STR markers we used *in-silico* folding analysis using online available tools.

5.5.1 *In-silico* analysis D2S1384

The seed sequence of the marker was obtained from UniSTS and hypothetical sequences were created by inserting or deleting the expected number of repeats in the seed sequence.

Table 5.44: D2S1384 Sequences analyzed for the folding

Category	Size	Sequence
Seed Sequence	149	aatagagggcccttgcttaaaatctatctatctatctatcaatcttatcttatctatctatctatctatct atctatctatctatctatctatctatctatcgagtgtgtgtgtgtacatattgcaaaataccttttatcccaaa
Seq with 9 repeats	133	aatagagggcccttgcttaaaatctatctatctatctatcaatcttatcttatctatctatctatctatct atctatctatctatcgagtgtgtgtgtgtacatattgcaaaataccttttatcccaaa
Seq with 13repeats	173	aatagagggcccttgcttaaaatctatctatctatctatcaatcttatcttatctatctatctatctatct atctatctatctatctatctatctatctatctatctatctatctatcgagtgtgtgtgtg tacatattgcaaaataccttttatcccaaa
(GATA) 133bp		(GATA) 173bp

Fig 5.30: M FOLD analysis of the hypothetical transcripts having expected number of repeats in D2S1384. The minimum (133bp) and the maximum (173bp) alleles are compared in terms of structural loops formed on folding. The numbered boxes represent the loops in the structure.

Chapter 5

5.5.2 *In-silico* analysis D2S2944

The seed sequence of the marker was obtained from UniSTS and hypothetical sequences were created by inserting the expected number of repeats in the seed sequence. The table below summarizes the sequences obtained using online database and the hypothetical transcripts which were further analyzed for their using the online available tools.

Table 5.45: D2S2944 Sequences analyzed for the folding

Category	Size	Sequence
Seed Sequence	124 bp	Tctgtcttagatggatgaatggacagatagatagatagatagatagatagatagatagata gatagatagatagatagacagacagatagatctacctaacttggaatcttttccatccatcc
Seq with 5 repeats	92 bp	Tctgtcttagatggatgaatggacagatagatagatagatagatagacagacagatagat ctacctaacttggaatcttttccatccatccc
Seq with 15 repeats	132 bp	Tctgtcttagatggatgaatggacagatagatagatagatagatagatagatagatagata gatagatagatagatagatagacagacagatagatctacctaacttggaatcttttcc atccatccc

(GATA) 92bp	(GATA) 132bp

Fig 5.31: MFOLD analysis of the hypothetical transcripts having expected number of repeats in D2S2944. The numbered boxes represent the loops in the structure. The minimum (92 bp) and the maximum (132 bp) alleles are compared in terms of structural loops formed on folding. The numbered boxes represent the loops in the structure.

Chapter 5

5.5.3 *In-silico* analysis D2S439

The seed sequence of the marker was obtained from UniSTS and hypothetical sequences were created by inserting the expected number of repeats in the seed sequence. The table below summarizes the sequences obtained using online database and the hypothetical transcripts which were further analyzed for their using the online available tools.

Table 5.46: D2S439 Sequences analyzed for the folding

Category	Size	Sequence
Seed Sequence	187 bp	Aggaacaaagtctcattctcttgttaaaaatatttggtgagaacagtctgttagttcttttaatt cctcattctatctatctatctatctatctatctatctatctatctatcaatatgtgtatgtgtgt gtagatacatatataatatataacctacacacatgtgttgtaaatgatatctgt
Seq with 5 repeats	163 bp	Aggaacaaagtctcattctcttgttaaaaatatttggtgagaacagtctgttagttcttttaatt cctcattctatctatctatctatctatcaatatgtgtatgtgtgtgtagatacatatataatatata cctacacacatgtgttgtaaatgatatctgt
Seq with 15 repeats	203 bp	Aggaacaaagtctcattctcttgttaaaaatatttggtgagaacagtctgttagttcttttaatt cctcattctatctatctatctatctatctatctatctatctatctatctatctatctatcaa tatgtgtatgtgtgtgtgtagatacatatataatatataacctacacacatgtgttgtaaatgatatct

(CTAT) 163bp	(CTAT) 203 bp

Fig 5.32 : MFOLD analysis of the hypothetical transcripts having expected number of repeats in D2S439. The minimum (163bp) and the maximum (203bp) alleles are compared in terms of structural loops formed on folding. Numbered boxes represent loop positions

Chapter 5

5.5.4 *In-silico* analysis D3S2398

The seed sequence of the marker was obtained from UniSTS and hypothetical sequences were created by inserting the expected number of repeats in the seed sequence. The table below summarizes the sequences obtained using online database and the hypothetical transcripts which were further analyzed for their using the online available tools.

Table 5.47: D3S2398 Sequences analyzed for the folding

Category	Size	Sequence
Seed Sequence	282	Agcctgagcaaaacagtgaaaccccatctctactataaatagaaaaattagctggtcgtgggggcatgcgc ctntagtcccanctactcaggaggctgaggcaggaggagactnggaggtggagattgtagtgagccaag atcgtgccactgcactccagcttgggtgacagagtaagaccctntctaaaatagatagatagatagatagata gatagatagatagatagatagatactatccctgatgctattttctaggggacccaatccgtaggtcttc
Sequence with 06 repeats	262	Agcctgagcaaaacagtgaaaccccatctctactataaatagaaaaattagctggtcgtgggggcatgcgc ctntagtcccanctactcaggaggctgaggcaggaggagactnggaggtggagattgtagtgagccaag atcgtgccactgcactccagcttgggtgacagagtaagaccctntctaaaatagatagatagatagatagata gatactatccctgatgctattttctaggggacccaatccgtaggtcttc
Sequenec with 15 repeats	298	Agcctgagcaaaacagtgaaaccccatctctactataaatagaaaaattagctggtcgtgggggcatgcgc ctntagtcccanctactcaggaggctgaggcaggaggagactnggaggtggagattgtagtgagccaag atcgtgccactgcactccagcttgggtgacagagtaagaccctntctaaaatagatagatagatagatagata gatagatagatagatagatagatagatagatagatactatccctgatgctattttctaggggacccaatc cgtaggtcttc
(GATA) 262bp		**(GATA) 298bp**

Fig 5.33: MFOLD analysis of the hypothetical transcripts having expected number of repeats in D3S2398. The minimum (262bp) and the maximum (298bp) alleles are compared in terms of structural loops formed on folding. The numbered boxes represent the loop positions

Chapter 5

5.5.5 *In-silico* analysis D3S3609

The seed sequence of the marker was obtained from UniSTS and hypothetical sequences were created by inserting the expected number of repeats in the seed sequence. The table below summarizes the sequences obtained using online database and the hypothetical transcripts which were further analyzed for their using the online available tools.

Table 5.48: D3S3609 Sequences analyzed for the folding

Category	Size	Sequence
Seed Sequence	171	Agctgggggaccagtctgncctccgtttacaaganttcctancacacacacacacacacacac acacacacacacacacacacaccccctcgtaggcaggcactctttcctaccacacaccagt gnccaggagggncttatgtctgtccttccaccgtacaagttactctcg
18 repeats	163	Agctgggggaccagtctgncctccgtttacaaganttcctancacacacacacacacacacac acacacacacacacaccccctcgtaggcaggcactctttcctaccacacaccagtgnccagga ggncttatgtctgtccttccaccgtacaagttactctcg
30 repeats	187	Agctgggggaccagtctgncctccgtttacaaganttcctancacacacacacacacacacac acacacacacacacacacacacacacacaccccctcgtaggcaggcactcttt cctaccacacaccagtgnccaggagggncttatgtctgtccttccaccgtacaagttactctcg

| (GATA) 163 bp | (GATA) 187 bp |

Fig 5.34: MFOLD analysis of the hypothetical transcripts having expected number of repeats in D3S3609. The minimum (163bp) and the maximum (187bp) alleles are compared in terms of structural loops formed on folding. The numbered boxes represent the loops positions.

Chapter 5

5.5.6 *In-silico* analysis D10S1237 (GATA repeat)

The seed sequence of the marker was obtained from UniSTS and hypothetical sequences were created by inserting the expected number of repeats in the seed sequence. The table below summarizes the sequences obtained using online database and the hypothetical transcripts were further analyzed for their using the online tools.

Table 5.49 D10S1237 Sequences analyzed for the folding

Category	Size	Sequence
NCBI Sequence	376	Cttgtcctgcctttggactaaaaagttggctctccttgggtctcaaacctgcaggctttcagacnggaactctnatngg ntntnctgntactcagacnttcagacnnaaactaganctacacnatnnactctntnaggtttcnagcttgcnagcca cagatccnaggacttctnagcctcnataatcntgtgggtnattncctnntangnaatctctncccaaataggnagata cctnggtatctcantatctagctngatagatagatagatagatagatagatagatagatagatagatagatagccggn acaancaganaaancagaccnaccnacnggnncngnnccncnggagaacctgactaggacagag
12 repeats	376	Cttgtcctgcctttggactaaaaagttggctctccttgggtctcaaacctgcaggctttcagacnggaactctnatngg ntntnctgntactcagacnttcagacnnaaactaganctacacnatnnactctntnaggtttcnagcttgcnagcca cagatccnaggacttctnagcctcnataatcntgtgggtnattncctnntangnaatctctncccaaataggnagata cctnggtatctcantatctagctngatagatagatagatagatagatagatagatagatagatagatagatagccggn acaancaganaaancagaccnaccnacnggnncngnnccncnggagaacctgactaggacagag
26 repeats	432	Cttgtcctgcctttggactaaaaagttggctctccttgggtctcaaacctgcaggctttcagacnggaactctnatngg ntntnctgntactcagacnttcagacnnaaactaganctacacnatnnactctntnaggtttcnagcttgcnagcca cagatccnaggacttctnagcctcnataatcntgtgggtnattncctnntangnaatctctncccaaataggnagata cctnggtatctcantatctagctngatagatagatagatagatagatagatagatagatagatagatagatagatagat agatagatagatagatagatagatagatagatagatagatagatagatagccggnacaancaganaaancagacc naccnacnggnncngnnccncnggagaacctgactaggacagag

(GATA) 376bp	(GATA) 432bp

Fig 5.35: MFOLD analysis of the hypothetical transcripts having expected number of repeats in D10S1237. The minimum (367bp) and the maximum (432bp) alleles are compared in terms of structural loops formed on folding. The numbered boxes represent the positions

Chapter 5

5.5.7 *In-silico* analysis D10S521 (GATA repeat)

The seed sequence of the marker was obtained from UniSTS and hypothetical sequences were created by inserting the expected number of repeats in the seed sequence. The table below summarizes the sequences obtained using online database and the hypothetical transcripts which were further analyzed for their using the online available tools.

Table 5.49 D10S521 Sequences analyzed for the folding

Category	Size	Sequence
NCBI Sequence	187 bp	Ctccagagaaaacagaccaataagatatgtctctctgtctccctctataatatggttagatagat agatagatagatagatagatagatagatagatagatagatagatagatagatagatagatagat agatagatagatagatagatgatttcaagtaacttgcctcagttgattgatggtagg
15 repeats	155	Ctccagagaaaacagaccaataagatatgtctctctgtctccctctataatatggttagatagat agatagatagatagatagatagatagatagatagatagatagatagatagatagatttcaagta acttgcctcagttgattgatggtagg
24 repeats	191	Ctccagagaaaacagaccaataagatatgtctctctgtctccctctataatatggttagatagat agatagatagatagatagatagatagatagatagatagatagatagatagatagatagatagat agatagatagatagatagatagatttcaagtaacttgcctcagttgattgatggtagg

(GATA) 155 bp	(GATA) 191 bp

Fig 5.36: MFOLD analysis of the hypothetical transcripts having expected number of repeats in D10S521. The minimum (262bp) and the maximum (298bp) alleles are compared in terms of structural loops formed on folding. The numbered boxes represent the loop positions

Chapter 5

Table 5.50: Summary of MFOLD results

S.NO	Marker	Folding	Remark	Table	Fig
1	D2S1384	No change	Only center loop increased in size	5.44	5.30
2	D2S2944	Yes	Addition of a single loop	5.45	5.31
3	D2S439	Yes	Loss of a loop at one end	5.46	5.32
4	D3S2398	Yes	Only end length increased	5.47	5.33
5	D3S3609	No	Only center loop size increased	5.48	5.34
6	D10S1237	Yes	Complete structure remodeled	5.49	5.35
7	D10S521	Yes	Repositioning of loops	5.50	5.36

Studies have shown that microsatellite motifs in the UTR form structural elements (stem-loops) and contribute to mRNA regulation. Our prediction using the MFOLD program for these seven markers have shown either a addition of a single loop or loss of a loop or complete structure remodeling or repositioning of stem-loop structures formed by sequences containing the microsatellite motifs but also favorable free energy level in sequences with increased number of repeats as compared with sequences minimum repeats (Table 5.51). Probably the change in the folding pattern of the sequence with the repeats may be responsible for the change in the interactions at mRNA levels. However this hypothesis still needs to be validated in a much larger cohort and with a bigger panel of STR markers.

Table 5.52: Summary of the Studied STR Markers and their association with Anthropometric markers

S.No.	Marker	Allele	T2DM	SBP	DBP	BMI	WHR
1	D2S1384	≥145bp	OR= 1.82, p=0.0001	t=2.595 p=0.01	NS	t = 6.08 p< 0.0001	t = 4.250 p<0.0001
		<145 bp		t = 2.455 p = 0.01	NS	NS	NS
2	D2S2944	≥124bp	OR= 1.34,p =0.0268	t = 4.588 p =0.0001	t = 3.132, p =0.0022	t = 4.61 p< 0.0001	t = 5.302 p<0.0001
		<124 bp		t=2.08 p = 0.0379	NS	t = 6.59 p< 0.0001	t = 3.25 p<0.0013
3	D2S439	≥195bp	OR=1.95,p<0.0001	t = 1.982 p =0.04	NS	t = 4.65 p< 0.0001	t = 6.003 p<0.0001
		<195bp		t = 4.110, p = 0.0001	t =3.279, p = 0.0012	t = 6.362 p< 0.0001	t = 5.208 p<0.0001
4	D3S3609	<171bp	OR=2.39,p< 0.0001	NS	NS	NS	NS
		≥171bp		t = 2.646, p =0.008	NS	NS	t = 5.525 p<0.0001
5	D3S2398	<286bp	OR =2.17,p<0.0001	NS	NS	NS	t = 2.865 P=0.0053
		≥286bp		NS	NS	t = 2.406 P = 0.0183	t = 5.970 P<0.0001
6	D10S1237	≥179bp	OR= 5.66,p<0.0001	NS	t =2.149, p = 0.03	t = 3.278 p =0.0012	NS
		<179bp		t = 2.327 p = 0.02	NS	t = 2.271 p =0.02	t = 4.424 p<0.0001
7	D10S521	≥416bp	OR = 3.01,p<0.0001	t = 2.587 p = 0.01	NS	t = 3.455 p =0.0008	t = 5.587 p<0.0001
		<416bp		NS	t =2.099 p = 0.03	t = 2.293 p =0.02	NS

BMI:Body Mass Index; WHR: Waist Hip Ratio; SBP/DBP: Systolic Blood Pressure/Diastolic Blood Pressure; NS: Non Significant

Table 5.53: Summary of the Studied STR Markers and their association with the Biochemical variables

S.No.	Marker	Allele	T2DM	FBG	TC	TG	HDL	LDL	VLDL
1	D2S1384	≥145bp	OR= 1.82, p=0.0001	t = 13.553 p< 0.0001	t = 1.770 p = 0.0788	t = 3.722 p=0.0003	t = 1.923 p = 0.0564		t = 3.721 p = .0003
		<145 bp		t = 12.407 p< 0.0001	NS	t = 2.803 p =0.006	NS	NS	
2	D2S2944	≥124bp	OR= 1.34,p=0.0268	t = 10.091 p< 0.0001	NS	t = 2.720 p=0.0077	NS	NS	t = 2.729 p=0.0075
		<124 bp		t = 14.768 p< 0.0001	t = 4.281 p<0.0001	t = 3.673 p=0.0003	t=2.923 p = 0.0039	NS	t = 3.603, p=0.0004
3	D2S439	≥195bp	OR=1.95,p<0.0001	t = 8.780 p< 0.0001	NS	t= 2.165 p =0.03	t =3.206 p =0.0018	NS	t = 2.191 p = 0.03
		<195bp		t = 13.493 p< 0.0001	t = 3.004 p=0.0031	t = 2.667 p = 0.008	t=2.717 p = 0.0073	NS	t = 2.555, p=0.0116
4	D3S3609	<171bp	OR=2.39,p< 0.0001	t = 3.419 p 0.0016	NS	NS	NS	NS	NS
		≥171bp		t = 16.630 p< 0.0001	NS	NS	NS	NS	NS
5	D3S2398	<286bp	OR =2.17,p<0.0001	t = 2.583 p = 0.011	t = 9.882 p< .0001	NS	NS	NS	NS
		≥286bp		NS	t = 15.134 p< 0.0001	NS	NS	NS	NS
6	D10S1237	≥179bp	OR= 5.66,p<0.0001	t = 16.054 p< 0.0001	t = 2.485 p =0.013	NS	t=2.261 p = 0.0271	NS	NS
		<179bp		t = 12.596 p<0.0001	NS	t = 2.027 p = 0.046	t=2.971 p = 0.03	NS	t =2.047 p= 0.04
7	D10S521	≥416bp	OR = 3.01,p<0.0001	t = 9.869 p< 0.0001	t = 3.361 p =0.0011	NS	t=3.699, p = 0.004	NS	NS
		<416bp		t = 18.171 p< 0.0001	NS	NS	t=2.177, p = 0.03	NS	NS
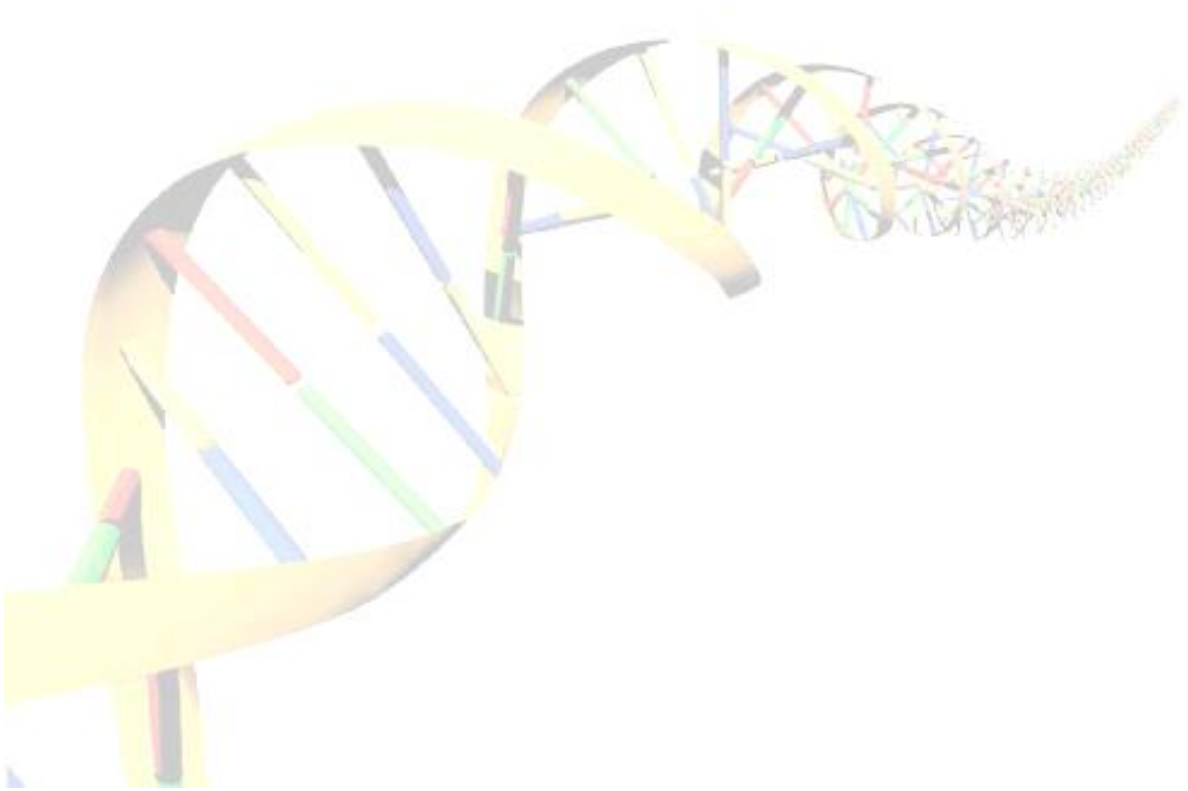
Discussion
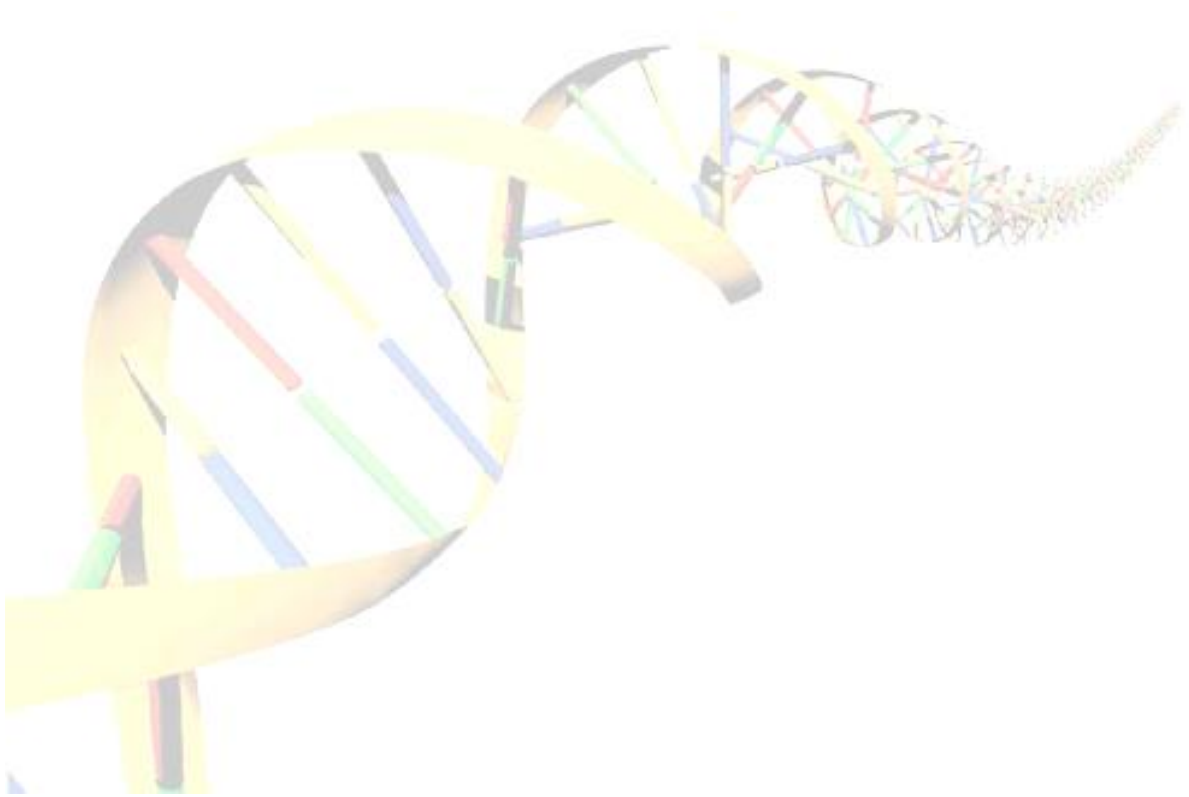
Chapter 6

Discussion

Chapter 6

Chapter 6

Discussion

The present study involves genotyping of SSLP markers in the region repeatedly found to host the disease haplotype for diabetes in several studies. The study assessed the association of the alleles/genotypes of each of the marker in a case control design, where the study cohort included unrelated diagnosed Type 2 Diabetes Mellitus (T2DM) cases and healthy volunteers with no history of diabetes and associated comorbidities. The relationship between alleles/genotypes of STR marker with anthropometric and biochemical /clinical markers was also accessed among cases and controls.

Obesity is the major risk factor for the development of pre-diabetes and type 2 diabetes mellitus. BMI is widely used as a surrogate measure of overall obesity. In the present study the cases and controls differ significantly for BMI. The overall obesity as estimated by BMI revealed that the mean BMI was higher among the cases as compared to the controls. Several studies investigating the association between obesity and diabetes incidence have been carried out in various groups including populations of European, Asian, and Indian descent, as well as among Mexican- and Japanese-Americans (Wang *et al.*, 1997; Bonora *et al.*, 2004; Bergstrom *et al.*, 1990; Hafner *et al.*, 1991; Lundgren *et al.*, 1989, Stolk *et al.*, 1993). The significance between BMI and T2DM was not lost even when segregated for gender. The results of present study were consistent with global reported associations between body mass index (BMI; in kg/m2) and type 2 diabetes in men (Hanson *et al.*, 1995; Wannamethee *et al.*, 2005) and women (Haffner *et al.*, 1990; Resnick *et al.*, 2000).

Abdominal obesity, measured by an elevated waist to hip ratio (WHR), is shown to be a strong risk factor for type 2 diabetes mellitus (Lahti-Koski *et al.*, 2000). The diabetics

Chapter 6

showed higher mean abdominal obesity, as reflected by altered WHR (marker for abdominal obesity), as compared to controls in this study. Prospective studies also support the association of various anthropometric indices of abdominal adiposity and the future development of diabetes (Ohlson *et al.*, 1985; Karter *et al.*, 1996). It has been suggested that abdominal adiposity is an independent predictor of alterations in the plasma lipid, lipoprotein and plasma glucose concentrations (Parikh *et al.*, 2002). The high abdominal obesity is also a well-known component of Insulin resistance, a key feature in T2DM. Insulin resistance and hyperinsulinemia are associated with lipoprotein lipase (LPL) deficiency, which causes elevation in the levels of free fatty acids and a reduction in high density lipoprotein cholesterol (HDL-C) levels. These elevated levels of free fatty acids may induce insulin resistance in the peripheral tissues and liver (Ohlson *et al.*, 1985; Karter *et al.*, 1996). Insulin resistance eventually produces sufficient glucose intolerance to result in frank diabetes (Parikh *et al.*, 2002). The results from the current study reemphasize the same association and reflect that both BMI and WHR are independent risk factors for T2DM.

The cases also had significantly higher mean values of systolic blood pressure (BP) as compared to controls, although the difference in mean diastolic blood pressure was found to be higher but non-significant. This significance was not lost even on segregation for gender. Infact it was evident from the data that the female participants had higher systolic blood pressure as compared to male participants. Elevated BP is a known risk factor for diabetes complications, and high BP before the onset of diabetes may also explain the observed high prevalence of cardiovascular disease at the time of diabetes diagnosis. The presence of hypertension may also be an indicator of the pathogenesis of type 2 diabetes. According to the "Common Soil Hypothesis" (Stern,

Chapter 6

1995), elevated BP could be an early sign of underlying insulin resistance, related to central adiposity. An alternative hypothesis is that elevated BP is a marker of endothelial dysfunction, which is itself a risk factor for the development of insulin resistance, type 2 diabetes, and vascular disease (Golden *et al.*, 2003). Our result is in concordance to the reported baseline blood pressure data from several recent trials indicating that, in diabetic subjects, there is nearly a fourfold excess in systolic over diastolic pressure with respect to the recommended systolic/diastolic target pressure of <130/80 mmHg. Additionally, systolic pressure was 2–3 mmHg higher and diastolic pressure was 1–3 mmHg lower in diabetic hypertensive than in non-diabetic hypertensive individuals, which adds ~4 mmHg to pulse pressure and also to the difference between the excess systolic and excess diastolic pressure (Osher *et al.*, 2008). The study results also show that cases had a significantly higher prevalence of hypertension as compared to the healthy control population, though it seems from the data that probably the population itself had a higher prevalence of hypertension. The fact that diabetes and high blood pressure tend to occur together could be due to the fact that both diabetes and hypertension seems to share certain common physiological traits like increased fluid volume, increased arterial stiffness and impaired insulin handling. Though these common biological traits only partially explain why diabetes and high blood pressure are such a common pair, in many cases, the two diseases are likely to occur together simply because they share a common set of risk factors. The prevalence of coexistent hypertension and diabetes varies across different ethnic, racial, and social groups. Importantly, hypertension in patients with diabetes causes a significant increase in the risk of vascular complications in this population, and together both conditions predispose to chronic kidney disease (Lago *et al.*, 2007). The overlap between hypertension and diabetes substantially increases the risk of ischemic

cerebrovascular disease, retinopathy, and sexual dysfunction. Diabetes mellitus is an independent risk factor for coronary artery disease, and the risk is markedly increased when hypertension is present. Also another alarming observation in the study was both hypertension and prehypertension were present at a very high rate in the general population recruited as control group in the study. This observation is also supported by the other reports across the globe showing an alarming increase in the prevalence of hypertension among the general population in the developing countries (Saeed *et al.*, 2011)

Further it was seen that the cases and controls differed significantly for mean FBG, which is an expected observation since FBG was one of the diagnostic criteria for segregating cases and controls. Hyperglycemia is a major cause of complications for individuals with diabetes. Chronic hyperglycemia at levels more than slightly above normal can produce a wide variety of serious complications over a period of time, including kidney damage, neurological damage, cardiovascular damage, damage to the retina or damage to feet and legs. Diabetic neuropathy may be a result of long-term hyperglycemia. In Type 2 Diabetes mellitus, hyperglycemia is usually caused by resistance to insulin at the cellular level. Low insulin levels and/or insulin resistance prevent the body from converting glucose into glycogen (a starch-like source of energy stored mostly in the liver), which in turn makes it difficult or impossible to remove excess glucose from the blood. With normal glucose levels, the total amount of glucose in the blood at any given moment is only enough to provide energy to the body for 20-30 minutes, and so glucose levels must be precisely maintained by the body's internal control mechanisms. When the mechanisms fail in a way that allows glucose to rise to abnormal levels, hyperglycemia is the result. It was also observed that females had significantly higher levels of mean FBG as compared to the male participants. Both the

Chapter 6

observations (Hyperglycemia in cases and FBG levels higher in females) are in line with the previously reported studies across the globe (Khan, 2006). Though there are mixed observations for FBG being higher in female diabetics (Khan, 2007; Joo 2007, Khan, 2006).

The cases also had significantly higher mean for lipid markers. The presence of dyslipidemia i.e. raised total cholesterol (TC) and decreased high density lipoproteins (HDL) is an established phenomenon in type 2 diabetes mellitus. Its relationship to the fasting blood sugar (FBG) level in both diabetics and non-diabetics has yet to be established. The characteristic features of diabetic dyslipidemia are a high plasma triglyceride concentration, low HDL cholesterol concentration and increased concentration of small dense LDL-cholesterol particles (Mooradian *et al.*, 2009). The present study also had the classic symptoms of diabetic dyslipidemia including significantly altered TG and HDL. The lipid changes associated with diabetes mellitus are attributed to increased free fatty acid flux secondary to insulin resistance. It is well known that defects in insulin action and hyperglycemia could lead to changes in plasma lipoproteins in patients with diabetes. Alternatively, especially in the case of type 2 diabetes, the obesity/insulin-resistant metabolic disarray that is at the root of this form of diabetes could, itself, lead to lipid abnormalities exclusive of hyperglycemia (Goldberg *et al.*, 2001). Infact several factors are likely to be responsible for diabetic dyslipidemia are: insulin effects on liver apoprotein production, regulation of lipoprotein lipase (LpL), actions of cholesteryl ester transfer protein (CETP), and peripheral actions of insulin on adipose and muscle.

Thus the study results evaluating the association of these endo-phenotypes with T2DM are in line with the previously reported observations worldwide. Majority of Indian type 2 diabetic patients recruited in our study seems to be dyslipidemic at baseline. The

Chapter 6

most common pattern of dyslipidemia is high LDL and low HDL among both males and females. Investigations from several groups have shown that Asian Indians are predisposed to develop type 2 diabetes, proatherogenic metabolic abnormalities (metabolic syndrome, insulin resistance syndrome) and CHD (Misra *et. al.*, 2004)

Case control association analysis

The case control association analysis in this study is twofold. First, we were able to establish the association of seven identified STR repeats in the three chromosomal loci with T2DM in Indians from Rajasthan. Second the study also looks at the association of these STRs with various endophenotypes reported to be associated with T2DM. Despite our relatively small cohort (as compared to the recent GWAS), we were able to establish in our population the baseline frequency of the studies STRs and also the association of these marker with T2DM.

Case control association analysis with alleles of D2S1384

We investigated the association between variations of the $(CTAT)_n$ repeats of microsatellite marker D2S1384 in T2DM cases and healthy controls. D2S1384 is flanked by gene coding inducible T-cell co-stimulator precursor (*ICOS*) at 5' side (402565 bp) and par-3 partitioning defective 3 homolog B Isoform b (*PARD3b*) at 3' side (183402 bp) in human build-37.1 of NCBI Map viewer (Fig 6.1).

Chapter 6

AFMa100zh5	gbsts	196.85	248.60	145.45	0.81	· · · · ● · ●	Info
AFMc013yd5	gbsts Gene	196.85	248.60	145.45	0.67	· · · · · · ●	Info
AFM318wf1	gbsts	196.85	248.60	145.45	0.63	· · · · · · ●	Info
AFM135xf12	gbsts	196.85	248.60	145.45	0.63	· · · · · · ·	Info
AFM234zh6	gbsts	196.85	248.60	145.45	0.81	· · · · · · ●	Info
AFMa289xc1	gbsts	196.85	248.60	145.45	0.20	· · · · · · ·	Info
GATA149B10	gbsts	196.85	248.60	145.45	0.66	· · · · · · ·	Info
Mfd36	gbsts	198.65	249.33	148.01	0.79	· · · · · ●	Info
AFM269zf9	gbsts Gene	198.65	248.99	148.01	0.61	· · · · · ●	Info
AFM234vg5	gbsts Gene	198.65	249.69	148.01	0.76	· · · · · ●	Info
AFM289vf5	gbsts Gene	198.65	249.69	148.01	0.57	· · · · · ●	Info
AFM064xh7	gbsts	198.65	248.68	148.01	0.75	· ● · · ●●	Info
GATA161E02	gbsts	198.65	249.69	148.01	0.70	· · · · · · ·	Info
AFMa203zf1	gbsts	199.18	250.76	148.01	0.63	· · · · · · ●	Info
AFM074xg9	gbsts	199.18	250.76	148.01	0.58	· ●● · ●●●	Info
GATA52A04	gbsts	200.43	253.29	148.01	0.79	· · ● · · ●●	Info
AFM304tb5	gbsts Gene	202.92	258.34	148.01	0.81	· ●● · · ●●	Info
AFMa050ya5	gbsts Gene	203.46	259.41	148.01	0.83	· · · · ● · ●	Info
AFM296xb9	gbsts Gene	204.53	261.55	148.01	0.74	· · · · · ●●	Info
AFM266vc5	gbsts Gene	204.53	261.55	148.01	0.81	· · · · · · ●	Info
UT430	gbsts Gene	204.53	261.55	148.01	0.58	· · · · · · ·	Info
AFMa351zd1	gbsts	205.06	262.62	148.01	0.86	· · · · · · ·	Info
AFMb297ya9	gbsts	205.06	262.62	148.01	0.63	· · · · · · ·	Info
ATC3E01	gbsts Gene	205.06	262.62	148.01	0.37	· · · · · · ●	Info
AFMc009wh1	gbsts	205.06	262.62	148.01	0.76	· · · · · · ·	Info
AFMa246xc5	gbsts Gene	205.06	262.62	148.01	0.88	● · ●●● · ●	Info
AFM157xg9	gbsts Gene	205.59	262.62	149.08	0.60	· · · · · · ●	Info
AFM212ze9	gbsts Gene	206.13	262.62	150.15	0.75	· · · · · · ●	Info
GATA29A06	gbsts Gene	206.13	262.62	150.15	0.72	· · ● · ●●●	Info
AFM262xe5	gbsts Gene	206.13	262.62	150.15	0.49	· · · · · · ·	Info
AFM210yc11	gbsts Gene	206.13	262.62	150.15	0.67	· ● · · · ●	Info
GATA52C06	gbsts	206.74	263.69	150.15	0.60	· · · · · · ●	Info

Fig 6.1 Cytogenetic position of D2S1384 marker
The ruler in the partial map of chromosome 2 shows the relative nucleotide position expressed in megabases (205Mb-209 Mb) and all the markers distances are according to NCBI: Mapviewer

In the present study significant difference in allele frequencies of D2S1384 was observed among cases verses controls. It was seen that the presence of the shorter base pair allele seems to protect the individuals from the disease. The longer base pair allele ($\geq$141bp) on the other hand was found to increase the risk for T2DM by a nominal significance (~2 fold). Also the presence of atleast single copy of 149 bp allele seems to increase the risk for the disease by nearly 3 folds.   When distribution of mean values of anthropometric and blood parameters were compared between T2DM cases and controls, mean values of both BMI and WHR differed in presence of one or both copies

of longer allele. It was observed that in the presence of shorter base pair allele the case and control had almost similar mean BMI and WHR levels as opposite to the longer base pair allele where in cases had significantly higher mean BMI and WHR levels. When the same allele was evaluated for the association of short and long allele with SBP it was seen that carriers of both long and short allele had significantly higher mean SBP in cases as compared to the compared controls. No such observation was seen in case of DBP, since the mean DBP levels were almost comparable among cases and controls. The observed association of this marker with the BMI and WHR levels was counter verified by literature where in reports have identified this locus (2q33.3) to harbor QTL for BMI (Daley *et al.*, 2003; Larkin *et al.,* 2004; Meulenbelt *et al.,* 2006; Ruchat *et al.,* 2008). The finding clearly indicates the importance of 2q33.3 loci in regulating BMI. D2S1384 seems to be a crucial STR marker when looking at the genetic contribution of BMI to occurrence of T2DM or metabolic syndrome. Studies have shown that microsatellite motifs in the UTR form structural elements (stem-loops) and contribute to mRNA regulation (Chen *et al*., 2007). However our prediction using the MFOLD program for D2S1384 though did not showed any stem loop formed by sequences containing the microsatellite motifs.

Case control association analysis with alleles of D2S2944

This region on chromosome 2 has been evaluated by Iwasaki *et al.* (2003) in the whole genome scan and they reported a marginal linkage near D2S2944 marker (LOD = 1.45) in an analysis of 164 Japanese families suffering from T2DM. In our study it was seen that the presence of longer base pair allele marginally imparts the risk for T2DM. It was also seen that the carriers of longer base pair allele seems to have significantly elevated mean DBP as compared to the controls with longer repeats. Also the group

having longer base pair alleles had higher difference in the mean SBP level as compared to those having shorter allele. In case of BMI the carrier for shorter allele had higher difference in the mean BMI among cases and controls as compared to the carriers of longer allele, whereas in case of WHR the difference in the mean WHR was higher in the longer allele carriers only. This marker after bimodal distribution into longer and shorter allele did not show any significance with total cholesterol, HDL and LDL. However the shorter base pair allele seems to be associated with the altered triglyceride levels. Fig 6.2 shows the cytogenic position of D2S2944.

Fig 6.2 Cytogenetic position of D2S2944 marker
The ruler in the partial map of chromosome 2 shows the relative nucleotide position expressed in megabases and the markers distances are according to NCBI: mapviewer

This marker was also evaluated using *in silico* tools to see the effect of the repeat expansion mutation on the folding and hence the probable regulation of the nearby genes. Infact it was interesting to see that our prediction using the MFOLD program for D2S2944 not only showed stem-loop structures formed by sequences containing the microsatellite motifs but also favorable free energy level in sequences with > (GATA)13 repeats as compared with sequences <(GATA)13 repeats. It can be envisaged that the stabilized structure of microsatellite repeat expansion could affect

and regulate transcription machinery. The observed association (marginal) with the T2DM or endophenotypes needs to be evaluated in terms of effect these repeat sequences have on the folding of the transcript and the resultant interaction with the neighboring genes/ Sequence(s)/sequence elements.

Case control association analysis with alleles of D2S439

We investigated the association between variations of the $(CTAT)_n$ tetra-nucleotide repeats of microsatellite marker D2S439 in our cohort. The marker according to human builds 27.1 provided by NCBI mapviewer is mapped to a non-annotated region as shown in fig 6.3. Based on the few currently existing reports 2q37 has been found to harbor the QTL's for visceral fat, waist hip ratio, percentage body fat, other related metabolic Syndrome (MetS) components (Rice *et al.*, 2002, Tang *et al.*, 2003) and obesity (Lee *et al.*, 1999). In the present study it was seen that the presence of the longer base pair allele seems to increase the risk for T2DM by nearly two fold. The carrier of longer base pair allele seems to have increased Waist hip ratio. In our cohort cases bearing the longer allele of D2S439 differs significantly in mean waist circumference from those bearing the shorter allele. Surprisingly majority of our cases possessing the longer allele display all four characteristics of proposed risk for Mets defined by IDF. Further the prediction using the MFOLD program for D2S2944 not only showed any extra stem-loop structures formed by sequences containing the microsatellite motifs but there was a clear loss of a stem loop in the longer base pair allele. This kind of loss may be associated with any of the gene regulation mechanism which still needs to be evaluated further

Marker	Type						
AFM127xb4	gbsts		225.67	293.72	156.55	0.60	
GATA23D03	gbsts Gene		227.00	293.72	160.82	0.79	
AFMb354we5	gbsts		227.00	293.72	160.82	0.84	
UT709	gbsts		227.00	293.72	160.82	0.43	
AFMa240xh1	gbsts Gene		227.00	293.72	160.82	0.75	
AFMa046we1	gbsts		227.54	293.72	161.88	0.80	
ATA24E12	gbsts Gene		228.01	293.74	162.44	0.57	
AFMa090zh5	gbsts Gene		228.61	294.78	162.95	0.52	
AFM218zg3	gbsts Gene		228.61	293.76	162.95	0.77	
AFMa140yg9	gbsts		229.14	294.78	164.02	0.61	
AFMa121za9	gbsts Gene		229.14	294.78	164.02	0.77	
GATA6E08	gbsts		231.27	296.92	166.15	0.54	
UT5193	sts		231.27	296.92	166.15	0.63	
AFMa196vd1	gbsts		231.27	296.92	166.15	0.52	
GATA43F06	gbsts		231.27	296.92	166.15	0.69	
AFM302vh9	gbsts		231.27	296.92	166.15	0.78	
UT506	gbsts Gene		231.27	296.92	166.15	0.40	
UT501	gbsts		231.27	296.92	166.15	0.64	
AFMa283yb1	gbsts Gene		231.81	296.92	167.25	0.75	
AFMa289ye1	gbsts		231.81	296.92	167.25	0.64	
AFMb329we9	gbsts Gene		232.36	296.92	168.34	0.87	
AFM277wg9	gbsts Gene		232.90	297.98	168.34	0.73	
AFM361ta5	gbsts Gene		232.90	297.98	169.05	0.83	
AFMc004xh5	gbsts Gene		233.62	297.98	169.44	0.81	
UT432	gbsts Gene		235.07	300.12	170.54	0.62	
AFM324vc9	gbsts Gene		235.07	300.12	170.54	0.79	
AFM248wc5	gbsts Gene		235.07	300.12	170.54	0.92	
AFMb298we9	gbsts Gene		236.70	302.26	171.63	0.37	
GATA12H10	gbsts Gene		236.70	302.26	171.63	0.76	

Fig 6.3 Cytogenetic position of D2S439 marker
The ruler in the partial map of chromosome 2 shows the relative nucleotide position expressed in megabases and all the markers distances are according to NCBI: Mapviewer

Chapter 6

Chromosome 2 and T2DM: Summary

Several studies have identified some adjacent 2q chromosomal regions to contain type 2 diabetes susceptibility loci and these regions are close to our 2q chromosomal region of interest. Importantly, Hanis et al. (1996) found strong evidence for linkage of type 2 diabetes on chromosome 2q37.3 near marker D2S125 (260 cM), called NIDDM1, in a Mexican-American population. Subsequently, Horikawa et al (2000) identified the calpain 10 (CAPN10) gene in the NIDDM1 region by positional cloning. Some potential replications of this finding have been reported in Caucasians also. Also, Li et al. (2004) found some evidence for linkage of fasting plasma glucose concentrations near the marker D2S434 (216 cM) region, which is not that far from our 2q chromosomal region of interest. Our region of interest on chromosome 2q overlaps with the region(s) reported to be linked to type 2 diabetes in other studies across various ethnicities, it should be noted that there is only a marginal linkage of type 2 diabetes with this region in our study also. Moreover the observed association seems to be more indirect (via anthropometric markers/endo-phenotypes) rather than being direct. To summarize Fig 6.4 shows the list of candidate genes which are associated with T2DM or associated risk factors and could be envisaged to be responsible for the marginal observed significance with T2DM. It is also interesting to note that the changes in the folding and hence the stem loop structure in certain markers seems to be big and this change in the folding pattern of the sequences still needs to be worked on to correlate the change in the folding (stem loop structure) with the risk for T2DM.
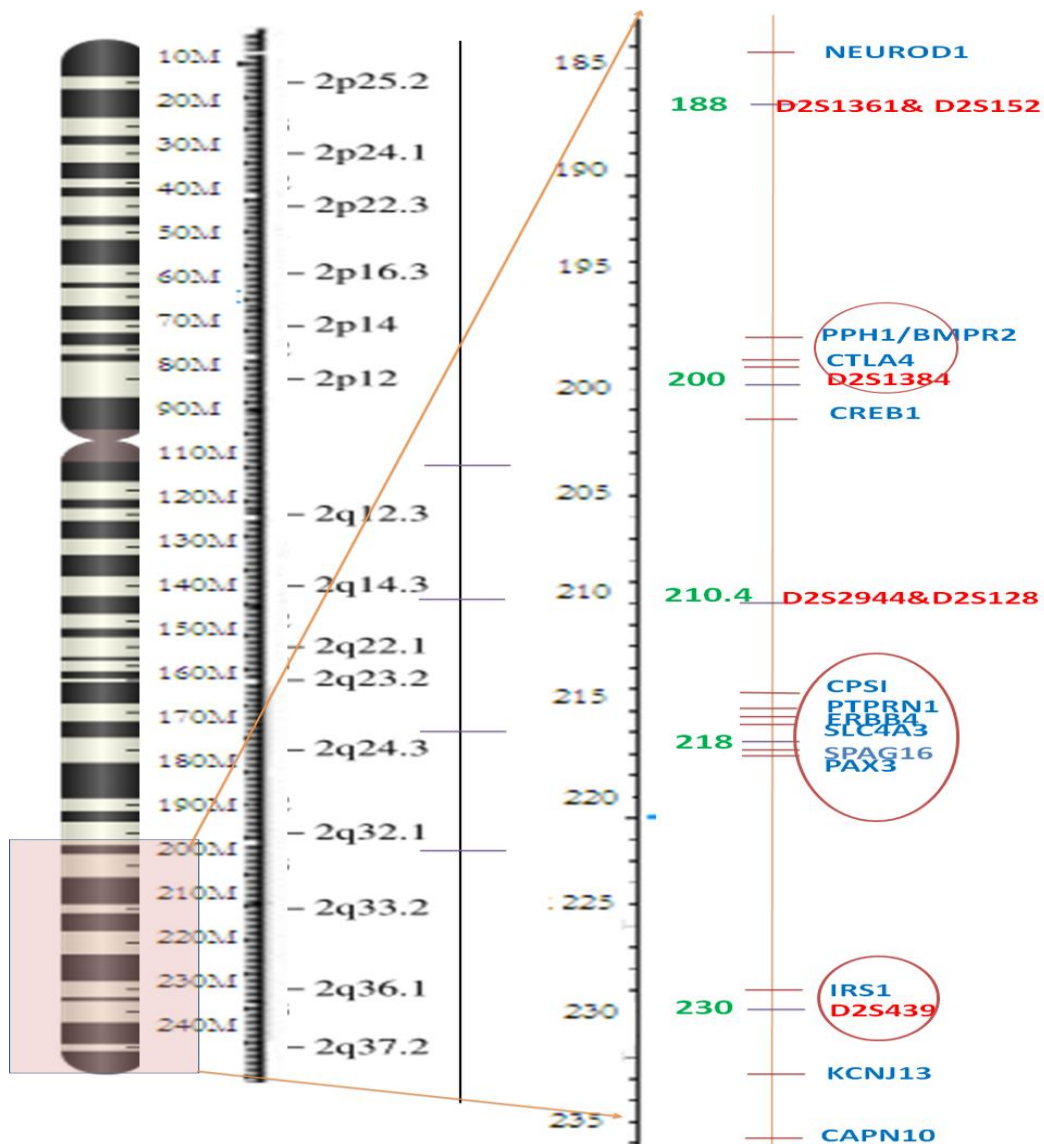
Fig 6.4 Studied STR markers and reported T2DM candidate genes on 2q32-27
All the markers distances are according to NCBI: Mapviewer. The positioning of the gene is done manually and is only a representation of the actual position.

Chapter 6

Case control association analysis with alleles of D3S3609

This region 3q22-29 is roughly 68M bp (130M-198M bp) in genetic distance which contains 692 genes according to NCBI (build 37.1). Genome-wide searches for this chromosomal loci associated with a modulation of serum adiponectin level have been performed in several populations. Kissebah *et al.* in 2000 showed the association of STR repeat D3S2398 with BMI ($p=0.016$ and OR=3.3), high cholesterol ($p=0.009$, OR=3.54), weight ($p=0.022$, OR=3.27) and insulin ($p=.032$, OR=3.01. However in our population this association could not be replicated. Infact the data clearly showed no significant difference among cases and controls in presence of absence of risk allele. Hence we could not establish any association (except for the risk for T2DM) of D3S3609 STR with the obesity markers.
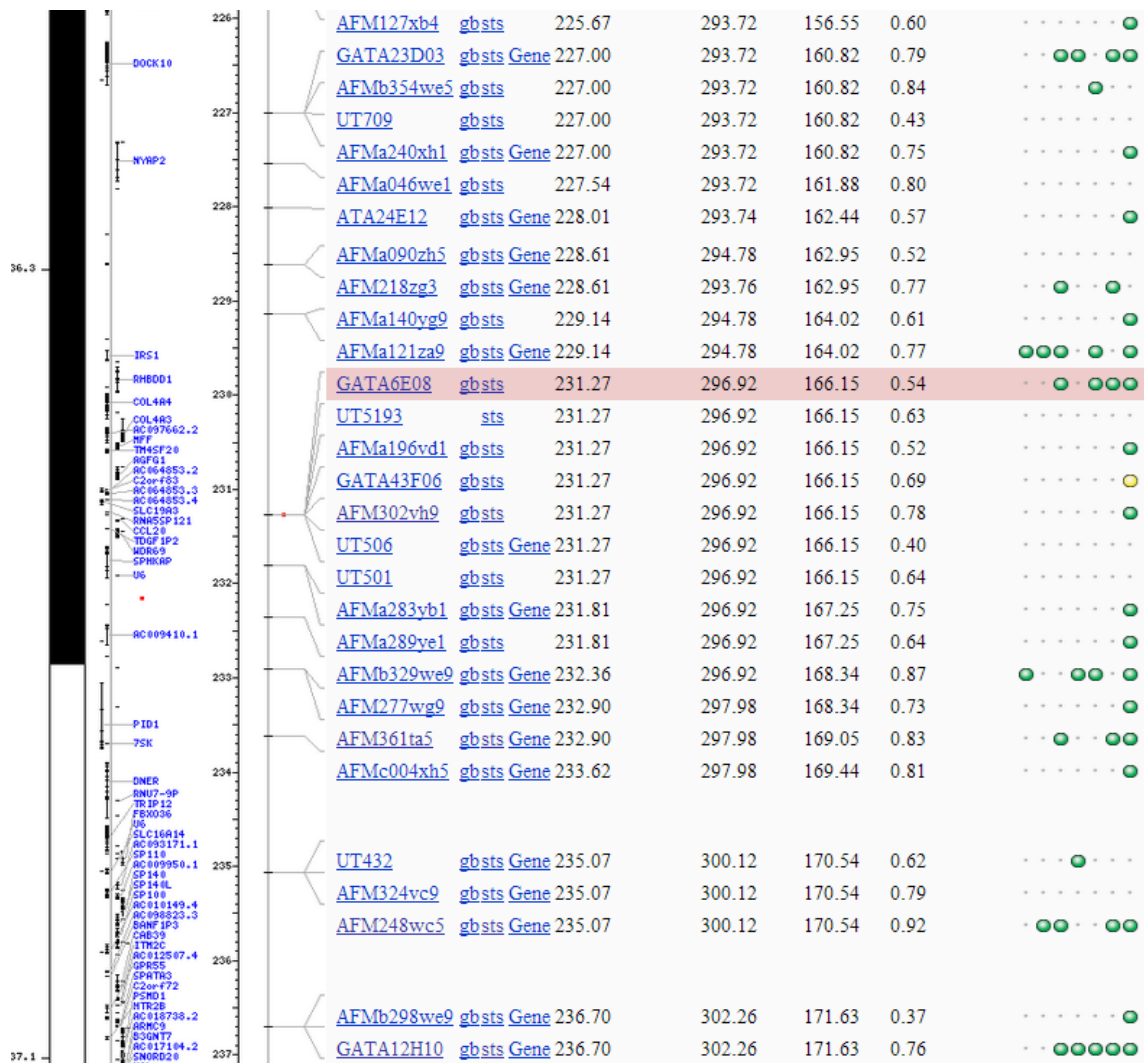

Fig 6.5 Cytogenetic position of D3S3609 marker
The ruler in the partial map of chromosome 3 shows the relative nucleotide position expressed in megabases and all the markers distances are according to NCBI:Mapviewer

Chapter 6

This region presents putative candidate genes, including GLUT2 (SLC2A2), apolipoprotein D (APOD), phosphatidylinositol 3-kinase catalytic p110α subunit (PI3KCA), and adiponectin (APM1 or GBP28). Two studies have reported a polymorphism in the apolipoprotein D locus to be associated with diabetes and obesity. Recently reports have shown that the polymorphisms in the APM1/adiponectin gene were associated with type 2 diabetes in both the Japanese and French populations. Our prediction using the MFOLD program for D3S3609 did not show any extra stem-loop structures formation by sequences containing the microsatellite motifs but there was only increase in the size of central loop.

Case control association analysis with alleles of D3S2398

We investigated the association between variations of the $(CTAT)_n$ tetra-nucleotide repeats of microsatellite marker D2S2398 in adult diabetics. The marker according to human build 27.1 provided by NCBI map viewer is mapped to a non-annotated region as shown in fig 6.3. In the present study it was seen that the presence of the longer base pair allele seems to increase the risk for T2DM by nearly two fold. The carrier of longer base pair allele seems to have increased Waist hip ratio. In our cohort cases bearing the longer allele of D2S2398 differs significantly in mean waist circumference from those bearing the shorter allele. Also there was significant difference observed in the mean BMI levels among carriers of longer base pair allele. As is discussed earlier the region seems to be strongly associated with the obesity parameters and hence the potential association seen with T2DM. The two potential candidate genes in this region are the solute carrier family 2 of the facilitated glucose transporter (GLUT2, at 3q26-q27); the catalytic a polypeptide of phosphoinositide 3-kinase (at 3q26.3); Kissebah *et al.* in 2000 proposed that this region encodes proteins thought to influence fat

partitioning, lipid homeostasis, and energy balance. These include the adipose tissue-secreted protein adiponectin (synonyms: AdipoQ; adipose most abundant gene transcript 1, or apM1; gelatin-binding protein of 28 kDa, or GBP28, at 3q27), the receptor protein known to bind to globular ''heads'' of the complement C1q (gC1qR, at 17p13.3), and the peroxisome proliferative-activated receptor a (at 17p12-p11.2). Our prediction using the MFOLD program for D3S2398 did not show any extra stem-loop structures formation by sequences containing the microsatellite motifs but there was only increase in the length of the motif at one end in the longer base pair allele.

Fig 6.6 Cytogenetic position of D3S2398 marker
The ruler in the partial map of chromosome 3 shows the relative nucleotide position expressed in megabases and all the markers distances are according to NCBI: Mapviewer

Case control association analysis with alleles of D10S521

Another promising region which was identified during our study was the region 10q25-26. The presence of the longer base pair allele of D10S521 in the studied cohort was significantly associated with T2DM. There was a highly significant association of this region seen with T2DM. Infact the presence of longer base pair allele seems to increase the risk for T2DM by nearly six folds. There was a very strong association seen

between the longer base pair allele and FBG. The presence of longer base pair allele seems to be significantly associated with the mean FBG level among cases. The longer base pair allele also seems to be associated with increased total cholesterol. This region seems to be strong and independently associated with T2DM. Our prediction using the MFOLD program for D10S521 showed a drastic change in the structure of the folded sequence of the longer base pair allele and this entire remodeled structure was energy stable. Such structures need to be evaluated in detail for their role in regulatory mechanism.
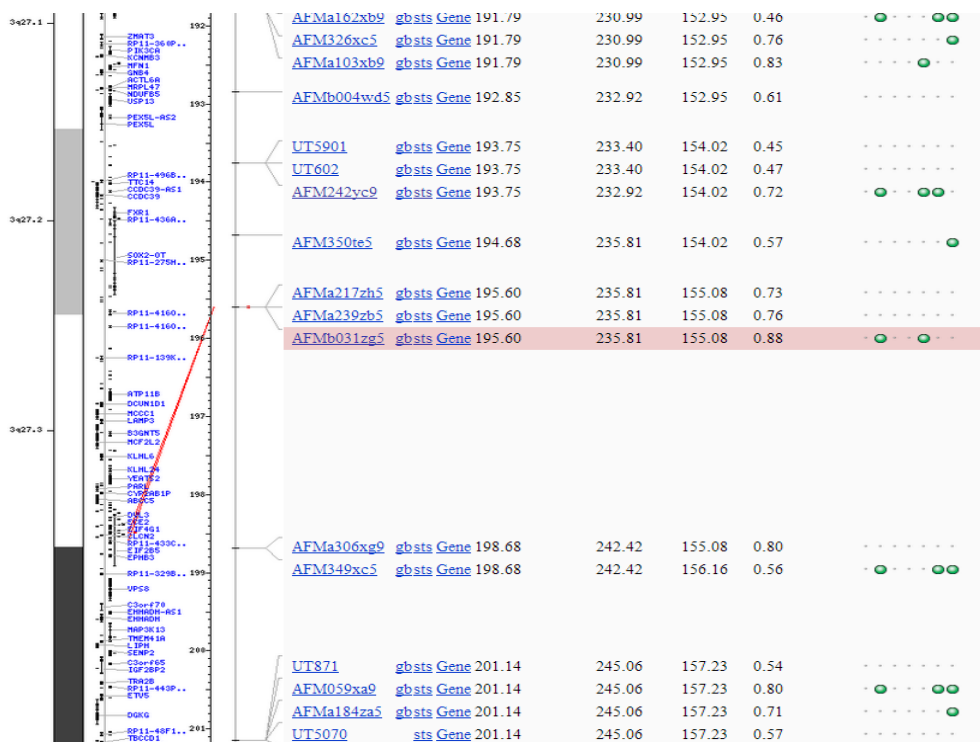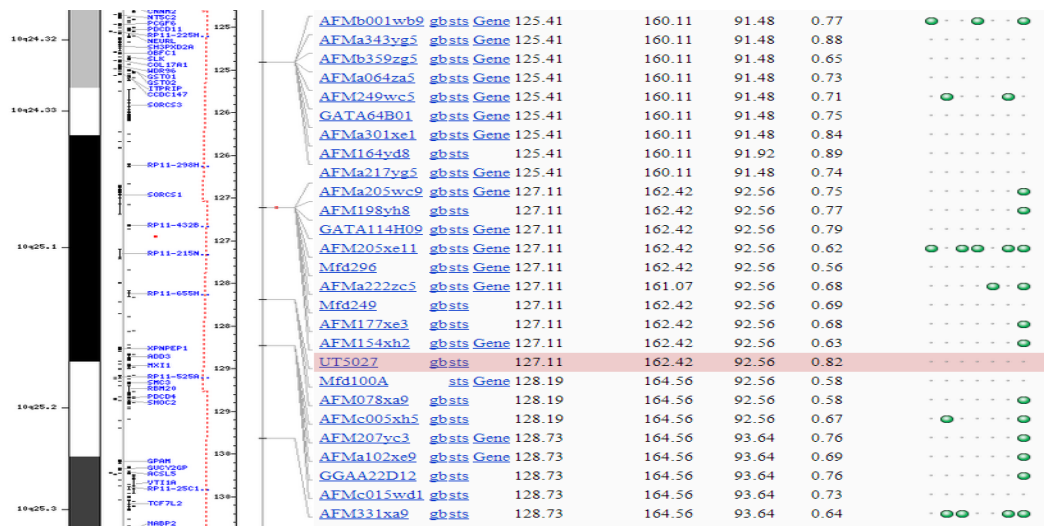

Fig 6.7 Cytogenetic position of D10S521 marker
The ruler in the partial map of chromosome 10 shows the relative nucleotide position expressed in megabases and all the markers distances are according to NCBI: Mapviewer

Case control association analysis with alleles of D10S1237

In this study we investigated the association between variations of the (CTAT)$_n$ tetra-nucleotide repeats of microsatellite marker D101237 in adult diabetics. Also the region seems to be associated with BMI and WHR. It was seen that the cases with the longer base pair allele had significantly higher BMI as compared to the cases with shorter base pair allele. A similar trend was seen with WHR also. The region also seems to have

marginal association with SBP and DBP. In case of biochemical parameters the association of the longer base allele could be only seen with FBG and the difference between cases and controls was lost in the cohort containing shorter base pair allele. Also it was seen that the longer allele harboring cohort seems to have significantly higher total cholesterol levels as compared to the shorter allele harboring cohort. This region also seems to be strong and independently associated with T2DM. Our prediction using the MFOLD program for D10S1237 showed a complete remodeling of the structure of the folded sequence of the longer base pair allele and this entire remodeled structure was energy stable.
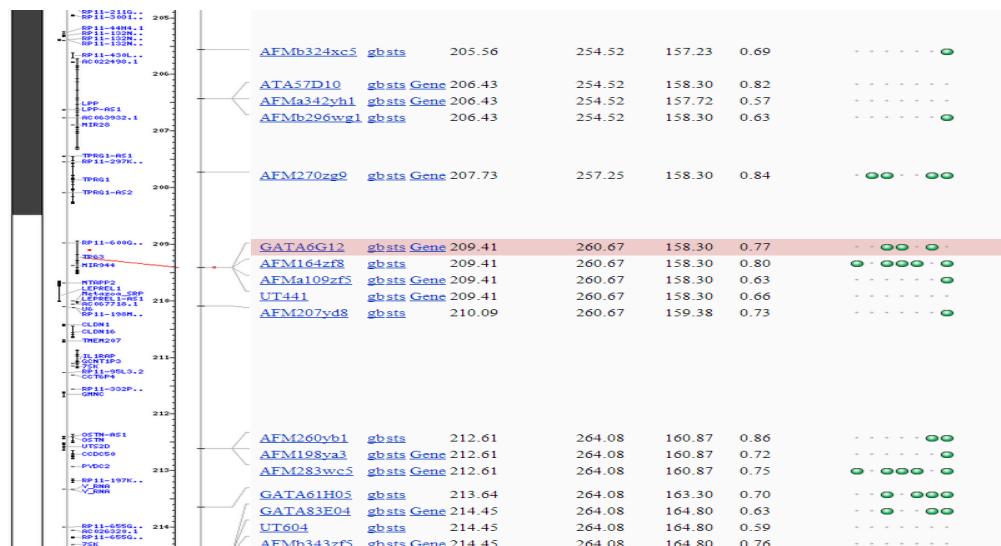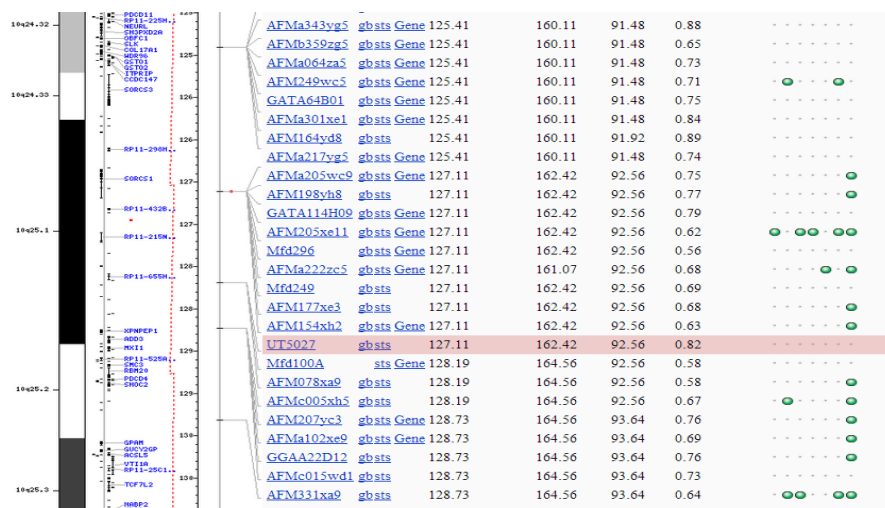

Fig 6.8 Cytogenetic position of D10S1237 marker
The ruler in the partial map of chromosome 10 shows the relative nucleotide position expressed in megabases and all the markers distances are according to NCBI: Mapviewer

Conclusions

The aim of this work was to reveal the genetic factors associated with Type 2 Diabetes mellitus and to look for the molecular lesion that could be responsible for the onset of the disease. We combined information from reports of familial linkage, genome wide association studies on diabetes, studies on metabolic and systemic disorders having diabetes as co-morbidity, associated QTLs, genome mapping and sequencing data from public databases along with phenotypic description of type 2 diabetes to achieve this goal. To start with, we selected highly polymorphic microsatellite marker previously shown to be a part of disease haplotype or reported to have a LOD>1.2 in more than 3 studies within the disease critical region. Second we explored the association of these microsatellite markers with anthropometric and clinical parameters and their role in the pathogenesis of the disease. This technical implementation required extensive data mining using both molecular and bioinformatics web based tools. We also evaluated the association between the observed differences of these UTR microsatellite marker frequencies among cases and controls using bioinformatics to predict their role in mRNA regulation

The thesis starts with a review of all genetic linkage/association studies on diabetes published in last one decade (Chapter 2 and 3). The review in chapter 3 included all linkage studies where the disease phenotype co-segregated with the chosen regions of interest 2q32-q37, 3q22-29 and 10q25-26 of the genome in large multigenerational affected families or in independent affected families or in GWAS. These loci were included because they were significant or were suggestive in at least one study and at least nominal in two or more studies. We summarized the genetic and physical position of all the markers in this region having a LOD >1.2 analyzed in studies meeting the

inclusion criteria laid down in the present study. This population-based study was done on a cohort of Two hundred and ninety one (291) subjects with Type 2 Diabetes Mellitus and around four hundred and ninety (475) unrelated healthy volunteers as controls without any history of diabetes and related comorbidities. Demographic data and other parameters were collected during sample collection done from 2008 to 2011. The work presented in this thesis focuses on the largest part on parameters governing body composition which was recorded according to standard techniques laid down by WHO and Joint National Committee (JNC VII) guidelines. Clinical parameters were analyzed in accordance to manufacturer's instruction. The results from the current study reemphasize the previously reported association of both central and abdominal obesity with T2DM and reflect that both BMI and WHR are independent risk factors for T2DM. Our preliminary investigation on demographic data in the first part of chapter 5 and 6 revealed the presence of higher systolic blood pressure in cases as compared to controls, although the difference in mean diastolic blood pressure was found to be higher but non-significant. This significance was not lost even on segregation for gender. Infact it was evident from the data that the female participants had higher systolic blood pressure as compared to male participants. Elevated BP is a known risk factor for diabetes complications, and high BP before the onset of diabetes may also explain the observed high prevalence of cardiovascular disease at the time of diabetes diagnosis. In the current study the presence of dyslipidemia was also been observed among cases as compared to controls. Presence of dyslipidemia, i.e. raised total cholesterol (TC) and decreased high density lipoproteins (HDL) is an established phenomenon in type 2 diabetes mellitus although as stated before its relationship to the fasting blood sugar (FBG) level in both diabetics and non-diabetics has yet to be established.

Later half of chapter 5 and 6 investigated and conveyed the genetic relationship between T2DM and selected endo-phenotypes. Microsatellite marker D2S1384 (200.43cM) was found to be tightly associated to QTL affecting BMI at 2q33.3. The longer base pair allele ($\geq$141bp) on the other hand was found to increase the risk for T2DM by a nominal significance (~2 fold). Also the presence of atleast single copy of 149 bp allele seems to increase the risk for the disease by nearly 3 folds. The next marker D2S2944 when evaluated in the case control cohort it was seen that the presence of longer base pair allele marginally imparts the risk for T2DM. It was also seen that the carriers of longer bp allele seems to have significantly elevated mean DBP as compared to the controls with longer repeats. Moreover the group having longer base pair alleles had higher difference in the mean SBP level as compared to those having shorter allele. In case of BMI the carrier for shorter allele had higher difference in the mean BMI among cases and controls as compared to the carriers of longer allele, whereas in case of WHR the difference in the mean WHR was higher in the longer allele carriers only. The results on D2S439 microsatellite located at 231.27cM revealed association of T2DM with visceral adiposity and several components of metabolic syndrome such as high triglyceride, low HDL-cholesterol and high blood pressure levels in presence of longer allele. In the present study it was seen that the presence of the longer base pair allele seems to increase the risk for T2DM by nearly two fold. The carrier of longer base pair allele of D2S439 seems to have increased Waist hip ratio. Surprisingly majority of our cases possessing the longer allele displayed all four characteristics of proposed risk for Mets defined by IDF. Another marker D3S3609 analysed in the study showed marginal association (~2) with T2DM. Infact the carriers of longer base pair allele seems to have increased mean WHR as compared to their respective controls. Also there was significant difference in the mean BMI levels

among carriers of longer base pair allele. Though the second marker evaluated in this region did not show any association with the T2DM or associated comorbidities. Another region on chromosome 10 which was evaluated for its association with T2DM showed highly significant association with T2DM. It was seen that the presence of the longer base pair allele of D10S521 in the studied cohort was significantly associated with T2DM. The presence of longer base pair allele seems to increase the risk for T2DM by nearly 6 folds. There was a very strong association seen between the longer base pair allele and FBG. The presence of longer base pair allele seems to be significantly associated with the mean FBG level among cases. The longer base pair allele also seems to be associated with increased Total cholesterol. This region seems to be strong and independently associated with T2DM. The second marker which was studied to span this region was D10S1237 and it was seen that the region seems to be associated with BMI and WHR. Cases with the longer base pair allele had significantly higher BMI as compared to the cases with shorter base pair allele. A similar trend was seen with WHR also. The region also seems to have only marginal association with SBP and DBP. The association of the longer base allele with FBG was quiet evident. Also it was seen that the longer allele harboring cohort seems to have significantly higher Total cholesterol levels as compared to the shorter allele harboring cohort.

Thus the study gave evidences of co-occurrence of symptoms of T2DM and selected endophenotypes resulting from the same genetic factors. Additionally these findings provided a clue that the obesity-T2DM and metabolic syndrome –T2DM relationships can be explained through risk alleles of these markers. All the three regions studied here showed association with T2DM; however the loci on chromosome 10 showed very strong and robust association with T2DM.

Taken together, the results support that the pathogenesis of T2DM is associated with clustering of genetic factors controlling the QTL's for body composition, lipid parameters and FBG. Also together the study also reports a strong direct association of specific loci on chromosome 10 with T2DM in North Indians. A healthier life style with weight control could only delay the onset but cannot entirely rule out the occurrence of T2DM in presence of genetic factors. The new genes or sequence elements underlying T2DM can be identified near the associated marker by combination analysis. Use of data related to cytogenetic localization, phenotypes and related expression patterns contained in different internet databases can restrict the number of genes to be analyzed in a chromosomal region indicated in positional cloning studies.

Specific Contributions

- The association of Type 2 Diabetes Mellitus with BMI and WHR is subjected to confounding factors such as Age, Gender, SBP, DBP etc

- The region from 200-229 Mb on chromosome 2q harboring the QTLs regulating the risk for both obesity and metabolic syndrome risk factors amplifies the genetic contribution to susceptibility to T2DM as well.

- The diabetics of north Indian origin (Asian phenotype) possessing the longer allele of D2S439 depict several symptoms of MetS. The association of T2DM and MetS has already been well reported.

- The 10q25-26 Locus of chromosome 10 seems to be the most important loci among the selected regions in the study and the association of this region with T2DM seems to be direct

- The association of the loci 2q32-27 and 3q22-26 with T2DM seems to be more indirect via obesity parameters (via BMI or WHR).

- The study establishes the baseline frequency for all the seven markers among the studied North Indian cohort and shows the marginal (chromosome 2 and 3) to Robust association (chromosome 10) of the selected loci with T2DM.

- The study also brings out that there is a change in the folding of the sequence on repeat expansion; however the correlation between the number of repeats, its folding and association with T2DM still needs to be validated in a bigger panel of STR markers.

Future Scope of work

Future studies should be continued to investigate the chromosomal regions detailed in this thesis by fine mapping. An increase in marker density of the current map would help locate genes associated with T2DM in region where the candidate gene could not be pinpointed with surety in the present study. Investigations must also focus on the other chromosomes to examine if they carry QTLs or genes with major affects related to initiation of lens opacity and disease progression.Nevertheless our findings on *the effect of the repeats expansion on the folding of the sequence also needs to be investigated.*As the etiology of T2DM is complex, it may be multi-factorial with respect to genetic factors also. Thus, itcould be associated with many genetic markers with modest contributions to disease susceptibility.Further investigations are needed to elucidate theserisk factors.

Limitation of the Study

To conclusively substantiate the role of loci or a gene or its variant in a particular disease, it is necessary to replicate the findings of an association study, in other populations or meta-analyses of all the available data. It may be possible that microsatellite polymorphism is associated with the altered expression of other genes. The susceptibility to T2DM due to joint occurrence of risk alleles of all markers studied could not be drawn due to insufficient sample size. The heterogeneous endo-phenotypes and clinical features of T2DM could not be associated with alleles of markers due to the same reason. The problem of subtle sub-stratification that might exist even when cases and controls are taken from the same population cannot be neglected either. The study needs to be extended to a bigger cohort in order to reconfirm the significances observed.

References

- A Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*. 1997 20(7):1183-97.

- Abdalla A. Saeed, Nasser A. Al-Hamdan, Ahmed A. Bahnassy, Abdelshakour M. Abdalla, Mostafa A. F. Abbas, and Lamiaa Z. Abuzaid, Prevalence, Awareness, Treatment, and Control of Hypertension among Saudi Adult Population: A National Survey *International Journal of Hypertension*, 2011 2011;8.

- Aljada A, Ghanim H, Mohanty P, Syed T, Bandyopadhyay A and Dandona P. Glucose intake induces an increase in activator protein 1 and early growth response 1 binding activities, in the expression of tissue factor and matrix metalloproteinase in mononuclear cells, and in plasma tissue factor and matrix metalloproteinase concentrations. *Am. J. Clin. Nutr*. 2004 80:51–57.

- American Diabetes Association (ADA). Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2012 35 Suppl 1:S64-71.

- American Diabetes Association (ADA) Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2010 33 Suppl 1:S62-69.

- Aronoffd SL, Berkowitz K and Want L. Glucose Metabolism and Regulation: Beyond Insulin and Glucagon .*Diabetes Spectrum*. 2004 17(3)183-90.

- Arslanian SA. Clamp techniques in paediatrics: what have we learned? *Horm Res*. 2005 3:16-24.

- Arya R, Blangero J, Williams K, Almasy L, Dyer TD, Leach RJ *et al*. Factors of insulin resistance syndrome--related phenotypes are linked to genetic locations on chromosomes 6 and 7 in non diabetic mexican-americans. *Diabetes.* 2002 51(3):841-47.

- Baier LJ, Permana PA, Yang X, Pratley RE, Hanson RL andShen GQ. A calpain-10 gene polymorphism is associated with reduced muscle mRNA levels and insulin resistance. *J Clin Invest*. 2000 106(7): 69–73.

- Balasubbu S, Sundaresan P, Rajendran A, Ramasamy K, Govindarajan G, Perumalsamy N and Hejtmancik JF. Association analysis of nine candidate gene polymorphisms in Indian patients with type 2 diabetic retinopathy. *BMC Med Genet*. 2010 11:158.

- Bartl J, Scholz CJ, Hinterberger M, Jungwirth S, Wichart I, Rainer MK *et al*. Disorder-specific effects of polymorphisms at opposing ends of the Insulin Degrading Enzyme gene. *BMC Med Genet*. 2011 12:151.

- Bergstrom RW, Newell-Morris LL, Leonetti DL, *et al*. Association of elevated fasting C-peptide level and increased intra-abdominal fat distribution with development of NIDDM in Japanese-American men. *Diabetes*. 1990; 39:104–111.

- Bonora E, Kiechl S, Willeit J, *et al*. Population-based incidence rates and risk factors for type 2 diabetes in white individuals: the Bruneck study. *Diabetes.* 2004; 53:1782–1789.

- Borkan GA, Hults DE, Gerzof SG, Robbins AH, Silbert CK. Age changes in body composition revealed by computed tomography. *J Gerontol*. 1983 38(6):673-77.

- Bossé Y, Vohl MC, Després JP, Lamarche B, Rice T, Rao DC, Bouchard C, Pérusse L. Heritability of LDL peak particle diameter in the Quebec Family Study. *Genet Epidemiol*. 2003 25(4):375-81.

- Bossé Y, Chagnon YC, Després JP, Rice T, Rao DC and Bouchard C *et al*. Compendium of genome-wide scans of lipid-related phenotypes: adding a new genome-wide search of apolipoprotein levels. *J Lipid Res*. 2004 45(12):2174-84.

- Bostrom MA, Freedman BI, Langefeld CD, Liu L, Hicks PJ, Bowden DW. Association of adiponectin gene polymorphisms with type 2 diabetes in an African American population enriched for nephropathy. *Diabetes*. 2009 58(2):499-504.

- Boushel R, Gnaiger E, Schjerling P, Skovbro M, Kraunsoe R and Dela F. Patients with type 2 diabetes have normal mitochondrial function in skeletal muscle. *Diabetologia*. 2007 50:790–96.

- Cassell PG, Jackson AE, North BV, Evans JC, Syndercombe-Court D, Phillips C, Ramachandran A *et al*. Haplotype combinations of calpain 10 gene polymorphisms associate with increased risk of impaired glucose tolerance and type 2 diabetes in South Indians. *Diabetes*. 2002 51(5):1622-28.

- Cauchi S, Meyre D, Durand E, Proença C, Marre M, Hadjadj S *et al*. Post genome-wide association studies of novel genes associated with type 2

diabetes show gene-gene interaction and high predictive value. *PLoS One*. 2008 3(5):e2031.

- Chan JC, Malik V, Jia W, Kadowaki T, Yajnik CS, Yoon KH, Hu FB. Diabetes in Asia: epidemiology, risk factors, and pathophysiology. *JAMA*. 2009 301:2129–40.

- Chauhan G, Kaur I, Tabassum R, Dwivedi OP, Ghosh S, Tandon N, Bharadwaj D. Common variants of homocysteine metabolism pathway genes and risk of type 2diabetes and related traits in Indians. *Exp Diabetes Res.* 2012:960318.

- Chen TM, Kuo PL, Hsu CH *et al*. Microsatellite in the 3' untranslated region of human fibroblast growth factor 9 (FGF9) gene exhibits pleiotropic effect on modulating FGF9 protein expression. *Human Mutation*. 2007 28:98.

- Cheung N, Wong TY. Obesity and eye diseases. *Surv Ophthalmol*. 2007 52(2):180-95.

- Chillaron JC, Hernandez-Valencia M, Reamer C, Fisher S, Joszi A, Hirshman M *et al*. Beta-cell secretory dysfunction in the pathogenesis of low birth weight-associated diabetes: a murine model. *Diabetes*. 2005 54(3):702-11.

- Chillaron JC and Patti ME. To catch up or not to catch up: is this the question? Lessons from animal models. *Curr Opin Endocrinol Diabetes Obes*. 2007 14(1):23-29.

- Dabelea D, Dolan LM, D'Agostino R Jr, Hernandez AM, McAteer JB, Hamman RF *et al*. Association testing of TCF7L2 polymorphisms with type 2 diabetes in multi-ethnic youth. *Diabetologia*. 2011 54(3):535-39.

- Daimon M, Oizumi T, Saitoh T, Kameda W, Yamaguchi H, Ohnuma H, Igarashi M, Manaka H, Kato T. Calpain 10 gene polymorphisms are related, not to type 2 diabetes, but to increased serum cholesterol in Japanese. *Diabetes Res Clin Pract*. 2002 56(2):147-52.

- Daley D, Edwards SR, Song Y, Baechle D, Puppala S, Schick JH, Olson JM, Goddard KA; Framingham Heart Study. Interaction of gender and body mass index (BMI) reveals evidence of linkage for hypertension in the Framingham Heart Study. *BMC Genet*. 2003 31;4,

- Dastani Z, Hivert MF, Timpson N, Perry JR, Yuan X, Scott RA *et al*. Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic

traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet*. 2012 8(3):e1002607.

- Davidson MB. The effect of aging on carbohydrate metabolism: a review of the English literature and a practical approach to the diagnosis of diabetes mellitus in the elderly. *Metabolism*. 1979 28(6):688-705.

- Deepa R, Sandeep S, Mohan V. Abdominal obesity, visceral fat and Type 2 diabetes - "Asian Indian Phenotype". In: Mohan V, Gundu HR Rao, editors. Type 2 diabetes in South Asians: Epidemiology, risk factors and prevention. New Delhi: *Jaypee Brothers Medical Publishers (P) Ltd*; 2006. 138-52.

- Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia: report of a WHO/IDF consultation. Geneva: World Health Organization. 2006

- Defronzo RA. Banting Lecture. From the triumvirate to the ominous octet: a new paradigm for the treatment of type 2 diabetes mellitus. *Diabetes*. 2009 58(4):773-95.

- Deng H and Shen H. Current Topics in Human Genetics: Studies in Complex Diseases Singapore World Scientific Publishing Co.Pt.Ltd, 2007.

- Diagnosis and Classification of Diabetes Mellitus: New Criteria Jennifer mayfield, Am Fam Physician. 1998 58(6):1355-62.

- Duesing K, Fatemifar G, Charpentier G, Marre M, Tichet J, Hercberg S *et al.* Evaluation of the association of IGF2BP2 variants with type 2 diabetes in French Caucasians. *Diabetes.* 2008 57(7):1992-96.

- Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010 42(2):105-16.

- Ehm MG, Karnoub MC, Sakul H, Gottschalk K, Holt DC, Weber JL *et al.*Genome wide search for type 2 diabetes susceptibility genes in four American populations. *Am J Hum Genet*. 2000 66(6):1871-81.

- Ellis KL, Frampton CM, Pilbrow AP, Troughton RW, Doughty RN, Whalley GA *et al*. Genomic risk variants at 1p13.3, 1q41, and 3q22.3 are associated with subsequent cardiovascular outcomes in healthy controls fand in established coronary artery disease. *Circ Cardiovasc Genet*. 2011 4(6):636-46.

- Enzi G, Gasparo M, Biondetti PR, Fiore D, Semisa M and Zurlo F. Subcutaneous and visceral fat distribution according to sex, age, and

overweight, evaluated by computed tomography. *Am J Clin Nutr*. 1986 44(6):739-46.

- Erdmann J, Grosshennig A, Braund PS, König IR, Hengstenberg C and Hall AS. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet*. 2009 41(3):280-82.

- Evans JC, Frayling TM, Cassell PG, Saker PJ, Hitman GA, Walker M *et al.* Studies of association between the gene for calpain-10 and type 2 diabetes mellitus in the United Kingdom. *Am J Hum Genet*. 2001 69(3):544-52.

- Ezzidi I, Turki A, Messaoudi S, Chaieb M, Kacem M, Al-Khateeb GM *et al.* Common polymorphisms of calpain-10 and the risk of Type 2 Diabetes in a Tunisian Arab population: a case-control study. *BMC Med Genet*. 2010 11:75.

- Fingerlin TE, Erdos MR, Watanabe RM, Wiles KR, Stringham HM, Mohlke KL *et al*. Variation in three single nucleotide polymorphisms in the calpain-10 gene not associated with type 2 diabetes in a large Finnish cohort. *Diabetes.* 2002 51(5):1644-48.

- Fonseca SG, Fukuma M, Lipson KL, Nguyen LX, Allen JR, Oka Y *et al*. WFS1 is a novel component of the unfolded protein response and maintains homeostasis of the endoplasmic reticulum in pancreatic β-cells. *J. Biol. Chem.* 2005 280(47):39609–615.

- Francke S, Manraj M, Lacquemant C, Lecoeur C, Leprêtre F, Passa P *et al.* A genome-wide scan for coronary heart disease suggests in Indo-Mauritians a susceptibility locus on chromosome 16p13 and replicates linkage with the metabolic syndrome on 3q27. *Hum Mol Genet*. 2001 10(24):2751-65.

- Frazer KA, Murray SS, Schork NJ and Topol EJ. Human genetic variation and its contribution to complex traits. *Nat Rev Genet.* 2009 10(4):241-51.

- Fried SK, Bunkin DA and Greenberg AS. Omental and subcutaneous adipose tissues of obese subjects release interleukin-6: depot difference and regulation by glucocorticoid. *J Clin Endocrinol Metab*. 1998 83(3):847-50.

- Furukawa Y, Shimada T, Furuta H, Matsuno S, Kusuyama A, DoiA*et al*. Polymorphisms in the IDE-KIF11-HHEX gene locus are reproducibly associated with type 2 diabetes in a Japanese population. J *Clin Endocrinol Metab*. 2008 93(1):310-14.

- Genuth S, Alberti KG, Bennett P, Buse J, Defronzo R, Kahn R, Kitzmiller J,

Knowler WC, Lebovitz H, Lernmark A, Nathan D, Palmer J, Rizza R, Saudek C, Shaw J, Steffes M, Stern M, Tuomilehto J, Zimmet P; Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care*. 2003 26(11):3160-7

- Gesta S, Tseng YH and Kahn CR. Developmental origin of fat: tracking obesity to its source. *Cell*. 2007 131(2):242-56.

- Golden SH, Wang NY, Klag MJ, Meoni LA, Brancati FL. Blood pressure in young adulthood and the risk of type 2 diabetes in middle age. *Diabetes Care*. 2003 26(4):1110-5

- Goldfine AB, Bouche C, Parker RA, Kim C, Kerivan A, Soeldner JS *et al*. Insulin resistance is a poor predictor of type 2 diabetes in individuals with no family history of disease. *Proc Natl Acad Sci U S A*. 2003 100(5):2724-29.

- Golla A, Strauch K, Dietter J, Baur MP. Quantitative trait linkage analysis of longitudinal change in body weight. *BMC Genet*. 2003 4(1):S7.

- Gómez LE, Cruz M, Martínez-Nava GA, Madrid-Marina V, Parra E,García-Mena J, Espinoza-Rojo M *et al*. A replication study of the IRS1, CAPN10, TCF7L2, and PPARG gene polymorphisms associated with type 2 diabetes in two different populations of Mexico. *Ann Hum Genet*. 2011 75(5):612-20.

- Graham TE, Yang Q, Blüher M, Hammarstedt A, Ciaraldi TP, Henry RR *et al*. Retinol-binding protein 4 and insulin resistance in lean, obese, and diabetic subjects. *N Engl J Med*. 2006 354(24):2552-63.

- Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J *et al*. Variant of transcription factor 7-like 2 (TCF7L2) genes confers risk of type 2 diabetes. *Nat Genet*. 2006 38(3):320-23.

- Gregor MF and Hotamisligil GS. Adipocyte stress: the endoplasmic reticulum and metabolic disease. *J. Lipid Res*. 2007 48(9):1905-14.

- Guo X, Saad MF, Langefeld CD, Williams AH, Cui J, Taylor KD *et al.* Genome-wide linkage of plasma adiponectin reveals a major locus on chromosome 3q distinct from the adiponectin structural gene: the IRAS family study. *Diabetes*. 2006 55(6):1723-30.

- Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 1992 48:361-372.

- Gupta V, Khadgawat R, Ng HK, Kumar S, Aggarwal A, Rao VR and Sachdeva MP. A validation study of type 2 diabetes-related variants of the

TCF7L2, HHEX, KCNJ11, and ADIPOQ genes in one endogamous ethnic group of north India. *Ann Hum Genet*. 2010 74(4):361-68.

- HaffnerSM, Hazuda HP, Mitchell BD*et al.* Increased incidence of type II diabetes mellitus in Mexican Americans. *Diabetes Care*. 1991; 14:102–108.

- Haffner SM, Stern MP, Mitchell BD, Hazuda HP, Patterson JK. Incidence of type II diabetes in Mexican Americans predicted by fasting insulin and glucose levels, obesity, and body-fat distribution. *Diabetes* 1990; **39**: 283–8.

- Hammond JB, Spanswick G, Mawn JA. Extraction of DNA from preserved animal specimens for use in randomly amplified polymorphic DNA analysis. *Anal Biochem* 1996 240(2):298-300.

- Hanson RL, Narayan KM, McCance DR, et al. Rate of weight gain, weight fluctuation, and incidence of NIDDM. *Diabetes* 1995; **44**: 261–6.

- Harder T, Rodekamp E, Schellong K, Dudenhausen JW, Plagemann A. Birth weight and subsequent risk of type 2 diabetes: a meta-analysis. *Am J Epidemiol*. 2007 165(8):849-57.

- Hennis A, Wu SY, Nemesure B, Li X, Leske MC. Barbados Eye Study Group. Diabetes in a Caribbean population: epidemiological profile and implications. *Int J Epidemiol*. 2002 31(1):234-39.

- Hertel JK, Johansson S, Raeder H, Midthjell K, Lyssenko V, Groop L, Molven A, Njolstad PR. Genetic analysis of recently identified type 2 diabetes loci in1,638 unselected patients with type 2 diabetes and 1,858 control participant from a Norwegian population-based cohort (the HUNT study). *Diabetologia.* 2008 51(6):971-77.

- Horikawa Y, Miyake K, Yasuda K, Enya M, Hirota Y, Yamagata K *et al*. Replication of genome-wide association studies of type 2 diabetes susceptibility in Japan. *J Clin Endocrinol Metab*. 2008 93(8):3136-41.

- Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet.* 2000 26(2):163-75.

- Horikawa Y, Oda N, Yu L, Imamura S, Fujiwara K, Makino M, Seino Y, ItohM,Takeda J. Genetic variations in calpain-10 gene are not a major factor in the occurrence of type 2 diabetes in Japanese. *J Clin Endocrinol Metab*. 2003 88(1):244-47.

- Hotamisligil GS, Shargill NS and Spiegelman BM. Adipose expression of tumor necrosis factor-α: direct role in obesity-linked insulin resistance. *Science*. 1993 259:87–91.

- Hovi P, Andersson S, Eriksson JG, Järvenpää AL, Strang-Karlsson S, Mäkitie O, Kajantie E. Glucose regulation in young adults with very low birth weight. *N Engl J Med*. 2007 356(20):2053-63.

- Huang Q, Yin JY, Dai XP, Pei Q, Dong M, Zhou ZG *et al*. IGF2BP2 variations influence repaglinide response and risk of type 2 diabetes in Chinese population. *Acta Pharmacol Sin*. 2010 31(6):709-17

- International Diabetes Federation. The IDF Diabetes Atlas. Fifth Edition. Brussels: *International Diabetes Federation*; 2012

- International Diabetes Federation. The IDF Diabetes Atlas. Fourth Edition. Brussels: *International Diabetes Federation*; 2009.

- Iwasaki N, Cox NJ, Wang YQ, Schwarz PE, Bell GI, Honda M, *et al*. Mapping genes influencing type 2 diabetes risk and BMI in Japanese subjects. *Diabetes*. 2003 52(1):209-13.

- Jacques PF, Moeller SM, Hankinson SE, Chylack LT Jr, Rogers G, Tung W *et al*. Weight status, abdominal adiposity, diabetes, and early age-related lens opacities. *Am J Clin Nutr*. 2003 78(3):400-5.

- Jin W, Patti ME. Genetic determinants and molecular pathways in the pathogenesis of Type 2 diabetes. *Clin Sci (Lond)*. 2009 116(2):99-111.

- Joshi SR. Metabolic syndrome--emerging clusters of the Indian phenotype. *J Assoc Physicians India*. 2003 51:445-46.

- Kahn SE. The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of type 2 diabetes. *Diabetologia* 2003; 46:3–19.

- Kallio P, Tolppanen AM, Kolehmainen M, Poutanen K, Lindström J, Tuomilehto J *et al*. Association of sequencevariations in the gene encoding insulin-like growth factor binding protein 5 with adiponectin. *Int J Obes (Lond)*. 2009 33(1):80-88.

- Karter AJ, Mayer-Davis EJ, Selby JV, D'Agostino RB, Haffner SM, Sholinsky P, Bergman R, Saad MF, Hamman RF. Insulin sensitivity and abdominal obesity in African-American, Hispanic and Non-Hispanic White

men and women. The Insulin Resistance and Atherosclerosis Study. *Diabetes* 1996; 45: 1547-55

- Kelley DE. He J, Menshikova EV and Ritov VB. Dysfunction of mitochondria in human skeletal muscle in type 2 diabetes. *Diabetes*. 2002 51:2944–50.

- Khan HA, Sobki SH, Khan SA. Association between glycaemic control and serum lipids profile in type 2 diabetic patients: HbA1c predicts dyslipidaemia. *Clin Exp Med*. 2007 7(1):24-9.

- Kirkland JL, Dobson DE. Preadipocyte function and aging: links between age-related changes in cell dynamics and altered fat tissue function. *J Am Geriatr Soc*. 1997 45(8):959-67.

- Kissebah AH, Sonnenberg GE, Myklebust J, Goldstein M, Broman K, James RG *et al*. Quantitative trait loci on chromosomes 3 and 17 influence phenotypes of the metabolic syndrome. *Proc Natl Acad Sci USA*. 2000 97:14478–83.

- Kooner JS, Saleheen D, Sim X, Sehmi J, Zhang W, Frossard P *et al*. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet*. 2011 43(10):984-89.

- Kuhl JE, Ruderman NB, Musi N, Goodyear LJ, Patti ME, Crunkhorn S *et al*. Exercise training decreases the concentration of malonyl-CoA and increases the expression and activity of malonyl-CoA decarboxylase in human muscle. *Am J Physiol Endocrinol Metab*. 2006 290(6):1296-303.

- Kurucz I, Morva A, Vaag A, Eriksson KF, Huang X, Groop L and Koranyi L. Decreased expression of heat shock protein 72 in skeletal muscle of patients with type 2 diabetes correlates with insulin resistance. *Diabetes*. 2002 51(4):1102–09.

- Kwak SH, Cho YM, Moon MK, Kim JH, Park BL, Cheong HS *et al*. Association of polymorphisms in the insulin-degrading enzyme gene with type 2 diabetes in the Korean population. *Diabetes Res Clin Pract*. 2008 79(2):284-90.

- Lago RM, Premranjan P Singh and Richard W Nesto Diabetes and hypertension. *Nature Clinical Practice Endocrinology & Metabolism* 2007 3, 667

- Lahti-Koski M, Pietinen P, Mannisto S, Vartiainen E. Trends in waist-to-hip ratio and its determinants in adults in Finland from 1987 to 1997. *Am J ClinNutr* 2000; 72: 1436-44

- Larkin EK, Elston RC, Patel SR, Tishler PV, Palmer LJ, Jenny NS, Redline S. Linkage of serum leptin levels in families with sleep apnea. *Int J Obes (Lond).* 2004 29(3):260-67.

- Laukkanen O, Lindström J, Eriksson J, Valle TT, Hämäläinen H, Ilanne-Parikka P *et al.* Polymorphisms in the SLC2A2 (GLUT2) gene are associated with the conversion from impaired glucose tolerance to type 2 diabetes: the Finnish Diabetes Prevention Study. *Diabetes.* 2005 54(7):2256-60.

- Leonard E, Egede, CE. Diabetes and depression: Global perspectives. Diabetes research and clinical practice. 2010 87:302-12.

- Levine JA, Lanningham-Foster LM, McCrady SK, Krizan AC, Olson LR, Kane PH *et al.* Interindividual variation in posture allocation: possible role in human obesity. *Science.* 2005 307(5709):584-86.

- Lee, Chong Do, Steven Blair, and Andrew Jackson Cardiorespiratory fitness, body composition, and all-cause and cardiovascular disease mortality in men. *American Journal of Clinical Nutrition* 1999 69: 373—80

- Li WD, Dong C, Li D, Garrigan C, Price RA. A quantitative trait locus influencing fasting plasma glucose in chromosome region 18q22-23. *Diabetes.* 2004 53(9):2487-91.

- Li WD, Dong C, Price A A Common Variant in the ABCA12 Gene Accounts for Linkage of Total Cholesterol to Chromosome Region 2q34 and is Associated with Obesity 2008**,**21

- Lin Y, Li P, Cai L, Zhang B, Tang X, Zhang X *et al.* Association study of genetic variants in eight genes/loci with type 2 diabetes in a Han Chinese population. *BMC Med Genet.* 2010 11:97.

- Lin Y, Sun Z. Current views on type 2 diabetes. *J Endocrinol.* 2010 204(1):1-11.

- Lipson KL, Fonseca SG, Ishigaki, S *et al.* Regulation of insulin biosynthesis in pancreatic β cells by an endoplasmic reticulum-resident protein kinase IRE1. *Cell Metab.* 2006 4: 245–54.

- Malecki MT, Moczulski DK, Klupa T *et al*. Homozygous combination of calpain 10 gene haplotypes is associated with type 2 diabetes mellitus in a Polish population. *Eur. J. Endocrinol*. 2002 146 (5): 695–99.

- ManavKapoor, SumanKapur and LC Dhaka. D2S2944 marker: A common marker for the obesity-depression associations. *Journal of Mental Health & Human Behavior*. 2010 24-30.

- Manson JE, Rimm EB, Stampfer MJ *et al*. Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. *Lancet*. 1991 338:774–78.

- Martin BC, Warram JH, Krolewski AS, Bergman RN, Soeldner JS, Kahn CR. Role of glucose and insulin resistance in development of type 2 diabetes mellitus: results of a 25-year follow-up study. *Lancet*. 1992 340(8825):925-29.

- Meulenbelt I, Min JL, van Duijn CM, Kloppenburg M, Breedveld FC, SlagboomPE.Strong linkage on 2q33.3 to familial early-onset generalized osteoarthritis and a consideration of two positional candidate genes. *Eur J Hum Genet*. 2006 14(12):1280-7

- Stumvoll, Barry J Goldstein, Timon W van Haeften. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet*. 2005 365:1333-46.

- Mlynarski W. Room for the candidate gene approach in the genome-wide association era: an example from GSTP1 gene polymorphism. *Arch Med Sci.* 2012 8(4):606-7.

- Mohan V, Deepa R, Deepa M, Somannavar S, Datta M. A simplified Indian Diabetes Risk Score for screening for undiagnosed diabetic subjects. *J Assoc Physicians India*. 2005 53:759-63.

- Mootha VK, Lindgren CM, Eriksson KF *et al*. PGC-1α-responsive genes involved in oxidative phosphorylation are co-ordinately downregulated in human diabetes. *Nat. Genet*. 2003 34:267–273.

- Mori Y, Otabe S, Dina C, Yasuda K, Populaire C, Lecoeur C *et al.* Genome-wide search for type 2 diabetes in Japanese affected sib-pairs confirms susceptibility genes on 3q, 15q, and 20q and identifies two new candidate Loci on 7p and 11p. *Diabetes*. 2002 51(4):1247-55.

- Morino K, Petersen KF, Dufour S *et al*. Reduced mitochondrial density and increased IRS-1 serine phosphorylation in muscle of insulin-resistant offspring of type 2 diabetic parents. *J. Clin. Invest*. 2005 115:3587–93.

- Mtiraoui N, Turki A, Nemr R, Echtay A, Izzidi I, Al-Zaben GS *et al*. Contribution of common variants of ENPP1,IGF2BP2, KCNJ11, MLXIPL, PPARγ, SLC30A8 and TCF7L2 to the risk of type 2 diabetes in Lebanese and Tunisian Arabs. *Diabetes Metab*. 2012 5:444-49.

- Nakatani K, Noma K, Nishioka J, *et al*. Adiponectin gene variation associates with the increasing risk of Type 2 diabetes in non-diabetic Japanese subjects. *Int J Mol Med*. 2005 15:173–77.

- Narayanan RP, Stephens R, Simon G Anderson, Kirk W Siddals, Rob L Oliver, Julie E Hudson *et al*. A SNP in the Promoter Region of the IGFBP-5 Gene Is Associated with Circulating Levels of IGFBP-2. *Endocr Rev*. 2011 32

- Narayanan RP, Bo Fu, Adrian H Heald, Kirk W Siddals, Robert L Oliver, Julie E Hudson, Antony Payton, Simon G Anderson, Anne White, William E R Ollier, and J Martin Gibson IGFBP2 is a biomarker for predicting longitudinal deterioration in renal function in type 2 diabetes. *Endocr Connect*. 2012 1:95-102.

- National Diabetes Data Group (NDDG). Diabetes in America. 2d ed. Bethesda, Md.: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, 1995NIH publication no. 95-1468;

- Nemr R, Echtay A, Dashti EA, Almawi AW, Al-Busaidi AS, Keleshian SH *et al*. Strong association of common variants in the IGF2BP2 gene with type 2 diabetes in Lebanese Arabs. *Diabetes Res ClinPract*. 2012 96(2):225-29.

- Nurnberger JI Jr, Foroud T, Flury L, Su J, Meyer MT, Hu K *et al*. Evidence for a locus on chromosome 1 that influences vulnerability to alcoholism and affective disorder. *Am J Psychiatry* 2001 158: 718-24.

- Oeppen J, Vaupel JW. Demography. Broken limits to life expectancy. *Science*. 2002 296(5570):1029-31.

- Ohlson LO, Larsson B, Svardsudd K, Welin L, Eriksson H, Wilhelmsen L, Bjorntorp and Tibblin G. The influence of body fat distribution on the incidence of diabetes mellitus: 13.5 years of follow-up of the participants in the study of men born in 1913. *Diabetes* 1985; 34: 1055-8

- Omori S, Tanaka Y, Horikoshi M, Takahashi A, Hara K, Hirose H *et al*. Replication study for the association of new meta-analysis-derived risk loci with susceptibility to type 2 diabetes in 6,244 Japanese individuals. *Diabetologia*. 2009 52(8):1554-60.

- Ong KL, Li M, Tso AW, Xu A, Cherny SS, Sham PC *et al*. Association of genetic variants in the adiponectin gene with adiponectin level and hypertension in Hong Kong Chinese. *Eur J Endocrinol*. 2010 163(2):251-57.

- Orho-Melander, Klannemark M, Svensson MK, Ridderstrale M, Lindgren CM, Groop L. Variants in the calpain-10 gene predispose to insulin resistance and elevated free fatty acid levels, *Diabetes*. 2002 51:2658–64.

- Osher E, Stern N. Diastolic pressure in type 2 diabetes: can target systolic pressure be reached without "diastolic hypotension"? *Diabetes Care*. 2008 Suppl 2:S249-54

- Patti ME, Butte AJ, Crunkhorn S *et al*. Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: potential role of PGC-1 and NRF1. *Proc. Natl. Acad. Sci. U.S.A*. 2003 100:8466–71.

- Perez-Martinez P, Perez-Caballero AI, Garcia-Rios A, Yubero-Serrano EM, Camargo A, Gomez-Luna MJ *et al*. Effects of rs7903146 variation in the Tcf7l2 gene in the lipidmetabolism of three different populations. *PLoS One*. 2012 7(8):e43390.

- Petersen JL, McGuire DK. Impaired glucose tolerance and impaired fasting glucose--a review of diagnosis, clinical implications and management. *Diab Vasc Dis Res*. 2005 2(1):9-15.

- Petersen KF, Dufour S, Befroy D, Garcia R and Shulman GI. Impaired mitochondrial activity in the insulin-resistant offspring of patients with type 2 diabetes. *N. Engl. J. Med*. 2004 350:664–71.

- Pezzolesi MG, Nam M, Nagase T, Klupa T, Dunn JS, Mlynarski WM *et al*. Examination of candidate chromosomal regions for type 2 diabetes reveals a susceptibility locus on human chromosome 8p23.1. *Diabetes*. 2004 53(2):486-91.

- Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--apopulation-based twin study. *Diabetologia*. 1999 42(2):139-45.

- Ramachandran A, Mary S, Yamuna A, Murugesan N and Snehalatha C. High prevalence of diabetes and cardiovascular risk factors associated with urbanization in India. *Diabetes Care*. 200831:893–98.

- Ramachandran A, Snehalatha C. Current scenario of diabetes in India. *J Diabetes*. 2009 1(1):18-28.

- Rasmussen SK, Urhammer SA, Berglund L, Jensen JN, Hansen L, Echwald SM *et al*. Variants within the calpain-10 gene on chromosome 2q37 (NIDDM1) and relationships to type 2 diabetes, insulin resistance, and impaired acute insulin secretion among Scandinavian Caucasians. *Diabetes.* 2002 51(12):3561-67.

- Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW *et al*. STAT4 and the risk of rheumatoid arthritis and systemic lupuserythematosus. *N Engl J Med*.2007 357(10):977-86.

- ResnickHE, Valsania P, Halter JB, Lin X. Relation of weight gain and weight loss on subsequent diabetes risk in overweight adults. *J Epidemiol Community Health* 2000; 54: 596–602.

- Richardson DK, Schneider J, Fourcaudot MJ, Rodriguez LM, Arya R, Dyer TD *et al.* Association between variants in the genes for adiponectin and its receptors with insulin resistance syndrome (IRS)-related phenotypes in Mexican Americans. *Diabetologia.* 2006 49(10):2317-28.

- Roup E, EoulouriA ,Sotiropoulou P , Makrinika E , Marneras X , Lahana E and Gourni I. Anxiety and depression in patients with Type 2 Diabetes Mellitus, depending on sex and body mass index. *Health Science Journal*. 2009 3:32-40.

- Ruchat SM, Després JP, Weisnagel SJ, Chagnon YC, Bouchard C, Pérusse L. Genome-wide linkage analysis for circulating levels of adipokines and C-reactive protein in the Quebec family study (QFS). *J Hum Genet.* 2008;53(7):629-36

- Rudovich N, Pivovarova O, Fisher E, Fischer-Rosinsky A, Spranger J, Möhlig M *et al*. Polymorphisms within insulin-degrading enzyme (IDE) gene determines insulin metabolism and risk of type 2 diabetes. *J Mol Med (Berl).* 2009 87(11):1145-51.

- Rung J, Cauchi S, AlbrechtsenA, Shen L, Rocheleau G, Cavalcanti-Proença C *et al.* Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet*. 2009 41(10):1110-15.

- Russell AP, Hesselink MK, Lo SK, Schrauwen P. Regulation of metabolic transcriptional co-activators and transcription factors with acute exercise. *FASEB J*. 2005 19(8):986-88.

- Salpea KD, Gable DR, Cooper JA, Stephens JW, Hurel SJ, Ireland HA, Feher MD, Godsland IF, Humphries SE. The effect of WNT5B IVS3C>G on the susceptibility to type 2 diabetes in UK Caucasian subjects. *Nutr Metab Cardiovasc Dis*. 2009 19(2):140-45.

- Schaumberg DA, Glynn RJ, Christen WG, Ajani UA, Stürmer T, Hennekens CH. A prospective study of blood pressure and risk of cataract in men. *Ann Epidemiol*. 2001 11(2):104-10.

- Schwarz PEH, Horikawa Y, Vcelak J, Selisko T, Rietzsch H, Bendlova B, Schulze J, Cox NJ. Genetic variation of CAPN10 affects susceptibility to type 2 diabetes in German and Czech population. *Diabetes*. 2001 50 Suppl 2:A232.

- Scott WK, Hauser ER, Schmechel DE, Welsh-Bohmer KA, Small GW, Roses AD, Saunders AM, Gilbert JR, Vance JM, Haines JL, Pericak-Vance MA. Ordered-subsetslinkage analysis detects novel Alzheimer disease loci on chromosomes 2q34 and15q22. *Am J Hum Genet*. 2003 73(5):1041-51.

- Sharma R, Prudente S, Andreozzi F, Powers C, Mannino G, Bacci S *et al.* The type 2 diabetes and insulin-resistance locus near IRS1 is a determinant of HDL cholesterol and triglycerides levels among diabetic subjects. *Atherosclerosis*. 2011 216(1):157-60.

- Shaw JE, Sicree RA and Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Res. Clin. Pract*. 2010 87:4–14.

- Shetty P. Public health: India's diabetes time bomb. *Nature*. 2012 485(7398):S14-16.

- Shoelson SE, Lee J, Goldfine AB. Inflammation and insulin resistance. *J Clin Invest*. 2006 116(7):1793-801.

- Silbernagel G, Renner W, Grammer TB, Hügl SR, Bertram J, Kleber ME *et al*. Association of TCF7L2 SNPs with age at onset of type 2 diabetes and Proinsulin/insulin ratio but not with glucagon-likepeptide 1. *Diabetes Metab Res Rev*. 2011 27(5):499-505.

- Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D *et al*. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 2007 445(7130):881-85.

- Sordi V, Bertuzzi F, Piemonti L. Diabetes mellitus: an opportunity for therapy with stem cells*? Regen Med*. 2008 3:377–97.

- Sreekumar R, Halvatsiotis P, Schimke JC and Nair KS. Gene expression profile in skeletal muscle of type 2 diabetes and the effect of insulin treatment. *Diabetes*. 2002 51:1913–20.

- Stambolian D, Ibay G, Reider L, Dana D, Moy C, Schlifka M *et al*. Genomewide linkage scan for myopia susceptibility loci among Ashkenazi Jewish families shows evidence of linkage on chromosome 22q12. *Am J Hum Genet*. 2004 75(3):448-59.

- Stern MP: Diabetes and cardiovascular disease: the "common soil" hypothesis. *Diabetes* 1995 44:369–374.

- Stolk RP, van Splunder IP, Schouten JS, et al. High blood pressure and the incidence of non-insulin dependent diabetes mellitus: findings in a 11.5 year follow-up study in The Netherlands. *Eur J Epidemiol*. 1993 9:134–139.

- Strissel KJ, Stancheva Z, Miyoshi H *et al*. Adipocyte death, adipose tissue remodeling, and obesity complications. *Diabetes*. 2007 56:2910–18.

- Sun HX, Zhang KX, Du WN *et al*. Single nucleotide polymorphisms in CAPN10 gene of Chinese people and its correlation with type 2 diabetes mellitus in Han people of northern China. *Biomed. Environ. Sci*. 2002 15(1):75–82.

- Takeuchi F, Serizawa M, Yamamoto K, Fujisawa T, Nakashima E, Ohnaka K, *et al.* Confirmation of multiple risk Loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. *Diabetes.* 2009 58(7):1690-99.

- Talmud PJ, Cooper JA, Gaunt T, Holmes MV, Shah S, Palmen J *et al*. Variants of ADRA2A are associated with fasting glucose, blood pressure, body mass index and type 2 diabetes risk: meta-analysis of four prospective studies. *Diabetologia*. 2011 54(7):1710-19.

- Tang Y, Han X, Sun X, Lv C, Zhang X, Guo W *et al*. Association study of a common variant near IRS1 with type 2 diabetes mellitus in Chinese Han population. *Endocrine*. 2012.

- Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M *et al*. Biological,clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010 466(7307):707-13.

- Thorsby PM, Midthjell K, Gjerlaugsen N, Holmen J, Hanssen KF, Birkeland KI, Berg JP. Comparison of genetic risk in three candidate genes (TCF7L2,

PPARG, KCNJ11) with traditional risk factors for type 2 diabetes in a population-based study--the HUNT study. *Scand J Clin Lab Invest*. 2009 69(2):282-87.

- Tsai HJ, Sun G, Weeks DE, Kaushal R, Wolujewicz M, McGarvey ST *et al*. Type 2 diabetes and three calpain-10 gene polymorphisms in Samoans: no evidence of association. *Am J Hum Genet*. 2001 69(6):1236-44.

- Vasseur F, Leprêtre F, Lacquemant C, Froguel P. The genetics of adiponectin. *CurrDiab Rep*. 2003 3(2):151-58.

- Vijan S. Type 2 diabetes. *Ann Intern Med*. 2010 2:152-55.

- Vionnet N, Hani EH, Dupont S, Gallina S, Francke S, Dotte S *et al.*Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27-qter and independent replication of a type 2-diabetes locus on chromosome1q21-q24. *Am J Hum Genet*. 2000 67(6):1470-80.

- Wagner R, Dudziak K, Herzberg-Schäfer SA, Machicao F, Stefan N, Staiger H, Häring HU, Fritsche A. Glucose-raising genetic variants in MADD and ADCY5 impair conversion of proinsulin to insulin. *PLoS One* 2011 6(8):e23639.

- Wang SL, Pan WH, Hwu CM, et al. Incidence of NIDDM and the effects of gender, obesity and hyperinsulinaemia in Taiwan. *Diabetologia*. 1997; 40:1431–1438.

- Weisberg SP, McCann D, Desai M, Rosenbaum M, Leibel RL and Ferrante, Jr AW. Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest*. 2003 112:1796–808.

- Wellen KE, Hotamisligil GS. Inflammation, stress, and diabetes. *J Clin Invest*. 2005 115(5):1111-19.

- WHO Expert Committee on Diabetes Mellitus: second report. World Health Organ Tech Rep Ser. 1980 646:1-80.

- Whiting DR, Guariguata L, Weil C, Shaw J. IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Res Clin Pract*. 2011 94(3):311-21.

- Wiltshire S, Hattersley AT, Hitman GA, Walker M, Levy JC, Sampson M. *et al*. A genomewide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees

provides independent replication of a susceptibility locus on chromosome 1q. *Am J Hum Genet*. 2001 69(3):553-69.

- World Health Organization. Diabetes Mellitus: Report of a WHO Study Group. Geneva: WHO, 1985.Technical Report Series 727.

- World Health Organization. Prevention of diabetes mellitus. Geneva: WHO, 1994.Technical Report Series no. 844.

- Xiang K, Fang Q, Zheng T, Jia W, Wang Y, Zhang R *et al* The impact of calpain-10 gene combined-SNP variation on type 2 diabetes mellitus and its related metabolic traits. 2001 18(6):426-30.

- Xu H, Barnes GT, Yang Q *et al*. Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J. Clin. Invest.* 2003 112:1821–30.

- Yatsu K, Mizuki N, Hirawa N, Oka A, Itoh N and Yamane T. High-resolution mapping for essential hypertension using microsatellite markers. *Hypertension.* 2007 49(3):446-52.

- Yiannakouris N, Cooper JA, Shah S, Drenos F, Ireland HA, Stephens JW *et al.* IRS1 gene variants, dysglycaemic metabolic changes and type-2 diabetes risk. *Nutr Metab Cardiovasc Dis*. 2011

- Zhou DZ, Liu Y, Zhang D, Liu SM, Yu L, Yang YF *et al*. Variations in/nearby genes coding for JAZF1, TSPAN8/LGR5 and HHEX-IDE and risk of type 2 diabetes in Han Chinese. *J Hum Genet*. 2010 55(12):810-15.

- Zubenko GS, Maher B, Hughes HB, Zubenko WN, Stiffler JS, Kaplan BB, Marazita ML. Genome-wide linkage survey for genetic loci that influence the development of depressive disorders in families with recurrent, early-onset,major depression. *Am J Med Genet Neuropsychiatr Genet.* 2003 15:123(1):1-18.

Online Resources

- ALFRED: http://alfred.med.yale.edu/
- BLAST: http://blast.ncbi.nlm.nih.gov/
- CEPH: http://www.ceph.org/
- http://lpgws.nci.nih.gov/html-chlc/ChlcMarkers.html
- EMBL: http://www.ebi.ac.uk/embl/
- http://www.gdb.org
- Gene bank: http://www.ncbi.nlm.nih.gov/genbank/
- Gene: http://www.ncbi.nlm.nih.gov/gene
- Genome: http://www.ncbi.nlm.nih.gov/genome
- Human Genome Database: http://nar.oxfordjournals.org/content/26/1/94.full
- HWE Calculator: http://www.tufts.edu/~mcourt01/Documents/Court%20lab%20-%20HW%20calculator.xls
- Mendalian Inheritance: http://www.ncbi.nlm.nih.gov/omim
- MFOLD: http://mfold.rna.albany.edu/?q=mfold
- NCBI Mapviewer: http://www.ncbi.nlm.nih.gov/mapview/
- NCBI: http://www.ncbi.nlm.nih.gov/
- Nucleotide: http://www.ncbi.nlm.nih.gov/nuccore
- PCR Primers: http://frodo.wi.mit.edu/
- PIC Calculator: http://www.liv.ac.uk/~kempsj/pic.html
- Promo: http://alggen.lsi.upc.es/cgi-in/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3
- Pubmed Central: http://www.ncbi.nlm.nih.gov/pmc/
- SNP: http://www.ncbi.nlm.nih.gov/snp
- TPMD: http://tpmd2.nhri.org.tw/tpmd/php-bin/index_en.php
- Unigene: http://www.ncbi.nlm.nih.gov/unigene
- UniSTS: http://www.ncbi.nlm.nih.gov/unists
- Wikipedia: http://www.wikipedia.org/
- http://en.wikipedia.org/wiki/Diabetes_mellitus_type_2
- http://www.drugs.com/health-guide/type-2-diabetes-mellitus.html
- http://sahely.com/medicine/normal_glucose_regulation.htm

Appendices I

CONSENT FORM
(CERTIFICATE BY INVESTIGATOR)

I certify that I have described all details about the study in the terms readily understood by the patient.

Date…………………………..   Signature……………….………………

                                                                  Name…………………….……………..

CONSENT FORM
(CERTIFICATE BY PATIENT)

I have been informed to my satisfaction by the attending physician, the purpose of the clinical study, including the laboratory investigations to be performed using my blood sample.

I am also aware of my right to opt out of the study at ant time during the course of the study without having to give the reasons for doing so.

I exercising my free power of choice hereby give my consent to be included as a subject in the study of genotyping of my DNA sample.

Date…………………………..   Signature……………………..…………….

                                                                  Name………………….………………….

                                                                  Sign of contact person………………………

                                                                  Name……..…………………..……………

                                                                  Relationship………………..………………

Appendices II

PATIENT DATA SHEET

Registration No……………...............

Name of patient…………………………………………………………………………………

Age ……………....years

Sex Male / Female

Father's name……………………………………………………………………………..

Year of formal training- ……………...years

Marital status married / unmarried

Residence urban / rural /semi-urban

Occupation unemployed / unskilled / skilled / student / housewife / professional / retired / business/others

Permanent address……………………………………………………………………..
  (With phone No.)
  ……………………………………………………………………..

 Mailing address…………………………………………………………………………
(With phone No.)
 …………………………………………………………………….

Contact person……………………………………………………………………………..

Duration of illness………………years

Age of onset …………………….years

Glucose level Fasting……..mg/ml, PP…….mg/ml Random …….mg/ml

Co morbidity (if any)……………………………………………………………………

Obesity………………………………………………………………………………………

Hypertension……………………………………………………………………………..

Cardiac complications (if any)……………………………………………………..………

………………………………………………………………………………….…………………

Eating habits………………………………………………………..……………………………

……………… …………………………………………………………………………………………

Disease history of blood relatives………………………………………………………………..

…….

S.No.	Name of drug	Dose	Duration
1.			
2.			
3.			
4.			

Other details………………………………………………………………………………………..

………

………

………

Attending Physician   **Laboratory Staff**

……………………………….   ………………………………

Follow up action
- Blood sample taken to be taken
- Parents to be contacted
- Siblings (blood relatives) to be contacted
- Grand parents to be contacted
- Nephews / Nieces to be contacted
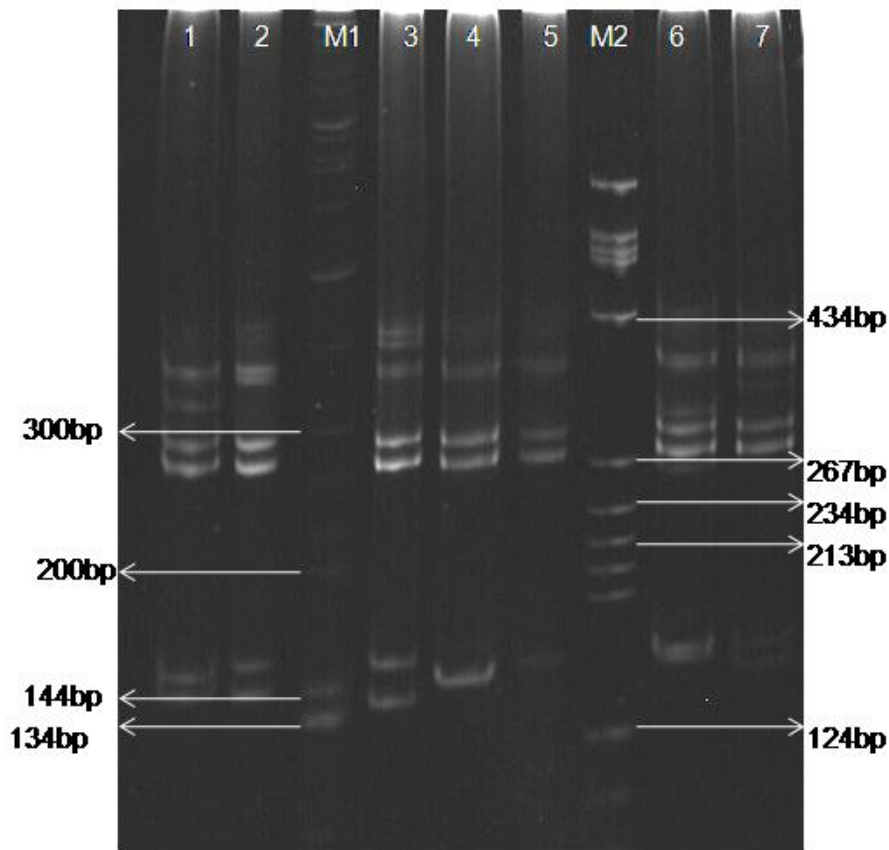
Appendices III:

Figure A1: Electrophoresis on 12% native Midi-PAGE resolving SSLP of D2S2359 and D2S1384 markers.

PCR product ranges from 227-271bp for D2S2359 and 133-161bp for D2S1384. Genotyping D2S2359:D2S1384 – Lane 1, 261/261bp: 141/149bp; Lane 2, 261/261bp:141/153bp; Lane 3, 263/263bp:153/141bp; Lane 4, 263/263bp:149/148bp; Lane5, 271/271bp:153/153bp; M1: 100bp ladder+ house-made markers, M2: pBR322 *Hae*III digested ladder
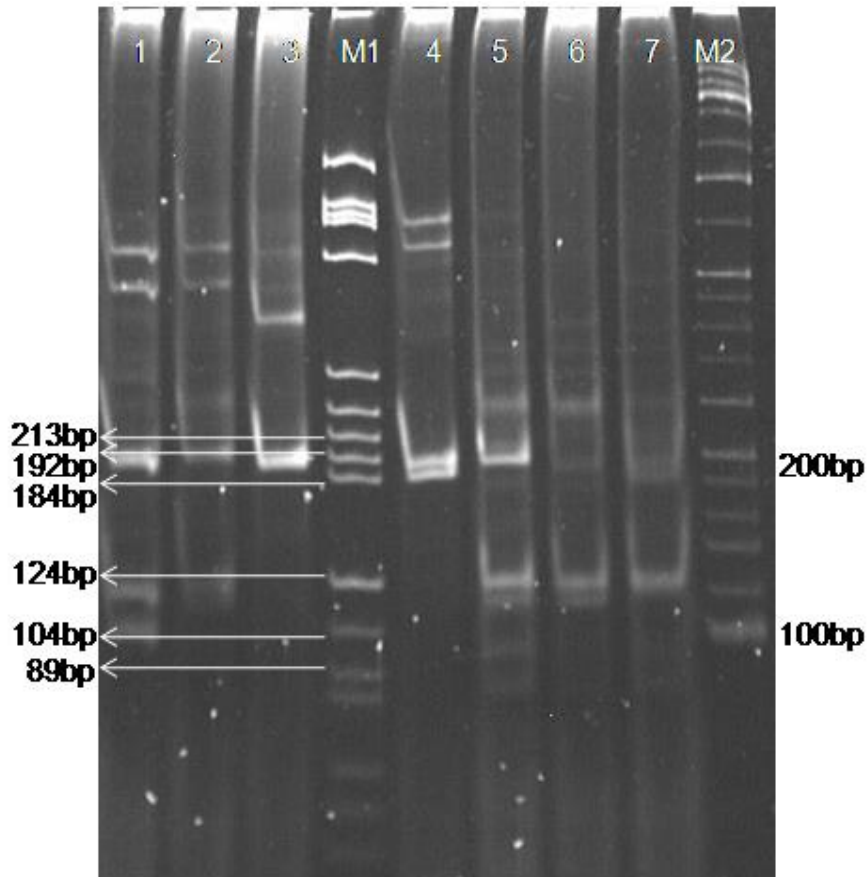
Figure A2: Electrophoresis on 12% native Midi-PAGE resolving SSLP of D2S439 and D2S2944 markers.

PCR product ranges from 167-203bp for D2S439 and 92-124bp for D2S2944. Genotyping D2S439:D2S2944 – Lane 1, 195/195bp: 104/120bp ; Lane 2, 199/199bp:116/124bp; Lane 3, 195/195bp:NB/NB; Lane 4, 183/191bp:NB/NB; Lane 5, 195/195bp:124/116bp; Lane 6, 191/191bp:116/124bp; Lane 7, 183/191bp:120/124bp; M1: pBR322 *Hae*III digested ladder; M2: 20bp ladder; NB: No-Band
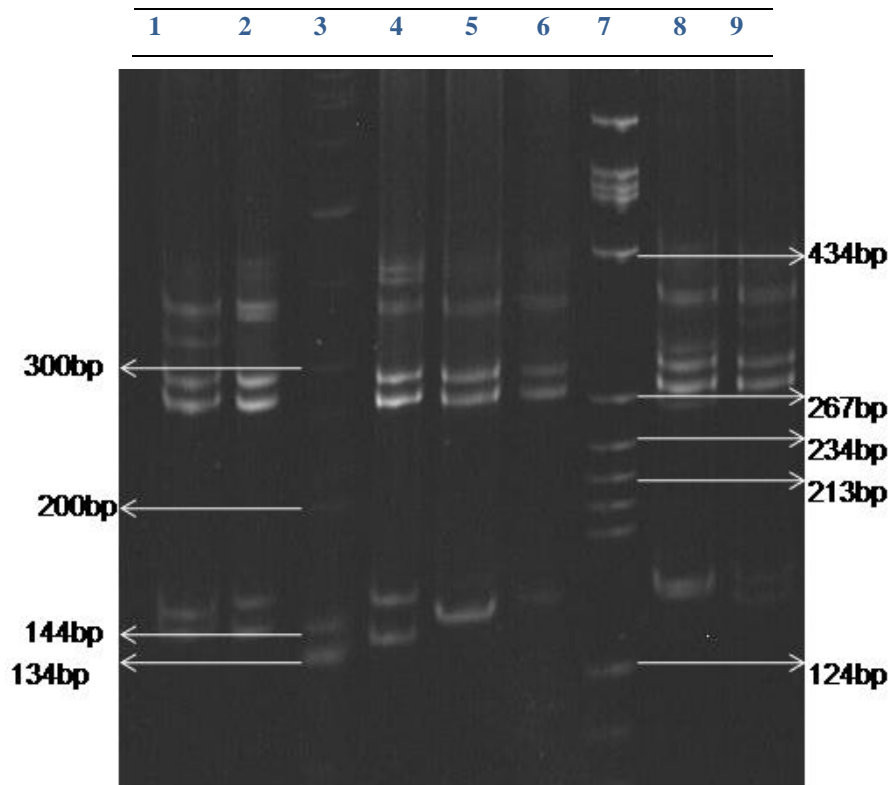
Appendices III:

Figure A1: Electrophoresis on 12% native Midi-PAGE resolving SSLP of D2S1384 markers.

PCR product ranges from 133-161bp for D2S1384. Genotyping 1384 – Lane 1, 141/149bp; Lane 2, 141/153bp; Lane 3, 153/141bp; Lane 4, 149/148bp; Lane5,153/153bp; M1: 100bp ladder+ in house made markers, M2: pBR322 *Hae*III digested ladder
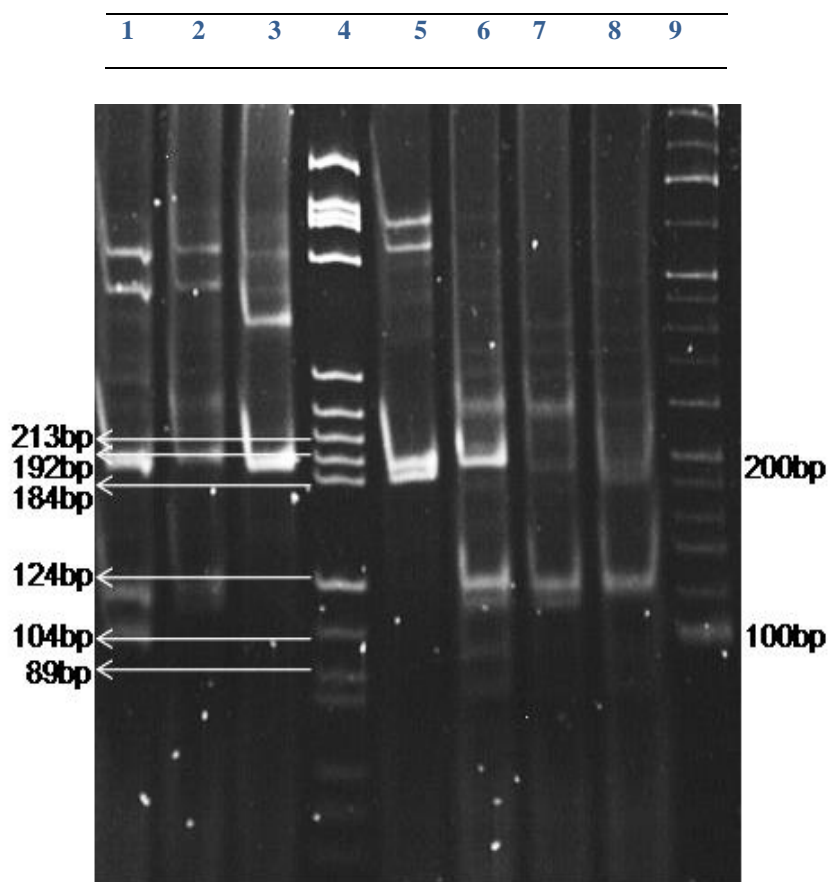
Figure A2: Electrophoresis on 12% native Midi-PAGE resolving SSLP of D2S439 and D2S2944 markers.

PCR product ranges from 167-203bp for D2S439 and 92-124bp for D2S2944. Genotyping D2S439:D2S2944 – Lane 1, 195/195bp: 104/120bp ; Lane 2, 199/199bp:116/124bp; Lane 3, 195/195bp:NB/NB; Lane 4, M1; Lane 5 183/191bp:NB/NB; Lane 6, 195/195bp:124/116bp; Lane 7, 191/191bp:116/124bp; Lane 8, 183/191bp:120/124bp; Lane 9: M2; M1: pBR322 *Hae*III digested ladder; M2: 20bp ladder; NB: No-Band
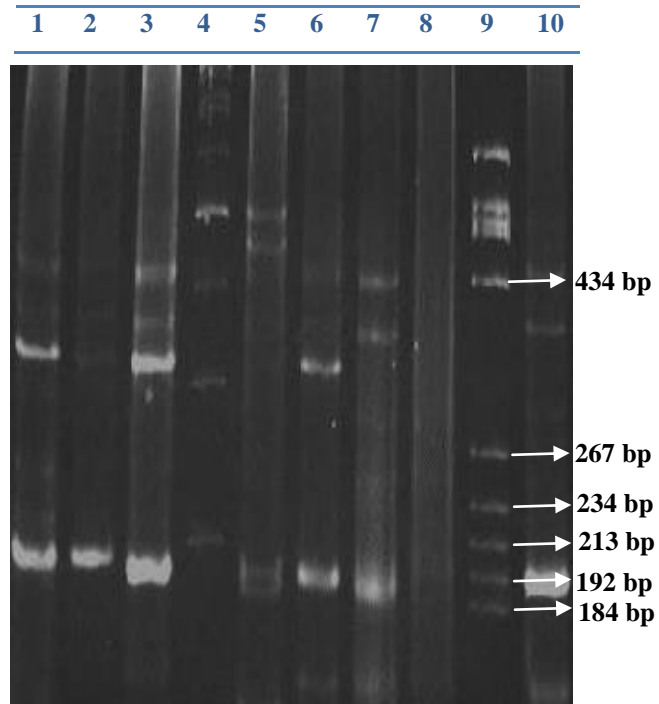
Figure A3: Electrophoresis on 12% native Midi-PAGE resolving SSLP of D3S3609 marker.

PCR product ranges from 163-185bp for D3S3609. Genotyping D3S3609 – Lane 1: 185; Lane 2: 185; Lane 3: 183 Lane 4: M1: 100bp Ladder; Lane 5: 185/181, Lane 6, 183bp; Lane 7: 179bp; Lane 8: NB ; Lane 9: M2: pBR322 *Hae*III digested ladder; Lane 10: NB
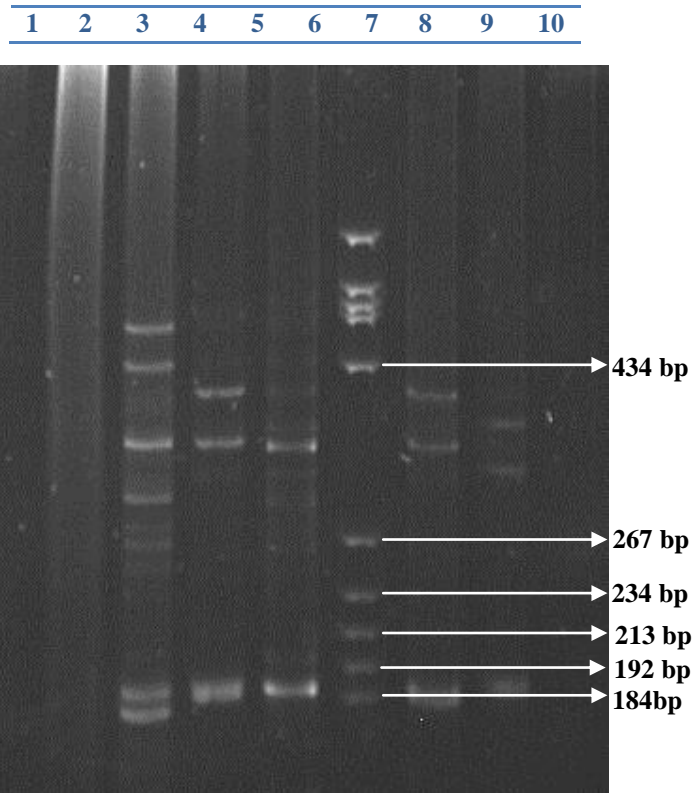
Figure A4: Electrophoresis on 12% native Midi-PAGE resolving SSLP of D10S1237 and D10S521 markers.

PCR product ranges from 376-432bp for D10S1237 and 155-189bp for D10S521. Genotyping D10S1237:D10S521 – Lane 1, NB;NB; Lane 2: 432/400, 183/171; Lane 3: 420/400, 183 Lane 4: 400,183; Lane 5: pBR322 *Hae*III digested ladder, Lane 6: 420/400 bp:183bp; Lane 7,408/392 bp: 187bp; NB: No-Band

List of Publications

- Evidence of marginal significance of region 2q32-37 with T2DM in North Indians (Under preparation)

- Suman Kapur, **Urvashi Dube**, Anuradha Pal, Lokesh Chaturvedi, Bhupendra Singh, "Are police personnel at higher risk for NCDs including multiple metabolic syndrome? A cohort study from Rajasthan India" Scandinavian Journal of work environment and Health (Under Review)

- Suman Kapur, Bhupendra Singh, Lokesh Chaturvedi, Anuradha Pal, **Urvashi Dube,** "Occupation related health risks: As observed in police personnel from Rajasthan", India Rakshin. 2011;84 - 100

- Hulme C, Akritopoulou-Zanze I, Dai WM, Beck B, Srivastava S, Wang W, Wang K, Czarna A, Holak TA, Meireles L, Camacho C, Raghavan B, Day BW, Dömling A, Qin C, Zhang R, Wang Q, Ren J, Tian L, Nikulnikov M, Krasavin M, Krasavin M, Parchinsky V, Shkavrov S, Bukhryakov K, Tsirulnikov S, Krasavin M, Bushkova E, Parchinsky V, Krasavin M, Kalinski C, Kysil V, Tsirulnikov S, Ivachtchenko A, Potapov VV, Kysil VM, Fetisova NA, Nikitin AV, Ivachtchenko AV, Potapov VV, Ilyn AP, Fetisova NA, Kravchenko DV, Ivachtchenko AV, Shilova O, Ilyin A, Ivachtchenko A, Shkirando AM, Kysil VM, Potapov VV, Ivachtchenko AV, Vaddula BR, Kumar D, Sharad S, **Dube U**, Kapur S. 2010, A.Multi-component reactions in drug discovery, 2010, Adv Exp Med Biol. 2011;699:75-106.

- Dalip Kumar, V. Bucchi Reddy, Shashwat Sharad, **Urvashi Dube**, Suman Kapur: "A facile one-pot green synthesis and antibacterial activity of 2-amino-

4H-pyrans and 2-amino-5-oxo-5,6,7,8-tetrahydro-4H-chromenes" 2009, Eur J Med Chem. 2009, 44(9):3805-3809.

- Manav Kapoor, Suman Kapur, Shipra Mehra, **Urvashi Dube**, Shashwat Sharad and Sharda Sidhu: "Genetic variation of D7S1875 dinucleotide repeats in the leptin gene increases risk for depression: A population-based case-control study in India". Depress Anxiety. 2009, 26(9):791-795.

- Jhanjaria A, Kapur Suman, Agrawal RP, Kapoor M, Sharad S, **Dube Urvashi**, Mohta N, Kochar DK, "First study of aryl hydrocarbon nuclear translocator gene in Indians: Correlation with resistance to diabetes in Raika community". Diabetes Res Clin Pract., 2008 82(1):e21-2.

Conference publications

- **Urvashi Dube**, Suman Kapur, R.P Agarwal, Liakat Gauri and P.K Sehgal,APOE & OB: Probable common genetic markers in metabolic disorders, , presented at 14th Human Genome meeting held at Montplier, France 2010.

- Jilowa Virendra, Gauri LA, Kapur Suman, Kapoor Manav, **Dube Urvashi** Study of ApoE and Ob Genes as Genetic Markers in Rheumatoid Arthritis, APICON-2010 held at Jaipur Rajasthan.

- **Urvashi Dube**, Suman Kapur Chromosomal region 2q32-27: susceptibility locus for T2DM and other metabolic disorders International Symposium on "Genetic and Epigenetic Basis of Complex Diseases" held at Hyderabad, India, 2009.

- Dalip Kumar, V. Bucchi Reddy, Shashwat Sharad, **Urvashi Dube**, Suman Kapur: A facile one-pot green synthesis and antibacterial activity of 2-amino-4H-pyrans and 2-amino-5-oxo-5,6,7,8-tetrahydro-4H-chromenes" Presented at 4th International Conference on Multi-Component Reactions and Related Chemistry held at Ekaterinburg, Russia, 2009.

- **Urvashi Dube** and Suman Kapur: Evaluating role of leptin and adiponectin as emerging mediators in chronic inflammatory disorders presented at National Conference on Emerging Trends in Life Sciences Research, 2009, held at BITS, Pilani, India.

- Shashwat Sharad, Manav Kapoor, Shipra Mehra, **Urvashi Dube**, Anuradha Pal, R P Agarwal, R P Pareek, Suman Kapur, Role of OPRM1 gene in T2DM susceptibility and its prevalence in North-West Part of Rajasthan, HGV2008, held at Ontario, Canada, from 15-17th Oct 2008.

- Suman Kapur, Shashwat Sharad, **Urvashi Dube**, Manav Kapoor, RP Pareek, RP Agarwal, APOE and CAPN gene and risk for the metabolic complications in diabetes HUGO's 13th Human Genome Meeting held at Hyderabad, India from 27-30 Sep 2008.

- Shashwat Sharad, Manav Kapoor, Shipra Mehra, **Urvashi Dube**, Anuradha Pal, R P Agarwal, R P Pareek, Suman Kapur, Role of OPRM1 gene in T2DM susceptibility and its prevalence in North-West Part of Rajasthan HUGO's 13th Human Genome Meeting held at Hyderabad, India from 27-30 Sep 2008.

- Manav Kapoor, Shipra Mehra, **Urvashi Dube**, Shashwat Sharad, Sharda Sidhu and Suman Kapur. Leptin (LEP) gene: The missing link between the Depression, Obesity and Metabolic disorders to be presented at HUGO's 13th Human Genome Meeting held at Hyderabad, India from 27-30 Sep 2008.

- **Urvashi Dube**, Manav Kapoor, Shashwat Sharad, Ashok Jhanjharia, RP Agarwal, Suman Kapur, Possible protective role of ARNT gene in low risk for Type 2 Diabetes Mellitus in Raica community of Rajasthan to be presented at HUGO's 13th Human Genome Meeting held at Hyderabad, India from 27-30 Sep 2008.

- Anuradha Pal, Suman Kapur, Shashwat Sharad, Manav Kapoor, **Urvashi Dube**, Anoop Saraya Role of alcohol metabolizing genes in alcohol induced Pancreatitis presented at HUGO's 13th Human Genome Meeting held at Hyderabad, India from 27-30 Sep 2008.

- Suman Kapur, Pareek, R. P., Aggarwal, R.P, Khinvasara, R.K, Manav Kapoor, Shaifali Gurjar, **Urvashi Dube** and Shashwat Sharad, APOE and Ob Gene: Correlation of Genotypes with Various Clinical and Anthropometric Parameters in Type 2 Diabetes Mellitus, presented at 5th Winter Symposium held at CMC, Vellore on 9th-11th January 2007.

Brief Biography of the Candidate

Urvashi Dube is a doctoral candidate in Department of Biological Sciences at BITS-Pilani. She holds a M.Tech and M.Sc in biotechnology from BITS-Pilani. After joining Ph. In 2007 she was awarded the CSIR, Senior Research fellowship and has successfully completed the same. In her thesis she aimed at identifying the genetic factors associated with Type 2 Diabetes Mellitus.

She has worked on several Research projects investigating the Genetic basis of Non communicable diseases like Obesity, depression, rheumatoid arthritis and pancreatitis. She has been pursuing the field of genetic epidemiology and her work has been published in different peer reviewed journals and presented in various national and international conferences. Her scientific exposure is of not only molecular biology but also genetic epidemiology and population genetics.

Her good command over statistical tools and data handling is due to the exposure she had in international workshops on "Ethical Issues in Human Subject Research and Biostatistics for Health Research" by Johns Hopkins University (JHU), Balitmore, USA held at Lucknow and Fogarty workshop "Introduction to Chronic Diseases Research Methods" organized by University of Pittsburgh and Share India, held at Hyderabad. Along with her research work she has been actively involved in organizing scientific workshops aimed at encouraging the young school students to open their mind to "Science, Imagination and Discovery". She has also been involved in various community based activities which are targeted at providing basic medical care to the grass root level people in India.

Bio Sketch of Dr. Suman Kapur

Dr. Suman Kapur joined BITS, Pilani on 17th July 2004 as Professor in the center for Biotechnology, department of Biological Sciences. She is presently working as the Dean, International Programmes and Collaboration Division since June 2012. She has earlier served in several senior administrative positions like Dean, Research & Consultancy at the Hyderabad Campus and Chief of Community Welfare and International Relations Unit at BITS-Pilani.

Dr. Kapur is a popular teacher at BITS, Pilani and has been instrumental in introducing several new courses, namely MPH G513, MPH G522, MPH G692, MPH G539, MPH G521, MPH G681,BIO G515 and has developed the curriculum for a new degree program "Master of Public Health", incorporating learning through field visits and interdisciplinary teaching. Several of her students have gone to make excellent careers for themselves as CEO's of start up companies and faculty at some of the best institutions in USA.

Dr. Kapur with her team of a dozen research scholars has been instrumental in building a state of the art Human Genomics laboratory from funds (Rs 362.19 lacs) received as Principal and/or Co-Investigator of now more than eighteen grants awarded since her joining BITS in 2004. As a mentor she has been able to motivate younger faculty to submit and execute independent grants in the form Women scientist (DST), Research Associate and senior research fellows (ICMR & CSIR). She has published more than 80 research articles in International and national journals.

Dr. Kapur's research interests lie in identifying biomarkers for unraveling the genetic basis of human diseases such as psychiatric disorders like depression, schizophrenia, addiction and Alzheimer's disease and metabolic disorders such as diabetes (T2DM), obesity, cataract and metabolic syndrome. Early and specific diagnosis is the backbone of effective treatment and reduction of both disease associated morbidity and mortality. Modern day integration of electronics and biological possibilities on an integrated chip can be successfully used to develop POC devices, especially suited for low-cost settings and our group has already developed two such devices which will be launched in 2013 and 2014. Attempts for early, non-invasive diagnosis of Urinary bladder and prostate cancer are underway.

In modern times several research approaches suggest that liability to complex inherited illnesses like obesity, diabetes, mental disorders and neurodegenerative diseases is influenced by several genes. Study of the involved genes will shed light on genetic architecture of these illnesses. More over the genetic profile of different populations for the complex disorders will serve as platform to diagnose the at risk individuals at an early age and help design strategies for early and timely intervention of the disease. The focus of the group is on genetic basis of human diseases with ongoing work in the field of Addictive disorder, Psychiatric disorder like Depression and Schizophrenia, Alzheimer's disease, Diabetes (T2DM) and Metabolic Syndrome. The genes being studied are APOE, CAPN, PPARgamma, ALDH2, ADH1B, ADH1C, Oprm1, OB, TPH, CRYGA, CRYGB, ADCY4, ADCY3, CCKAR, CCKBR, CF508, SPNK-1, PS-1, CYP2E1, CTSB, HSP70, TNFalpha, PRSS-1, CRP, HLA, and several micro-satellite markers on chromosome segments 1, 2, 3, 6, 7, 10 and 12. **Ours is the first group to show that in the Indian population a mutation in the mu opiate receptor is linked to risk for addiction to opiates, a mutation in the Ob (leptin)**

gene may be linked to hypertension in depressed individuals and similarly mutations in CRYGA & B, SPHK and SPAG 16 genes are linked to Cataract.

Revival of research on Traditional Medicine/Herbal Remedies with a locally-relevant evidence-based, disease-oriented approach is particularly relevant for India. Her group is also involved in developing clonal variants of Indian Medicinal plants and screening natural products for anti-diabetic, anti-inflammatory and anti-obesity activities in specific animal models for these diseases. Several industry sponsored projects are also underway for bio-conversion, -remediation & effluent treatment using consortia of microbial populations.