Faculty of Medicine

University of Helsinki

Helsinki, Finland

# Delineation of genetic variants modulating quantitative metabolic traits in an Arab population

**Prashantha Hebbar Kiradi**

Doctoral Programme in Population Health,

University of Helsinki

AND

Dasman Diabetes Institute

Kuwait

DISSERTATION

Doctoral thesis, to be presented for public examination with the permission of the Faculty of Medicine of the University of Helsinki, in lecture hall 2 (40), Haartman Institute, Helsinki, on the 29th of November 2021 at 12 o'clock.

ISBN 978-951-51-7487-1 (paperback)

ISBN 978-951-51-7488-8 (PDF)

http://ethesis.helsinki.fi

Unigrafia Oy

Helsinki 2021

The Faculty of Medicine uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations.

Supervised By:

**Professor Emeritus Jaakko Tuomilehto**
Department of Public Health,
University of Helsinki, Helsinki, Finland
and
National Institute for Health and Welfare (THL),
Helsinki, Finland

**Adjunct Professor Janne Pitkäniemi**
Finish Cancer Society,
Helsinki, Finland
and
Department of Public Health,
University of Helsinki, Helsinki, Finland

**Dr. Alphonse Thanaraj Thangavel**
Population genomics and genetic epidemiology
Dasman Diabetes Institute, Kuwait

Reviewed by:

Professor Johannes Kettunen
Genetics and Biostatistics,
Faculty of Medicine,
University of Oulu, Oulu, Finland

Assistant Professor Juan Rodriguez-Flores
Department of Genetic Medicine,
Weill Cornell Medical College, New York, USA

Opponent:

Professor Reedik Mägi
Institute of Genomics,
Estonian Genome Center,
University of Tartu, Tartu, Estonia

***I dedicate this work to the participants of this study and patients distressed with metabolic health complications.***

# Abstract

**BACKGROUND:** The highly inbred Arab population of the Middle East region has been experiencing complex metabolic health catastrophe compared with populations from other continents, ever since adopting a change in lifestyle in the rich post-oil era. Genetic risk variants of metabolic health featuring biological and pharmacological properties have been typically established for European and other ethnicities, but such data are scarce for the Arab population. Characterization of genetic risk variants by genome-wide profiling of genetic variations and phenotypes of metabolic health in this population would help understand population specific disease etiology. This would pave the way for early diagnosis and disease management strategies to mitigate cardio-metabolic disorders.

**AIM:** This study employs single nucleotide variant association analysis to 1) discover variant association with 13 different quantitative metabolic traits using genome-wide genotype data, 2) characterize runs of homozygous regions and examine the association of variants harbored in such regions with metabolic traits, 3) impute untyped variants and assess their association with quantitative metabolic traits of the Arab population.

**MATERIALS AND METHODS:** The study is focused on two cohorts, namely the Kuwait Obesity Genetics Project (KOGP) and the Kuwait Diabetes Epidemiology Program (KDEP), who comprise inhabitants of Kuwait. The genotype and phenotype profiles for anthropometry, blood pressure, glycemic and lipid levels of participants were obtained from blood DNA and plasma/serum specimens, respectively. Overall, the cohorts comprised 1965 individuals (KDEP; genotyped genome-wide using Illumina HumanCardio-MetaboChip array or genotyped by TaqMan targeted SNP Genotyping Assay) and 1350 individuals (KOGP; genotyped genome-wide using Illumina HumanOmniExpress array). The respective genome-wide variant data obtained were subjected to quality control measures and used to analyze variants association with 13 different quantitative traits. Quality control, statistical association tests, and genotype imputation were conducted using GenomeStudio and PLINK, PLINK or RVTests, and Michigan Imputation Server, respectively. Moreover, runs of homozygous regions were identified using PLINK. Several secondary analyses were performed for inferring the functional implications of identified association signals.

**RESULTS:** Examination of variant association using genome-wide genotyping and/or imputing revealed several variants modulating metabolic traits. Upon using only the genotyped variants, risk variants from the following genes were revealed at genome-wide significant p-values ($P \leq 5E-08$): RPS6KA1, LAD1, Or5v1, [CTTNBP2-LSM8], PGAP3, [RP11-191L9.4-CERK], ST6GALNAC5, [SPP2-ARL4C], NPY1R, [LINC00911-FLRT2], [CDK12-NEUROD2], and STARD3 for serum triglycerides (TG); TCN2 for waist circumference; and RPS6KA1, CADPS, and [VARS and VWA7] for fasting plasma glucose (FPG). All these genes, except TCN2, were located in regions of runs of

homozygosity. Furthermore, imputing untyped genotypes and combining genome-wide association (GWA) signals from both the study cohorts revealed 70 unique variants from 9 gene loci (7 genomic regions) at genome-wide significant p-values. Among these variants, 63 were associated with HDL-C, 1 with TG, 1 with LDL-C, 3 with SBP, and 2 with each of FPG and DBP. The gene loci included CETP, [INTS10, LPL], and [LOC105377613, LOC105377614] for HDL-C; CSMD1 for elevated FPG levels; [DYRK1A, LOC105372798] and RTN4 for SBP; RTN4 for DBP, BUD13 for TG, and [INTS10, LPL] for LDL-C. Fine mapping of these genomic regions identified 8 lead variants with most plausible causal variant sets. Moreover, 83 variants (from 42 genes) were found to replicate with exact metabolic traits reported in the catalog of NHGRI-EBI GWA studies (GWAS Catalog). Analysis using GTEx-eQTL data and literature reports on genotype-tissue specific expression in human or animal corroborated that many of these variants affect gene expression of harboring or neighboring genes and are involved in metabolic pathobiological pathways/diseases, respectively.

**CONCLUSION:** This study comprehensively evaluated the Kuwaiti Arab population-specific distributions of genetic risk variants associated with metabolic traits. Many of the variant associations identified were not previously reported in global GWAS studies. Hence, these provide new insights into the Arab specific etiology of the metabolic syndrome and could contribute to early intervention, prevention, and management of cardio-metabolic disorders more effectively in this population.

# Tiivistelmä

**TAUSTAA:** Elintapojen muuttumisen myötä Lähi-idän vauraiden öljyvaltioiden arabiväestöä, jossa esiintyy lähisukulaisten välisiä avioliittoja enemmän kuin monissa muissa väestöissä, on kohdannut vakava ja monimuotoinen metaboliasairauksien epidemia. Eurooppalaisten ja muiden etnisten ryhmien keskuudessa on tunnistettu metaboliseen terveyteen vaikuttavien geenien riskivariantteja, joihin liittyy biologisia ja farmakologisia erityispiirteitä, mutta arabiväestön osalta vastaavaa tietoa on olemassa vain vähän. Riskivarianttien tunnistaminen metaboliseen terveyteen vaikuttavien geenien ja fenotyyppien variaation genominlaajuisen tutkimuksen avulla voi auttaa ymmärtämään kyseisessä väestössä tällaisten yleisesti esiintyvien sairauksien etiologiaa. Näin voidaan edistää varhaista diagnostiikkaa ja kardiometabolisen sairastavuuden vähentämiseen tähtääviä toimenpiteitä.Tässä tutkimuksessa arvioitiin laajasti Kuwaitin arabiväestössä tavattavien riskivarianttien esiintymistä ja niiden yhteyttä metabolisiin sairauksiin.

**TAVOITTEET:**Tutkimuksessa käytettiin yhden nukleotidin variaation assosiaatioanalyysiä, jonka avulla 1) selvitettiin varianttien yhteyttä 13 metaboliseen muuttujaan koko genomin kattavassa genotyyppidatassa, 2) tunnistettiin homotsygoottisia jaksoja sisältäviä alueita ja tarkasteltiin kyseisillä alueilla esiintyvien varianttien yhteyttä metabolisiin muuttujiin, 3) imputoitiin tyypittämättömiä variantteja ja arvioitiin niiden yhteyttä kvantitatiivisiin metabolisiin muuttujiin arabiväestössä.

**AINEISTO JA MENETELMÄT:** Tutkimusaineisto koostui kahdesta kohorttiaineistosta, *Kuwait Obesity Genetics Project* (KOGP) ja *Kuwait Diabetes Epidemiology Program* (KDEP), joihin kuului Kuwaitissa asuvia ihmisiä. Tutkittavien antropometriaa, verenpainetta sekä veren glukoosi- ja lipiditasoja koskevat geno- ja fenotyyppiprofiilit saatiin verestä otetuista DNA-näytteistä ja plasma-/seeruminäytteistä. Kohortteihin kuului yhteensä 1965 (KDEP, koko genomin genotyypitys Illumina HumanCardio-MetaboChip -menetelmällä tai kohdennettu SNP-genotyypitys TaqMan-menetelmällä) ja 1350 henkilöä (KOGP, koko genomin genotyypitys Illumina HumanOmniExpress -menetelmällä). Tälle koko genomia koskevalle varianttidatalle tehtiin laadunvarmistustoimenpiteitä, ja datan avulla analysoitiin varianttien yhteyttä 13 kvantitatiiviseen metaboliseen muuttujaan. Laadunvarmistukseen käytettiin GenomeStudio- ja PLINK-ohjelmia, tilastollisiin assosiaatiotesteihin PLINK- tai RVTests-ohjelmia ja genotyyppien imputointiin Michigan Imputation Server -palvelinta. Homotsygoottisia jaksoja sisältävien alueiden tunnistamiseen käytettiin PLINK-ohjelmaa. Lisäksi tunnistettujen assosiaatiosignaalien merkityksen selvittämiseksi tehtiin useita sekundäärisiä analyysejä.

**TULOKSET:** Koko genomin genotyypitykseen ja/tai imputointiin perustuvassa varianttien assosiaatiotutkimuksessa löydettiin useita metabolisiin muuttujiin vaikuttavia variantteja. Pelkkiä genotyypitettyjä variantteja käyttämällä tunnistettiin seuraavien geenien riskivariantit, joiden p-arvot olivat tilastollisesti merkitseviä koko genomin kattavassa tutkimuksessa ($P \leq 5E$-

08): RPS6KA1, LAD1, Or5v1, [CTTNBP2-LSM8], PGAP3, [RP11-191L9.4-CERK], ST6GALNAC5, [SPP2-ARL4C], NPY1R, [LINC00911-FLRT2], [CDK12-NEUROD2] ja STARD3 seerumin triglyseridien (TG); TCN2 vyötärönympäryksen; sekä RPS6KA1, CADPS ja [VARS ja VWA7:n ylävirta-alueen 2 Kb] plasman paastoglukoosin (FPG)osalta. Kaikki kyseiset geenit, lukuun ottamatta geeniä TCN2, paikallistettiin alueille, joilla esiintyi homotsygoottisia jaksoja. Lisäksi imputoimalla tyypittämättömiä genotyyppejä sekä yhdistämällä molemmista tutkimuskohorteista genominlaajuisessa assosiaatiotutkimuksessa (GWA) saatuja signaaleja löydettiin 70 uniikkia varianttia 9 lokuksesta (7 genomialueelta) p-arvojen ollessa tilastollisesti merkitseviä koko genomin kattavassa tutkimuksessa. Näistä varianteista 63 assosioitui HDL-kolesteroliin (HDL-C), 1 TG:iin, 1 LDL-kolesteroliin (LDL-C), 3 systoliseen verenpaineeseen, 2 FPG:iin sekä 2 diastoliseen verenpaineeseen. Geenilokukset olivat CEPT [INTS10, LPL] ja [LOC105377613, LOC105377614] HDL-C:n, CSMD1 kohonneiden FPG-tasojen, [DYRK1A, LOC105372798] ja RTN4 systolisen verenpaineen, RTN4 diastolisen verenpaineen, BUD13 TG:n sekä [INTS10, LPL] LDL-C:n osalta. Näiden genomialueiden tarkemmassa tutkimuksessa paljastui 8 tärkeintä varianttia, joilla on todennäköisimmin kausaalinen rooli. Lisäksi 83 variantin (42 eri geeniä) havaittiin replikoituvan tarkasti niiden metabolisten tekijöiden kanssa, jotka on raportoitu NHGRI-EBI GWA -tutkimusten yhteydessä (GWAS Catalog). GTEx-eQTL-datan analysoinnin sekä genotyypin ja kudoksen ekspressiota ihmisellä tai eläimillä koskevien kirjallisuustietojen perusteella monet näistä varianteista vaikuttavat kyseisten geenien tai viereisten geenien ilmentymiseen ja ovat siten osallisia metabolisten sairauksien kehittymisen biologiassa.

**JOHTOPÄÄTÖKSET:** Useita tässä assosiaatiotutkimuksessa tunnistettuja variantteja ei ole raportoitu aiemmissa muualla tehdyissä GWAS-tutkimuksissa. Näin ollen ne tarjoavat uutta tietoa arabiväestössä esiintyvän metabolisen oireyhtymän etiologiasta ja voivat auttaa kardiometabolisten sairauksien varhaisessa hoidossa, ennaltaehkäisyssä ja hallinnassa.

**Table of contents** **Page**

# 1. Publications

## 1.1. List of Original publications

This thesis is based upon already published research

1. Hebbar P*, Elkum N*, Alkayal F, John SE, Thanaraj TA†, Alsmadi O†., Genetic risk variants for metabolic traits in Arab populations., Sci Rep. 2017 Jan 20;7:40988. doi: 10.1038/srep40988. [PMID: 28106113]
2. Hebbar P*, Alkayal F*, Nizam R, Melhem M, Elkum N, John SE, Abufarha M, Alsmadi O†, Thanaraj TA†., The TCN2 variant of rs9606756 [Ile23Val] acts as risk loci for obesity-related traits and mediates by interacting with Apo-A1., Obesity (Silver Spring). 2017 Jun;25(6):1098-1108. doi: 10.1002/oby.21826. [PMID: 28417558]
3. Hebbar P, Nizam R, Melhem M, Alkayal F, Elkum N, John SE, Tuomilehto J, Alsmadi O†, Thanaraj TA†., Genome-wide association study identifies novel recessive genetic variants for high TGs in an Arab population., J Lipid Res. 2018 Oct;59(10):1951-1966. doi: 10.1194/jlr.P080218. [PMID: 30108155]
4. Hebbar P, Abu-Farha M, Alkayal F, Nizam R, Elkum N, Melhem M, John SE, Channanath A, Abubaker J, Bennakhi A, Al-Ozairi E, Tuomilehto J, Pitkaniemi J, Alsmadi O†, Al-Mulla F†, Thanaraj TA†., Genome-wide association study identifies novel risk variants from RPS6KA1, CADPS, VARS, and DHX58 for fasting plasma glucose in Arab population. Sci Rep 10, 152 (2020) doi:10.1038/s41598-019-57072-9. [PMID:31932636]
5. Hebbar P, Abubaker JA, Abu-Farha M, Alsmadi O, Elkum N, Alkayal F, John SE, Channanath A,Iqbal R, Pitkaniemi J, Tuomilehto J, Sladek R, Al-Mulla F†, Thanaraj TA†., Genome-wide landscape establishes novel association signals for metabolic traits in the Arab population,Human Genetics, 2020 Sep 9. doi: 10.1007/s00439-020-02222-7. [PMID: 32902719]

* Contributed equally to the work as first authors

† Contributed equally to the work as corresponding authors

## 1.2. List of publications not included in the Doctoral thesis

i. Hammad M*, Abu-Farha M*, Hebbar P*, Cherian P, Al Khairi I, Melhem M, Alkayal F, Alsmadi O, Thanaraj TA†, Al-Mulla F†, Abubaker J†, MC4R Variant rs17782313 Associates With Increased Levels of DNAJC27 Which Mediates Obesity by Modulating cAMP Formation, Frontier's in Endocrinology, 2020, Jul 7;11:437.doi: 10.3389/fendo.2020.00437. [PMID: 32733386]

ii. Hebbar P*, Abu-Farha M*, Mohammad A, Alkayal F, Melhem M, Abubaker JA†, Al-Mulla F†, Thanaraj TA†., FTO variant rs1421085 associates with increased body weight, soft lean mass and total body water through interaction with ghrelin and apolipoproteins in Arab population., Frontier's in Genetics, 2020 Jan 31;10:1411, doi: 10.3389/fgene.2019.01411. [PMID: 32076432]

iii. Hebbar P, Abubaker J, Abu-Farha M, Tuomilehto J, Al-Mulla F†, Thanaraj TA†., A perception on genome-wide genetic analysis of metabolic traits in Arab populations., Frontier's in Endocrinology, 2019, Jan 28;10:8. doi: 10.3389/fendo.2019.00008. [PMID: 30761081]

iv. John SE, Antony D, Eaaswarkhanth M, Hebbar P, Alkayal F, Tuomilehto J, Alsmadi O†, Thanaraj TA†., Genetic variants associated with warfarin dosage in Kuwaiti population., Pharmacogenomics. 2017 Jun;18(8):757-764. doi: 10.2217/pgs-2017-0020. [PMID: 28592190]

v. Alsmadi O, Melhem M, Hebbar P, Thareja G, John SE, Alkayal F, Behbehani K, Thanaraj TA., Leptin in association with common variants of MC3R mediates hypertension., Am J Hypertens. 2014 Jul;27(7):973-81. doi: 10.1093/ajh/hpt285. [PMID: 24487982]

vi. Alsmadi O, Melhem M, Hebbar P, Channanath A, Thareja G, John SE, Alkayal F, Behbehani K, Thanaraj TA., Response to leptin and nitric oxide in blood pressure regulation in humans., Am J Hypertens. 2014 Nov;27(11):1429-30. doi:10.1093/ajh/hpu178. [PMID: 25273037]

* Contributed equally to the work as first author

† Contributed equally to the work as corresponding author

## 2. Abbreviations

| | |
|---|---|
| ACC/AHA | American College of Cardiology and American Heart Association |
| ADA | American Diabetes Association |
| ALK | Arabs Living in Kuwait |
| BH-FDR | Benjamini and Hochberg False Discovery Rate |
| BMI | Body Mass Index |
| BP | Blood pressure |
| CAD | Coronary Artery Disease |
| CNS | Central nervous system |
| CNV | Copy number variation |
| CVD | Cardiovascular disorder |
| DBP | Diastolic Blood Pressure |
| DKD | Diabetic Kidney Disorder |
| DNA | Deoxyribonucleic Acid |
| DYSIS | Dyslipidemia International Study |
| ECG | Electrocardiogram |
| FBAT | Family-based association test |
| FFA | Free Fatty Acid |
| FPG | Fasting Plasma Glucose |
| FTO | Fat mass and obesity |
| GCC | Gulf Cooperation Council |
| GME | Greater Middle Eastern |
| GWAS | Genome-Wide Association Study |
| GTEx | Genotype-Tissue Expression |
| HbA1c | Hemoglobin A1c |
| HDL-C | High Density Lipid Cholesterol |
| HGDP | Human genome diversity project |
| HOMA-IR | Homeostatic Model Assessment of Insulin Resistance |
| HWE | Hardy-Weinberg Equilibrium |
| IBD | Identity by descent |
| IGT | Impaired glucose tolerance |
| KDEP | Kuwait Diabetes Epidemiology Program |
| KOGP | Kuwait Obesity Genetics Project |
| 1KGP | 1K genome project |
| LD | Linkage Disequilibrium |
| LDL-C | Low Density Lipid Cholesterol |
| LMM | Linear Mixed Model |
| LOD | Logarithm of the odds |
| LOH | Loss of heterozygosity |
| MAF | Minor Allele Frequency |
| Mb | Mega bases |

| | |
|---|---|
| MENA | Middle East and North Africa |
| MODY | Maturity onset diabetes in Young |
| NAFLD | Non-alcoholic fatty liver disease |
| NCD | Non-communicable disease |
| NPL | Nonparametric linkage |
| OOA | Out of Africa |
| OGTT | Oral Glucose Tolerance Test |
| OPEC | Organization of the Petroleum Exporting Countries |
| PNDM | Permanent neonatal diabetes mellitus |
| RFLP | Restriction fragment length polymorphism |
| ROH | Runs of Homozygosity |
| RNA | Ribonucleic acid |
| RT-PCR | Real Time-Polymerase Chain Reaction |
| SBP | Systolic Blood Pressure |
| SKAT | Sequence kernel association test |
| SLM | Soft lean mass |
| SNP | Single Nucleotide Polymorphism |
| STR | Short tandem repeat |
| TBW | Total body water |
| TC | Total Cholesterol |
| TDT | Transmission disequilibrium test |
| TF | Transcription factor |
| TG | Triglyceride |
| TNDM | Transient neonatal diabetes mellitus |
| T1DM | Type 1 diabetes mellitus |
| T2DM | Type 2 diabetes mellitus |
| QC | Quality control |
| QT | Quantitative trait |
| QTDT | Quantitative transmission disequilibrium test |
| QTL | Quantitative Trait Loci |
| VNTR | Variable number tandem repeats |
| WC | Waist Circumference |
| WcHtR | Waist Circumference to Height Ratio |
| WES | Whole Exome Sequencing |
| WGS | Whole genome sequencing |

## 3. List of tables

analysis. R-regular correction: Corrected for age, sex, and the top 10 principal components that resulted from the principal component analysis of the genotype data; LM: Corrected for lipid medication in addition to the regular correction.

Table 15: Results of the statistical association tests from the discovery, replication, and meta-analysis phases with TG (p-values approaching borderline to genome-wide significance at >6.12E-09 &<5.0E-08) upon using recessive and additive models. [#], Recessive Model; [@], Additive Model

Table 16: List of the four identified risk variants associated with FPG either at genome-wide significant p-values (<1.8E-08) or at nominal p-values of <1.0E-06 and association of the same variants with HbA1c. EffectSize, Effect size represents beta value for discovery and replication phases, and Z-score for meta-analysis. R-regular correction: Corrected for age, sex, and the top 10 principal components; DM: Corrected for diabetes medication in addition to the regular correction; #, association was observed under recessive mode of inheritance; @, association was observed under additive mode of inheritance.

Table 17: Summary of ROH mapping of 4 variants associated with FPG and their proximity to ROH regions identified from two different methods (for details see relevant section of study III and IV).

Table 18. Summary statistics of 8 lead SNVs from 9 associations with p-value ≤ 5E-08.

# 4. List of figures

Figure 12: Distribution of total ROHs & ROH segments, and ROH vs. ROH segments in three Kuwaiti groups; KW1 (Persian) and KW2 (Saudi) show more total ROH and ROH segments, whereas KW3 (Bedouin) shows very less, suggesting KW1 and KW2 were a highly inbred population.

Figure 13: A) regional plot of rs12440118 (ZNF106, W > R) at 1 Mb region center to it, along with LD variants chr15:40433466, $r^2$ = 0.61; chr15:40471747, $r^2$ = 0.53; chr15:40490502, $r^2$ = 0.33; chr15:40497290, $r^2$ = 0.62; chr15:40497469, $r^2$ = 0.64; chr15:40500754, $r^2$ = 0.48; chr15:40503873, $r^2$ = 0.62; chr15:40506209, $r^2$ = 0.64; chr15:40507580, $r^2$ = 0.64; chr15:40510699, $r^2$ = 0.63; chr15:40513902, $r^2$ = 0.76; chr15:40514171, $r^2$ = 0.52; chr15:40514188, $r^2$ = 0.52; chr15:40517485, $r^2$ = 0.34; chr15:40528928, $r^2$ = 0.52. B) Variation observed in effect allele frequency of reported SNVs among the three subgroups of Kuwait Arab population.

Figure 14: A) MAF distribution of rs9606756 among the three subgroups of Kuwaiti Arab population; B and C) Interaction between WC vs. Apo-A1 and HDL-C, respectively, at three genotype levels of rs9606756.

Figure 15: Minor allele frequencies of associated variants from lipid and glycemic traits analysis in subgroups of Kuwait Arab population.

Figure 16: Bubble plots illustrating the distributions of observed established variants (A), gene loci (B), and publications (C) reporting for the observed associations in GWAS Catalog with the study-specific metabolic traits with direct, indirect, or broad relationships.

Figure 17: Distribution of ethnic transferable SNP associations for 13 metabolic traits among global populations. A) represents SNP counts from “direct” and B) represents SNP count from “indirect” relationship.

Figure 18: Effect allele frequency distribution among subgroup population for 42 variants selected from 163 genotyped variants upon clumping at $r^2$ = 0.1.

# 5. Introduction

The fortuitous discovery of oil and thriving prosperity in Middle Eastern (ME) countries such as Bahrain, Iraq, Iran, Kuwait, Oman, Qatar, Saudi Arabia, and UAE rapidly altered the lifestyle of the Arab population. Consequently, an increase in the prevalence of chronic metabolic disorders has been observed. Recent surveys conducted among the adult population of Qatar and Kuwait have shown that 35% & 36% of men and 45% & 48% of women, respectively, are obese (Bener et al. 2004). The prevalence of obesity is similar in Saudi Arabia with an annual increase by 4.1% in men and 1.5% in women (Ng et al. 2011). Similar disturbing statistics were also noted among young adolescents (Al-Sendi et al. 2003; Ng et al. 2011). Conspicuously, contemporary pharmacological treatments for metabolic disorders particularly to lower glycemic and lipid levels have been noticed to be less efficacious in Arabs as compared to their European counterparts (Al Sifri et al. 2014; Qaddoumi et al. 2019). Several candidate gene studies have examined variants found in association with the European and Asian populations for obesity and type 2 diabetes mellitus (T2DM) (e.g., PPARγ, KCNJ11, TCF7L2, SLC30A, ABCC8, HHEX, CDKN2A, IGF2BP2, CDKAL1 and FTO) in Arab population. They observed that these associations were not completely transferable to the Arab population or that they do not exhibit the same high effect size as the European population (Al-Daghri et al. 2014b; O'Beirne et al. 2016). This inconsistency could be attributable to differences in constellation of risks and protective variants, or causal and lead variants, or environmental exposure in the non-European populations. Therefore, population-wide differences raise significant concerns about the present medication regimen and warrant a comprehensive genetic and molecular level research in the Arab population.

From the perspective of human evolution, Arabian Peninsula is regarded as the prime landmass of modern human migration from Africa to elsewhere and contact point for introgression of Homo sapiens with Neanderthals (Rodriguez-Flores et al. 2016). Historically, this region also witnessed recent Europeans (or Caucasus related ancestry) in conjunction with Greek (during the 3$^{rd}$ century BC), Roman and Turkish conquests (Haber et al. 2020; Jonathan Wallace 1979). Furthermore, African and South Asian admixtures due to slavery of Africans (during the 2$^{nd}$ century BC) and business engagement with South Asians respectively, have sculpted a distinct genetic architecture in the region (Lovejoy 2011; Segal 2002). Admirably, the contemporary Arabs are diverse in their ethnic, religious, and political behavior but still descend from a common linguistic and cultural heritage. Their proclivity toward consanguineous marriage is widespread with an estimated prevalence of over 50% (Al-Awadi et al. 1985; al-Gazali et al. 1997; el-Hazmi et al. 1995). Since consanguinity is known to increase inbreeding signatures (such as runs of homozygosity and identity by descent), and the recessive deleterious mutations (Teeuw et al. 2010), the region also evidenced excess of recessive genetic disorders (Tadmouri et al. 2006; Tadmouri et al. 2014). Due to these preponderances, the Arabian Peninsula offers a novel population for the study of genetic metabolic diseases.

Since the first successful GWA study in 2005 on age-related macular degeneration (Klein et al. 2005), GWAS has been the recommended tool for identification of genetic signatures associated

with many common disorders. These studies have uncovered many significant associations between genetic variants and biological traits and helped in identifying biological and pharmacological impactful risk loci in different populations. However, so far, GWA studies were predominantly carried out among the European population (about 88%) with 72% of discoveries in participants recruited from several countries including the USA, UK, Finland, and Iceland (Mills and Rahal 2019). Interestingly, only 25% of the variants found associated with BMI, T2DM, and lipid levels in European-Americans were trans-ethnic and such associations vary in at least one out of the five non-European ancestry populations (Popejoy and Fullerton 2016). Hence, GWAS that is inclusive of the understudied populations enhance insights of complex architecture of metabolic traits and provide comprehensive picture of variants involved in etiology of metabolic disorders. Since the Arab population is highly inbred and rich in high frequencies of rare variants, utilizing such populations in GWAS benefits the population in question and the global populations as well, in whom rare variants are insufficient to detect associations with clinically important phenotypes.

Despite high prevalence of metabolic disorders in the Arab region, not many GWA studies have been conducted. A thorough literature search on PubMed revealed 21 candidate variant/gene studies (Jamaluddine et al. 2016), 4 family based studies (Al Safar et al. 2013; Shalata et al. 2013; Verge et al. 1998; Zadjali et al. 2013) and 5 genome-wide or exome-wide association analysis studies in relation to metabolic traits/disorders (Dajani et al. 2017; Ghassibe-Sabbagh et al. 2014; Saade et al. 2011; Yousri et al. 2018) were identified. Interestingly, most of these studies were tested using case-control status of metabolic disorders, but not their metabolic quantitative phenotypes. Since binary traits are usually the summarization of the trait derived from several quantitative traits, directly using quantitative traits would evaluate the biological mechanism of diseases more precisely.

In this study we genotyped approximately 2700 individuals with Arab ancestry in two dense SNV array platforms and evaluated their association with 13 metabolic quantitative traits relating to anthropometry, blood pressure, lipid, and glycemic levels using additive and recessive mode of inheritance models. Additionally, we identified runs of homozygous regions (ROH) and mapped variants showing association with metabolic traits onto ROH regions. Furthermore, we separately imputed missing variants from both SNV arrays and performed association tests with inverse normal transformed 13 metabolic traits. Meta-analysis and fine mapping were carried out to prioritize variants.

# 6. Review of literature

## 6.1 Introduction to Genetics and concepts of Genetic analysis

Genetics is a study of fundamental units of heredity and its transmission, particularly deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) in organisms. An Augustinian prelate, Gregor Johann Mendel (1822-1884) conducted a series of experiments using nearly 30,000 pea plants and showed that the traits are passed down from parents to offspring in a predictable manner. Each trait in the plant is controlled by a pair of factors (i.e., genes) and members of a gene pair separate from each other during gamete formation. This breakthrough established a strong foundation for 'Genetics', now referred to as the laws of Mendelian inheritance.

Advances in microscopy paved the way for better understanding of chromosomes, their behavior during cell division and their ploidy across different species. Most eukaryote species were discovered to have a characteristic number of chromosomes, designated as the diploid number (2n) in their cells and existed in homologous pairs. Transmission of chromosome number varies depending on the type of cell division, either mitosis or meiosis. In mitosis, each daughter cell receives a diploid set of chromosomes identical to those in the parental cell. Whereas in meiosis, cells receive only one chromosome from each chromosome pair and the resulting number of chromosomes is called the haploid number (n). Hence, an offspring arising from the fusion of egg and sperm maintain the constant number of chromosomes characteristic of their parents and other members of their species (William S Klug 2018).

**Genetic variation**

Genetic variation refers to differences in genome sequences between individuals within a population. Genetic variation arises from mutation and recombination. An error during DNA replication when not repaired by DNA repairing enzymes, *de novo* mutations occur (**Figure 1A**). External agents such as viruses, radiation, and chemical mutagens, may also induce changes in the DNA sequence. Mutations may be beneficial, deleterious or neutral regarding the fitness of the organism. Similarly, recombination of homologous DNA strands (or crossover) with shuffling of maternal and paternal DNA introduces new combinations of variants in the daughter germ-cells (**Figure 1B**). Such processes cause permanent change in the DNA sequence. Genetic variation occurs both in germ (i.e., egg and sperm) and somatic (i.e. all other cells) cells. But only the variation occurring in germ cells are inherited from parents to offspring, resulting in the changes in population dynamics. The beneficial mutations ultimately lead towards the path of evolution (William S Klug 2018).

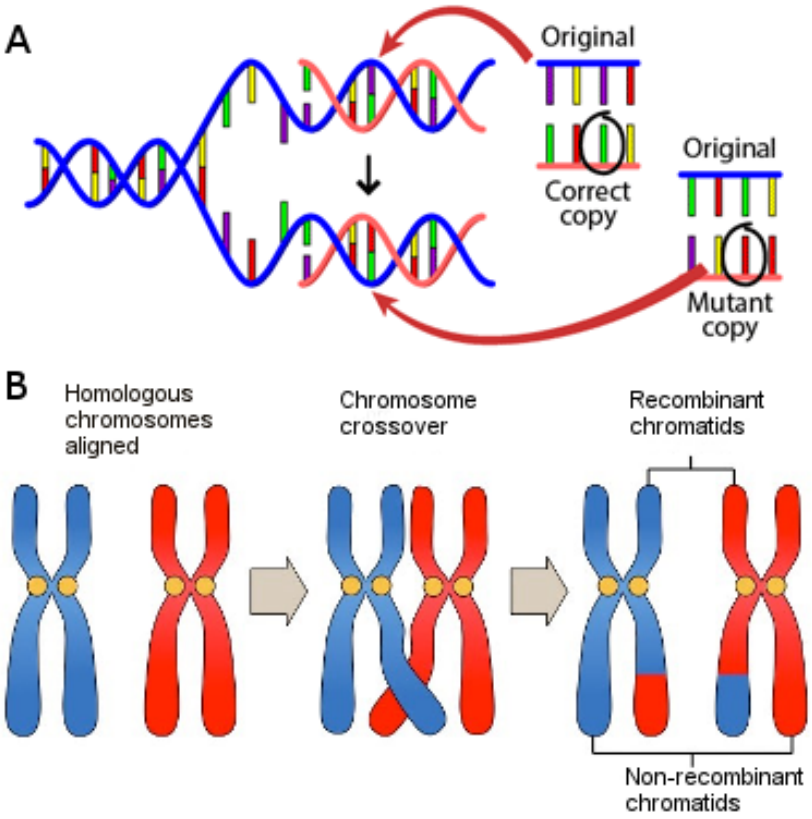


**Figure 1: Illustrate introduction of A) a mutation in DNA replication B) genetic variations in recombination process to constellation of DNA.**

**Image source:** A) https://evolution.berkeley.edu/evolibrary/article/evo_20 and B) https://ib.bioninja.com.au/standard-level/topic-3-genetics/33-meiosis/crossing-over.html

**Forms of genetic variation**

In the genome, genetic variations exist in many forms. They span across a spectrum of sizes from single nucleotides to mega-bases. Some of these genetic variations are depicted in **Figure 2**.

- Single nucleotide variations (SNVs) involve a change in a single nucleotide at a particular locus in the DNA sequence, which includes substitutions, single nucleotide insertions, and single nucleotide deletions. Substitution can be either transition (i.e., interchange of the purine such as Adenine/Guanine or pyrimidine such as Cytosine/Thymine nucleic acids) or transversion (i.e., interchange of a purine and pyrimidine bases).

- Indels (short insertions or deletions) are a single stretch of DNA sequence that can range from 2 to 50 bp (base-pairs) in length.
- Restriction fragment length polymorphisms (RFLPs) are single nucleotide substitutions. These variations are specifically recognized in the DNA sequences by restriction enzymes. The earlier disease gene mapping process utilized them as genetic markers. The genetic linkage map of RFLPs was constructed in the 1980s (Liu et al. 2017).
- Tandem repeats or microsatellites are types of polymorphisms in which genomic elements are prone to repetition in several numbers. Microsatellites are highly polymorphic in human populations. They have higher allelic states (as each repeat is considered as one allelic state) and are more informative than the bi-allelic genetic markers, such as SNPs. They are of two types- Short tandem repeats (STR) and Variable number tandem repeats (VNTR). STR are the more widely distributed microsatellites (>100,000) in the human genome. They have higher allelic diversity than RFLPs. Hence, they are commonly exploited as the genetic markers in linkage studies for monogenic disorders and complex diseases. Similarly, STRs also out-perform VNTRs in terms of their numbers, where there are only a few thousand VNTRs in the human genome. The availability of the genetic linkage map of microsatellites has resulted in immense success of linkage studies in identifying genes for monogenic disorders (Weissenbach et al. 1992). Although tandem repeats are more informative than SNPs at the individual marker level, their number is far less than the several million SNPs in the human genome. Thus, tandem repeats are not ideal genetic markers for applications that require high marker density or resolution, such as genome-wide association studies (GWASs) (Ku et al. 2010).
- Copy number variations (CNV) are another type of large polymorphisms with additions or deletions in the number of copies of a particular segment of DNA when compared with a reference genome sequence. The distinction between CNVs and indels is obscure. But typically, deletions and duplications/insertions larger than 50 bp are classified as CNVs (MacDonald et al. 2014). They are estimated to contribute 4.8–9.5% of the human genome (Zarrei et al. 2015).
- Inversions and translocations: Like duplication and deletion events, inversion and translocation events can be caused by breakage of DNA double helices in the genome at two different locations, followed by a re-joining of the broken ends to produce a new chromosomal arrangement of genes, which is different from the gene order of the chromosomes before they were broken. Sometimes these breaks will occur within genes and disrupt gene function. For both events, no chromosomal material is gained or lost, there is simply a change in the relative locations of genes on the rearranged chromosomes (Puig et al. 2015).
- Loss of heterozygosity (LOH): A change in polymorphic sites from a heterozygous state in the germline to an apparently homozygous state in somatic DNA lead to loss of the normal, functional allele at a heterozygous locus. This event is termed as loss of heterozygosity. It may originate as a consequence of a multi-locus chromosomal event

such as deletion, mitotic recombination, or non-disjunctional chromosome loss (i.e., failing to separate chromosomes at anaphase stage) with or without reduplication; a locus-restricted event, such as gene conversion or point mutation; or even an epigenetic allelic inactivation. As a result of these events the locus may change to (1) homozygous through mitotic recombination, gene conversion, or chromosome loss with reduplication, (2) hemizygous through deletion or chromosome loss without reduplication, (3) compound heterozygous through the introduction of a different point mutation in the second allele, or (4) may remain heterozygous, in terms of the DNA nucleotide sequence, if one allele is inactivated epigenetically. This phenomenon is often seen in inherited cancers. LOH is known to disrupt function of tumor suppressing gene product and normal growth of cell, either due to cell-cycle deregulation, defective DNA repair, or altered cell-to-cell communication (Tischfield 1997).

- Runs of homozygosity (ROH) or Autozygosity: ROHs are contiguous regions of the genome with homozygous segments across all polymorphic sites. This arises when parents have a common ancestor and pass shared chromosomal segments on to their progeny. It means that when identical by descent (IBD) segments from both the parents are passed on to an individual, it results in homozygous segments in the genome of the offspring. However, homozygosity may originate from either consanguineous or non-consanguineous parentage (Peripolli et al. 2017).

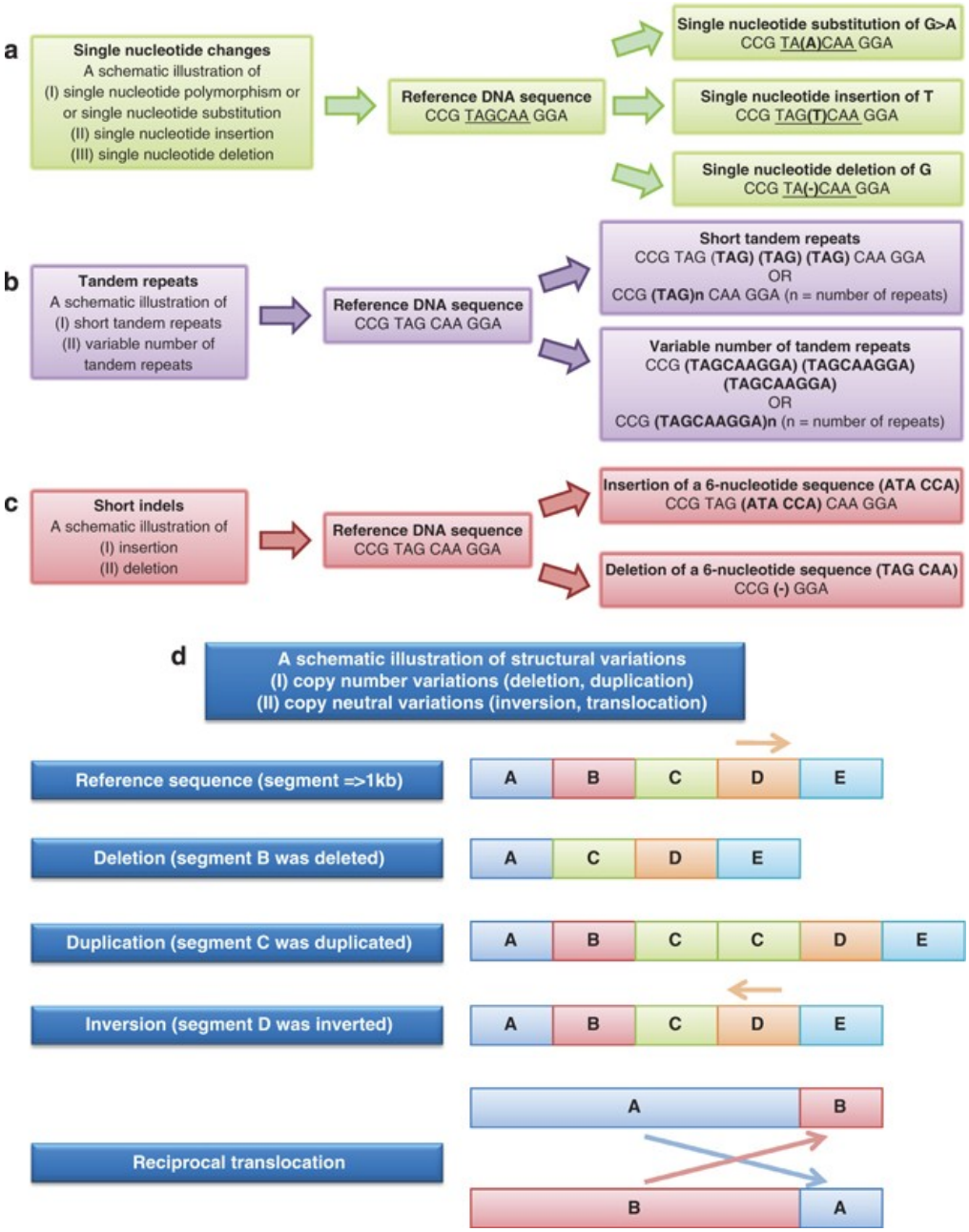


**Figure 2: A schematic illustration of different types of genetic variations in human genome. With an exception to SNVs, all the genetic variations are broadly classified as structural variations. Image source (Ku et al. 2010)**.

**Natural selection, genetic drift, and gene flow**

Mechanisms such as natural selection, genetic drift, and gene flow cause changes in allele frequency of genetic variations over time. Natural selection refers to the process through which individuals within a population adapt and change. Genetic drift refers to change in frequency of alleles within a population due to random sampling (i.e., not by adaptation). Genetic drift in a population can be either due to the bottle effect (for example, reduction of individuals carrying a particular trait allele due to catastrophic events) or the founder effect (for instance, when few members of a population breakaway and create their own group; the allele frequency of the new group can dramatically change due to selection pressure). Gene flow refers to transfer of alleles or gametes from one population to another. For example, when a population migrates or becomes geographically isolated. Natural selection is influenced by changes in environmental conditions while genetic drift is random and serendipitous.

**Linkage disequilibrium**

Linkage disequilibrium (LD) is an assortment of alleles in a chromosome more often than expected by chance (i.e., non-random) due to their physical proximity. LD patterns indicate genetic history of the population and helps inferring a population's effective size (*Ne*). LD is also an effective tool to determine the power and precision of disease association mapping in humans (Qanbari 2019).

**Inbreeding**

Inbreeding refers to a close relative mating system in which mating partners carry alleles that originated from a common ancestor. Hence, inbreeding increases the chances of receiving a deleterious recessive allele inherited from a common ancestor. It is measured by 'inbreeding coefficient' (proposed by Sewell Wright in 1922), which represents the probability of an offspring receiving an allele from each parent that is a copy of a single shared ancestral allele (Rousset 2002).

**Hardy-Weinberg Equilibrium (HWE)**

Hardy-Weinberg principle describes a hypothetical situation where allele and genotypic frequencies remain unchanged in a randomly mating, infinitely large population in the absence of selection, migration, or new mutations (Edwards 2008). If these assumptions are true,

- Allele frequencies remain constant from one generation to the next generation.
- After one or more generations of random mating, the genotype frequencies are in the proportions such that $p^2_{(AA)} : 2pq_{(Aa)} : q^2_{(aa)}$. The relation between frequencies of alleles and frequencies of genotypes under HWE is shown in **Figure 3**.
- For a population to be in HW equilibrium, the observed genotype frequencies must match to those predicted by the equation $p^2+2pq+q^2$.

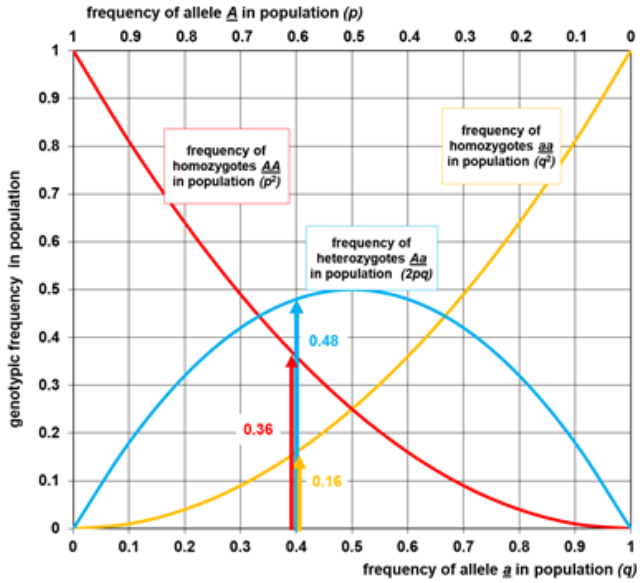


**Figure 3: Showing relation between frequencies of alleles and frequencies of genotypes under HWE.** Image source: https://www.wikilectures.eu/w/Genetic_aspects_of_populations,_Hardy-Weinberg_equilibrium

### 6.2 Quantitative genetics and other related concepts

A great diversity of phenotypic variation, from morphology to disease susceptibility is seen in organisms living in natural populations. This variation in phenotype is contributed by multiple interacting loci and their allelic sensitivity to the environmental conditions experienced by each individual. Phenotypes can be either dichotomous (e.g., disease status) or continuous (e.g., quantitative traits such as Height, BMI etc). Quantitative phenotypes can vary among individuals, over a wide range, to produce a continuous distribution of phenotypes. This distribution of the phenotypes set limitations in the analysis, as many of the statistical methods assume a normal distribution. If quantitative phenotype is clearly not following a normal distribution, adequate transformations can be invoked. Usually, binary traits are the summarized information of quantitative traits (for example, dyslipidemia status is derived by using a certain threshold to continuous phenotypes of lipids). The loss of information in this transformation process leads to a reduced potential of binary traits in genetic analysis (Duggirala et al. 1997).

The time of origin of quantitative genetics is similar to that of Mendelian genetics. Mendel applied quantitative data analysis to his pea plant experimental results and showed that traits are passed from parents to offspring in specific patterns. The results of his pea plant crosses involved seven different characteristics- plant height, seed texture, seed color, flower color,

peapod size, peapod color, and flower position, each with two contrasting traits. However, the mathematical modeling of inheritance was introduced by Sir Francis Galton (1822–1911). He introduced the concepts of regression and correlation to study continuous variation in humans while analyzing inheritance pattern in parent and offspring height (Galton 1989). Galton, Karl Pearson (1857–1936), and Walter Weldon (1860–1906) together founded the scientific journal '*Biometrika'*, to publish articles on the biometrical theory of inheritance, in 1901. A British statistician, Ronald A. Fisher (1890–1962), introduced several statistical methods and concepts, including the distribution of correlation coefficient, maximum likelihood, analysis of variance and Fisher information (Fienberg Stephen E 1980) related to genetics research. He postulated that genetic variation of many loci with small individual allelic effects and random environmental variation simultaneously contributed to continuous phenotypic variation. Unlike Mendelian inheritance, according to which the action of dominant or recessive genes contributed to phenotypic variation, Fisher assumed a more general model of gene action at a single locus that could account for any relationship between homozygous and heterozygous allelic effects and derived expectations of the magnitude of genetic variance contributed by many such loci.

**Mapping of Quantitative Trait Loci**

Quantitative Trait Loci (QTL) refers to the genomic region associated with the expression of a quantitative trait. QTL analysis links phenotypic data to genotypic data to discover the genetic basis of variation affecting complex traits (Miles 2008). The key feature of this statistical method is that it identifies phenotypic variation primarily affected by a few loci with large effects, or by numerous loci each with minute effects.

The basic principle of quantitative trait mapping is that phenotypic value of an individual (P) is determined by combination of a genetic value (G), an environmental value (E) and their interaction (G × E), is expressed as

$P = G + E + G \times E$

Fisher's concept of quantitative genetics laid the foundation for partitioning of genetic variance for quantitative traits in terms of additive ($V_A$), dominance ($V_D$), and inter-locus interaction (epistatic) variance ($V_I$) components. The additive effect describes the cumulative effect of individual genes. The dominance component refers to dominant gene action in a locus where the effect of one allele dominates the effect of the other. Hence, the phenotypic variance of the trait is variance components of the corresponding genetic (A+D), environmental, and genotype × environment interaction, expressed as

$V_P = V_A + V_D + V_E + V_{G\times E}$

**Heritability**

Heritability is defined as that proportion of the variability of a phenotype that can be accounted for by genetic factors or $h^2 = V_G/V_P$, where $V_G$ and $V_P$ are the genetic and phenotypic variabilities, respectively.

If $h^2 = 0$, then none of the variations are genotypic, they are all due to variation in the environment.

If $h^2$ is small, then the trait is strongly influenced by the environment.

If $h^2$ is large, then the trait is only slightly influenced by the environment.

Heritability depends on population-specific factors, such as allele frequencies, the effects of gene variants, and variation due to environmental factors. In humans, estimation of heritability has been applied to diseases and behavioral phenotypes to evaluate proportion of disease risk variation that is genetic in origin.

**Biometric model for single bi-allelic QTL**

A QTL locus with two alleles can be defined with alleles A and B, where allele B encompasses all of the non-A alleles with similar genetic effects. Let *p* be the frequency of the allele A, and *q=1-p* be the frequency of the allele B. Let *a* be the mean genotypic value for individuals with genotype AA, let *d* be the genotypic value for individuals with AB, and let *-a* be the genotypic value of BB individuals (Falconer 1989). The value of *d* under additive genetic effect is 0, under dominant genetic effect is *a*, under recessive genetic effect is *-a* (**Figure 4**), and under partial dominant or partial recessive genetic effect is $-a<d<a$; but $d \neq 0$ respectively.

The additive genetic variance of this locus is $V_A = 2pq[a+(q-p)d^2]$ and the dominant genetic variance is $V_D=(2pqd)^2$ (Falconer 1989). The total genetic variance resulting from this QTL is $V_G=V_A+V_D$.

Since, heritability is a proportion of cumulative genetic effects from the total phenotypic variance, the heritability $h^2$ results from this QTL is $h^2=V_G/(V_G+V_E)$ (Deng et al. 2000).

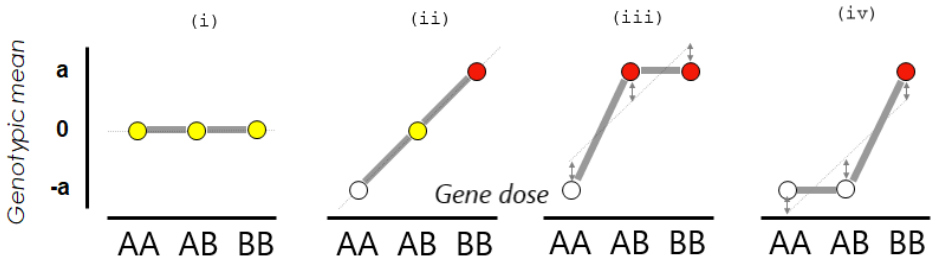


**Figure 4: Graphical representation of the relations between gene dose of homozygotes and heterozygotes and the genotypic effect in a one-locus model with two alleles across (i) Null, (ii) Additive, (iii) Dominant, and (iv) Recessive inheritance models. Image source: https://genepi.qimr.edu.au/staff/manuelF/talks/20080304_biometrical.ppt**

**Generalized linear model**

Association between genotype and a quantitative trait with additional covariates can be tested using multivariate linear regression models. The primary advantage of multivariate model is that it accounts for multiple potential confounders and effect modifiers appropriately.

A simple linear regression model is expressed by y = Xβ + ∈. The same model for a quantitative trait is given as,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}; X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ 1 & x_n \end{bmatrix}; \in = \begin{bmatrix} \in_1 \\ \in_2 \\ . \\ . \\ . \\ \in_n \end{bmatrix}$$

The scalar formulation of this simple linear regression model is given by,

$$y_i = \beta_0 + \beta_1 x_i + \in_i$$

Where, *i = 1,..,n* indicates individual. In this model, $\beta_0$ is intercept, the measure of association between *x* and *y* is given by the parameter $\beta_1$, defined as the amount of change in *y* that occurs with one-unit change in *x*. For example, if *x* is an indicator for the presence of a variant allele at a given SNP locus and *y* is BMI, then $\beta_1$ is the difference in mean BMI between individulas with and without this variant allele.

The least squares estimate of $\beta_0$, the overall mean, and $\beta_1$, the association parameter, are respectively given by,

$$\hat{\beta}_0 = \left( \sum_i y_i - \hat{\beta}_1 \sum_i x_i \right) / n$$

and

$$\hat{\beta}_1 = \frac{n \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n \sum_i x_i^2 - (\sum_i x_i)^2}$$

The multivariable linear model is a generalization of the simple linear model in which additional variables are included on the right-hand side of the equation. For example, suppose we have m covariates, given by *zi1,….,zim* for the *i*th individual, where *zi1* may be gender and *zi2* may represent smoking status for individual *i*. Such a model fitting is given by,

$$y_i = \beta_0 + \beta_1 x_i + \sum_{j=1}^{m} \propto_j z_{ij} + \in_i$$

$\beta_0$ is intercept and $\beta_1$ is the measure of association between the genotype and trait. However, estimation and testing of this parameter takes into account the additional variables in the model. These additional variables may be confounders or may help to explain the variability in the trait; including such variables is important in drawing a valid conclusion about the effect of genotype on the trait (Foulkes 2009).

### 6.3 Linkage mapping

The first and second genetic principles of Mendel- postulated segregation of alleles and an independent (or random) assortment of non-alleles respectively; the third principle: non-random or dependent assortment, a phenomenon termed as '*genetic linkage*' by Thomas Hunt Morgan (1866-1945), together establish foundation of transmission genetics. When an offspring derived from two inbred lines show a detectable difference in their traits, a certain inheritance tests (or breeding tests) may be employed to detect the genetic origin and transmission of that difference. These tests include,

1. Transmission tests to check whether the new phenotypic change is hereditary or not.

2. Segregation test to establish forms of alleles and segregation ratios.
3. Linkage tests to verify assortment of new trait is independent of or dependent with already known traits.
4. Mapping of identified genes on to the chromosome with relative positions of previously mapped genes of the same chromosome.
5. Association tests for fine mapping of the genes to the traits of difference.

Linkage means propensity of genes to inherit together due to their physical proximity in the same chromosome (**Figure 5**). Linkage analysis detects the locations of loci affecting a trait of an unknown position in a family, by using co-segregation of genetic polymorphisms of known position in the genome. Different types of genetic markers are used in studying Linkage. Early Linkage studies used microsatellite markers for mapping. Linkage can be performed either by testing single marker at a time or by several markers together from a region against the trait of interest. The later approach increases linkage detection power since it incorporates haplotype information from multiple markers.

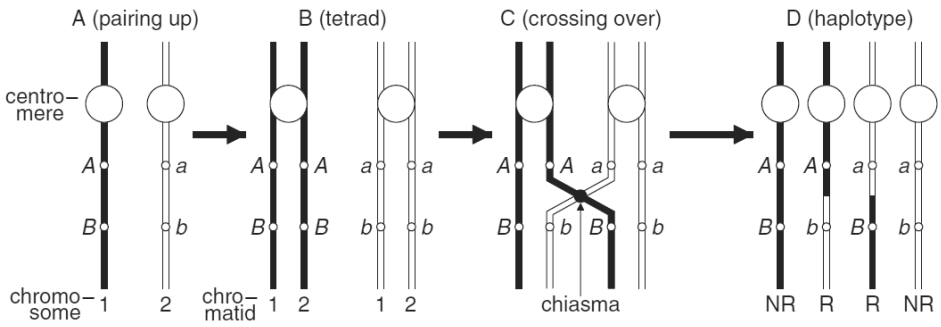


**Figure 5: Illustrating formation of cross over leading to recombinant (or repulsion) types indicated as, *Ab/aB*. This new assortment of non-alleles (i.e, *A vs B, A vs b, a vs B, or a vs b*) at different loci along a chromosome is called Linkage phase. Image source: (Rongling Wu 2007).**

There are two types of Linkage analysis methods, Parametric and Nonparametric. The parametric linkage method requires a genetic model that describes the relationship between the phenotype and the genotype, such that the mode of inheritance (dominant, recessive, or additive), disease allele frequency and the pattern of penetrance (i.e., proportion of individuals with a particular genetic change who exhibit specific trait), need to be specified to infer the disease locus genotype of all individuals from their phenotype. This method is commonly applied in pedigrees with high penetrance mutations, clear clinical phenotypes and distinct hereditary patterns such as Mendelian disorders. The basic principle is that, in families it tests whether the trait locus co-

segregates with genetic markers of known chromosomal location by estimating the recombination fraction (denoted by θ) between the trait locus and a given genetic marker. It is a likelihood ratio test that compares the likelihood under the null hypothesis of no linkage ($\theta_0$ = 0.5), $LH_0$, to the likelihood under the alternative hypothesis of linkage ($\theta_1$< 0.5), $LH_1$. It is usually expressed as logarithm of the odds (LOD) score, $Z(\theta) = \log_{10}(LH_1/LH_0)$. A higher LOD score suggests greater evidence for linkage. The classical threshold scores to declare significant linkage and to exclude linkage are LOD scores ≥3 (which is equivalent to a p-value of 1E-04) and ≤ -2 respectively (Morton 1955, 1998).

In Nonparametric linkage (NPL) analysis, no specific assumptions are made with regards to the trait model but genotyped marker locus model must be described. This method utilizes sib-pairs (Penrose 1935) or is affected by it (Bishop and Williamson 1990) and evaluates Identity by Descent (IBD) sharing among affected individuals without any particular model assumptions. IBD is a measure of shared DNA segments in two or more individuals when they have inherited it from a common ancestor without recombination (i.e., the segment has the same ancestral origin in these individuals). For linkage analysis of quantitative traits, Haseman-Elston modified the NPL method by regressing squared trait value difference (i.e., change of phenotypic covariance) against the IBD estimate of each locus (Haseman and Elston 1972). The mode of inheritance is generally assumed to be additive, so that each causal allele contributes equally to the trait.

In linkage analysis, the approximation of the correct genetic model is a fundamental issue. When the specified model for analysis is close enough to the true mode of inheritance that governs the trait, then the parametric linkage analysis method overpowers the nonparametric method. This is noticed in diseases that generally, exhibit Mendelian patterns of inheritance. Whereas while studying complex traits, approximation of genetic model is increasingly difficult as a number of factors including genetic and environmental are implicated in the etiology of the disease. Hence, a model-free approach of nonparametric analysis is beneficial to study complex traits. In general Linkage analysis lacks precision in mapping of genes, i.e., often genes are poorly localized to the region covering many megabases (Mb). Hence, in analyzing complex traits to precisely map the disease-causing genes, Linkage analysis may be ideal for the initial detection of the promising linked region and followed by targeted association testing of candidate genes in the promising regions.

### 6.4 Association analysis

Genetic association studies analyze the correlation between disease status and genetic variation to identify candidate genes or genome regions that contribute to that specific disease. A higher frequency of an allele or genotype in individuals with affected status can be interpreted as the tested variant increases the risk of a specific disease. The genetic association analysis can be carried out in either families or 'unrelated' singleton individuals (population-based). This analysis method is based on linkage disequilibrium. Population-based association studies are

comparatively easier to implement than family-based studies with flexible pedigree-free recruitment design/approach of subjects. Nevertheless, both the designs have their strengths and weaknesses. While a population-based design is suitable to detect the effect of multiple rare variants, a family-based design can potentially enrich the sample in rare variants for which the effect would be concealed at the population level (Kazma and Bailey 2011). The major difference between association and linkage analysis is that association testing assumes shared common ancestral haplotype regardless of population-based or family-based design, while Linkage on the other hand assumes shared haplotype allele within each family to delineate disease causing alleles. The term genome-wide association study (GWAS) refers to association testing of single nucleotide variants (SNV) across the genome by genotyping samples on a dense genome-wide SNP panel, where each SNP is correlated against the phenotype of interest. Commercial companies like Affymetrix, Agilent, Illumina, and others manufacture pre-designed or custom-made SNP panels for genome-wide genotyping experiments. Despite the increasing popularity of next-generation sequencing, relatively low cost of SNP genotyping arrays and recent introduction of exome arrays continue to make GWAS the better alternative for genetic research.

Previously family-based studies utilized high-density SNP genotyping panels and modern analytical algorithms for quantifying phenotypic change associated with SNV to understand the metabolic disease etiology in genome-wide scale approach. However, the current practice is to use whole-exome sequencing (WES) techniques coupled with high-density genotyping panels to make the experiments cost efficient. Family-based design also offers several advantages like better-quality control and curtails the quantity of samples required as compared to population-based studies. Particularly, there is no requirement of population stratification as they possess a common genetic background and uniform exposure to environmental factors is expected among family members (Hebbar et al. 2019b). However, when families of admixed population are used, adjusting for family sub-stratification using Principal Components (PCs) demonstrated to improve statistical power and reduce type-I error (Mersha et al. 2015). Familial studies such as twins (Newman et al. 1987; Sung et al. 2009), siblings (Feng et al. 2008), and parent–offspring (Adams et al. 1993; Benrahma et al. 2011; Park et al. 2006) have shown a strong link of familial aggregation for obesity, T2DM, and metabolic syndrome, since parents influence their children genetically and environmentally. The transmission disequilibrium test (TDT) (Ewens and Spielman 1995) is a special type of association study conducted in related individuals. TDT assumes that an allele of a given polymorphism contributes to a trait by analyzing the frequency with which affected individuals inherit the allele from a heterozygous parent. Suppose the allele contributes to the trait or disease in question, the probability of the affected person inheriting the allele from a heterozygous parent should vary from the expected Mendelian ratio 50:50. TDT is a valuable technique; however, only heterozygous individuals are informative for the test. Hence, it reduces the sample size and limits statistical power like linkage analysis. Later, TDT is generalized for other family designs, other modes of inheritance, quantitative and time-to-onset phenotypes with family-based association test (FBAT) (Horvath et al. 2001). Furthermore, advancement in the field allowed combination of linkage and association data. Variance-components based, combined

linkage and association analysis (where the association is modelled as a mean effect and linkage is modelled as the covariance structure) for quantitative traits in sib pairs is extended to general pedigrees (quantitative transmission disequilibrium test, QTDT). This approach accommodated data not only from parents and siblings, but also from all available relatives (Abecasis et al. 2000).

Population-based GWAS offers a reliable, cost-effective, and user-friendly platform to identify novel genes, underlying biological mechanisms, and drug response for complex traits in human populations. GWAS provides comprehensive picture of genomic distribution of the variants affecting complex traits and the distribution of their frequency and effect sizes. Genome-wide genotyped data has several applications, including identification of population structure, and inbreeding signatures (ROH and IBD), imputation of missing variants, examination of variant association in cohorts along with well phenotyping of clinical or molecular traits, the heritability of SNP with the trait and gene x environmental interaction, and from genotype intensity signals-detection of copy number variations (CNV) in the genome, etc.

Based on the type of phenotype, GWA studies can be classified as binary trait association analysis (e.g., obesity and diabetes status) and quantitative trait (QT) association analysis (e.g., BMI and fasting plasma glucose- FPG). From the point of view of power calculation, analysis of binary traits often requires considerably more samples than quantitative traits to attain 80% study power at type 1 error > 0.05. Additionally, binary traits typically summarize the trait derived from several quantitative traits. For example, a binary trait representing different hyperlipidemia status is derived from assessing levels of four lipid quantitative traits such as total cholesterol (TC), high-density lipoprotein-cholesterol (HDL-C), low-density lipoprotein-cholesterol (LDL-C), and triglyceride (TG). Hence, examining genetic signature against these quantitative traits make provisions for definitive biological mechanism of hyperlipidemia, rather than testing one summary binary trait-hyperlipidemia status. When using quantitative trait in analysis, additional care is required because QT often violates normality assumption. Normality can be achieved by subjecting QT to transformation using either natural logarithm or rank based inverse normal transformations (Beasley et al. 2009). Such transformations can sometimes mask the true phenotype variance, thereby misleading the outcome of the analysis. Therefore, it is advised to perform association tests using transformed trait and validate its result with that of raw trait (Hebbar et al. 2019b).

The success of GWAS depends on rigorous quality control (QC). Numerous quality control methods, statistical association test analysis tools have already been earmarked for GWAS (Agler et al. 2019). Common QC methods include evaluating batch quality by looking for disparities in statistics between data batches, filtering SNPs based on allele frequency, missing call rate, departing from HWE, and excess heterozygosity. Statistical methods such as logistic regression (for dichotomous traits) or linear regression (for quantitative traits) are mostly used for examining association of variants with traits. A genetic association is interpreted as significant association when either the genotyped SNP is the true causal variant conferring disease susceptibility (direct association) or an SNP in LD with the true causal variant is associated

(indirect association). A false-positive result due to systematic confounding, such as population stratification can be identified by adjusting for such confounding factors. A common paradigm to identify population substructure is through a principal component analysis (PCA), which estimates ancestral genetic differences between samples and then that could be further used to correct for the sample genetic differences in the association analysis (Price et al. 2006). Software tools such as PLINK (Purcell et al. 2007), GenABEL (Aulchenko et al. 2007) project suite and GWAS Tools (Gogarten et al. 2012) offer flexibility in implementing QC and statistical association analysis procedure. Furthermore, the linear mixed model tools which account for sample structure and relatedness in association analysis are increasingly being employed in recent studies. These include EMMAX (Kang et al. 2010), RvTests (Zhan et al. 2016), and FaST-LMM (Lippert et al. 2011) etc.

From the inception of GWAS in 2005 (Klein et al. 2005) to the recent SARS-COV-2 GWAS study (Ellinghaus et al. 2020), it has been well tested in several different settings and a standard protocol has been established to suit the need of the study-specific research questions (Marees et al. 2018). A key challenge encountered in the GWAS is the translation of GWAS identified statistical associations to the functional demonstration of complex traits and disease risk (Edwards et al. 2013). There are two major complicating factors due to which GWAS requires additional maneuvers to harvest definite insights into underlying biological mechanisms. They are 1) GWAS is not sensitive enough to distinguish the gene-variant signal effect from other variants that are in complex LD with the lead variant. Variants in high LD are often located in intergenic regions or non-coding DNA regions, indicating that any of these could be the actual causal variant. 2) More than 90% of disease-associated variants are in non-protein coding regions of the genome (such as introns and intergenic regions) and not in the 1.5% coding part of the human genome (Hindorff et al. 2009; Maurano et al. 2012; Schaub et al. 2012). The effect of non-coding causal variants can be cell-type-, context-, or disease-specific (Andersson et al. 2014). Additionally, these variants may locate in non-coding DNA regulatory regions such as enhancers & promoters and could affect the binding of transcription factor (TF) proteins and regulate gene expressions of proximal genes (Gallagher and Chen-Plotkin 2018; Haberle and Stark 2018; Maurano et al. 2012; Schaub et al. 2012). Hence, the causal variant is not essentially the strongest GWAS signal, but rather a variant in strong LD with the lead variant's effect located in an active enhancer region.

Nevertheless, deciphering the functional role of GWAS associated risk variants requires a comprehensive computational and experimental functional analysis strategy. A computational approach called "fine mapping" was developed to identify true causal variants and genes associated with the trait. The vast data generated by a combination of biochemical assays and massively parallel sequencing initiatives such as - the Encyclopedia of DNA Elements (ENCODE) project (Consortium 2012), Functional Annotation of the Mammalian Genome (FANTOM5) project (Consortium et al. 2014), 1000 Genomes project (Genomes Project et al. 2010),the Haplotype reference consortium (HRC) (McCarthy et al. 2016), the Epigenome Roadmap (Bernstein et al. 2010), and Genotype-Tissue Expression (GTEx) project (Consortium et al. 2017)

have enabled systematic fine mapping analysis with remarkable choice of tools such as PAINTOR (Fachal et al. 2020), CAVIAR (Hormozdiari et al. 2015), CAVIARBF (Chen et al. 2015), and FINEMAP (Benner et al. 2016) etc.

Importantly, GWASs have been exploited as a systemic tool to interrogate several quantitative traits such as anthropometric (Akiyama et al. 2017; Locke et al. 2015; Wood et al. 2014), lipid (Surakka et al. 2015; Willer et al. 2013), kidney-related (Okada et al. 2012; Pattaro et al. 2016), hematological (Astle et al. 2016; Kamatani et al. 2010), and blood pressure traits (Liu et al. 2016; Sivakumaran et al. 2011; Surendran et al. 2016). The relation between the genetics of quantitative traits and methods for linking them to disease end points has been established through approaches like pleiotropy (Han et al. 2016; Sivakumaran et al. 2011), genetic correlation (Bulik-Sullivan et al. 2015; Lee et al. 2012), and Mendelian randomization (Davey Smith and Hemani 2014). Most of these approaches have been tried and tested in European or Asian cohorts. However, findings from these cohorts are only partially (approximately 25%) trans-ethnic and such associations vary in at least one of five non-European ancestry populations (Popejoy and Fullerton 2016). Therefore, due to the increasing prevalence of metabolic disorders in every corner of the world, there is a simultaneous need to create a comprehensive genetic landscape with additional studies of non-European populations with the focus on the investigations using a wide range of clinical measurements.

## 6.5 Metabolic traits and their modulation in metabolic disorders

Metabolic phenotypes are clinically observable characteristics or traits of an individual's metabolic health including morphology, development, biochemical, and physiological properties etc. These phenotypes help in diagnosing metabolic disease, assess progression, and determine the potential targets necessary for successful therapies. Since they result from gene products and are influenced by environmental factors, the metabolic traits are extensively utilized in genetic studies of metabolic disorders to unravel the genetic origin of disease etiology. Some important basic clinical phenotypes to determine obesity are weight, body mass index (BMI), waist circumference (WC), hip circumference, waist-to-hip-ratio, body fat, total body water (TBW), and soft lean mass (SLM). In type 2 diabetes, following clinical phenotypes are measured - fasting plasma glucose, glycated hemoglobin A1c (HbA1c), fasting insulin, c-peptide, and insulin resistance (IR). In dyslipidemia - triglyceride (TG), total cholesterol (TC), low density lipid cholesterol (LDL-C), and high-density lipid cholesterol (HDL-C) are determined. In hypertension-systolic blood pressure (SBP) and diastolic blood pressure (DBP) are recorded. Whereas, in cardiovascular disorders along with electrocardiogram (ECG) readings, C-reactive proteins, and lipid traits etc are usually measured from the individuals.

**Anthropometric traits as measure for obesity**

BMI is an instant and crude measure of obesity, which is obtained by dividing weight (in kilograms) with square of height (in meters) of an individual. According to World Health

Organization guidelines, an adult with a BMI of <18.5 Kg/$m^2$ is underweight, 18.50-24.99 Kg/$m^2$ is Normal, 25-29.9 kg/$m^2$ is said to be overweight, whereas ≥30 Kg/$m^2$ is considered to be obese. Based on the BMI range the obese category is further classified into three classes, 30-34.99 as Obese class I, 35-39.99 as Obese class II and ≥40 Kg/$m^2$as Obese class III. In addition, waist-to-hip ratio is deemed the best measure for abdominal fat. Obesity measured by bioelectric impedance analysis, underwater weighing, computed tomography, and magnetic resonance imaging are more laborious but accurate.

**Blood pressure traits for hyper/hypotension**

Blood pressure (BP) is a variable hemodynamic phenomenon, usually measured on the arm over the brachial artery using sphygmomanometer (i.e., auscultation), semi-automated or automated devices. BP is measured using two readings, systolic blood pressure (SBP) and diastolic blood pressure (DBP). Systolic blood pressure indicates the pressure exerted by circulating blood against arterial walls when the heart beats. Diastolic blood pressure indicates the pressure exerted by circulating blood against arterial walls while the heart is resting between beats. According to the American Heart Association, SBP of <120 mmHg and DBP of <80 mmHg is considered to be normal. If SBP is 120-129 and DBP>80, they are said to be elevated. Hypertension is further classified as Stage 1: 130-139 or higher for SBP and 80-89 for DBP, Stage 2: 140 or higher SBP or >=90 DBP, whereas >180 for SBP and/or >120 for DBP is regarded as Hypertensive crisis. Additional measures with relation to BP can be calculated using SBP and DBP. An important measure of pulsatile hemodynamic stress and arterial stiffness termed as pulse pressure can be calculated by subtracting SBP by DBP. Other measures such as Mid-BP (the average of SBP and DBP) and mean arterial pressure (often approximated for individuals with normal heart rate as 1/3 SBP+2/3 DBP or DBP+1/3 pulse pressure) provide overall estimates of arterial BP during a complete cardiac cycle (Muntner et al. 2019).

**Glycemic traits in type 2 diabetes**

Glycemic traits such as fasting blood glucose (FBG), post-challenge glucose (or Oral Glucose Tolerance Test- OGTT), insulin measures, and glycated hemoglobin (HbA1c) are used to diagnose and monitor diabetes. Also, more sophisticated glycemic measures help in disease stratification. For example, fasting insulin (FI) and insulin resistance by homeostasis model assessment [HOMA-IR] for obesity mediated diabetes, the insulin suppression test or euglycemic clamp or proinsulin measures to diagnose beta-cell stress. Usually, for primary diagnosis either HbA1C, or FBG or OGTT test is performed. In order to measure FBG, the individual is required to be on fasting for at least 8 hours. The HbA1C test does not require the individual to be on fasting; it is a measure of average blood sugar levels for the past two to three months. On the other hand, in 2 hour-glucose test or OGTT, blood sugar level is measured before and after two hours of consumption of a special sweet drink to understand how body processes the sugar.

According to the American Diabetes Association (ADA) guidelines, each of these tests use certain threshold to classify an individual as Normal, pre-diabetic, or diabetic (**Table 1**). Pre-diabetic

individuals have blood sugar levels that are higher than normal but not high enough to be designated as diabetes. In such cases, impaired glucose tolerance (IGT) or impaired fasting glucose (IFG) test is done. Pre-diabetes is accepted as a risk factor for developing type 2 diabetes and cardiovascular diseases (Wheeler et al. 2017).

| Diagnosis status | Glycemic test thresholds |
|---|---|
| FBG | |
| Normal | <100 mg/dl |
| Pre-diabetic | 100 mg/dl-125 mg/dl |
| Diabetic | ≥126 mg/dl |
| HbA1C | |
| Normal | <5.7% |
| Pre-diabetic | 5.7% to 6.4% |
| Diabetic | ≥6.5% |
| OGTT | |
| Normal | <140 mg/dl |
| Pre-diabetic | 140 mg/dl to 199 mg/dl |
| Diabetic | ≥200 mg/dl |

**Table 1: Displaying thresholds of glycemic test for diagnosis of diabetes as per ADA guidelines.**

**Lipid traits in dyslipidemia and Cardiovascular diseases**

Lipids are fatty molecules and their decomposition provides the energy necessary for various life processes. In conjunction with proteins, lipids form the most important structural elements of cells and cellular organelles. However, when certain lipids (such as cholesterol) accumulate excessively in the body, they become a risk factor for a multitude of metabolic diseases.

Lipids are largely constituted of fatty acids, cholesterols, triglycerides and phospholipids, which circulate in the blood. The lipid metabolism process comprises digestion, uptake, intracellular re-synthesis and packaging them into lymphatic system to blood stream is a complex process. The digestion of fat (consist of triacylglycerols and phospholipids) begins in the mouth by salivary digestive enzyme- lingual lipase. The physical action of chewing makes the fat to become more accessible to the digestive enzyme to break them as tiny fat droplets and separate them from watery components. The fat entering the gastrointestinal tract (include triacylglycerol, cholesteryl esters and phospholipids) gets digested by gastric lipase and breaks down triacylglycerols into diglyceride and fatty acids. Further, along with other endogenous lipids (include phospholipids and cholesterol) from bile and shedding process of intestinal epithelial cells, gets emulsified due to the action of bile in small intestine. The action of pancreatic lipase in small intestine, emulsified fat gets broken down into free fatty acids and monoglycerides. These molecules form micelles with the action of bile salts to allow efficient transportation into

intestinal microvillus. Inside the intestinal cells, the monoglycerides and free fatty acids reassemble themselves into triacyclglycerols. Then, triacylglycerols, cholesterol and phospholipids conjugate with a protein carrier to form lipoproteins. The different lipoproteins in order of increasing density include 1) chylomicrons consisting triglycerides as major lipid and ApoB-48, ApoC, ApoE, ApoA-I, A-II, A-IV as major apolipoproteins, 2) chylomicron remnants consisting triglycerides, cholesterol as major lipid and ApoB-48, ApoE as major apolipoproteins, 3) very-low-density lipoprotein (VLDL) consisting triglycerides as major lipid and ApoB-100, ApoE, ApoC as major apolipoproteins, 4) intermediate-density lipoprotein (IDL) consisting triglycerides, cholesterol as major lipid and ApoB-100, ApoE, ApoC as major apolipoproteins, 5) low-density lipoprotein (LDL) consisting Cholesterol as major lipid and ApoB-100 as major apolipoproteins, 6) high-density lipoprotein (HDL) consisting Cholesterol, Phospholipids major lipid and ApoA-I, ApoA-II, ApoC, ApoE as apolipoproteins, and 7) Lipoprotein (a) consisting Cholesterol as major lipid and ApoB-100, Apo(a) as major apolipoproteins (Feingold 2000). Lipoproteins which are packaged in endoplasmic reticulum of enterocytes, later enter the lymphatic system via Golgi apparatus and exocytosis into lamina propria. From lymph these re-synthesized lipids secreted into circulation and transport to specific destinations such as liver and other body tissues for produce of energy (Zimmerman M 2020). Lipoproteins play a key role in the absorption and transport of dietary lipids by small intestine and in the transport of lipids from the liver to peripheral tissues. In addition, they transport the lipids from peripheral tissues to the liver and intestine.

Cholesterol is an important component of cell membrane and plays a vital role in absorption of fats, and fat-soluble vitamins. The body uses cholesterol in synthesis of vitamin D and hormones such as estrogen, testosterone, and cortisol. Cholesterols are poorly absorbed when compared to phospholipids and triacylglycerols. However, high fat diet increases its absorption. Hence, a high intake of fiber is recommended to decrease blood cholesterol. Foods high in fiber such as fresh fruits, vegetables, and oats can bind bile salts and cholesterol, preventing their absorption and carrying them out of the colon. Triglycerides are another type of fat. Triglycerides are derived from esterification of glycerol and three fatty acids in intestine and liver and the excess fats are stored as triglycerides in liver cells. Later, when the body requires fatty acids as an energy source, hormone glucagon signals the breakdown of these triglycerides by hormone-sensitive lipase to release free fatty acids (Zimmerman M 2020). According to American College of Cardiology and American Heart Association (ACC/AHA), a desired level of total cholesterol is <200 mg/dL, LDL cholesterol <100 mg/dL, HDL cholesterol ≥60 mg/dL, and triglyceride <150 mg/dL in adults (Grundy et al. 2019; Program 2002). Abnormal levels of lipids (dyslipidemia) are commonly observed in obese individuals. Approximately 60%-70% of the obese individuals are dyslipidemic with elevated serum triglyceride, VLDL, apolipoprotein B, and non-HDL-C levels. Dyslipidemia is a leading risk factor for cardiovascular disease in individuals comorbid with type 2 diabetes (Mooradian 2009). The characteristic features of lipid profile with high plasma TG, low HDL-C, and increased LDL-C are considered as a hallmark of T2DM risk in CVD patients (Lee et al. 2017).

### 6.6 Obesity as a central risk factor to metabolic disorders

Obesity is excessive accumulation of adipose tissue due to an imbalance between food intake and energy expenditure. Adipose tissue, beyond storing excess energy (or triacylglycerols), it also serves as an endocrine organ. It synthesizes several biologically active compounds called adipokines and reactive oxygen species that regulate metabolic homeostasis. They include adiponectin, Leptin, Apelin, Angiotensins, tumor necrosis factor-α (TNF-α), plasminogen activator inhibitor 1 (PAI-1), Omentin, Chemerin, Visfatin, Resistin, Nitric oxide, Hydrogen sulfide, palmitic acid methyl ester, and Hydrogen peroxide etc., (Coelho et al. 2013; Saxton et al. 2019). A substantial amount of evidence shows that adipokines have multidimensional effects; such that while regulating energy homeostasis by appetite regulation, adipokines also regulate glucose, vascular homeostasis, and inflammatory response (Saxton et al. 2019).

The effects of over-nutrition increase the size of Adipocytes. However, when it exceeds storage capacity, lipid contents spillover to ectopic storage (i.e., non-adipose tissues), and subsequently adipose dysfunction leads to lipotoxicity (Montgomery et al. 2019). Further, excess in adiposity and adipocyte dysfunction dysregulate secretion of adipokines which can contribute to the development of other metabolic pathological conditions such as insulin resistance, T2DM, dyslipidemia, hypertension, coronary heart disease, stroke, and non-alcoholic fatty liver disease (NAFLD) in conjunction with altered glucose and lipid homeostasis as well as inflammatory responses (Halberg et al. 2008; Hauner 2005). **Figure 6** shows multifaceted physiological effects of over-nutrition induced by adipokine dysregulation.

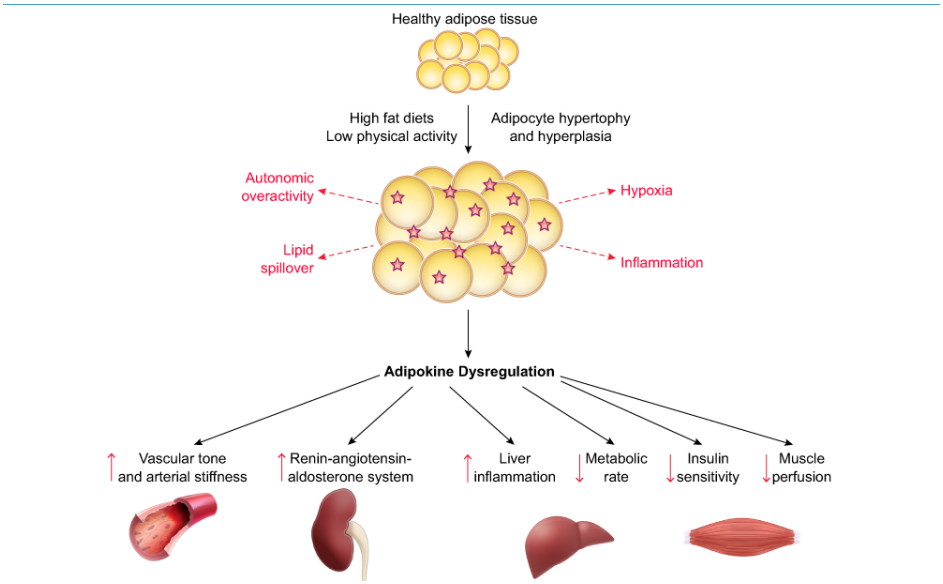


**Figure 6: Showing multi-dimension physiological effects of fatty adipose tissue led adipokine dysregulation (Saxton et al. 2019). With permission from the American Physiological Society.**

## 6.7 Genetics in mapping pathophysiology of complex metabolic disorders

A prevailing abnormality in metabolic traits is indicative of a metabolic disorder. Of the metabolic disorders, Obesity is believed to be one of the most predominant risk factors for other metabolic disorders and cancer. Obesity is a multi-factorial disorder that is caused by the interaction between several genes and environmental factors such as food choices, and physical activity. The genetic approaches to study obesity are progressing rapidly. Few large-scale GWA studies have identified more than 300 genes or loci responsible in the pathophysiology of obesity (Akiyama et al. 2017; Locke et al. 2015). It is an important risk factor for T2DM and cardiovascular diseases. Multiple independent GWA studies have reported common variants having strong LD with an intronic variant rs9939609 of fat mass and obesity associated (FTO) gene playing a crucial role in obesity and pathogenesis of T2DM. Its association with T2DM in the behest of obesity was confirmed when it was proven that T2DM signal is abolished upon adjusting for BMI (Frayling et al. 2007; Rees et al. 2011). The involvement of FTO in controlling food intake, energy homeostasis & expenditure, and its interaction with proximal genes such as RPGRIP1L and IRX3 are well studied in mouse models (Church et al. 2009; Speakman 2015). Interestingly, our recent investigation of variant rs1421085 from FTO with body composition traits showed its association with total body water (TBW), soft lean mass (SLM), interaction with satiety gene (or biomarker) ghrelin and plasma lipoproteins complex biomarkers such as ApoA1 and ApoB48, indicating a complex interplay of FTO in modulating obesity and lipid traits via genes expressed in the central nervous system (CNS) (Hebbar et al. 2019a).

Morbid obesity during early childhood is often associated with rare single gene (monogenic) defects. Therefore, severely obese children are screened for leptin, POMC, and MC4R deficiencies. Monogenic genetic studies have explained energy balance and storage mechanisms in adipose tissues and their relationship with severe obesity. The satiety hormone leptin, known to regulate the energy output by stimulating the sympathetic activation of brown adipose tissue, binds to hypothalamic receptors in the arcuate nucleus. This induces, among other effects, enhanced synthesis and secretion of α-MSH (which is formed from POMC). The α-MSH binds to MC4R in the paraventricular nucleus of hypothalamus and inhibits the effectors of food intake. Mutations in these genes have been implicated as a causal factor in severe obesity (Clement et al. 1998; Funcke et al. 2014; Hilado and Randhawa 2018; van den Berg et al. 2011). Numerous genetic syndromes (at least 79) have been reported in the literature (Kaur et al. 2017) and syndromic forms of obesity, such as Prader-Willi, Cohen, Alstrom, and Bardet-Biedl (BBS), have been genetically mapped (Rankinen et al. 2002). Astoundingly, of the 79 syndromes, 19 have been entirely genetically annotated, 11 have been partially annotated, 27 have been mapped to a chromosomal region, and for the remaining 22, neither the gene(s) nor the chromosomal location(s) have yet been identified (Kaur et al. 2017).

Several genome-wide association studies have suggested that quantitative traits are typically highly polygenic (Turchin et al. 2012), suggesting pleiotropy is evident in metabolic disorders. Most co-morbidities occur due to shared risk factors. A few genes with specific variants such as

ADAMTS9 (rs6795735), ADCY3 (rs10182181), ADCY3/POMC (rs713586), BDNF (rs11030104), FTO (rs9939609), GIPR (rs11671664), GIPR (rs10423928), MC4R (rs6567160), MC4R (rs12970134), PCSK1 (rs6234/rs6235), PEMT (rs4646404), POMC (rs1561288), SH2B1 (rs7359397), SH2B1 (rs7498665), and TBX15-WARS2 (rs2645294) are known to be associated with obesity and also show pleotropic association with cardiovascular risk traits such as T2DM, dyslipidemia, and hypertension (Sanghera et al. 2019).

The other important and most prevalent metabolic disorder worldwide is T2DM. T2DM is the most common form of diabetes and is influenced by age, pregnancy, and lifestyle-related factors such as obesity. Several studies for more than 30 years have established insulin resistance as the main pathophysiological factor in T2DM. Later, the role of beta cell dysfunction in T2DM was also reported (Chakravarthy and Semenkovich 2007). The current consensus suggests both the factors are essential and interlinked in disease pathogenesis (Cerf 2013). Although both insulin resistance and beta-cell failure result in hyperglycemia and increased demand for insulin, intrinsically, beta-cell dysfunction is a condition that overtakes the insulin resistance to induce diabetes (i.e., pre-diabetes to diabetes). The main difference between these two conditions is that beta cell dysfunction results from inadequate glucose sensing capacity to stimulate insulin secretion by beta cells. This eventually leads to hyperglycemia, whereas hyperglycemia due to insulin resistance is by damage of insulin signaling in glucose recipient tissue cells such as adipose tissue.

Large scale genetic studies have identified more than 500 genetic variants contributing to the risk of T2DM (Mahajan et al. 2018a; Mahajan et al. 2018b; Vujkovic et al. 2020). McCarthy and colleagues studied the molecular insights of beta cell dysfunction and insulin resistance through genome-wide genetic and epigenetic approaches (Mahajan et al. 2018a; Mahajan et al. 2018b; Thurner et al. 2018). These studies delineated risk loci into insulin secretion (indicating beta cell failure), insulin action (indicating insulin resistance), impaired lipid metabolism, and adiposity.

An increasing number of rare monogenic forms of single-gene mutations have been characterized in diabetes. Molecular genetic advances have identified at least eight genetic subgroups for maturity onset diabetes in young (MODY), three genetic subgroups of transient neonatal diabetes mellitus (TNDM), four genetic subgroups of permanent neonatal diabetes mellitus (PNDM), and five genetic syndromes that include developmental delay, epilepsy, and neonatal diabetes (DEND). Depending on the mutations in genes such as KCNJ11 (DEND, PNDM, TNDM, MODY); ABCC8 (DEND, PNDM, TNDM); INS (PNDM, MODY); GCK (PNDM, MODY2); SCLC2A2 (PNDM); PDX1 (PNDM, MODY4); GLIS3 (syndromic); FOXP3 (syndromic); EIF2AK3 (syndromic); PTF1A (syndromic); Rfx6 (syndromic); 6q24 abnormality (TNDM); TCF2 (TNDM, MODY5); HNF4A (MODY1); TCF1 (MODY3); and NEUROD1 (MODY6)(McCarthy and Hattersley 2008) discrete clinical differences in diabetes were also observed.

Hypertension is a key risk factor for renal, cerebro-vascular, and cardiovascular disorders. Both heritable and lifestyle factors combined with several physiological pathways and mechanisms ultimately influence blood pressure levels. Susceptible loci for hypertension were genetically characterized using candidate (Alsmadi et al. 2014b; Newton-Cheh et al. 2009), linkage

(Kristjansson et al. 2002; Wallace et al. 2006), and GWA studies (Franceschini et al. 2013; Ganesh et al. 2014). Large scale GWA studies (Ehret et al. 2016; Hoffmann et al. 2017; International Consortium for Blood Pressure Genome-Wide Association et al. 2011; Warren et al. 2017) alone identified more than 120 risk loci for hypertension. Family genetic studies have also implicated many genes, including CYP11B1, CYP11B2, WNK1, WNK4, KLHL3, CUL3, SCNN1B, SCNN1G, CYP17A1, HSD11B2, NR3C2, and KCNJ5 for Mendelian forms of monogenic hypertensive syndromes (Ehret and Caulfield 2013).

Dyslipidemia is an important risk condition among metabolic diseases with a complex etiology. It can also be caused by single gene defects or complex polygenic framework, resulting in overproduction or defective clearance of TGs and LDL-C, or underproduction or excessive clearance of HDL-C. Elevated LDL-C and TG levels and low HDL-C levels are well-known risk factors for CAD. In case of monogenic diseases, mutations in genes LDLR, ApoB, PCSK9, and LDLRAP1 have demonstrated evidence of association with familial hypercholesterolemia and genes ABCG5 and ABCG8 with phytosterolemia (Garcia-Giustiniani and Stein 2016). Similarly, mutations in APOA5, APOC2, APOE, GPD1, GPIHBP1, HNF1A, LMF1, LPL, and SLC25A40 genes known to cause familial hypertriglyceridemia or familial chylomicronemia syndrome (FCS) with increased levels of TGs as the main phenotype (Brown et al. 2020). Mutations in APOA1, ABCA1, LCAT, SAR1B, and ABCG1 genes associated with low HDL-C of monogenic etiologies including familial hypoalphalipoproteinemia (ApoA1), Tangier disease (ABCA1), fish-eye disease (LCAT), and chylomicron-retention disease (SAR1B) (Garcia-Giustiniani and Stein 2016).

Conversely, polygenic dyslipidemia results from multiple common variants with each contributing a small effect to an aggregated large effect on lipid traits. The Global Lipid Genetics Consortium (GLGC), an initiative to produce the largest GWAS meta-analysis in the lipid field (Kathiresan et al. 2009; Willer et al. 2013), has recognized more than 150 common variants with small effects on lipid traits. Interestingly, one-third of the significantly associated genes and loci identified by GLGC were already well-established, but a majority of them had no prior links with lipid metabolism. Some of the loci GLGC initiatives discovered include ABCG8, MAFB, HNF1A, and TIMD4 with LDL-C; ANGPTL4, FADS1-FADS2-FADS3, HNF4A, LCAT, PLTP, and TTC39B with HDL-C; and AMAC1L2, FADS1-FADS2-FADS3, and PLTP with TG.

Another important metabolic disorder is CVD (include coronary artery disease (CAD), heart failure, stroke and hypertension) with 523 million cases in total and 18.6 million deaths as of 2019 (Roth et al. 2020). Many monogenic studies have contributed to deciphering the role of rare mutations in CVD etiology. Dilated cardiomyopathy (DCM) is a cause of heart failure. Studies of families with DCM have identified rare variants in more than 33 genes (Hershberger and Siegfried 2011; Wells et al. 2013). A loss of function variants in the sodium channel gene, SCN5A is known to cause *Brugada syndrome*, a monogenic heart disorder characterized by an abnormal electrocardiogram and risk of ventricular fibrillation and sudden death. Over 160 lead variants have been discovered for CAD through individual and consortium based GWA studies (Erdmann et al. 2018). Risk alleles confer the strongest effects on CAD risk at chromosomal locus 9p21

(INK4), which are carried by almost 75% of the global population, excluding black Africans (Holdt and Teupser 2012). Some of these reported genes including NOS3, GUCY1A1, PDE5A, PDE3A, and MRVI1 are found to be involved in NO/cGMP signaling pathway suggesting a role of Nitric oxide in pathogenesis of CAD.

Although, cardio-metabolic disorders appear to be phenotypically different from each other; there is an overlap between biomarkers associated with coronary heart disease, T2DM, obesity, and hypertension. Hence, it is no surprise that obesity associated traits such as central adiposity, dyslipidemia, hypertension, and insulin resistance are major risk factors for T2DM and cardiovascular diseases. An estimation of genetic correlation among 24 traits, including anorexia, obesity, anthropometric traits, lipid traits, CAD, T2DM, and psychiatric traits, discovered a high genetic correlation among obesity, T2DM, CAD, and lipid traits using cross-trait LD Score regression test (Bulik-Sullivan et al. 2015). However, while several reports have evaluated individual disease conditions, limited information exists on shared genetic risk between metabolic disorders.

### 6.8 Relevance of genetic studies for mapping metabolic disease genes in the Arab population

Metabolic diseases are caused by the interaction between multiple genetic factors and environmental exposure. An estimation suggests, intra-population differences among individuals account for 93% to 95% of genetic variation and only 3% to 5% of genetic variation accounts for differences among major population groups (Rosenberg et al. 2002). Considering this fact, a substantial portion of knowledge existing on pathophysiology of metabolic diseases by de facto should be transferable to Arab population or any general population. However, approximately 25% of the variants found associated with BMI, T2DM, and lipid levels in European-Americans are trans-ethnic and such associations vary in at least one out of the five non-European ancestry populations (Popejoy and Fullerton 2016). Hence, environmental factors can be a prime contributor in metabolic disease etiology. Given the fact that, Arabs are exposed to a unique desert environment and culture; these exposures are exclusive to Arabs as compared to any other population living in different geographic regions. Novel selection pressure from unique cultural practices, food habits, lifestyle, and exposures to infections & toxins may contribute to population specific modulation of metabolic traits in conjunction with population specific genetic variation in Arabs. On the other hand, accumulating evidence suggests that complex metabolic traits are highly heritable and genetically complex; most variations in traits arise due to differences at numerous genetic loci (i.e., polygenic) in the genome with small contributions by each locus to phenotypic variation (Sella and Barton 2019). Since, consanguineous and polygamous marriages are prevalent among Arabs, novel mutations which are rare elsewhere is expected to be common in Arabs. This type of marriage from successive generations cumulatively has increased the inbreeding levels and, consequently, increased the distribution and frequency of founder mutations among the Arab population. Our own recent analysis of Whole Exome Sequences (WES) of Arabs from Kuwait showed 21.7% personal variants, 12% novel and 36% Arab

population-specific from the total SNV's identified (John et al. 2018). Hence, genetic studies in understudied Arab population provides opportunities to enhance insights of complex architecture of metabolic traits and provide comprehensive picture of variants involved in etiology of metabolic disorders in global population, while specifically benefiting the Arab population by allowing to devise newer therapeutic approaches after identification of genetic risk loci resulting from genetics and genomic studies.

## 6.9 Genetic history of Arabs

The “Arabian Peninsula” is a geographical landmass bordered by the Red Sea in the west, the Arabian Sea in the south, the Persian Gulf in the east, and a vast open *steppe* toward the Mediterranean with no major geographic boundary. It encompasses countries such as Saudi Arabia, Jordan, Yemen, Oman, United Arab Emirates, Qatar, Bahrain, and Kuwait- ethnic people of this region are called the Arabs. While many tribal Arab ethnic groups exist within the Arab population, they descend from a common linguistic and cultural heritage. Shortly after the introduction of Islam in 610 AD, tribes in this region united as an Islamic state and grew up to the current Middle-East North Africa (MENA) regions. Given many conquests in the region after introduction of Islam, stretching from Mauritania (West Africa) to the western China border (East Asia) and historical polygamous marriage practices in the region; the contemporary Arabs share cultural as well as genetic relationships with these ethnic groups. Until recently, the distinct genetic ancestry of these tribal populations was unknown to the scientific community.

Arabian Peninsula was the fundamental bridge for human migration between Africa, Europe, and Asia. Studies suggest that recurrent climate change from arid to humid and vice versa in the Arabian Peninsula during the quaternary period triggered these human migrations and transformed this region as a bridge connecting Africa to Eurasia (Michael D. Petraglia 2010). **Figure 7** illustrates contemporary global genetic diversity mimicking pattern of human migration; prominent Middle East region as the juncture of bifurcation for further migration to Europe, Asia, and Australia using 1K genome project (1KGP) and samples of Arabs Living in Kuwait (ALK). Nevertheless, archaeological and genetic evidence suggests human migration from “Out of Africa” (OOA) to have occurred in two main pathways, “Nile corridor-terrestrial Levantine” path between Africa and West Asia (Macaulay et al. 2005; Marta Mirazon Lahr 1994; Quintana-Murci et al. 1999) and through East Africans upon crossing Red Sea, reaching south of Arabian Peninsula (or Yemen) and migrating eastward (Armitage et al. 2011). Furthermore, excavation remains from archaeological sites in Arabia suggest that 4 human species, *Homo heidelbergensis* (evidence from the presence of Acheulean technologies)*, Homo neanderthalensis* (evidence from fossils in foothills of Zagros mountains in Iraq)*, Homo helmei* (evidence from a specific type of assemblages), and *Homo sapiens* [evidence based on macro-haplogroup L3 marker in mtDNA (Cabrera et al. 2018; Rosa and Brehem 2011; Soares et al. 2012) and modern human remains excavated at Skhul and Qafzeh Caves in the Levant] were inhabiting between 200 and 100 kiloannum (Ka).

Early 20$^{th}$ century excavations in the Arabian Peninsula attempted to reconstruct early civilizations from 1200 BC using available artifacts and text materials. These efforts identified many civilizations and kingdoms that ruled Arabian Peninsula. The southern part was ruled by three successive civilizations built by authentic Yemeni tribes: Mineans, who established their capital Karna (now known as Sadah) (1200–650 BC), Sabaeans in Marib (1000 BC—570 AD), and the Himyarite (2$^{nd}$–6$^{th}$ centuries AD) in Dhafar (Oman). During these times, this region served as an important hub for spice and aromatic trade. The Sabaean civilization survived for about 14

centuries; along with spice trade, they gave importance to agriculture and built a dam at Marib (ancient capital of Yemen) in the 8th century to irrigate farmlands (Korotayev 1995). Similarly, central Arabia was ruled by Kingdom of Kinda (4th–6th AD), eastern Arabia was founded by Dilmum civilization, East African Kingdom ruled by Aksum, later extended up to Yemen and Western Saudi Arabia (in 3rd century), the Lakhimds (of Yemen origin) ruled part of Iraq and Syria (from 300–600 AD). Further, the Arab Christian Ghassanids of Southern Arabia migrated to Jordan (in 3rd century) and established a kingdom comprising Syria and Yathrib (now known as Saudi Arabia) (Korotayev 1996; Munro-Hay 1991). Remains of artifacts excavated from the bay of Kuwait suggest that troops of "Alexander the Great" inhabited *Ikaros* (now known as Failaka Island) in 3rd–4th century BC. This settlement of ancient Greeks is believed to have lasted for approximately 200 years in the Arabian Peninsula (Jonathan Wallace 1979). Parts of the region was also ruled by Romans (Arthur Goldschmidt Jr. 2006) and the Turkish Ottoman Empires (Rogan 2015). These historical conquests have also contributed to Caucasus-related ancestry admixture to the region (Haber et al. 2020). Overall, the shift of kingdoms and civilizations reflects frequent migratory movements between the regions of the Middle East and hence witnesses an admixed genetic architecture at various degrees in Arabs.

Although research on genetic makeup of early humans of Arabian Peninsula is challenged by unfeasibility to recover ancient DNA material (due to non-favorable weather conditions of the region to find quality DNA in excavation dirt), some recent SNP studies using contemporary Arab samples have established that introgression with Neanderthals occurred in this region immediately after OOA and such introgression contributed similar, low levels of genetic signatures in both Levantine and southern Arabian populations when compared with any other non-Africans population (Vyas and Mulligan 2019). In support of this, three genetic sub-groups identified from Qatari population SNP data found 1% to 2.6% of Neanderthal signatures; distinctly, Q1-beduine-qatari sub-group (2.6%) and Q2-persian-qatari sub-group (2.4%) with higher proportion and Q3-african-qatari with lower proportion of Neanderthal signatures (Rodriguez-Flores et al. 2016).

Furthermore, studies from Kuwait and Qatar have identified three genetically well-defined sub-populations that are relatively isolated from one another from the contemporary population of Arabian Peninsula. Both the studies identified sub-groups comprising higher proportions of Arab ancestry (Q1=84%, KWS=69%), higher proportions of Persian or West Asian ancestry (Q2=45%, KWP=56%), and relatively lower proportions of African ancestry (Q3=37.6%, KWB=17%). Distinctively, Q3 of Qatar displayed relatively higher proportions of African ancestry than KWB of Kuwait (Alsmadi et al. 2013; Rodriguez-Flores et al. 2016). Recently sequenced WGS of two Emirati samples highlighted a very different admixture composition with approximately 50%–55% Central or South Asian, 35% Middle Eastern, and 2%–3% Sub-Saharan ancestries (AlSafar et al. 2019) .These studies suggest that the population of Arabian Peninsula is predominantly structured by three to four genetic sub-groups (**Figure 8A**).

Another study comprising a sample set of good representation from the Middle East region identified six ancestral admixture components with differential proportions of West-African, East-African (minute proportion), North-African, Levantine/Caucasian, European and South Asian admixtures. Arabian/North African component was seen in higher proportions in the west of the Peninsula and diminished gradually toward the south-east. The proportion of this component was seen 65% in Saudi, 61% in Yemen & Bedouin, 36% in Oman, and 32% in the UAE. The next highest component seen was a Levant/Caucasian component with 72% (highest) in Druze, 28% in the UAE & Oman, 20% in Yemen, 18% in Bedouin, and 16% in Saudi Arabia. Whereas, European admixture known to be higher in Levantine population with approximately 8%–17%; was seen with 1.5%–3.3% in the Arabian Peninsula. Similarly, sub-Saharan African ancestry was 11% in Yemen, the UAE & Oman and then between 4%–7% in Saudi Arabia. The South Asian component was especially seen frequently in the UAE and Oman with 23%–26% (Fernandes et al. 2019).

Additionally, using Human Genome Diversity Population (HGDP), admixture of following six ancestral components in Kuwaiti subgroups were identified: Negev Bedouin (denoting Arabian ancestry), Yoruba (from sub-Saharan Africa), Brahui tribe (from Pakistan), Druze (from Israel), Kalash tribe (from Pakistan), and French Basque (from Europe). The proportional assignment of these ancestries varied among three subgroups. The subgroup KWP (or KW1) is seen dominantly with West Asian (Brahui, Druze, and Kalash) ancestry at 56% and European ancestry at 12%; KWS (or KW2) subgroup is found predominantly comprising Arabian ancestry at 69%, and KWB (or KW3) is also found to predominantly comprise Arab ancestry at 40%, but with 17% of African ancestry along with it (Alsmadi et al. 2013). This suggests that KWP group is largely of Eurasian (Indo-Persian) origin and KWS and KWB groups are largely of Arabian ancestry with KWB group having a distinct African ancestry (**Figure 8B**).

The mtDNA variations and haplogroups are maternally inherited ancient polymorphisms, which help classify the geographical origins of populations. Theoretically, mtDNA never undergoes recombination and haplogroups are never interrupted by LD. Hence, mtDNA is crucial in exploring the maternal lineages and genetic diversity of Arab populations. Comparative analysis of mtDNA hypervariable region I (HVR I) of Saudi Arabian, Iranian, or Bedouin ethnic individuals suggested that the combined Kuwaiti population has a high frequency of haplogroup R0 (17%), J (12%), U (12%), M (8%), N or R (15%), and H (9%). Furthermore, the detection of sub-haplogroup L (L1 and L2) (2%) suggested contemporary African gene flow and the presence of haplogroup M and N or R indicated Asian gene flow into Kuwait (Theyab et al. 2012). The combined Kuwaiti population showed the absence of L3, thus reflecting no earlier African migration into Kuwait. But distinctively in Saudi Arabian population, high frequency of haplogroup L (10.5%) and the presence of L3 (Abu-Amero et al. 2007; Theyab et al. 2012) was observed, suggesting high frequency of contemporary African gene flow and early African migration in Saudi Arabia.

Conversely, regions of Y-chromosome except pseudo-autosomal regions, are transmitted from father to generations with few modifications. These regions are also helpful in discerning recent events of human migration in Arab population. The subgroup Bedouin is traditionally asserted to

descent from two main lineages: from Adnan (sons of Ishmael) - the Adnani lineage and from Qahtan- the Qahtani (or Joktani) lineage. Autosomal STR and Y-STR DNA from the Adnani tribes of Al-Aniza, Mutran, & Awazim (a Suluba tribe) and the Qahtani tribes of Ajman, Shimar, & Murrah, have identified haplogroups J1 (84%); R1a1(6.75%); E3b3 (6.00%); G2 (3.40%); R1b3 (1.35%); and K2, E3b1, Q*, & R2 (each 0.67%) from Y-STR; suggesting strong evidence of genetic isolation and drift while no evidence of segregation into the two male lineages (Mohammad et al. 2009). A similar study using Y-STR polymorphisms has shown that Arab populations share a substantial part of the genetic pool and a pronounced genetic similarity among Arabian Peninsula populations (Triki-Fendri et al. 2010).

In summary, ancient and recent migrations have shaped the genetic structure of the contemporary Arab population of the Arabian Peninsula. The proportional genetic ancestry assignment varied from western to eastern sides of the Peninsula with the dominant sub-Saharan African component in the west and stronger South Asian and Levantine/European component in the east of the Peninsula, which is distinctively classified as three subgroups of ancestries: Persian (European and South Asian), Saudi Arabian tribe (or Negev Bedouin, descendants of the first Eurasians), and nomadic Bedouin (with varying proportions of African ancestry).

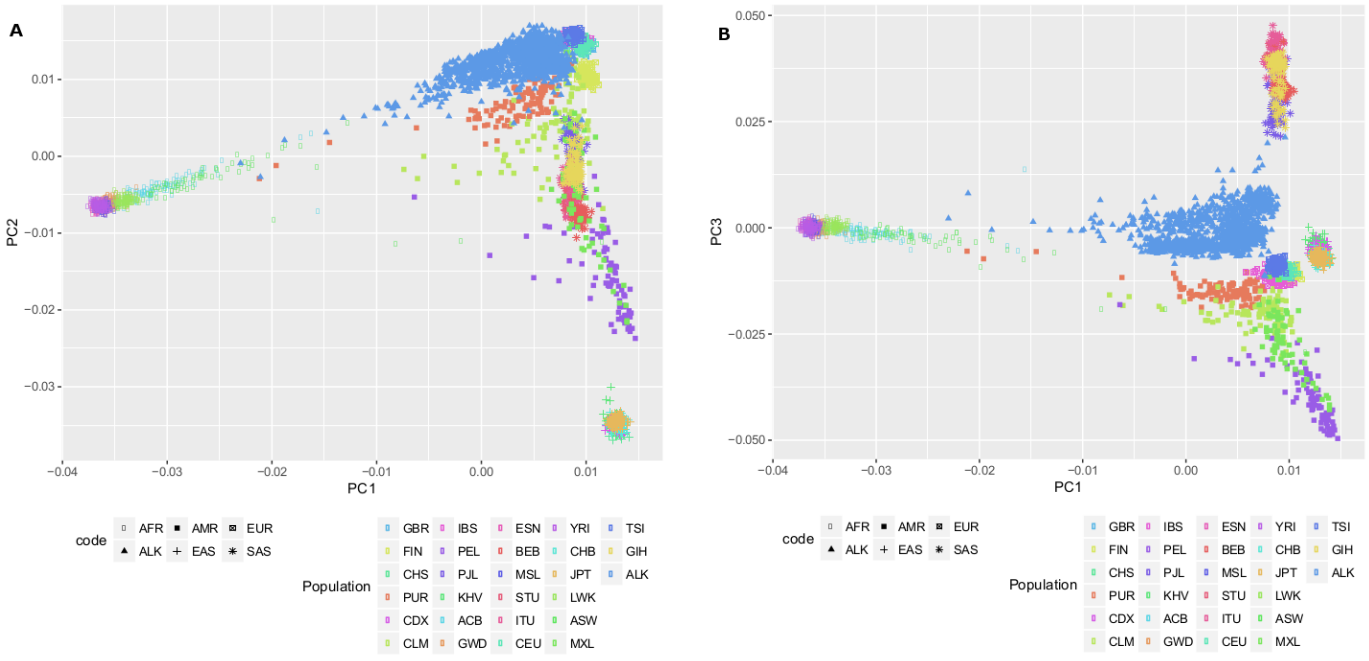


**Figure 7: Principal component-based scatter plots (A: PC1-PC2 and B: PC1-PC3) depicting pattern of global human genetic diversity due to early human expansion from Africa; genetic signatures of contemporary humans from the Middle East (Arabs are shown in blue triangles) hints that the initial spread was through the Middle East region and the region served as a bifurcating juncture for populations further north into Europe, east across Asia, and south to Australasia. With permission from Springer Nature.**

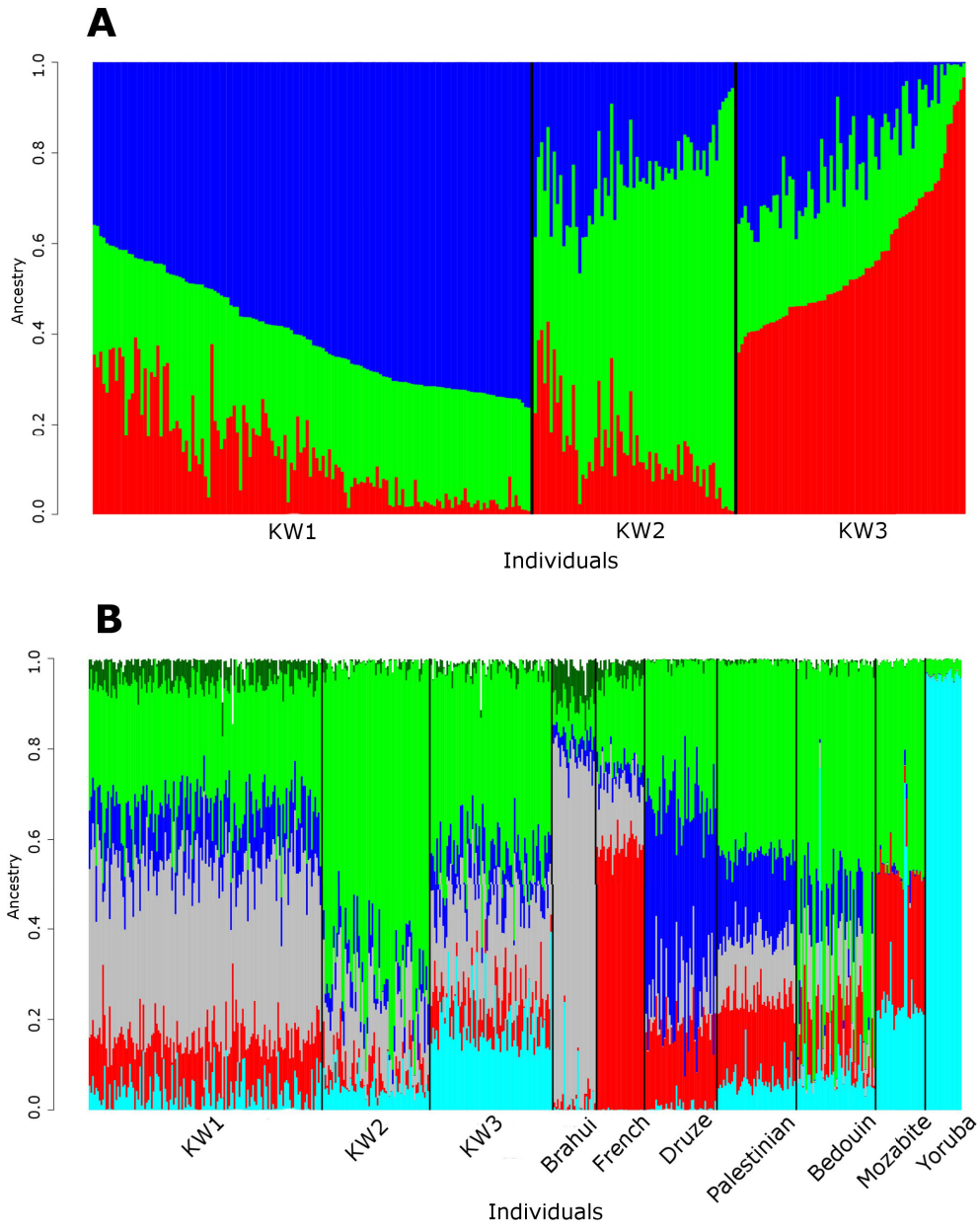


**Figure 8: Genetic admixture proportions observed in contemporary Arabs using structure program. A) Three subgroups were observed with KW1 being dominant with Eurasian (Indo-Persian) origin; KW2 and KW3 were largely of Arabian ancestry with KW3 group having a**

**distinct African ancestry as well. B) Corroboration of HGDP data with our data identified six ancestral elements in the three Kuwaiti groups; the admixture proportional assignment is as follows: West Asian (Brahui, Druze, and Kalash) ancestry is dominant in Kuwait 1 group (56%), European ancestry is seen more in Kuwait 1 group (12%), Arabian ancestry is seen more in Kuwait 2 group (69%) and Kuwait 3 group (40%), and African ancestry is seen more in Kuwait 3 group (17%) (Alsmadi et al. 2013). With the permission from PLOS.**

## 6.10 Pre-Oil era lifestyle of Arabs

In the pre-oil era, people in the Arabian Peninsula lived in a simple rural village with a nomadic lifestyle. Villages were often built around water resources (or oases) or near the coastal side. Settled people practiced regular farming and cultivated grains and palm trees in the oases. Nomadic communities, traditionally known as "Bedouin," were also seen who pastured camels, sheep, and goats using the limited water resources of the desert. Fishermen lived along the coast (Albert Hourani 2010). Village housing structures were built of mud; tree trunks supported roof covered with palm tree branches, and woven palm tree branches were used for flooring. Daily living needs were imported from Indian subcontinent and parts of Africa in exchange for dates, pearls, and oyster shells (Albert Hourani 2010; Barakat 1993; Hitti 1914). People mostly consumed pita (or khubz, a form of bread) or rice along with fish or other seafood. Archeological and textual data survey estimates that rice was a cultivar in the Middle East region from the 12$^{th}$ century BC up early Islamic period and was consumed as bread, porridge, or cake, much like barley and wheat (Muthukumaran 2014). A status mentioned in the old Arab folk saying "What do the people of paradise eat? — rice in butter" (Richard Tapper 1996) indicates that rice was a luxury and choice of wealthy people. Pitas were traditionally accompanied by hummus, baba ganoush, tzatziki, tabbouleh, lamb, steak, falafel, and chicken. In summary, pre-oil era of Arabian Peninsula was physically active and led a more natural lifestyle.

## 6.11 Post-Oil discovery: Urbanization led lifestyle change in Arabs

Since the discovery of oil in South-west Persia in 1908, many oil niches have been discovered and used in the Arabian Peninsula region to fulfill global consumer demands. An intergovernmental organization called "The Organization of the Petroleum Exporting Countries (OPEC)" including countries such as Iran, Iraq, Kuwait, Saudi Arabia, and Venezuela was formed at the Baghdad Conference in 1960 to coordinate in securing fair and stable prices for petroleum producers. Today, countries in this region have become the world's top oil or gas exporting countries. According to the U.S. Energy Information Administration report of 2019, the net oil export revenues of Middle Eastern countries in the OPEC of Iraq, Iran, Kuwait, Saudi Arabia, and UAE are respectively $91, $67, $61, $237 and $74 billion (EIA 2019). Most of this wealth is used to urbanize regions of the Arabian Peninsula and provide health & food security for the population.

According to the report by World Urbanization Prospects (WUP-2018), the proportion of urbanized regions in the total landmass area in Arabian countries by 2018 were as follows: Bahrain- 89%, Iraq- 70%, Jordan- 91%, Kuwait- 100%, Qatar- 99%, Saudi Arabia- 84%, UAE- 87%, and Oman- 85%. At the same time, there was a drastic increase in inhabited population of these countries by two times in Iraq, Kuwait & Saudi Arabia, three times in Bahrain, four times in Oman, and six times in Qatar & UAE. This increase in population was mainly due to the induction of a large-scale expatriate population for skilled and unskilled labor (DESA-UN 2018).

Although economic growth brought great opportunities for infrastructure development, it also invited the burden of high dependence on mechanization, access to cheap migrant labor, and proliferation of western fast-food diets. This led to a dramatic transformation in dietary patterns and the active lifestyle of native wealthy Arab population (Klautzer et al. 2014). Predominant dietary regimes such as dates, milk, fresh vegetables, fruit, rice, whole wheat bread, and fish changed to foods enriched with high saturated fats and refined carbohydrates with low fibers. A report from Saudi Arabia indicates that the average meat consumption had increased by 2.2%, whereas cereal consumption declined by 0.1% per individual from 1993 to 2003 (Kotilaine 2010).

Similarly, a study on eating behavior in Emirati adolescents (with medium or high family affluence) from UAE reported a significant change in dietary regimes. Only 28% of the study cohort met the recommended daily intake of fruits and vegetables (Makansi et al. 2018). Studies from Qatar also assessed dietary regime of Qatari adults and its effect on metabolic health (Al-Thani et al. 2017; Al Thani et al. 2018). *Al Thani M et al.,* reported that 83% of adults did not meet daily recommendations of vegetables, fruits, whole grains, legumes, and fiber intakes. Moreover, 50%–72% adults were reported to be consuming sweetened beverages & sweets and 47% were reported to frequently consuming fast food. Also, majority of the adults were found to have a metabolic syndrome (Al Thani et al. 2018).

Low physical activities aggravate the impact of nutritional changes. The physical environment plays a key role in this; however, excessive urbanization in the Arabian Peninsula offers limited physical activities and exercise facilities. Additionally, unbearably hot and dusty (due to sandstorms) climatic conditions prevent people from performing strenuous outdoor exercise. The decrease in physical activity is further instigated by easy access of automotive and mechanic appliances, cheap migrant workers from neighboring countries, leisure-time sitting, extended periods of TV viewing, as well as computers, and computer games in the region. Among women, sedentary lifestyle is augmented by cultural barriers to perform strenuous physical exercise and sports activities. Restrictions on women to be accompanied by a male family member and stern rules to wear traditional dresses like abayas when going outside pose significant barriers to undertake physical activities. The drastic reduction of domestic chores due to availability of cheap work force has led to a marked reduction in physical activity of Arab women.

In summary, exceptional growth in prosperity has brought rapid changes in lifestyle, resulting in a significant rise in chronic diseases in the Arabian Peninsula. The number of people diagnosed with diabetes has increased strikingly. The excessive consumption of high-energy food has

contributed to increase in body weight and various non-communicable disease-related morbidities and mortalities.

## 6.12 Prevalence of metabolic disorders in the Arabian Peninsula

The sudden rise of urbanization in the Arabian Peninsula has resulted in major demographic and epidemiological transitions with obesity, diabetes, hypertension, fatty liver, and chronic diseases becoming the leading causes of morbidity and mortality. A recent worldwide survey to evaluate mortality in the Gulf nations due to non-communicable diseases (NCD) suggests that 27%–41% of the total NCD-related deaths are attributed to cardiovascular diseases and 2%–14% to diabetes. **Figure 9** illustrates country-wise proportional mortality due to NCDs and major chronic disorders. The same survey also reported key risk factors for NCD, among which obesity and physical inactivity represented a significant proportion. **Figure 10** shows the Gulf country-wise proportion of risk factors involved in NCD mortality. In addition, recent International Diabetes Federation (IDF) survey indicates that the prevalence of age-adjusted diabetes in the Middle East and North Africa (MENA) regions is highest (12.2%) in the world and estimated to reach 13.9% by 2045 (IDF 2019). It is estimated that 21%–78% of women in Saudi Arabia are expected to be obese by 2022 (Al-Quwaidhi et al. 2014).

Studies from the region have observed that the prevalence of metabolic syndrome is 10%–15% higher than in western countries and is higher in women as compared to men (Alzaabi et al. 2019; Mabry et al. 2010b; Ng et al. 2014). Moreover, women outnumber men in diabetes with an incidence ratio of 57:43 (Channanath et al. 2013). Similarly, the estimated prevalence of obesity exceeds by 50% among women of Kuwait and Qatar (Ng et al. 2014). Nearly all the Gulf nations that have faced similar economic transitions show high prevalence of chronic metabolic disorders. **Tables 2 and 3** list the country-wise prevalence of major chronic metabolic disorders and dyslipidemia respectively, in Arab nations. Interestingly, similar differences are also observed in behavioral risk factors. For example, studies based on self-reported physical activity have shown low prevalence of physical activity in adults of Gulf Cooperation Council (GCC), ranging from 39% to 42% for men & 26% to 28% for women, compared with western populations where it is 50% & 60% for men, and 47% & 54% for women in the USA & Australia, respectively (Centers for Disease and Prevention 2007; Mabry et al. 2010a; Tim Armstrong 2000).

The current pharmacological approach and treatments have shown inadequate efficacy among Arabs. Nearly two-thirds of the dyslipidemia patients treated with statins and diabetes patients undergoing treatment in the Gulf region have inadequately controlled lipid and glycemic levels, accordingly (Al Sifri et al. 2014; Qaddoumi et al. 2019). An intriguing fact is that the Dyslipidemia International Study (DYSIS) of Middle East found that despite chronic statin treatment, target LDL-C levels were not attained in 61.8% of dyslipidemia patients, as compared to their counterparts in Canadian and European cohorts where 48.2% of patients had not reached target LDL-C levels (Al Sifri et al. 2014).


**Figure 9: Statistics of mortality due to NCDs in Middle Eastern countries. A) Proportion of mortality due to NCDs among total mortality rate. B) Proportional mortality rate of major class of disorders in NCDs observed in the Gulf countries.** Note: data shown here is obtained from Non-communicable diseases country profiles 2018 of World Health Organization.

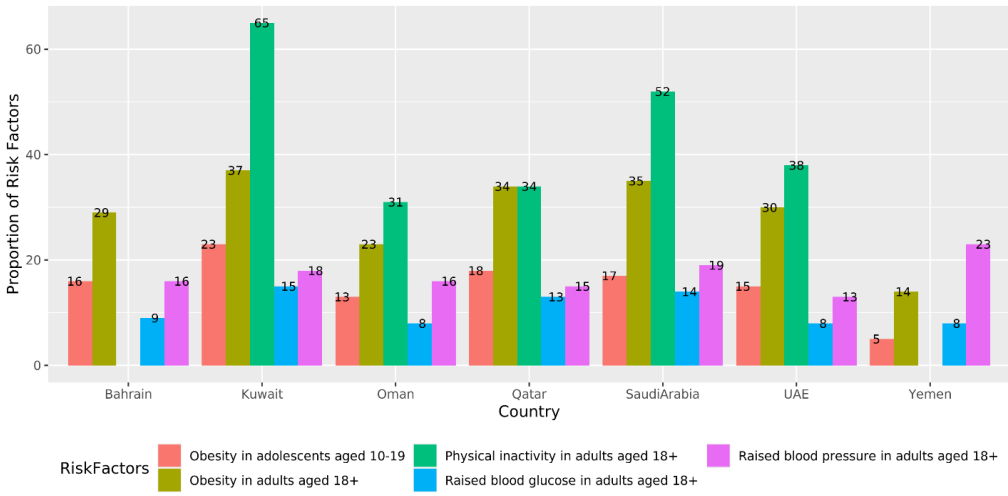


**Figure 10: Proportion of metabolic risk factors for non-communicable disease in seven countries in the region.** Note: data shown here is obtained from Non-communicable diseases country profiles 2018 of World Health Organization.

| Disorder | Country | Prevalence among men | Prevalence among women | Prevalence among children | Study |
|---|---|---|---|---|---|
| Obesity | Saudi Arabia | 30 | 44.4 | 12.7 - 24.1 | (Alzaabi et al. 2019; Farrag et al. 2017; Hammad and Berry 2017; Ng et al. 2014; Raja'a and Bin Mohanna 2005; Zayed et al. 2016) |
| | Jordan | 27.5 | 45.6 | 6.5 - 9.7 | |
| | Yemen | 4.1 | 24.7 | 1.8 | |
| | Oman | 20.6 | 36.9 | 9.9 | |
| | UAE | 27.1 | 33.2 | 8 | |
| | Qatar | 44.0 | 54.7 | 9.3 – 21.2 | |
| | Bahrain | 31.0 | 42.9 | 10.5 | |
| | Kuwait | 43.4 | 58.6 | 12.4 - 22.2 | |
| Type 2 Diabetes and pediatric diabetes | Kuwait | 20.4 | 17.4 | 44.5* | (Alkandari et al. 2018; Channanath et al. 2013; Moussa et al. 2008; Saraswathi et al. 2019) |
| | UAE | 21 | 23 | - | (Hamoudi et al. 2019) |
| | Saudi Arabia | 12.9 | 11.4 | 33.5* | (Bahijri et al. 2016; Saraswathi et al. 2019) |
| | Oman | 13.9-17.9 | 12.3-21.1 | 2.7* | (Al-Lawati et al. 2015; Saraswathi et al. 2019) |
| | Yemen | 9.75 | - | - | (Gunaid 2002) |
| | Bahrain | 33.6 | - | 2.7* | (Meo et al. 2019; Saraswathi et al. 2019) |
| Hypertension | Kuwait | 30.7 | 32.0 | - | (Akl et al. 2017) |
| | Saudi Arabia | 28.6 | 23.9 | - | (Al-Nozha et al. 2007) |
| | Oman | 46.5 | 50.7 | - | (Al Riyami and Afifi 2002) |
| | Yemen | 17.1 | - | - | (Gunaid 2002) |
| | UAE | 26.2 | 29.4 | - | (Akl et al. 2017) |
| Metabolic Syndrome | Oman | 19.5 | 23 | - | (Al-Lawati et al. 2003) |
| | UAE | 26 | 14 | 3.7-8.9 | (Haroun et al. 2018; Malik and Razig 2008; Rodhan Khthir 2014; Shah et al. 2020) |
| | Saudi Arabia | 31.4 | 25.2 | 9.4 | (Al-Daghri 2010; Al-Nozha et al. 2005; Aljohani 2014) |
| | Kuwait | 34.2-41.7 | 37.7-40.1 | 21.3 | (Al Zenki et al. 2012; Boodai et al. 2014) |
| Fatty liver | Saudi Arabia | 49.1 | 46.3 | - | (Alswat et al. 2018) |

**Table 2: Prevalence of major chronic metabolic disorders in the Gulf countries that have faced similar socio-economic transitions, Units in percentage; except for * shown in incidence per 100,000.**

| Country: Prevalence for dyslipidemia | Individual Lipid levels | Prevalence |
|---|---|---|
| UAE: 72.5% (Mahmoud and Sulaiman 2019) | high total cholesterol | 42.8% |
| | high triglyceride | 29% |
| | low high-density lipoprotein cholesterol | 42.5% |
| | high low-density lipoprotein cholesterol | 38.6% |
| Jordan: (Abujbara et al. 2018) | high total cholesterol | 44.3% |
| | high triglyceride | 41.9% |
| | low high-density lipoprotein cholesterol | 59.5% |
| | high low-density lipoprotein cholesterol | 75.9% |
| Saudi Arabia: (Yasser Taher Al-Hassan 2018) | high total cholesterol | 13.8% |
| | high triglyceride | 17.0% |
| | low high-density lipoprotein cholesterol | 40.0% |
| | high low-density lipoprotein cholesterol | 12.85% |
| Oman: (Hilal Al-Sabti 2010) | high total cholesterol | 29% |
| | high triglyceride | 34% |
| | low high-density lipoprotein cholesterol | 51% |
| | high low-density lipoprotein cholesterol | 20% |

**Table 3: Prevalence of dyslipidemia in Arab nations.**

## 6.13 Co-morbidities among complex metabolic disorders

Being overweight or obese increase the risk of co-morbidities such as metabolic syndrome, T2DM, cardiovascular diseases, and fatty liver disease. The coexistence of other metabolic disorders complicates their treatment. A study from Kuwait (Qaddoumi et al. 2019) showed that among patients with diabetes (with HbA1C ≥7%): 72.2% were dyslipidemic, 70.6% were hypertensive, 70.8% were nephropathic, 78.9% of were neuropathic, 74.2% were retinopathic, and 67.4% had Coronary heart diseases.

Understanding the effects of rare disorders on traits of obesity or diabetes is still a “Cinderella” in metabolic syndrome research in the Arabian Peninsula. Studies provide hints at nexus of hypogonadism with T2DM in Arab consanguineous families (Al-Gazali and Hamamy 2014; Bo-Abbas et al. 2003). A large cohort by *Al Hayek AA et al.,* showed that 36.5% men with T2DM had low serum testosterone levels and 17% of these patients had primary hypogonadism, whereas the remaining had secondary hypogonadism (Al Hayek et al. 2013). Furthermore, G6PD

deficiency was found to be a common X-linked disorder in Iranian (11.5%) (Usanga and Ameen 2000) and Saudi (10%) (Warsy and El-Hazmi 2001) population. Incidentally, G6PD deficiency shares a common link with diabetes (Cappai et al. 2011; Gaskin et al. 2001; Heymann et al. 2012; Lee et al. 2011). The prevalence of G6PD deficiency in men was more than twice when compared with women (Al-Riyami and Ebrahim 2003). Similarly, Wolfram syndrome is an autosomal recessive neuro-degenerative disorder found to manifest as early onset diabetes mellitus and progressive optic atrophy (Pallotta et al. 2019).

Rare disorders such as thalassemia, cystic fibrosis, Huntington's disease, and Friedreich's ataxia are known to increase patient's predisposition to diabetes (Cutting 2010; De Sanctis et al. 1988; Hu et al. 2014; Podolsky et al. 1972; Ristow 2004). Astonishingly, the genetic risk loci of many rare recessive disorders, including hepatic lipase deficiency (LIPC), Pancreatic agedness 1 (PDX1), Hypophosphatemic rickets- also associated with obesity (ENPP1), Wolfram syndrome 1 (WFS1), and SLC2A2 (Fanconi-Bickel syndrome) were also known to overlap with T2DM. A study by *Blair et al.* (Blair et al. 2013) discovered several Mendelian variations contributed to complex disease risk from large hospital medical records, and GWAS results revealed that the relative risk of Mendelian disorders likely contributed to the risk of a subset of complex diseases (**Figure 11**).

In summary, with this evidence in mind, paramount importance to account pleiotropic effect of rare disorders is needed when researching complex disorders. Identifying trait overlaps with metabolic disorders from already published literature on co-morbid disorders and extensive phenotyping related to such co-morbid disorders may clarify molecular links between them. Using such additional phenotypes as covariates in the statistical analysis may also help in identifying such overlapping markers.

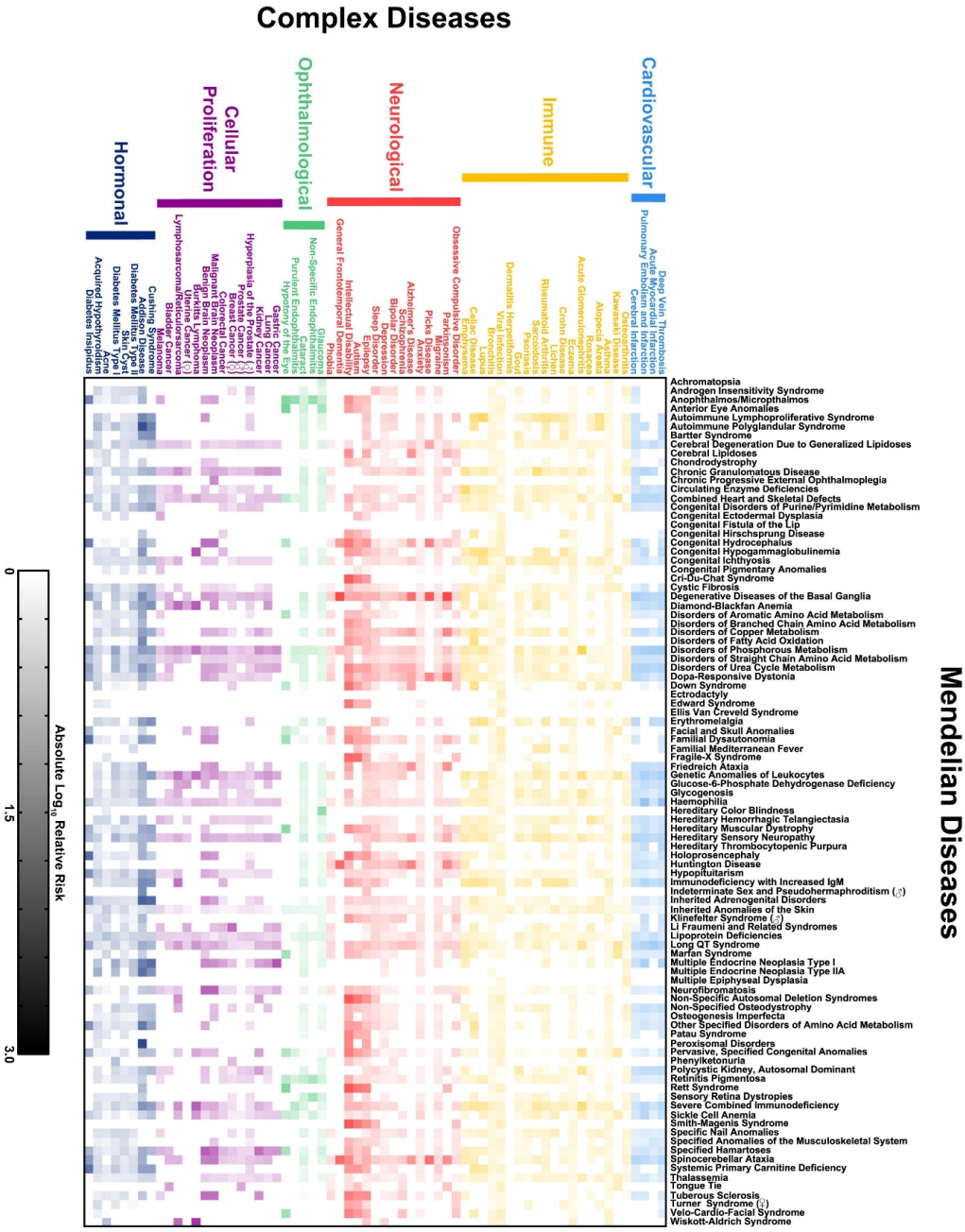


**Figure 11: Illustrating relative risk of Mendelian diseases among complex diseases. Image source: with the permission from Elsevier Science.**

## 6.14 Cultural barriers in Arabs and adverse effect on health

Consanguineous (in Latin, *con* means "shared" and *sanguis* means "blood") marriages are firmly embedded in Arab society with a prevalence of 54.3% in Kuwait (Al-Awadi et al. 1985), 50.5% in UAE (al-Gazali et al. 1997), and 57.7% in Saudi Arabia (el-Hazmi et al. 1995). This type of marriage is culturally favored for numerous reasons. It is widely believed that consanguinity strengthens family ties, reduces chances of hidden uncertainties within families in terms of health and wealth, ensures the properties of the family, and increases compatibility of women with the husband and his family members. However, research has shown that consanguinity could affect reproductive health (Fareed and Afzal 2014b; Tadmouri et al. 2009) (including postnatal mortality, congenital malformation, stillbirths, low birth weight, preterm delivery), cognitive impairments (Afzal 1988; Badaruddoza and Afzal 1993; Fareed and Afzal 2014a), cardiovascular risks (Badaruddoza et al. 1994; Fareed and Afzal 2016), and several other complex disorders (Bittles and Black 2010). The association of consanguineous marriages with late onset complex diseases such as T2DM, obesity, and cardiovascular disorders is not yet precisely linked with the metabolic risk effect. However, an increased risk of T2DM has been observed among the offspring of such consanguineous marriages in Saudi Arabia and Qatar (Bener and Hussain 2006; Gosadi et al. 2014). The familial clustering of T2DM has been reported in Arab populations from Morocco (Benrahma et al. 2011), Tunisia (Arfa et al. 2007), Oman (Al-Sinani et al. 2014a), and Qatar (Bener et al. 2013). Furthermore, many rare Mendelian and familial genetic disorders were observed in the Arab populations. Also, the disorders that are rare in other populations such as hypogonadism, G6PD deficiency, but common in Arab population, exhibit pleiotropic effects on complex metabolic disorders (Al-Gazali et al. 2006).

Another important form of marriage practice that exists in the Arabian Peninsula is polygamy, with 8%–12.5% prevalence in Kuwait (Chaleby 1985). The Islam religion permits men to practice polygamy (providing men have the financial and physical means to treat each of them equally, then men can marry up to four wives). This practice continues to exist, and some governments provide social benefits to such marriages. For example, UAE provides social benefits for the second wife in terms of housing. However, this is believed to be done to resolve a high number of spinsterhood from weakening social structure of the country. Although the ill effects of polygamy on health from the perspective of genetics are unclear, it does reduce the number of males contributing to gene pool, thereby leads to reduction of genetic diversity in the population. A study from Saudi Arabia consisting of patients with coronary artery disease (CAD) in polygamous marriages showed a significant association of CAD, multi-vessel disease (MVD), and left main disease (LMD) with the number of the wives (Daoulah et al. 2017). However, the underlying mechanism of association with these disorders is unknown, but psychosocial factors, stress of polygamy, and having less time for physical activity may attribute to the cause.

Arab marriages are predominantly consanguineous with a subtype called "*Bintamm*" (i.e., between a man and father's brother's daughter) (Al-Awadi et al. 1985). This form of marriage is traditional and socially well respected in the community. Balefully, the trend of consanguineous

marriages has still not declined in the Arab society (Warsy et al. 2014). Even when the marriage is outside close kin, it is rare that it is outside the tribe, thus leading to living in isolation of the community. These marriage practices from successive generations cumulatively increased the inbreeding levels and consequently unequal distribution of founder mutations was observed among the population.

## 6.15 Signatures of inbreeding in Arabs

In Arab population, high inbreeding edifies paramount importance for the evaluation of ROH and IBD, as these are vital signatures of inbreeding. ROH and IBD indicate ancient and recently shared common ancestry, respectively. Both help to understand population demography and recessive components of Mendelian and complex phenotypes. Stretches of homozygous regions along the human genome, called runs of homozygosity (ROH) are important for understanding human health and population history. Parental relatedness among others is a major mechanism for ROH and serves as a good measure for consanguinity. ROH arising from recent inbreeding tend to be long segments in the genome, whereas shorter ones indicate ancient inbreeding (Kirin et al. 2010). Outbred populations rarely show longer segments of ROH (>15Mb) with possible exceptional cases reported elsewhere (Nalls et al. 2009). IBD, on the other hand, gives insights about shared common ancestry and sheds light on population history from a different perspective. IBD segments from recent common ancestor tend to be long (~10cM) and IBD segments from an ancient common ancestor are short (~2cM).

The levels of inbreeding significantly vary between sub-groups of Arab population. The percentage distribution of inbreeding coefficient observed at various ranges in three subgroups of Kuwaiti Arab population is presented in **Table 4**. Among these, Saudi and Persian tribes exhibit a high level of inbreeding coefficient (0.04226 and 0.025742, respectively), indicating positive assertive mating/endogamy, whereas Bedouins in Kuwait exhibit less inbreeding coefficient (0.00274) in contrast to Negev Bedouins (from the Mediterranean region whose consanguinity is 44.8% and inbreeding is 0.0238) (Na'amnih et al. 2014)), indicating negative assertive mating/heterogamy (Alsmadi et al. 2013). This suggests that Saudi and Persian tribes may be at increased risk for recessive disorders because of the expression of autosomal recessive gene mutations inherited from a common ancestor in each sub-group. This genetic heterogeneity between the sub-groups warrants sub-group specific genetic association analysis with large sample size, particularly for subgroups with more inbreeding.

| Range of Inbreeding coefficient | Percentage distribution of individuals in subgroups | | |
|---|---|---|---|
| | KW1(P group) | KW2(S group) | KW3(B group) |
| (-0.05 to 0.00) | 23.91% | 0% | **72.22%** |
| (0.00 to 0.05) | **53.62%** | **68.25%** | 16.67% |
| (0.05 to 0.01) | 17.39% | 19.05% | 9.72% |
| (0.01 to 0.15) | 3.62% | 11.11% | 0% |
| (0.15 to 0.20) | 1.45% | 1.59% | 1.39% |

**Table 4. Percentage distribution of inbreeding coefficient observed at different ranges among three subgroups of Kuwaiti Arab population. With permission from PLOS.**

Similar to inbreeding coefficient, the differential distribution of ROH was also observed among the three subgroups of Kuwaiti Arabs (Alsmadi et al. 2013). This distribution is in alignment with the distribution of inbreeding coefficients, i.e., KW2 (Kuwait Saudi) group shows the highest total ROH and ROH segments, whereas KW3 (Kuwait Bedouin) shows the least total ROH and ROH segments (**Figure 12**).

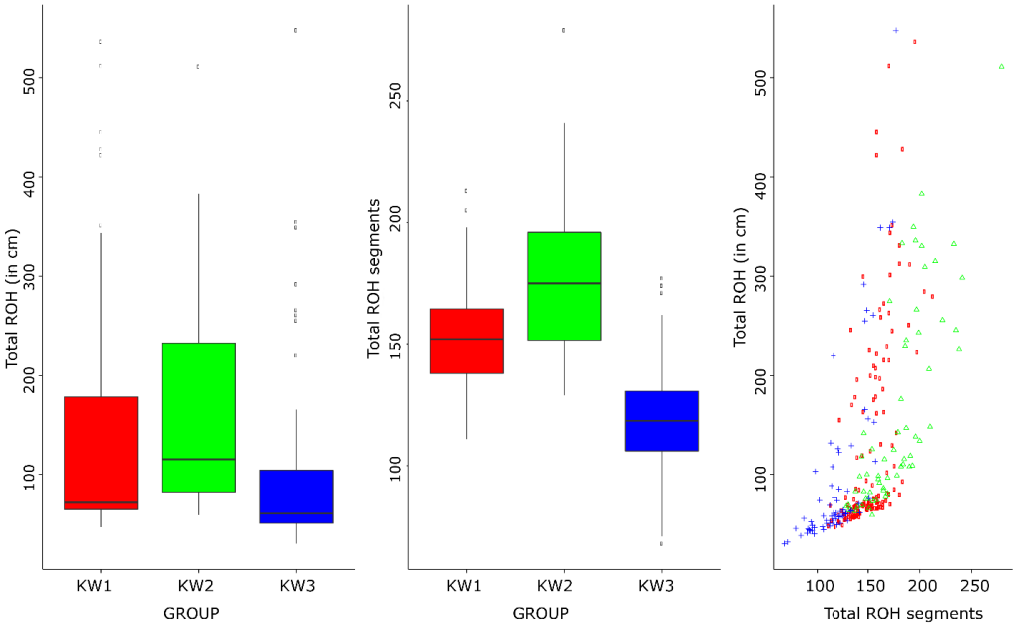


**Figure 12: Distribution of total ROHs & ROH segments and ROH vs. ROH segments in three Kuwaiti groups: KW1 and KW2 show higher total ROH and ROH segments, whereas KW3 shows a much lower segments, suggesting KW1 and KW2 are a highly inbred population. With permission from PLOS**.

## 6.16 Current status of genetic studies on metabolic disorders in Arabs

Recently, several large-scale genetic studies have been conducted in this region, such as the dissection of genetic sub-groups among Arabs (Alsmadi et al. 2013; Rodriguez-Flores et al. 2016) and whole genome sequence projects from Qatar (Fakhro et al. 2016), Kuwait (Alsmadi et al. 2014a; John et al. 2015; Thareja et al. 2015), UAE (AlSafar et al. 2019), and Saudi Arabia (Project Team 2015) to explore the distribution of Arab specific variants, development of curated database on regional inherited disorders and gene mapping the "Catalogue for Transmission Genetics in Arabs" database (Tadmouri et al. 2006), development of Greater Middle Eastern (GME) variation (Scott et al. 2016) and ALMENA (Koshy et al. 2017) databases to get insight on variant distribution in Arabs, biannual conferences from (Pan Arab Human Genetics Conference and Medical Genetics) UAE and Kuwait to provide a platform for regional and international genetic researchers to increase the knowledge of genetic architecture of Arab population as a whole. Nevertheless, despite these worthy endeavors, comprehensive genetic research for dissecting the molecular etiology of metabolic disorders is lacking. Despite the high prevalence of metabolic disorders, genetic studies on these are limited. Of the total research articles based on genetic research on non-communicable disorders, much heritable metabolic disorders like CVD (12.3%), diabetes (10.9%), obesity (3.70%), hypertension (2.80%), and metabolic syndrome (0.40%) were under-investigated in Arab region (Jamaluddine et al. 2016). This could be due to challenges in building facilities to conduct high throughput research, harnessing qualified manpower, and convincing public to participate in research programs through health promotion activities in the Middle East. A detailed viewpoint on challenges of conducting genetic research activities in the Arab population has been discussed by *Hebbar P et al.,* (Hebbar et al. 2019b).

## 6.17 Examination of ethnic transferability variants by candidate variants studies

Replication of the most significantly associated polymorphisms in multiple populations with distinct genetic backgrounds and lifestyles is crucial for understanding the pathophysiology of a multifactorial disease. Examining variants from candidate genes that are related or previously reported in a population for disease in question is a quick and cost-effective way for understanding the role of variants in disease etiology. Such variants are examined using two strategies: (1) variant-based replication, in which only variants from nucleotide sites are genotyped & followed up (2) sequence-based replication, in which the gene region is sequenced in the replication sample, and both known & novel variants are tested.

A few studies from Arab region have attempted to replicate variants associated with European population for metabolic traits or disease status. However, only few variants were replicated in Arabs. A possible reason for failure to replicate these variants could be that the examined variants may simply lead variants, not essentially causal variants in European population or they are less frequent in Arabs than in Europeans or these associations are causally linked to different

environmental exposures in these two populations. A thorough literature survey has shown that 11 studies on T2DM, 6 on obesity, 2 on Diabetic Kidney Disorder (DKD), and 1 on hypertension have attempted to examine variants found in association with European or East Asian population for respective metabolic disease traits. The list of candidate SNP studies on metabolic risk loci from the MENA region and examination of established metabolic risk loci in Arab population are shown in **Table 5**. These studies have shown that 52 variants of T2DM and 27 variants of obesity were transferable in Arabs. Collectively, 52 candidate variants from 35 genes (with rs7903146 of TCF7L2 replicating in 7 studies; rs5219 of KCNJ11 and rs13266634 of SLC30A8 in 3 studies; rs1470579 of IGF2BP2, rs10811661 of CDKN2B-AS1, rs231362 of KCNQ1, rs4506565 of TCF7L2 in studies) have been shown to replicate for T2DM (shown in **Table 6**). Similarly, 27 variants from 24 genes (with rs1421085 of FTO replicating in 2 studies) have been shown to replicate for obesity traits (**Table 7**) in the Arab population. Importantly, of the total T2DM and obesity SNVs studied, only 35% of the SNPs could be replicated in Arab population through candidate variant/gene studies.

| Candidate studies | Study details | Genes or SNPs included | SNVs that replicated at P<0.05 |
|---|---|---|---|
| (Li-Gao et al. 2018) | Tested 122 **T2DM** risk variants (from 84 gene loci) in **Saudi Arabian** population using 659 T2DM cases and 919 controls. | None of the tested markers was replicating at p-value corrected for Bonferroni threshold (0.05/122). At p-value<0.05, 11 variants from TCF7L2, CRY2, PROX1, IRS1, SLC30A8, FTO, G6PC2/ABCB11, FITM2-R3HDMLHNF4A, ZFAND3 were seen replicating. | rs7901695, rs4506565, rs7903146, rs11605924, rs340874, rs2943641, rs13266634, rs11642841, rs560887, rs6017317, rs9470794 |
| (Osman et al. 2018a) | Tested established loci (BMI:87; WC:58; **obesity** with **T2DM**:145) in **UAE Arab** population using 880 BMI cases; 455 WC cases; 464 T2DM cases and 415 controls. | Replicates BMI: FTO, LOC284260, USP37; WC: RFX7, MYEOV; Height: NSD1, MFAP2, AP4E1, ARHGEF12, NPPC, ZDHHC7, CYTL1, FNDC3B (for obesity) and TCF7L2, NLRC3, FLJ30838, KCTD15 (for T2DM). Also reports novel associations *KCNK3* and *RARB* for T2DM. | BMI: rs1421085, rs7239883, rs492400; WC: rs8030605, rs11607976; Height: rs28710014, rs2284746, rs2306335, rs4938802, rs749052, rs2326458, rs4689977, rs1039027; T2D: rs7903146, rs6567160, rs11126666, rs6804842, rs2077235, rs1016287, rs29941 |
| (Cauchi et al. 2012) | Tested 44 variants from 37 established loci for **T2DM** in **North African Arabs;** 1193 T2DM cases and 1055 controls from Morocco, 1446 T2DM cases and 942 controls from Tunisia. | 13 of 37 established gene loci confirmed in Moroccans and Tunisians. *BCL11A, ADAMTS9, IGF2BP2, WFS1, CDKAL1, TP53INP1, CDKN2A/B, TCF7L2, KCNQ1, HNF1A, FTO, MC4R,* and *GCK.* | rs243021, rs4607103, rs1470579, rs10010131, rs7756992, rs10811661, rs564398, rs7903146, rs231362, rs7957197, rs1421085, rs17782313, rs1799884 |
| (O'Beirne et al. 2016) | Tested 37 variants from 29 established **T2DM** gene loci and an additional 27 tag SNPs in **Qatari** population using 1124 T2DM cases and 590 controls. | Only *TCF7L2* of the tested 29 loci was seen replicating in Qatari population. | rs7903146, rs4506565 |
| (Al-Daghri et al. 2014b) | Tested 28 established **T2DM** loci in **Saudi Arabian** Population using 1166 T2DM cases and 1235 controls. | Replicates 8 SNV of the 37 established T2DM variants of *TCF7L2, WFS1, JAZF1, CDKN2A/B, KCNQ1, HNF4A,* and *DUSP9*. | rs1801214, rs849134, rs10965250, rs7903146, rs231362, rs163184, rs4812829, rs5945326 |
| (Tomei et al. 2015) | Tested 23 established **obesity**-related loci in **Qatari** population using 804 individuals. | Could identify only two (*TFAP2B* and *GNPDA2*) of the tested 23 obesity loci. | rs987237, rs10938397 |

| | | | |
|---|---|---|---|
| (Almawi et al. 2013) | Tested 19 SNPs in/near 15 established **T2DM** loci in **Lebanese Levant** population using 995 T2DM cases and 1076 controls. | 4 (*COL8A1, KCNQ1, ALX4, HNF1B*) of the 15 established loci were seen associated with T2DM in Levants. The authors have shown in their previous works replicability of IGF2BP2, CDKAL1, and TCF7L2. | rs792837, rs2237892, rs2237895, rs729287, rs4430796 |
| (Al-Sinani et al. 2015) | Tested 10 variants from 9 established **T2DM** loci - *KCNJ11, TCF7L2, CDKAL1, CDKN2A/B, FTO, IGF2BP2, SLC30A8, CAPN10, HHEX* in **Omani** Arabs using992 T2DM cases and 294 controls. | Only four of the 9 tested loci could be replicated - *KCNJ11, TCF7L2, CDKAL1, CDKN2A/B*. However, could not replicate IGF2BP2 (rs4402960), SLC30A8 (rs13266634), CAPN10 (rs3792267), and HHEX (rs1111875). | rs5219, rs7903146, rs10946398, rs10811661 |
| (Mtiraoui et al. 2012) | Tested 7 established **T2DM** loci - *ENNP1, IGF2BP2, KCNJ11, MLXIPL, PPARγ, SLC30A8* and *TCF7L2* in Levant Arabs from **Lebanese (**751 T2DM cases and 918 controls**) and North African Arabs** from Tunisia (1470 T2DM cases and 838 controls). | *TCF7L2* was replicating in both; *IGF2BP2* and *PPARγ* were replicating in Lebanese and not in Tunisia; *KCNJ11*, and *SLC30A8* were replicating in Tunisia but not in Lebanese. Neither *ENNP1* nor *MLXIPL* was seen replicating in Lebanese or Tunisians. | TCF7L2 (rs7903146) in both Lebanese and Tunisians; IGF2BP2 (rs1470579) and PPARγ (rs1801282) in Lebanese alone; KCNJ11 (rs5219) and SLC30A8 (rs13266634) in Tunisians alone. |
| (El Hajj Chehadeh et al. 2016) | Investigated the association between the *MTHFR* SNPs (C677T and A1298C) and **T2DM** in **Emirati Arabs.** | *MTHFR* gene polymorphisms are not related to T2DM in the Emirati population. | None of the SNPs replicated |
| (Al-Safar et al. 2015) | Tested the **T2DM** risk loci - *TCF7L2 (*rs10885409*)* and *PPARγ2 (*rs1801282*)* in **Arab Emirati** population using 169 T2DM cases and 209 controls. | Confirms the association of the *TCF7L2* variant but not *PPARγ2* as a T2DM loci in UAE Arabs. | rs10885409 |
| (Khan et al. 2018) | Tested an established **obesity** locus - *FTO* (rs9939609) and *VDR* (rs1544410), in **UAE** population using 201 obese, 115 overweight, and 98 normal subjects. | Replicates the *FTO* marker as adult obesity risk loci in UAE. *VDR* could not be replicated. | rs9939609 |
| (Osman et al. 2018b) | Previously reported SNPs in association with **CKD** and DKD were examined in **Emirati Arabs** for the renal functional indices such as blood urea, serum | Variants from genes such as BCAS3, RSPO3, STC1-ADAM28, KRT8P26-AP5B1, KBTBD2, SHROOM3, WDR72, MED1, WDR72, DLGAP3, SDCCAG8, CASR, GC, CYP2R1, LPIN2-MYOM1 replicated for examined traits. | Variants replicated for blood urea: rs11868441, rs1892172, rs4644087, rs4382293, rs2489629; Serum creatinine: rs6999484, rs1705699, rs2828785, rs11227279, rs7785065, rs4859682, |

| | | | |
|---|---|---|---|
| | creatinine, estimated glomerular filtration rate, Vitamin D (25-OH cholecalciferol), and albumin:creatinine ratio145 T2DM with DKD patients and 265 normal subjects. | | rs10317551, rs7740534; estimated glomerular filtration rate: rs2168785, rs12452509, rs4776168, rs10518733, rs7541937, rs10032549, rs2484639; Vitamin D: rs1801725, rs1155563, rs12794714, rs10500804; Albumin: Creatinine ratio: rs4528660 |
| (Cyrus et al. 2018) | Assessed established **CKD** locus of Eight genetic variants in four genes (SHROOM3, MYH9, SLC7A9, and CST3) in **Saudi Arabian** population using 160 CKD patients and 189 ethnicity-matched healthy controls. | Only rs4821480 (MYH9) was replicated with increased risk of development of CKD. Other variants were not replicated. | rs4821480 |
| (Alsmadi et al. 2014b) | Examined the established **hypertension** loci from LEP, MC3R and MC4R gene in Kuwaiti Arab population using 332 individuals. | MC3R variant rs3827103 found significant association with SBP. | rs3827103 |
| (Hebbar et al. 2019a) | Examined an established **obesity** variant from FTO (rs1421085) gene in 278 Kuwait Arab individuals with body composition traits. | FTO variant found replicated for body composition traits such as weight, total body water and soft lean mass. | rs1421085 |
| (Al-Daghri et al. 2014a) | Association 36 variants previously reported for T2DM in Europeans were examined for **obesity traits** in **Saudi** subjects using 975 obese, 825 overweight, and 423 lean controls and of these 927 had a history of T2DM. | Eleven SNPs of 36 shown significant association either with Additive or recessive or dominant genetic models; only 11 SNPs replicated. | rs10440833, rs7578326, rs7178572, rs2028299, rs11642841, rs7903146, rs13081389, rs1552224, rs7957197, rs8042680, rs163184 |
| (Alharbi et al. 2014) | Examination of 11 reported variants for **obesity** in367 **Saudi** individuals. | Eleven variants including rs10767664(BDNF), rs3751812, rs9939609, rs9941349(FTO), rs10938397(GNPDA2), rs571312(MC4R), rs2815752(NEGR1), rs713586(RBJ), rs543874(SEC16B), rs7359397(SH2B1), rs2867125(TMEM18) were examined; only two replicated. | rs10767664 (BDNF), rs3751812 (FTO) |

| | | | |
|---|---|---|---|
| (Bakhashab et al. 2020) | To examine PPARγ rs1801282, FTO rs9939609, and MC4R rs2229616 variants on the risk of **T2DM** among the western **Saudi** population using 415 diabetic cases and 323 controls. | Only MC4R rs2229616 variant found associated with T2DM. | rs2229616 |
| (Al-Sinani et al. 2014b) | A family-based candidate SNP study replicated the results of GWAS of **T2DM**in a cohort of 232 members of a large consanguineous **Omani Arab** pedigree consisting 27 diabetics and 50 pre-diabetics. | TCF7L2 (rs7903146, rs7901695), CDKN2A/B (rs10811661), PPARG (rs1801282), FTO (rs8050136, rs9939609), IGF2BP2 (rs4402960), KCNJ11 (rs5219), HHEX (rs1111875), SLC30A8 (rs13266634), CDKAL1 (rs10946398), CAPN10 (rs41266971, rs2975760, rs3792267). | KCNJ11 (rs5219), IGF2BP2 (rs4402960), SLC30A8 (rs13266634), CAPN10 (rs2975760), FTO (rs8050136), FTO (rs9939609) |
| (Osman et al. 2019) | Three loci previously reported in the European and six reported in the Japanese population for **DKD** were examined in 410 T2DM **Emirati** individuals of which 145 were DKD. | Examined 9 variants from GABRR1, UMOD, PRKAG2, FTO, PRCD, RAD51B, TRABD2B, CHAT, CCNH-TMEM161B, LRP8 with T2DM and observed 2 variants significantly associated for DKD. | rs1421086, rs17817449 |

**Table 5: Candidate SNP studies on metabolic risk loci from the MENA region and comparison with established metabolic risk loci.**

| SNP | Gene | Number of T2DM studies replicated variant | Population |
|---|---|---|---|
| rs7903146 | TCF7L2 | 7 | Saudi Arabian (Al-Daghri et al. 2014b; Li-Gao et al. 2018), Emirati (Osman et al. 2018a), Tunisian, Moroccan (Cauchi et al. 2012), Qatari (O'Beirne et al. 2016), Omani (Al-Sinani et al. 2015), Lebanese (Mtiraoui et al. 2012), Tunisian (Mtiraoui et al. 2012) |
| rs5219 | KCNJ11 | 3 | Tunisian (Mtiraoui et al. 2012), Omani (Al-Sinani et al. 2014b; Al-Sinani et al. 2015) |
| rs13266634 | SLC30A8 | 3 | Saudi Arabian (Li-Gao et al. 2018), Tunisian (Mtiraoui et al. 2012), Omani (Al-Sinani et al. 2014b) |
| rs1470579 | IGF2BP2 | 2 | Tunisian, Moroccan (Cauchi et al. 2012), Lebanese (Mtiraoui et al. 2012) |
| rs10811661 | CDKN2B-AS1 | 2 | Tunisian, Moroccan (Cauchi et al. 2012), Omani (Al-Sinani et al. 2015) |
| rs231362 | KCNQ1 | 2 | Tunisian (Cauchi et al. 2012), Saudi Arabian (Al-Daghri et al. 2014b) |
| rs4506565 | TCF7L2 | 2 | Saudi Arabian (Li-Gao et al. 2018), Qatari (O'Beirne et al. 2016) |
| rs7901695 | TCF7L2 | 1 | Saudi Arabian (Li-Gao et al. 2018) |
| rs11605924 | CRY2 | 1 | Saudi Arabian (Li-Gao et al. 2018) |
| rs340874 | PROX1 | 1 | Saudi Arabian (Li-Gao et al. 2018) |
| rs2943641 | NYAP2-IRS1 | 1 | Saudi Arabian (Li-Gao et al. 2018) |
| rs11642841 | FTO | 1 | Saudi Arabian (Li-Gao et al. 2018) |
| rs560887 | G6PC2 | 1 | Saudi Arabian (Li-Gao et al. 2018) |
| rs6017317 | FITM2-R3HDML | 1 | Saudi Arabian (Li-Gao et al. 2018) |
| rs9470794 | ZFAND3 | 1 | Saudi Arabian (Li-Gao et al. 2018) |
| rs6567160 | RPS3AP49-RNU4-17P | 1 | Emirati (Osman et al. 2018a) |
| rs11126666 | KCNK3 | 1 | Emirati (Osman et al. 2018a) |
| rs6804842 | RARB | 1 | Emirati (Osman et al. 2018a) |
| rs2077235 | NLRC3 | 1 | Emirati (Osman et al. 2018a) |
| rs1016287 | LINC01122-LOC105374754 | 1 | Emirati (Osman et al. 2018a) |
| rs29941 | KCTD15 | 1 | Emirati (Osman et al. 2018a) |
| rs243021 | BCL11A | 1 | Tunisian, Moroccan (Cauchi et al. 2012) |
| rs4607103 | ADAMTS9-AS2 | 1 | Tunisian, Moroccan (Cauchi et al. 2012) |
| rs10010131 | WFS1 | 1 | Tunisian, Moroccan (Cauchi et al. 2012) |
| rs7756992 | CDKAL1 | 1 | Tunisian, Moroccan (Cauchi et al. 2012) |
| rs1081161 | CDKN2B-AS1 | 1 | Tunisian, Moroccan (Cauchi et al. 2012) |
| rs7957197 | HNF1A | 1 | Tunisian, Moroccan (Cauchi et al. 2012) |
| rs1169288 | HNF1A | 1 | Moroccan (Cauchi et al. 2012) |

| | | | |
|---|---|---|---|
| rs1421085 | FTO | 1 | Moroccan (Cauchi et al. 2012) |
| rs17782313 | MC4R | 1 | Moroccan (Cauchi et al. 2012) |
| rs1799884 | GCK | 1 | Moroccan (Cauchi et al. 2012) |
| rs2283228 | KCNQ1 | 1 | Tunisian (Cauchi et al. 2012) |
| rs1801214 | WFS1 | 1 | Saudi Arabian (Al-Daghri et al. 2014b) |
| rs849134 | JAZF1 | 1 | Saudi Arabian (Al-Daghri et al. 2014b) |
| rs10965250 | CDKN2B-AS1 | 1 | Saudi Arabian (Al-Daghri et al. 2014b) |
| rs163184 | KCNQ1 | 1 | Saudi Arabian (Al-Daghri et al. 2014b) |
| rs4812829 | HNF4A | 1 | Saudi Arabian (Al-Daghri et al. 2014b) |
| rs5945326 | DUSP9 | 1 | Saudi Arabian (Al-Daghri et al. 2014b) |
| rs792837 | COL8A1 | 1 | Lebanese (Almawi et al. 2013) |
| rs2237892 | KCNQ1 | 1 | Lebanese (Almawi et al. 2013) |
| rs2237895 | KCNQ1 | 1 | Lebanese (Almawi et al. 2013) |
| rs729287 | ALX4 | 1 | Lebanese (Almawi et al. 2013) |
| rs4430796 | HNF1B | 1 | Lebanese (Almawi et al. 2013) |
| rs10946398 | CDKAL1 | 1 | Omani (Al-Sinani et al. 2015) |
| rs1801282 | PPARG | 1 | Lebanese (Mtiraoui et al. 2012) |
| rs10885409 | TCF7L2 | 1 | Emirati (Al-Safar et al. 2015) |
| rs4821480 | MYH9 | 1 | Saudi Arabian (Cyrus et al. 2018) |
| rs1421086 | FTO | 1 | Emirati (Osman et al. 2019) |
| rs17817449 | FTO | 1 | Emirati (Osman et al. 2019) |
| rs2229616 | MC4R | 1 | Saudi Arabian (Bakhashab et al. 2020) |
| rs4402960 | IGF2BP2 | 1 | Omani (Al-Sinani et al. 2014b) |
| rs2975760 | CAPN10 | 1 | Omani (Al-Sinani et al. 2014b) |
| rs8050136 | FTO | 1 | Omani (Al-Sinani et al. 2014b) |
| rs9939609 | FTO | 1 | Omani (Al-Sinani et al. 2014b) |

**Table 6: List of 54 SNVs that replicated for T2DM traits in Arab population of MENA region through candidate variants study efforts.**

| SNPs | Gene | Number of obesity studies replicated variant | Population |
|---|---|---|---|
| rs1421085 | FTO | 2 | Emirati (Osman et al. 2018a), Kuwaiti (Hebbar et al. 2019a) |
| rs7239883 | LINC00907 | 1 | Emirati (Osman et al. 2018a) |
| rs492400 | USP37 | 1 | Emirati (Osman et al. 2018a) |
| rs8030605 | RFX7 | 1 | Emirati (Osman et al. 2018a) |
| rs11607976 | LOC102724265-LINC02747 | 1 | Emirati (Osman et al. 2018a) |
| rs28710014 | PRELID1, RAB24 | 1 | Emirati (Osman et al. 2018a) |
| rs2284746 | MFAP2 | 1 | Emirati (Osman et al. 2018a) |
| rs2306335 | AP4E1 | 1 | Emirati (Osman et al. 2018a) |
| rs4938802 | ARHGEF12 | 1 | Emirati (Osman et al. 2018a) |
| rs749052 | NPPC-DIS3L2 | 1 | Emirati (Osman et al. 2018a) |
| rs2326458 | LINC02176-ZDHHC7 | 1 | Emirati (Osman et al. 2018a) |
| rs4689977 | CYTL1 | 1 | Emirati (Osman et al. 2018a) |
| rs1039027 | FNDC3B | 1 | Emirati (Osman et al. 2018a) |
| rs9939609 | FTO | 1 | Emirati (Khan et al. 2018) |
| rs10440833 | CDKAL1 | 1 | Saudi Arabian (Al-Daghri et al. 2014a) |
| rs7578326 | LOC646736 | 1 | Saudi Arabian (Al-Daghri et al. 2014a) |
| rs7178572 | HMG20A | 1 | Saudi Arabian (Al-Daghri et al. 2014a) |
| rs2028299 | AP3S2 | 1 | Saudi Arabian (Al-Daghri et al. 2014a) |
| rs11642841 | FTO | 1 | Saudi Arabian (Al-Daghri et al. 2014a) |
| rs7903146 | TCF7L2 | 1 | Saudi Arabian (Al-Daghri et al. 2014a) |
| rs13081389 | SYN2-PPARG | 1 | Saudi Arabian (Al-Daghri et al. 2014a) |
| rs1552224 | ARAP1 | 1 | Saudi Arabian (Al-Daghri et al. 2014a) |
| rs7957197 | OASL | 1 | Saudi Arabian (Al-Daghri et al. 2014a) |
| rs8042680 | PRC1 | 1 | Saudi Arabian (Al-Daghri et al. 2014a) |
| rs163184 | KCNQ1 | 1 | Saudi Arabian (Al-Daghri et al. 2014a) |
| rs10767664 | BDNF | 1 | Saudi Arabian (Alharbi et al. 2014) |
| rs3751812 | FTO | 1 | Saudi Arabian (Alharbi et al. 2014) |

**Table 7: List of 27 SNVs that replicated for obesity traits through candidate gene/variants study efforts in Arab population.**

## 6.18 Family based genetic studies for metabolic diseases

Due to large family structures of the Arabs, well documented genealogy, and high rates of consanguinity, Arabian Peninsula provides a unique platform for understanding population history and disease gene identification process. Although Arab families are consanguineous and

well suited for family-based studies, not many familial studies for diabetes or obesity risk loci have been reported. Hitherto, only four family-based genome-wide studies have described Arab families in the context of metabolic disorders. These include,

i) A study from the UAE (Al Safar et al. 2013) on an extended pedigree (consisting of 178 Arab individuals) implicated the association of KCTD8, COX7B2, and GABRA4 genes with T2DM for the first time; the implicated markers have not yet been replicated in any other population.

ii) Linkage analysis and positional gene cloning identified a truncating mutation (at telomeric end of 3q29) in a highly evolutionarily conserved gene CEP19 from a consanguineous multi-generational Israeli Arab family (consisting of 13 affected and 31 unaffected family members) affected by autosomal-recessive morbid obesity. Additionally, evaluating this finding in targeted knockout (KO) of CEP19 revealed markedly increased obesity in mice exhibiting hyperphagia, decreased energy expenditure, impaired whole-body fat oxidation, altered hepatic insulin signaling, and impaired glucose and insulin tolerance (Shalata et al. 2013).

iii) Another linkage study identified a locus contributing to type 1 diabetes in an extended Bedouin Arab family (108 individuals, including 16 affected family members) mapped to the long arm of chromosome 10 (10q25). All affected relatives of the family carried one or two high-risk HLA-DR3 haplotypes that were rarely found in other family members (Verge et al. 1998).

iv) Using a family-based association analysis comprising 380 individuals of a family from Oman, variant rs266729 (SNP-11377CNG) of ADIPOQ promoter region was found to be associated with obesity and its traits (Zadjali et al. 2013).

List of family-based studies, with study description and variants/genes discovered are shown in **Table 8**. Interestingly, the four abovementioned studies confirmed known, well-established gene loci and did not implicate any novel genetic signatures for diabetes. Hence, there is an imminent need to increase the representation of family-based studies for complex metabolic disorders using Arab families. Additional family-based studies should be planned with appropriate study designs as family data continue to provide important information in the search for trait loci.

| Family based studies | Examination of established loci in study population | Study replicated already known marker or identified novel marker | Replicated SNPs at P<0.05 |
|---|---|---|---|
| (Al Safar et al. 2013) | Performed family based GWAS for T2DM in UAE Arabs using 178 (66 cases, 112 controls) individuals in discovery; 315 (116 cases and 199 controls) individuals in replication cohort. | **Novel loci:** Identified novel associations (*KCTD8*, *PRKD1*, *GABRA2*, *GABRA4* and *GABRB1*). | rs4407541, rs10517178/rs1372491, rs10144903, rs7662743 |
| (Zadjali et al. 2013) | Family-based study to investigate the association of SNPs (rs17300539 and rs266729) from adiponectin gene *ADIPOQ* with obesity traits using 328 Arab individuals from one large extended family of Oman. | Showed family-based evidence for association of one (rs266729) of the two tested SNPs from *ADIPOQ* defining obesity in Arab population. | rs266729 |
| (Verge et al. 1998) | 19 T1DM affected individuals were observed to investigate a locus contributing to type 1 diabetes using an extended Bedouin Arab family from Israel comprising 108 individuals including 16 living T1DM affected family members. | A locus contributing to T1DM was mapped to the long arm of chromosome 10 (10q25). | All affected relatives carried one or two high-risk HLA-DR3 haplotypes that were rarely found in other members of the family. |
| (Shalata et al. 2013) | A multi-generation Arab family from Israel was considered to investigate a locus contributing to morbid obesity. Family included 13 affected and 31 unaffected family members. | Homozygosity mapping localized the disease locus to a region in 3q29. A homozygous nonsense mutation in CEP19 in all affected family members from sequence analysis. | Loss of the ciliary protein CEP19 in humans and mice was observed. |

**Table 8: Family based GWA studies on metabolic disorder and identification of risk loci.**

## 6.19 Population based GWA study

Before GWAS, candidate gene studies assessed several genes, but only few were confirmed, including MC4R (Stutzmann et al. 2007) and BDNF (Gunstad et al. 2006) as risk loci for obesity. Global GWAS studies have successfully discovered more than 300 variants for BMI and other obesity-related traits (Goodarzi 2018). However, only 27 risk loci such as FTO, LINC00907, USP37, RFX7, LOC102724265-LINC02747, PRELID1, RAB24, MFAP2, AP4E1, ARHGEF12, NPPC-DIS3L2, LINC02176-ZDHHC7, CYTL1, FNDC3B, CDKAL1, LOC646736, HMG20A, AP3S2, TCF7L2, SYN2-PPARG, ARAP1, OASL, PRC1, KCNQ1, and BDNF have been replicated for obesity traits in Arabs (Variants are shown in **Table 7**). The recent meta-analysis findings by *Mahajan et al.,* (Mahajan et al. 2018a) of European ancestries have identified over 240 genetic risk loci (with ~400 independent association signals) for T2DM, but thus far only 54 risk loci (ADAMTS9, ALX4, BCL11A, CDKAL1, CDKN2A/B, COL8A1, DUSP9, FTO, GCK, GNPDA2, HMG20A, HNF1A, HNF1B, HNF4A, IGF2BP2, JAZF1, KCNJ11, KCNQ1, MC4R, PPARγ, SLC30A8, TCF7L2, TFAP2B, TP53INP1, and WFS1) have been replicated in the Arab populations for T2DM (Variants are listed in **Table 6**) (Hebbar et al. 2019b).

Genetic studies have revealed that a great proportion of phenotypic variance for complex traits are from rare variants (Cohen et al. 2004; Cohen et al. 2006; Ji et al. 2008; Romeo et al. 2009) and effective study of such rare variants requires a large sample size (Zuk et al. 2014). Although hitherto population-based studies from the Peninsula are of small sample size, few studies have attempted to explore the association of rare variants with complex traits using gene-based analysis. A study comprising 864 Qatari subjects (574 T2DM cases, 290 controls) using sequence kernel association test (SKAT) approach identified low frequency deleterious risk variants for T2DM from CTNNB1, KIF12, DVL1, EPB4IL3, DTNB, and DLL1 genes (O'Beirne et al. 2018). Another study of 996 Qatari subjects (included T2DM and Obese individuals) identified 21 common and 12 rare variants associations (shown in Table 8) using gene-based and single variant analysis from understudied whole metabolite profiles (Yousri et al. 2018), thus connected genetic variants to metabolic diseases.

Importantly, excluding publications from our genome-wide investigation initiatives, only 5 GWAS metabolic related studies using Arab cohorts were found in the public domain (**Table 9**). Of these, 4 studies were based on the case-control (binary trait) approach and 1 study was based on the quantitative trait (with metabolites, not using basic quantitative clinical parameters of obesity or diabetes). A detailed investigation of genome-wide variants in relation to quantitative traits of anthropometric, blood pressure, and glycemic and lipid levels have not been conducted in Arab population.

| Population based GWA studies | Genome-wide examination of variants in Study population | List of identified genes | List of identified variants |
|---|---|---|---|
| (Ghassibe-Sabbagh et al. 2014) | Performed GWAS (on genotyped and imputed markers) in 3286 (1902 controls and 1384 T2DM cases) Lebanese participants. | **Partial overlap and Novel loci:** Only two established loci (*CDKAL1* and *TCF7L2*) surfaced. | rs7766070, rs34872471 |
| (Saade et al. 2011) | Performed GWAS on 2002 CAD patients (425 with no stenosis and 1524 with stenosis) in Lebanese population. | rs4977574 was significant with risk for MI and rs6922269 was found significantly protective for CAD after Bonferroni correction. | CDKN2A-CDKN2B_rs4977574, CXCL12_rs1746048, CDKAL1_Rs9295489, PTPRD_rs10115782 |
| (Dajani et al. 2017) | Performed T2DM study using 144 (34 cases and 109 controls) Chechen and 140 (33 cases and 105 controls) Circassian population of Jordan. | JAG1, MLXIP. | rs6134031 (at genome-wide p-value), rs4758690 (at suggestive p-value) |
| (O'Beirne et al. 2018) | Exome sequenced 864 Qatari subjects (574 T2DM cases, 290 controls). | SKAT analysis identified 6 risk genes such as KIF12, DVL1, EPB41L3, DTNB, | KIF12 (c.134T>C), DVL1 (c.1934A>C), EPB41L3 (c.2531T>C, rs8082898, rs117900256), DTNB (c.1672A>C,rs562264712), DLL1 (rs200861263), CTNNB1 (rs77750814) |

| | | | |
|---|---|---|---|
| | | DLL1, CTNNB1 for T2DM. | |
| (Yousri et al. 2018) | Discovery cohort: whole-exome sequencing data of 614 Qataris; Replication cohort: imputed array data of 382 Qataris. Analysed 1303 metabolites to identify metabolomics quantitative trait loci (mQTLs). | 21 common variants and 12 novel functional rare variants were identified from gene-based association and single variant analysis. | **Common variants**: rs4646257 [NAT2] with 5-acetylamino-6-amino-3-methyluracil/1-methylxanthine; rs4921913, rs1799958 [ACADS] with 5-acetylamino-6-amino-3-methyluracil Ethylmalonate; rs13538 [NAT8] with 2-aminooctanoate/X-12511 and N-acetylcitrulline; rs34109652 [TMPRSS11E] with X-11491 (deoxycholic acid glucuronide or isomer); rs4149056 [SLCO1B1] with glycochenodeoxycholate glucuronide (1); rs2147896 [PYROXD2] with N-methylpipecolate; rs3756669 [UGT3A1] with X-24348/pregnsteroid monosulfate and X-24348; rs28456 [FADS2] with 1-(1-enyl-palmitoyl)-2-arachi-donoyl-GPC(P-16:0=20:4)/X-24438(PC(P-16:0/20:3)); rs174560 [FADS2] with X-24438(PC(P-16:0/20:3)); rs37370 [AGXT2] with 3-aminoisobutyrate; rs181856093 [PHYHD1] with X-22145 (2'-O-methyluridine); rs6690449 [THEM4] with X-18921/X-23680 rs2999534 X-23293; rs78461713 [UGT1A1] with bilirubin (E, E); rs62129970 [SULT2A1] with X-11440(tentatively steroid)/4-androsten-3alpha; 17alpha-diolmonosulfate(2); rs78176967 [SLC22A24] with X-22379(androsterone glucuronide)/21-hydroxypregnenolone disulfate; rs61285056 [SLC22A24] with X-22379(androsterone glucuronide); rs2069258 [SPTLC1P4] with X-23293/cis-4-decenoyl carnitine; rs117135869 [TTC38] with X-22162/X-24513; rs117135869 [TTC38] with X-22162; rs274554 [SLC22A5] with Tryptophan betaine; rs7530513 [CCBL2] with Imidazole lactate; rs1165196 [SLC17A1] with X-12824(hexanoylglutamine)/X-16087; rs776746 [CYP3A5] with X-12063; c15p90683852 [SEMA4B] with Undecanedioate<br>**Rare variants**: rs3796543, rs34228795, rs34543011 [AASDH] with Thyroxine; rs199581976, rs75289684, rs115687886 [METTL7B], rs372117452 [OR6C6], rs144983062 [ITGA7] with Androsterone sulfate; c12p11506114 [PRB1] wit Mannose; rs150988100 [ACAN] with X-12844 (glucuronidated steroid); rs56332208 [OTOF] with Retinol (Vitamin A); rs150988100 [ACAN] with X-09789, c18p48256030 [MAPK4] with X-21365 (N-trimethyl 5-aminovalerate); rs377301648 [ZNF133] with 3-methyl-2-oxovalerate; rs61733595 [MRGPRX3], c18p48256030 [MAPK4] with Tryptophan. |

**Table 9: Population based GWA studies on metabolic disorders and identification of risk loci.**

# 7. Aims of the study

The objective of this doctoral thesis was to delineate genetic determinants of metabolic disorders by examining genome-wide association of SNVs with quantitative metabolic traits in Arab population. Specifically, this objective was explored by following three main approaches in present study.

1. To discover SNPs and their linkage disequilibrium variants engaged in causal pathway of metabolic disorders using genome-wide SNP genotyped data.
    a. Study I: The study aims to discover SNVs from 13 metabolic traits using genotype data from Illumina Cardio-MetaboChip, a chip comprising SNVs chosen based on GWAS meta-analyses of 23 metabolic related traits.
    b. Study II: The study aims to find SNVs that show association with obesity-related anthropometric traits using Illumina HumanOmniExpress chip and an independent replication cohort.
2. To identify associations of SNVs underlying ROH regions with metabolic traits.
    a. Study III: The study aims to examine GWA SNV association with lipid traits and identify ROH signatures overlapping on associated loci.
    b. Study IV: The study aims to examine GWA SNV association with glycemic traits and identify ROH signatures overlapping on associated loci.
3. To impute high density SNP data with 1KG reference data and identify SNP associations with metabolic traits.
    a. Study V: The study aims to constellate two approximate Arab whole genomes from two different arrays of data by separate imputation and then by performing statistical association analysis of 13 metabolic traits with each cohort, followed by a meta-analysis.

# 8. Materials and Methods

## 8.1 Study subjects, collection of blood, and vital signs

The study participants were recruited through two major initiatives of Dasman Diabetes Institute (DDI), Kuwait Obesity Genetics Project (KOGP) and Kuwait Diabetes Epidemiology Program (KDEP).

KOGP: The participants were largely Arab patients seeking tertiary medical care for diabetes, pre-diabetes, or related disorders at DDI clinics; visitors to open day events (which offer various diagnostic services) at DDI; or visitors to our campaigns at primary health centers and blood banks in each of the six governorates of Kuwait.

KDEP: The participants included a random representative sample of adults of Arab ethnicity across the six governorates of Kuwait. A stratified random sampling technique was used to select participants from the computerized register of the Public Authority of Civil Information, which maintains records of personal information on citizens of Kuwait as well as expatriates.

During participant recruitment in both programs, nationality and ethnicity were confirmed through a questionnaire. Data on age, sex, illness (e.g., diabetes and cardiovascular complications), and medication were documented and vital signs such as height, weight, waist circumference (WC), and blood pressure readings were recorded. Participants were excluded if any of the following criteria were met: (a) age less than 18 years (b) afflicted with serious chronic diseases, disorders resultant of diabetes/hypertension, mental illness, or cognitive limitation (c) pregnant or (d) undergoing weight reduction surgery or enrolled in fitness programs.

Informed consent was obtained from each participant. Upon confirming that the participants were fasting overnight, blood samples were collected. Lipid profile and glucose measurements were recorded. Blood sample collection and vital sign assessment were conducted in accordance with guidelines of DDI Institutional Ethical Review Committee. Data from all participants were collected in accordance with ethical guidelines and the Helsinki Declaration. For a detailed description of clinical characteristics of subjects of each cohort, relevant sections of study I and study II research articles can be referred.

## 8.2 Genotyping of discovery and replication subjects

Among the study subjects, KOGP largely contributed to the discovery cohort, whereas KDEP largely contributed to the replication cohort. Genotyping was performed in three different ways, which included

i. Genome-wide genotyping using two types of Illumina based microarray chips (namely, HumanCardio-MetaboChip and HumanOmniExpress).

ii. Genotyping of replication of candidate SNPs using RT-PCR based TaqMan SNP genotyping assay.
iii. Genome-wide imputation using both genome-wide genotyped datasets to fill-in genome-wide missing variants.

**Discovery phase:**

Microarray Platform I: Samples were genotyped using the Illumina HumanCardio-MetaboChip array (comprised ~200K variants chosen from GWAS meta-analysis of 23 traits) (Voight et al. 2012) utilizing the Infinium HD Assay Ultra genotyping assay methods. Overall, 2440 samples were distributed into 27 batches (each homogenized for sex and diabetes status) and then genotyping was conducted.

Microarray platform II: Samples were genotyped using the Illumina HumanOmniExpress array utilizing the Infinium HD Assay Ultra genotyping assay methods. Illumina HumanOmniExpress-12v1-Multi_H (730,525 variants) was used to genotype 1,097 individuals into 20 batches, HumanOmniExpress-12v1-1_B (719,665 variants) to genotype 336 individuals into 5 batches, and HumanOmniExpress-24v1-0_a (716503 variants) to genotype 480 individuals into 6 batches.

Both arrays used the same assay protocol. DNA was extracted using the GentraPuregene® kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocols and quantified using both Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies, NY, USA) and Epoch Microplate Spectrophotometer. The genotyping assay included whole genome amplification, fragmentation, hybridization, staining, and imaging of microarray chips using the Illumina iSCAN system.

**Replication phase:**

The candidate SNP genotyping for the replication phase was performed using TaqMan® SNP Genotyping Assay (Applied Biosystems, Foster City, CA, USA) and ABI 7500HT Real-Time PCR System (Applied Biosystems). The genotypes ascribed by real-time PCR were validated by direct Sanger sequencing of the PCR products for selected cases of homozygotes and heterozygotes. The sequencing reaction was performed using the BigDye™ terminator cycle sequencing FS Ready Reaction Kit (Applied Biosystems) following the manufacturer's recommendations on an ABI- PRISM 3130 genetic analyzer (Applied Biosystems).

**Genome-wide genotyping by imputation:**

The Illumina HumanOmniExpressBeadChip (OmniExpress) dataset and Illumina HumanCardio-MetaboBeadChip (Cardio-MetaboChip) genotyped datasets were separately used for imputation of untyped genotypes. Genotype imputation was performed using Michigan Imputation Server

(MIS) with the 1000G ALL reference panel from 1000G Project Phase 3 V5 for imputation and Eagle v2.3 and Minimac3 were used for phasing and imputation, respectively. The imputation quality was quantified using the Rsq score. Imputation was performed separately for genotypes from OmniExpress and Cardio-MetaboChip datasets.

## 8.3 Quality assessments and statistical analysis of studies (I – V)

The gist of quality assessments and statistical analysis methods used in all five studies emulate a common theme but slightly differ due to study specific aims and use of tools. A brief note on methods used in each study is given below. For a detailed method, relevant sections of specific study article may be referred. Furthermore, an assessment of allele frequency of variants reported for three subgroups of Kuwait (such as KWP, KWS and KWB) was performed only in Study IV. Hence, as additional data, allele frequency of variants reported at genome-wide and nominal p-value for subgroups was assessed in the rest of the studies. Each of the subgroup's effect allele frequency (EAF) was compared with the study EAF using proportional test and presented in the results section.

**Study I:**

General QC assessments such as duplicate samples, sample call rates <97%, gender mismatching, cryptic-related, and ancestry-mismatching samples (assessed using ADMIXTURE tool) were removed. Further QC at variants, such as SNPs with a call rate of <98%, of low intensity (AB R Mean ≤ 0.25), poor cluster separation (Cluster Sep < 0.3) with heterozygote clusters very close to homozygotes (AB T Mean ≤ 0.2 or ≥0.8), excess of heterozygotes (Het Excess ≥ 0.2), and those with fewer than expected heterozygotes (Het Excess ≤−0.3) were evaluated using Genome Studio software. Further, missingness per individual (--mind 0.1), allele frequency (--MAF 0.01), missingness per marker (--geno 0.1), and Hardy-Weinberg equilibrium (HWE < $10^{-3}$) were removed using PLINK version 1.07 (Purcell et al. 2007).

As Cardio-MetaboChip was designed based on variants associated with 23 traits of GWAS meta-analysis, over 60% of the SNPs of the chip were clustered in 1.5% of the genome. Therefore, Cardio-MetaboChip poses a challenge to standard analysis methodology. It requires special care to control genomic inflation by stringent LD pruning ($r^2$> 0.3) (Voight et al. 2012), which resulted in 45793 tag SNPs. After these QC steps, 1965 samples were used for interrogation of statistical association of 45793 SNPs. However, a separate analysis was performed using all markers (128143) to describe the effect of SNPs, which were in LD with associated indexed SNP.

Statistical association analysis was performed on 13 metabolic traits (height, weight, BMI, WC, WcHtR, TC, HDL-C, LDL-C, TG, SBP, DBP, FPG, and HbA1C) using linear regression adjusted for age, sex, principal components, and medication status. Tests were performed in both additive and recessive mode of inheritance. The genome-wide significance threshold of p-value was set to ≤3.41E-08. Since, the study did not have a separate replication cohort, Bonferroni correction and

Benjamini–Hochberg FDR correction procedures were performed to avoid false positive associations. Further, trans-ethnic frequency correlation analysis, GTEx portal (eQTL), and pathway analysis [using Ingenuity Pathway Analysis software (IPA, QIAGEN, Redwood City)] were performed to improve biological interpretation of the findings.

**Study II:**

The QC assessments undertaken in this study were analogous with Study I. After sample level QC, 1353 samples genotyped in Illumina HumanOmniExpress chip were used for statistical association analysis. An independent replication cohort of 1176 unrelated individuals genotyped using the TaqMan assay was used to evaluate the replication of top hits appeared in the discovery phase.

Further, association analysis was performed on WC, weight, BMI, and WcHtR using linear regression adjusted for age, sex, principal components, and medication status considering additive mode of inheritance. The genome-wide significance threshold of p-value was set to ≤5E-08. Haplotype association analysis was performed using "proxy-glm" setting from PLINK. Gene-based tests for association were performed using Versatile Gene-Based Association Study (VEGAS) implementation of cpvSNP tool (Liu et al. 2010), available as R package. This was followed by pathway analysis using Ingenuity Pathway Analysis kit. Interactions between SNPs and biomarkers (mainly Apo-A1 and HDL-C) were evaluated using linear regression adjusted for age and sex.

**Studies III and IV:**

Studies III and IV followed similar statistical analysis methods and QC as Study II (additionally strand designations were corrected to forward and REF/ALT designations were corrected using array-specific version of design files from Illumina HumanOmniExpress BeadChip). Variants association with lipid (study III) and glycemic (study IV) traits were assessed using linear regression method under additive and recessive genetic models. The genome-wide significance threshold of p-value for the study III and IV were set to ≤6.12E-09 and ≤1.8E-08, respectively. An independent replication cohort genotyped using TaqMan assay was used for examination of associated loci in discovery phase. Identification of ROH was carried out with two different parameter settings using PLINK 1.9. Then, overlapping ROH regions with associated loci were identified. Novel ROH signatures in Arabs were determined using previously identified ROH data from global population by *Pemberton et al.,* (Pemberton et al. 2012). As an additional value to study IV, 283 samples used in the replication cohort were randomly selected for C-peptide measurement. Then, using C-peptide values, other insulin resistance traits were derived. Furthermore, interactions between (TG, FPG, HbA1c) and insulin resistance traits with respect to the genotypes at risk variants were assessed using linear regression-based interaction models.

**Study V:**

This study was conducted using genome-wide imputation of 1298 samples genotyped in Illumina HumanOmniExpress (considered as discovery cohort) and 1434 samples genotyped using HumanCardio-MetaboChip arrays (considered as replication cohort). Inter- and intra-cohort cryptic-related samples were removed. This study was unique when compared to previous studies since it involved the imputation of genome-wide missing variants from two datasets and used transformed traits in variant association tests. Since both cohorts had a high proportion of obese, hypertensive, and diabetic individuals on medication; glycemic, lipid, and blood pressure traits were carefully pre-adjusted (details of pre-adjustments are described in method section of the article). All the traits were adjusted for age, age$^2$, sex, and first four principal components (obtained from each of the two study cohorts). The resultant raw residuals of each trait were then subjected to inverse normal transformation; except for TG, FPG, and HbA1c, for which log inverse transformation was performed as it yielded a better Gaussian distribution. Overall, 13 metabolic traits namely BMI, height, weight, WC, TC, HDL-C, LDL-C, non-HDL, TG, FPG, HbA1c, SBP, and DBP were used to test variant association.

The genotyping QC procedures included the following: (i) samples with >95% call rate were retained, (ii) samples were checked against relatedness and ancestry mismatch: one randomly selected representative of each set of related samples was retained and samples with ancestry mismatch were removed, (iii) samples with heterozygosity >median + 3 * inter-quartile range were excluded, and (iv) SNPs with >98% call rate, Hardy–Weinberg equilibrium (HWE) >$10^{-6}$, and MAF >1% were retained. Next, samples common to both the study cohorts were removed from one of the sample sets and finally, 1298 of OmniExpress and 1434 of Cardio-MetaboChip samples were retained. Imputation was performed separately for genotypes from OmniExpress and Cardio-MetaboChip. Genotypes were converted to the Genome Reference Consortium Human Build 37 (hg19) of National Center for Biotechnology Information to ensure consistent SNP phasing for each genotyping array. Strand designations were corrected to the forward strand and allele designations were corrected using PLINK2 & the design files for OmniExpress and Cardio-MetaboChip. Genotype imputation was performed on the MIS (Das et al. 2016) using the 1000G ALL reference panel from 1000G Project Phase 3 V5 for imputation, Eagle v2.3 for phasing, and Minimac3 as the algorithm for imputation (Loh et al. 2016). SNV's associations with inverse transformed 13 quantitative metabolic traits were separately tested using RVTESTS software (Zhan et al. 2016); for each imputed array platform with all genotyped and imputed SNPs that passed the QC threshold metrics (Rsq>0.05 and MAF ≥ 5%). Variants with p-values < 0.05 in both the platforms were combined using METAL software (Willer et al. 2010) and assessed with NHGRI-EBI GWAS Catalog for reported association signals and variants. Additionally, variants were prioritized by fine mapping using FINEMAP tool (Benner et al. 2016) and plausible causal variant configurations from a given genomic region were identified. Furthermore, functional mapping was performed using SuRFR (Ryan et al. 2014) and genotype-tissue expression data were examined using GTeX v8 (https://www.gtexportal.org). The genome-wide significance threshold of p-value for meta-analysis was set at ≤5E-08.

# 9. Results

## 9.1 Study I: Identification of genetic risk variants for 13 different metabolic traits by genotyping and analyzing Cardio-MetaboChip array data

The demographic and clinical characteristics of the 1965 study participants and the statistical association tests used have been described in detail in the study (Hebbar et al. 2017b) and may be referred for further details. The examination of 128,143 variants (of which 45,793 were LD independent) using linear regression adjusted for age, sex, and 10 principal components led to the identification of chr15:40531386/rs12440118 [ZNF106] with HbA1c at "close to genome-wide significant" p-value (≤3.41E-08). Moreover, five other risk variants emerged at nominal p-value threshold (3.41E-08 > p ≤5.45E-07), including rs7144734 [OTX2-AS1,RPL3P3] with FPG and rs17501809 [PLGRKT], rs11143005 [LOC105376072], rs900543 [THSD4,NR2E3], and Chr12:101494770 [IGF1] with TG. These variants continued to associate when models were adjusted for medication and disease status and additionally qualified through BH-FDR test with p-value < 0.05 (shown in **Table 10**). All these 6 markers were identified considering inheritance model as recessive. The regional association plot for the top variant rs12440118 [ZNF106 (W > R)] association along with LD variants is shown in **Figure 13A**.

In addition, 30 more variants (from 33 associations) were found to be associated at suggestive p-value, of which 16 variants (such as with weight- rs10005556 [GAPDHP56, LOC105377421], WcHtR- rs17117722 [TRA, TRD], WC- rs10005556 [GAPDHP56, LOC105377421]& rs9390649 [UST], HbA1c- rs3767494 [C1orf106], TG- rs12722856 [RPS3AP9, GAPDHP75], rs10014125 [LOC105377567], rs17073574 [LAMA4], rs11777524 [LY6D, GML], rs11602685 [MICAL2], chr11:45779819 [DKFZp779M0652, SLC35C1], rs17569297 [LOC105369738, LOC105369739], and rs7342999 [LOC105372082, LOC105372084]; and FPG- rs7729384 [LOC101927421], rs3799125 [RGS17], rs747486 [ANKRD11], and rs4764409 [PIK3C2G]) qualified through BH-FDR test. The remaining 14 variants did pass FDR p-value (such as with height- rs3959929 [LOC105376567, LOC105376570]; SBP- rs10497520 [TTN]; HbA1c- chr1:16038119 [LOC105371466], rs17716285 [KSR1], and rs17639988 [MDGA1,ZFAND3]; weight- rs4559034 [LOC102467224] and rs1184476 [TEX29]; HDL-C- rs1800775 [CETP]; LDL-C- chr2:43355981 [THADA]; TC- chr2:43355981 [THADA] and rs10935794 [RPL32P9, LINC01213]; TG- rs9326246 [BUD13], rs925530 [TMEM120B]; FPG- rs10247084 [LOC105375159, LOC105375160] and rs883431 [SLC28A3]). However, examination of NHGRI-GWAS Catalog showed that 2 of these "suggestive" associations (rs1800775 [CETP] with HDL-C and rs9326246 [BUD13] with TG) exhibited evidence at genome-wide significance in previous studies of Euro-centric populations.

Evaluating these 36 variants in the Genotype-Tissue Expression (GTEx) data suggested that 7 of these variants had the potential to differentially regulate either its own or proximal genes. For instance, rs12440118 [ZNF106; W> R] downregulates LRRC57 in cells-transformed fibroblast, rs11143005 upregulates LOC105376072 and PGM5 in whole blood, rs1800775 downregulates NLRC5 in cells-transformed fibroblast, rs9326246 downregulates RP11-109L13.1 in skin exposed

to sun, rs17716285 upregulates NOS2 in muscle skeletal, rs11777524 downregulates LOC101928087 in testis, and rs10497520 downregulates FKBP7 in cells-transformed fibroblasts.

Searching the canonical pathways to which the identified genes overlap indicated that 5 of the identified gene loci (KSR1, PIK3C2G, MICAL2, CETP, and UST) are involved in metabolic pathways such as ceramide signaling (KSR1 and PIK3C2G), pregnenolone biosynthesis (MICAL2), ERK/MAPK signaling (KSR1 and PIK3C2G), histidine degradation VI (MICAL2), and LPS/IL-1 mediated inhibition of RXR function (CETP and UST). Evidence from literature also supported the involvement of all reported genes in metabolic trait-related processes. For instance, the variation at rs12440118 (associated with HbA1c) was a non-synonymous amino acid change (W103R, aromatic to basic residue) in ZNF106. Interestingly, this protein happens to be a product of the SIRM gene and shares complete homology except at 3'-UTR. Therefore, it is expected to share some properties of SIRM. Moreover, the SIRM protein has been described as a novel insulin-regulated SH3 binding protein that associates with Grb2 and FYN (Salvatore et al. 1998). Another gene OTX2-As1 is a transcription factor of OTX2, which is involved in regulating gonadotrophin releasing hormone (GnRH) in hypogonadism. A study from the region observed that 36.5% of individuals with T2DM had hypogonadism (Al Hayek et al. 2013) and another study using mice OTX2 knockout experiments indicated that OTX2 is crucial for GnRH expression (Diaczok et al. 2011). Therefore, rs7144734 (associated with FPG) could be a plausible link for the observed pleiotropy. Similarly, PLGRKT regulates catecholamine release, which in turn regulates TG metabolism (Bai et al. 2011; Iriyama et al. 1984). Likewise, almost all variants are directly or indirectly associated with metabolic traits-related processes, as comprehensively explained in the original article.

The allele frequency variation of subgroups and study variants found near genome-wide or nominal p-value is shown in **Figure 13B**. A proportional test for EAF of study variants with each EAF of subgroup variants suggested rs900543 in Bedouin subgroup (EAF = 0.196) displayed a significant difference from GWAS (EAF = 0.086) with a p-value < 0.05. Hence, the association of rs900543 at the study cohort may differ in Bedouin subgroup.

In summary, this study identified 36 unique variants at p-value < 1.0E-05 from 13 metabolic traits after analyzing 1,965 subjects of Arab ethnic background. Of these, 1 variant associated at close to genome-wide p-value threshold, 5 variants at a nominal p-value threshold, and 30 (from 33 associations) variants between nominal to suggestive p-value threshold. Of the latter 30, 16 variants passed the BH-FDR test and 2 have been replicated in global population studies. Examination of these variants in GTEx revealed gene expression regulation from 7 variants. The evidence from canonical pathway analysis and literature survey together entwined to connect study results to metabolic processes or disorders.

| Phenotype | Variant: effect allele | Gene | Model | MAF | Beta[$] | P-value[$] | Beta[#] | P-value[#] | Bonferroni p-value | Benjamini – Hochberg FDR p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| **A. Association at close to genome-wide significance of p-value ≤3.41E-08** | | | | | | | | | | |
| HbA1c | chr15:40531386-rs12440118: G | *ZNF106* (W>R) | Recessive | 0.1059 | 2.006 | 7.07E-08 | 1.498 | 2.70E-05 | 0.003 | 0.0032 |
| **B. Associations at nominal significance of p-value ≤5.45E-07** | | | | | | | | | | |
| FPG | rs7144734: A | *OTX2-AS1* (upstream) | Recessive | 0.2036 | 1.465 | 2.82E-07 | 1.361 | 4.31E-07 | 0.012 | 0.012 |
| TGL | rs17501809: A | *PLGRKT* (intronic) | Recessive | 0.0593 | 1.807 | 1.04E-07 | 1.74 | 3.29E-07 | 0.0047 | 0.0016 |
| TGL | rs11143005: A | *LOC105376072* (intronic) | Recessive | 0.2829 | 0.4196 | 4.03E-07 | 0.4225 | 3.21E-07 | 0.018 | 0.0030 |
| TGL | chr12:101494770/ rs10860880: C | *IGF1* (downstream) | Recessive | 0.05575 | 1.596 | 2.077-07 | 1.603 | 1.74E-07 | 0.0095 | 0.0019 |
| TGL | rs900543: A | *THSD4*, *NR2E3* (intergenic/ downstream to *NR2E3*) | Recessive | 0.08696 | 1.625 | 1.27E-07 | 1.607 | 1.67E-07 | 0.0058 | 0.0016 |

**Table 10: List of variants identified from study I at genome-wide and nominal significance of p-value; these variants passed BH-FDR p-value. [$],** Association tests were adjusted for age, sex, and the first 10 principal components; **[#],** tests with anthropometric and lipid traits were adjusted for age, sex, principal components, and lipid lowering medication status; tests with HbA1c and FPG were adjusted for age, sex, principal components, and diabetes medication status; tests with SBP and DBP were adjusted for age, sex, principal components, and hypertension medication status.

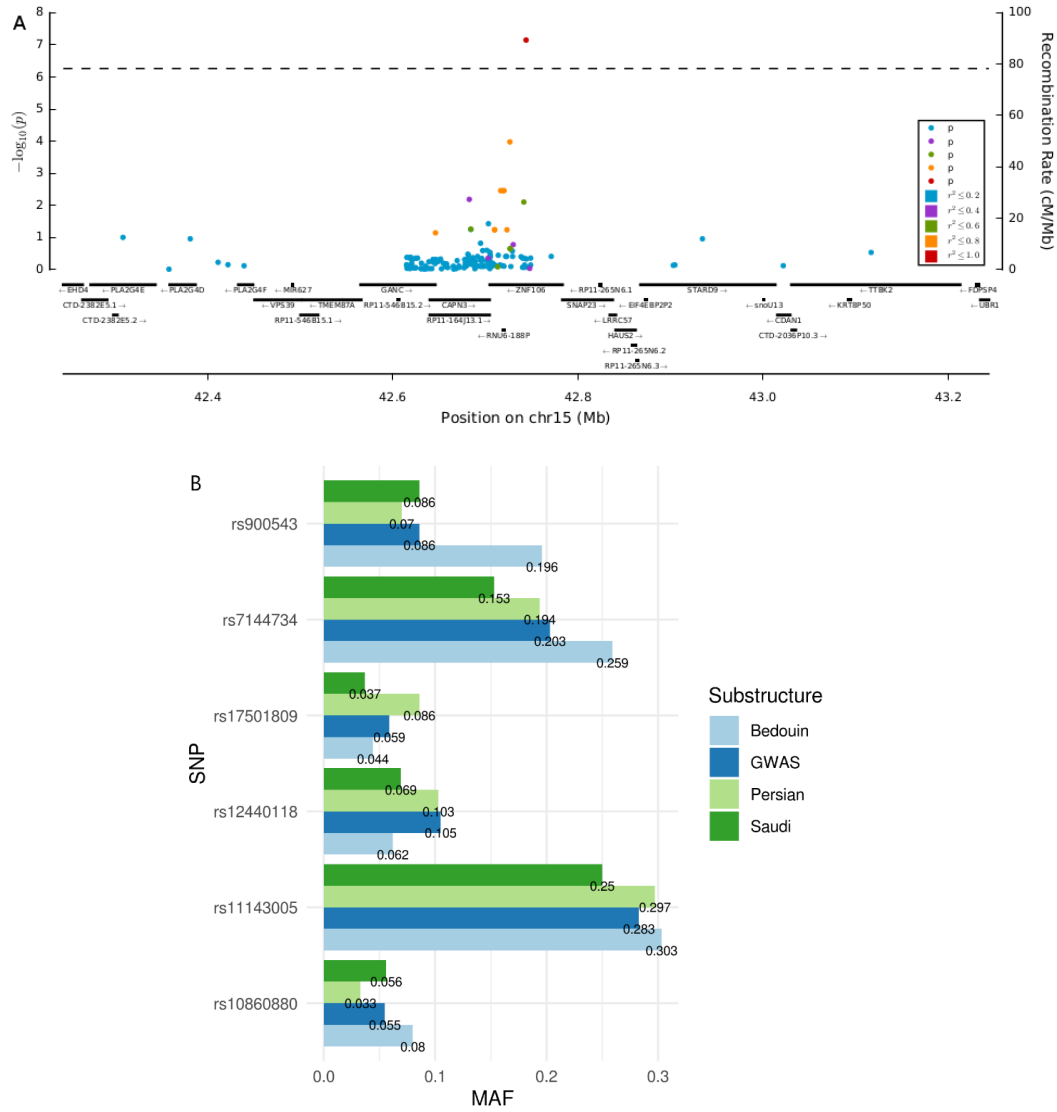


**Figure 13: A) regional plot of rs12440118 (ZNF106, W>R) at 1 Mb region center to it, along with LD variants chr15:40433466, $r^2$ = 0.61; chr15:40471747, $r^2$ = 0.53; chr15:40490502, $r^2$ = 0.33; chr15:40497290, $r^2$ = 0.62; chr15:40497469, $r^2$ = 0.64; chr15:40500754,$r^2$ = 0.48; chr15:40503873,$r^2$ = 0.62; chr15:40506209,$r^2$ = 0.64; chr15:40507580,$r^2$ = 0.64; chr15:40510699, $r^2$ = 0.63; chr15:40513902,$r^2$ = 0.76; chr15:40514171,$r^2$ = 0.52; chr15:40514188,$r^2$ = 0.52; chr15:40517485,$r^2$ = 0.34; chr15:40528928,$r^2$ = 0.52. B) Variation observed in effect allele frequency of reported SNVs among the three subgroups of Kuwaiti Arab population.**

## 9.2 Study II: Examination of genome-wide variants association with anthropometric traits implicate rs9606756 of TCN2 to be associated with obesity and reveals a plausible mechanism of mediation via interaction with Apo-A1

Upon QC, 1353 participants in the discovery and 1176 participants in the replication phase were assessed for 634055 SNPs. The clinical characteristics of participants in both the cohorts have been described in the original article. From anthropometric traits (such as WC, WcHtR, BMI, and weight) association tests, 9 SNVs (including the association of rs9606756 [TCN2, I > L] with all 4 anthropometric traits, rs12347078 [DOCK8] with WC, WcHtR, and BMI; rs4910469 [ZNF143] with WC, BMI, and Weight; rs6887349 [FYB] with BMI; rs9479923 [CNKSR3, SCAF8] with BMI; rs736100 [MGAT5B] with weight; rs9419689 [GLRX3, PPP2RD] with weight; rs2578461 [LOC105369608, LOC105369611] with WC and WcHtR; rs12366294 [CCDC91, FAR2] with WC, WcHtR, and BMI) were identified at p-value < 1E-05. After evaluating the effect size direction and p-value, 3 variants: rs9606756 [TCN2, I > L], rs12347078 [DOCK8], and rs4910469 [ZNF143] were selected for replication. Intriguingly, after adjusting for diabetes and medication status along with regular correction, the discovery p-value of rs9606756 reached 8.92E-08 with WC, a value close to the standard GWA study p-value (5.0E–08). In addition, from replication analysis, only rs9606756 [TCN2, I > L] was found to be replicated and joint analysis was conducted (**Table 11**). A proportional test to assess rs9606756 allele frequency [EAF = 0.114] difference between the study and each of the subgroup population suggested no significant difference between each subgroup and study allele frequencies. **Figure 14A** shows the distribution of MAF of rs9606756 among the three subgroups of Kuwaiti Arab population.

A key point is that separate lipid trait analysis showed that the same TCN2 variant rs9606756 was also found to be associated with decreased levels of HDL-C at p-value 3.87e-05 (**Table 11**), thus suggesting a plausible lipid mediated causal link for obesity. Hence, additional biomarkers were examined, including Apo-A1, Apo-B1, homocysteine, leptin, vitamin B12, and methylmalonic acid from the plasma of 200 randomly selected samples from replication cohort. These biomarkers were tested for interaction with anthropometric traits using linear regression adjusted for regular correction. Interesting enough, only ApoA1 and HDL-C consistently showed significant interaction with all 4 anthropometric traits with the effect of unit increase in anthropometric traits with unit decrease in Apo-A1 and HDL-C traits at AG genotypes of rs9606756. **Table 12** shows the summary statistics of interaction between Apo-A1 and HDL-C with WC at genotype levels, when the reference AA genotype was considered as an intercept in linear regression analysis. **Figure 14B and 14C** illustrate the interaction between WC vs. Apo-A1 and HDL-C, respectively, at three genotype levels.

Examination of rs9606756 in GTEx data showed the downregulation of SLC35E4 and DUSP18 in skin unexposed to sun, adipose subcutaneous, and whole blood, as well as downregulation of TCN2, RP1-56J10.8, and DUSP18 in Testis. Although TCN2 functionally encodes vitamin B12 binding protein family, levels of vitamin B12 measured from plasma were not significantly different among the genotype groups of rs9606756 (although vitamin B12 is a product of TCN2).

This could be explained by tissue-specific expression of TCN2 or the common practice of vitamin supplementation in subjects (physicians generally prescribe vitamin supplements to people of this region either because of diabetic patients on Metformin prescription, which is known to lower vitamin B12 levels or for common lack of sun exposure in the region due to extreme hot weather).

Upon observing high LD and number of genes in 500 Kb centric to rs9606756 region, haplotypes were identified, and haplotype association analysis was performed. Haplotypes GGAA, AGAA, and AGGA involving rs11912660, rs9606756, rs9621049, and rs5994329 SNVs were found to be associated with increased WC with p-value 0.018, 0.021, and 0.0028, respectively (**Table 13**). Moreover, to examine the contributions of genes neighboring to TCN2 (from a 2Mb region centered at identified risk variant) to the causal pathway, gene-based tests for association were performed using VEGAS based algorithm, which successfully combined GWA study SNP p-value using LD between SNPs and assigned corrected p-values for each gene in the region. SNVs from six candidate genes (SLC14L6, GAL3ST1, PES1, SLC35E4, DUSP18, and OSBP2) were involved in addition to TCN2 SNVs in this analysis. These p-values were used for canonical pathway analysis applying IPA software. Next, the protein–protein interactions were explored and TP53 (tumor P53 protein) was found to be a key element in connecting the network components. Literature supports that TP53 is involved in brown adipogenic differentiation and protection against diet-induced obesity (Molchadsky et al. 2013). Furthermore, TP53 regulates insulin resistance (Minamino et al. 2009) and lipid metabolism (Wang et al. 2013).

In summary, rs9606756 [TCN2, I > L] displayed a reliable association with increased WC, BMI, weight, and WcHtR traits in cohorts of discovery and replication comprising 1353 and 1176 individuals, respectively. SNV rs9606756 [TCN2, I > L] also showed an association with lower levels of HDL-C in both cohorts. An analysis further employed to discover the interaction between anthropometric traits and ApoA1, showed a unit increase in anthropometric traits with unit decrease in Apo-A1 and HDL-C traits at AG genotypes of rs9606756, thus implicating a plausible lipid-mediated biomarker for obesity. Haplotypes from the 500 Kb region of rs9606756 encompassing rs11912660 [GAL3ST1, upstream], rs9606756 [TCN2, I>L], rs9621049 [TCN2, S>F], rs5994329 [SLC35E4, intronic] variants also showed an association with anthropometric traits. Subsequently, gathering significant genes from 2 Mb region centric to rs9606756 by gene-based association tests and feeding these genes to the canonical pathway analysis revealed a complex gene network with TP53 as a central gene component.

| Trait | SNP: effect allele | Phase | Beta | P-value | Beta | P-value |
|---|---|---|---|---|---|---|
| WC | rs9606756: G | Discovery | 4.815 | 1.46e-07 | 4.887 | 9.462e-08 |
| | | Replication | 1.845 | 0.0132 | 1.842 | 0.0134 |
| | | Joint | 3.452 | 1.00E-08 | 3.496 | 6.15e-09 |
| WcHtR | rs9606756: G | Discovery | 0.0271 | 1.28e-06 | 0.02754 | 9.46e-07 |
| | | Replication | 0.0098 | 0.0302 | 0.00982 | 0.0305 |
| | | Joint | 0.0180 | 9.53e-07 | 0.0183 | 6.26e-07 |
| BMI | rs9606756: G | Discovery | 1.858 | 1.22e-05[e] | 1.899 | 7.646e-06 |
| | | Replication | 0.5974 | 0.0911 | 0.5971 | 0.0915 |
| | | Joint | 1.22 | 1.70e-05 | 1.247 | 1.05e-05 |
| Weight | rs9606756: G | Discovery | 5.535 | 7.16e-06 | 5.664 | 4.17e-06 |
| | | Replication | 1.947 | 0.0569 | 2.074 | 0.0458 |
| | | Joint | 3.747 | 5.77e-06 | 3.824 | 3.51e-06 |
| HDL-C | rs9606756: G | Discovery | -0.0822 | 3.87e-05 | -0.08606 | 6.81e-05 |
| | | Replication | 0.0081 | 0.6670 | 0.00825 | 0.6619 |
| | | Joint | -0.0398 | 0.00558 | -0.04117 | 0.00403 |

**Table 11: Summary statistics of rs9606756 (Ile23Val) [EAF = 0.114] variant of TCN2 with waist circumference, waist circumference to height ratio, BMI, weight, and HDL-C traits in each of the three phases of the study.**

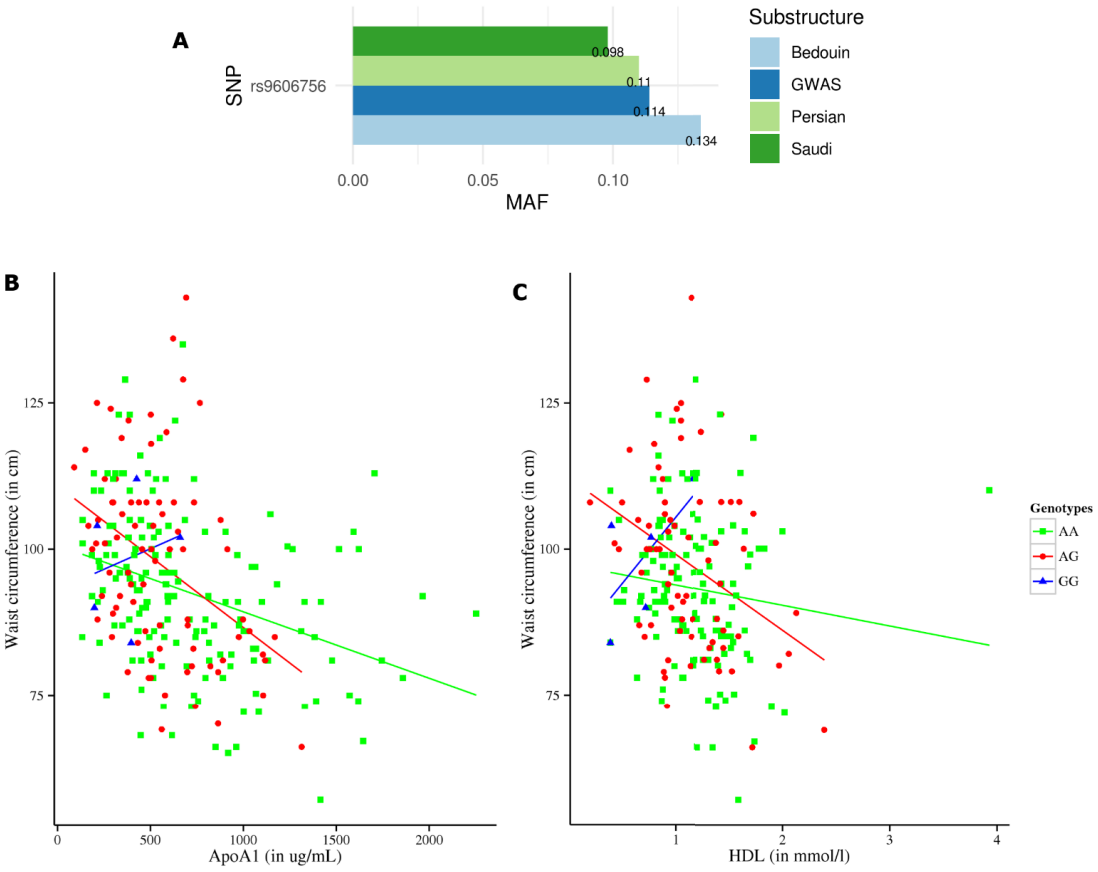


**Figure 14: A) MAF distribution of rs9606756 among the three subgroups of Kuwaiti Arab population; B) and C) Interaction between WC vs. Apo-A1 and HDL-C, respectively, at three genotype levels of rs9606756.**

| Trait (or response variable) | Genotype and/or interacting trait | Estimate | Std. Error | P-value | Adjusted $R^2$ of model |
|---|---|---|---|---|---|
| WC | GA | 8.76 | 3.55 | 0.0143 | 0.32 |
| | GG | -2.33 | 13.42 | 0.862 | |
| | GA*ApoA1 | -0.01 | 0.005 | 0.046 | |
| | GG*ApoA1 | 0.014 | 0.032 | 0.645 | |
| | | | | | |
| HDL-C | GA | 13.574 | 5.094 | 0.008 | 0.28 |
| | GG | -8.876 | 14.29 | 0.535 | |
| | GA*ApoA1 | -9.708 | 4.213 | 0.022 | |
| | GG*ApoA1 | 20.19 | 18.88 | 0.286 | |

**Table 12: Interactions of WC and HDL-C with ApoA1 at rs9606756 risk genotypes.**

| Haplotypes | Frequency | Effect size | P-value |
|---|---|---|---|
| GGAA | 0.054 | 3.11 | 0.018 |
| AGAA | 0.026 | 4.33 | 0.021 |
| AGGA | 0.021 | 6.1 | 0.0028 |
| AAGA | 0.052 | -0.81 | 0.545 |
| AAAC | 0.021 | 0.414 | 0.841 |
| GAGC | 0.020 | -3.22 | 0.13 |
| AAGC | 0.789 | -2.12 | 0.0034 |

**Table 13: Haplotypes from rs11912660 (distance to rs9606756: −34.5 Kb, LD-$R^2$:0.398), rs9606756, rs9621049 (distance to rs9606756: 6.56 Kb, LD-$R^2$:0.587), and rs5994329 (distance to rs9606756: 42.2 Kb, LD-$R^2$:0.536) and their association with WC. Haplotypes GGAA, AGAA, and AGGA have shown an association with increased WC, whereas haplotype AAGC has shown an association with decreased WC.**

## 9.3 Studies III and IV: Genome-wide association study of lipid and glycemic traits in corroboration with homozygosity mapping implicate recessive metabolic risk signatures in Arab genome

Studies III and IV were conducted on the same discovery (n = 1353) and replication (n = 1176) cohorts as study II. These studies differed in the types of traits; study II was conducted on anthropometric traits, whereas studies III and IV were conducted on lipid and glycemic traits, respectively. Therefore, clinical characteristics of the study subjects remained the same across all three studies. However, in study IV, while FPG measurements were available for all samples of the discovery cohort, HbA1c values were available for only 750 samples. Consequently, variants associated with both FPG and HbA1c were taken for further analysis and variants associated with HbA1c alone were excluded from it.

From the discovery cohort, 632,375 SNVs were interrogated in association tests with lipid and glycemic traits, using linear regression adjusted for age, sex, pc (regular correction), and medication status. Following examination of test results, 32 variants with p-value < 1E-$10^{-5}$ were short-listed for replication phase from lipid traits (of which 19 were associated with TG, 8 with HDL-C, 4 with TC, and 1 each with TC and LDL-C). Nonetheless, upon replication phase analysis, only 12 variants associated with TG at either the genome-wide significant (p-value <6.12E-09) or borderline p-values (6.12E-09 > p <5.0E-08) under recessive model were replicated. Moreover, 1 variant from TG (under recessive) and 3 variants associated with HDL-C at nominal p-values (under additive model) also passed the replication p-value threshold of <0.05. Six risk variants were associated with high TG (namely rs1002487 [RPS6KA1], rs11805972 [LAD1], rs7761746 [Or5v1], rs39745 [CTTNBP2-LSM8], rs2934952 [PGAP3], and rs9626773 [RP11-191L9.4-CERK]) at genome-wide significance (**Table 14**), and another six variants (namely rs10873925 [ST6GALNAC5], rs4663379 [SPP2-ARL4C], rs10033119 [NPY1R], rs17709449 [LINC00911-FLRT2], rs11654954 [CDK12-NEUROD2], and rs9972882 [STARD3]) were associated at borderline significance (**Table 15**). The summary statistics of both phases were combined by meta-analysis. Additionally, 22 variants from glycemic traits with p-value < 1E-$10^{-5}$ were short-listed for replication phase. Of these, in replication phase 4 variants were found replicated which include, variants rs1002487 [RPS6KA1] at genome-wide p-value <1.8E-08 and rs487321 [CADPS] at nominal p-value (i.e., 1.84E-08> p <E-05) under recessive model, rs707927 [VARS, VWA7] and rs12600570 at nominal p-value under additive model from discovery phase. These 4 variants from both the phases were combined by meta-analysis (**Table 16**). Along with regular and medication (lipid or glucose lowering) corrections, both lipid and glycemic traits were also corrected for obesity and diabetes status. A large proportion of participants of discovery cohort (59% and 45%) and replication cohort (46% and 38%) were obese and diabetic. After corrections, p-value remained significant in associated variants. But, it required careful re-examination of associations with FPG, as a concern arises that FPG values measured in participants receiving glucose-lowering medication represent “naturally” observed values in the population. Hence, associations were further scrutinized using two approaches: one, by pre-adjusting FPG by adding

a value of 2.5 mmol/L (an average FPG value that was observed in an in-house clinical database of diabetic patients) to the FPG values of the participants taking diabetes medication & then performing the association tests and second, by assessing the identified associations in sub-cohorts of entirely diabetic or entirely non-diabetic participants. Subsequently on pre-adjusting the FPG values, p-value remained significant in 4 associated variants. Similarly, p-values of CADPS, VARS, and DHX58 variants remained significant in diabetic and diabetes-free cohorts, whereas variant from RPS6KA1 showed a significant p-value with the diabetic cohort alone.

Interaction (or correlation) analysis of TG, FPG, and HbA1c with insulin resistance traits, namely, HOMA-IR, HOMA-β, C-peptide, and HOMA-S at genotypes of 4 risk variants from the third cohort of 283 samples revealed a significant correlation between the risk genotypes. With rs1002487 of RPS6KA1, genotypes homozygous for the risk allele showed significant correlation between TG, FPG, and HbA1c vs. insulin resistance traits (HOMA-β, C-peptide, HOMA-S) at a multiple testing significance threshold of p-value <0.003. rs707927 of [VARS, VWA7], at genotypes heterozygous or homozygous for the risk allele, interactions between FPG and insulin resistance traits (HOMA-β and HOMA-S) were observed at a multiple testing significance threshold of <0.003. With a heterozygous genotype, the interactions between HbA1c and insulin resistance traits (HOMA-IR and HOMA-S) were observed at a multiple testing significance threshold of <0.003. rs487321 of CADPS, shown with a heterozygous genotype, interaction between TG and HOMA-IR were observed at a multiple testing significance threshold of <0.003. Interaction between FPG and HOMA-β with the heterozygous genotype was observed at a p-value < 0.003; rs12600570 of DHX58 shown with genotypes homozygous for the risk allele, TG and FPG were seen to be associated with (HOMA-IR and C-peptide levels) and HOMA-β, respectively, at a multiple testing significance threshold of <0.003.

ROH analysis results suggested that all 12 identified risk variants (with genome-wide or borderline p-value) of TG (**Table 15 and Table 16**), 4 risk variants of FPG (**Table 17**), and rs1002487 being a common variant between TG and FPG, in total 15 of these variants were harbored in ROH segments in our study cohort. Comparison of these ROH segments (harboring risk variants) with data by *Pemberton et al.* revealed that 4 (namely rs11805972 [LAD, P > Q], rs7761746 [OR5V1], rs4663379 [SPP2, ARL4C], and rs10033119 [NPY1R]) of the 15 risk variants harbored were novel ROH segments.

The role of these variants in the regulation of gene expression was evaluated using GTEx and 6 of the 12 risk variants of TG were found to regulate their proximal or own genes in various tissues. SNP rs2799686 in LD with rs11805972 regulates LAD1; rs39745 regulates LSM8; rs2934952 and rs11654954 regulate PGAP3 and STARD3; rs12897409 in LD with rs17709449 regulates FLRT2; rs9972882 regulates STARD3, PGAP3, and RP11-690G19.3. In conjunction with this, all the 4 risk variants of FPG showed gene expression regulation: rs1002487 of RPS6KA1 regulates DHDDS; rs487321 regulates CADPS; rs707927 regulates LY6G5B, GPANK1, AIF1, C6orf25, SAPCD1-AS1, and TNXA; rs12600570 regulates DHX58, KCNH4, HSPB9, and RAB5C.

The EAF variation in the three subgroups for 12 variants observed at genome-wide and borderline p-value with TG and 4 variants at genome-wide and nominal p-value with FPG (of which rs1002487 is common variant between TG and FPG) is shown in **Figure 15**. None of these variants showed significant differences between EAF of study and each of subgroup variants. Therefore, association of these variants may not differ in each subgroup.

In summary, the analysis of lipid traits revealed 6 variants at genome-wide, 6 variants at borderline, and 3 variants at nominal p-value for elevated TG. Likewise, 3 variants at nominal p-values were identified for low HDL-C. ROH analysis identified that 4 (namely rs11805972 [LAD, P > Q], rs7761746 [OR5V1], rs4663379 [SPP2, ARL4C], and rs10033119 [NPY1R]) among the 15 risk variants harbored were novel ROH segments. Analysis of glycemic traits revealed 1 variant at genome-wide and 3 variants at nominal p-values. All the 4 variants were harbored on known ROH regions. The tests for interactions of TG, FPG, and HbA1c with insulin resistance traits at genotype levels showed a significant correlation at genotypes heterozygous or homozygous for the risk allele of all 4 variants.

| SNP: Effect Allele: EAF | Gene: functional consequences | Phase | EffectSize$^{R}$ | P-value$^{R}$ | EffectSize$^{LM}$ | P-value$^{LM}$ | Mean±SD of ROH groups | Distance to SNP from mean±SD window (in Mb) | Presence of SNP in ROH regions identified in global population |
|---|---|---|---|---|---|---|---|---|---|
| rs1002487:C: 0.0594 | RPS6KA1: intron | Discovery | 3.337 | 7.17E-11 | 3.344 | 6.69E-11 | 1: 24917436-33009426 | Overlapping | Yes |
| | | Replication | 1.036 | 0.01805 | 1.045 | 0.0165 | | | |
| | | Meta | 6.517 | 7.17E-11 | 6.197 | 5.76E-10 | | | |
| rs11805972:T: 0.072 | LAD1: P>Q | Discovery | 2.481 | 8.55E-11 | 2.48 | 8.79E-11 | 1: 195716208-206342586 | Overlapping | No |
| | | Replication | 1.893 | 4.03E-08 | 1.878 | 4.50E-08 | | | |
| | | Meta | 8.485 | 2.16E-17 | 8.351 | 6.75E-17 | | | |
| rs7761746:A: 0.076 | OR5V1: synonymous | Discovery | 2.447 | 1.89E-09 | 2.443 | 2.08E-09 | 6: 26954921-31677127 | Overlapping | No |
| | | Replication | 0.8277 | 9.97E-03 | 0.8443 | 7.88E-03 | | | |
| | | Meta | 6.066 | 1.31E-09 | 6.229 | 4.68E-10 | | | |
| rs39745:T: 0.12 | CTTNBP2, LSM8:intergenic | Discovery | 1.485 | 3.63E-09 | 1.486 | 3.56E-09 | 7: 114686349-122109936 | Overlapping | Yes |
| | | Replication | 0.4809 | 3.69E-02 | 0.4769 | 3.95E-02 | | | |
| | | Meta | 5.643 | 1.67E-08 | 5.619 | 1.92E-08 | | | |
| rs2934952:G: 0.245 | PGAP3: intron | Discovery | 0.773 | 3.17E-09 | 0.7722 | 3.38E-09 | 17:35214542-40789744 | Overlapping | No, but SNPs in LD with it are present (rs2941503 and rs1565922). |
| | | Replication | 0.3243 | 6.87E-03 | 0.3167 | 7.94E-03 | | | |
| | | Meta | 6.086 | 1.16E-09 | 6.054 | 1.41E-09 | | | |
| rs9626773:A: 0.051 | RP11-191L9.4,CERK: intergenic | Discovery | 2.848 | 1.42E-09 | 2.857 | 1.29E-09 | 22:40225588-44582123 | 3.6 | Yes |
| | | Replication | 3.395 | 8.60E-07 | 3.315 | 1.54E-06 | | | |
| | | Meta | 7.776 | 7.47E-15 | 7.855 | 4.00E-15 | | | |

**Table 14: Results of the statistical association tests from three phases with TG (p-values approaching genome-wide significance at ≤6.12E-09) and ROH mapping of associated variants.** Note: Effect size represents beta value for discovery &replication phases, and Z-score for meta-analysis. R-regular correction: Corrected for age, sex, and the top 10 principal components that resulted from the principal component analysis of the genotype data; LM: Corrected for lipid medication in addition to the regular correction.

| SNP: Effect Allele | Gene: functional consequences | Phase | EffectSize$^{R}$ | P-value$^{R}$ | EffectSize$^{LM}$ | P-value$^{LM}$ | Mean±SD of ROH groups | Distance to SNP from mean±SD window (in Mb) | Presence of SNP in ROH regions identified in global population |
|---|---|---|---|---|---|---|---|---|---|
| rs10873925: G$^{\#}$ : 0.30 | ST6GALNAC5: intron | Discovery | 0.633 | 4.11E-08 | 0.6031 | 4.19E-08 | 1:75058792-81502412 | Overlapping | Yes |
| | | Replication | 0.2035 | 0.0427 | 0.2092 | 0.03631 | | | |
| | | Meta | 5.364 | 8.12E-08 | 5.184 | 2.18E-07 | | | |
| rs4663379: G$^{\#}$: 0.105 | SPP2, ARL4C: intergenic | Discovery | 1.841 | 8.38E-09 | 1.842 | 8.39E-09 | 2:235304942-239225206 | 0.095 | No |
| | | Replication | 0.5718 | 0.005562 | 0.5861 | 0.0042 | | | |
| | | Meta | 6.078 | 1.22E-09 | 6.051 | 1.44E-09 | | | |
| rs10033119: G$^{\#}$: 0.065 | NPY1R: 3'-utr | Discovery | 2.698 | 8.79E-09 | 2.695 | 9.24E-09 | 4:162229756-174103686 | Overlapping | No |
| | | Replication | 1.362 | 0.00064 | 1.336 | 0.00078 | | | |
| | | Meta | 6.446 | 1.15E-10 | 6.583 | 4.60E-11 | | | |
| rs17709449: T$^{\#}$ :0.144 | LINC00911, FLRT2: intergenic | Discovery | 1.173 | 5.12E-08 | 1.172 | 5.36E-08 | 14: 78521481-89540460 | Overlapping | Yes |
| | | Replication | 0.3883 | 0.02078 | 0.3937 | 0.0185 | | | |
| | | Meta | 5.447 | 5.12E-08 | 5.481 | 4.24E-08 | | | |
| rs11654954: A$^{\#}$: 0.174 | CDK12, NEUROD2: intergenic | Discovery | 0.9881 | 2.18E-08 | 0.9915 | 2.00E-08 | 17: 35214542-40789744 | Overlapping | No, But LD SNPs present (rs4795369, rs879606, rs907094) |
| | | Replication | 0.3692 | 1.98E-02 | 0.3544 | 2.59E-02 | | | |
| | | Meta | 5.602 | 2.13E-08 | 5.555 | 2.77E-08 | | | |
| rs11654954: A$^{@}$:0.174 | CDK12, NEUROD2: intergenic | Discovery | 0.3311 | 3.75E-08 | 0.3312 | 3.74E-08 | | | |
| | | Replication | 0.1289 | 1.47E-02 | 0.1149 | 2.89E-02 | | | |
| | | Meta | 5.612 | 2.00E-08 | 5.432 | 5.57E-08 | | | |
| rs9972882: A$^{\#}$ :0.245 | STARD3: intron | Discovery | 0.7284 | 1.81E-08 | 0.7287 | 1.80E-08 | 17: 35214542-40789744 | Overlapping | No, But LD SNPs present (rs879606, rs907094, rs1877031, rs931992) |
| | | Replication | 0.3311 | 4.43E-03 | 0.3168 | 5.73E-03 | | | |
| | | Meta | 5.99 | 2.11E-09 | 5.918 | 3.25E-09 | | | |

**Table 15: Results of the statistical association tests from the discovery, replication, and meta-analysis phases with TG (p-values approaching borderline to genome-wide significance at >6.12E-09 &<5.0E-08) upon using recessive and additive models.** $^{\#}$, Recessive Model; $^{@}$, Additive Model.

| SNP: Effect Allele: Trait | Gene: functional consequences | Phase | Effect Size$^{R}$ | P-value$^{R}$ | Effect Size$^{DM}$ | P-value$^{DM}$ |
|---|---|---|---|---|---|---|
| rs1002487: C$^{\#}$, FPG | *RPS6KA1*: intronic | Discovery | 8.315 | 1.64E-08 | 8.297 | 1.58E-08 |
| | | Replication | 3.442 | 3.7E-04 | 3.509 | 2.15E-04 |
| | | Meta | 6.551 | 5.72E-11 | 6.652 | 2.89E-11 |
| rs1002487: C$^{\#}$, HbA1C | *RPS6KA1*: intronic | Discovery | 7.367 | 4.91E-08 | 7.186 | 9.649E-08 |
| | | Replication | 1.811 | 2.71E-03 | 1.875 | 0.00115 |
| | | Meta | 5.784 | 7.27E-09 | 5.896 | 3.71E-09 |
| rs487321: A$^{\#}$, FPG | *CADPS*: intronic | Discovery | 6.133 | 1.53E-07 | 6.161 | 1.23E-07 |
| | | Replication | 3.955 | 2.25E-06 | 3.88 | 3.033E-06 |
| | | Meta | 7.047 | 1.83E-12 | 7.031 | 2.054E-12 |
| rs487321: A$^{\#}$, HbA1C | *CADPS*: intronic | Discovery | 2.387 | 2.47E-03 | 2.38 | 2.44E-03 |
| | | Replication | 1.893 | 2.77E-04 | 1.826 | 3.82E-04 |
| | | Meta | 4.723 | 2.32E-06 | 4.569 | 3.18E-06 |
| rs707927: G$^{@}$, FPG | *VARS (intronic), VWA7 (2Kb upstream)* | Discovery | 0.9453 | 8.24E-06 | 0.9262 | 1.19E-05 |
| | | Replication | 0.6375 | 8.25E-05 | 0.6503 | 3.18E-05 |
| | | Meta | 5.928 | 3.074E-09 | 6.033 | 1.61E-09 |
| rs707927: G$^{@}$, HbA1C | *VARS (intronic), VWA7 (2Kb upstream)* | Discovery | 0.5632 | 5.43E-04 | 0.5502 | 6.96E-04 |
| | | Replication | 0.3689 | 1.63E-04 | 0.3799 | 8.33E-05 |
| | | Meta | 5.088 | 3.61E-07 | 5.181 | 2.21E-07 |
| rs12600570: T$^{@}$, FPG | *DHX58*: intronic | Discovery | 0.8166 | 7.49E-06 | 0.8374 | 4.11E-06 |
| | | Replication | 0.3892 | 4.67E-03 | 0.3682 | 5.65E-03 |
| | | Meta | 5.142 | 2.715E-07 | 5.186 | 2.15E-07 |
| rs12600570: T$^{@}$, HbA1C | *DHX58*: intronic | Discovery | 0.31 | 2.82E-02 | 0.3344 | 1.76E-02 |
| | | Replication | 0.194 | 1.98E-02 | 0.1805 | 2.81E-02 |
| | | Meta | 3.179 | 1.47E-03 | 3.178 | 1.48E-03 |

**Table 16: List of the four identified risk variants associated with FPG either at genome-wide significant p-values (<1.8E-08) or at nominal p-values of <1.0E-06 and association of the same variants with HbA1c.** EffectSize, Effect size represents beta value for discovery and replication phases and Z-score for meta-analysis. R-regular correction: Corrected for age, sex, and the top 10 principal components; DM: Corrected for diabetes medication in addition to the regular correction; #, association was observed under recessive mode of inheritance; @, association was observed under additive mode of inheritance.

| SNP: gene | ROH group and the method used to identify the ROH[@] | Consensus ROH region | Distance to SNP from consensus ROH (in Mb) | Number of individuals in ROH group | Length of consensus ROH (in Kb) | Number of SNPs in consensus ROH region | Mean±SD of ROH groups | Distance to SNP from mean±SD window (in Mb) | Presence of SNP in ROH regions identified in global population |
|---|---|---|---|---|---|---|---|---|---|
| rs1002487: *RPS6KA1* | S1818[1] | 1:28864435-29062427 | 1.99 | 51 | 197.99 | 11 | 24917436-33009426 | Overlapping | Yes |
| | S1557[2] | 1:28056342-28084571 | 1.19 | 44 | 28.23 | 5 | 27723540-28417372 | 0.85 | |
| rs487321: *CADPS* | S7177[1] | 3:62647115-63435226 | 0.143 | 29 | 788.11 | 226 | 3:55304385-70777955 | Overlapping | Yes |
| | S4114[2] | 3:61981197-62189189 | 0.60 | 31 | 207.99 | 85 | 3:56659352-67511033 | Overlapping | |
| rs707927: [*VARS*, *VWA7*] | S1706[1] | 6:31001421-32989521 | 0.744 | 53 | 1988.10 | 1077 | 6:26827255-36745549 | Overlapping | No, But LD SNP rs805267 ($r^2$=0.69) is present |
| | S687[2] | 6:29569045-29593788 | 2.176 | 71 | 24.74 | 24 | 6:26617526-32545306 | Overlapping | |
| rs12600570: *DHX58* | S5153[1] | 17:39980819-40041676 | Overlapping | 34 | 60.858 | 8 | 17:36532212-43490282 | Overlapping | No, But LD SNP rs2074158 ($r^2$=0.56) is present. |
| | S1741[2] | 17:40041676-40063083 | 0.219 | 43 | 21.408 | 5 | 17:39559717-40545041 | Overlapping | |

**Table 17: Summary of ROH mapping of 4 variants associated with FPG and their proximity to ROH regions identified from two different methods (for details see relevant sections of study III and IV).**

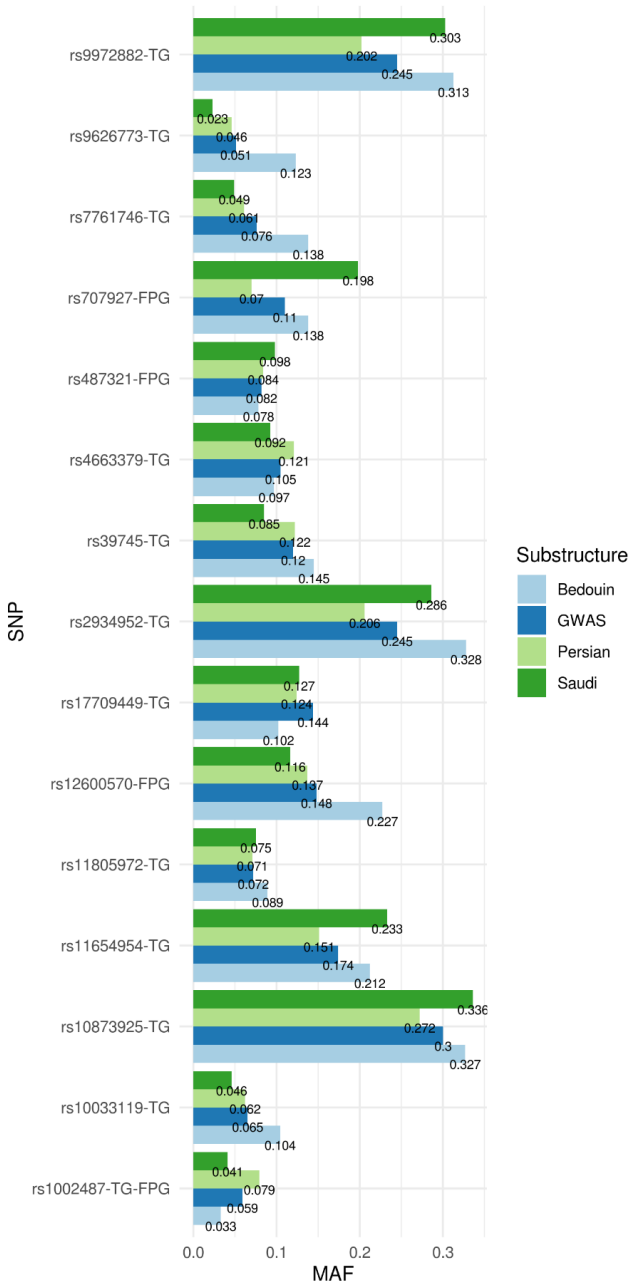


**Figure 15: Minor allele frequencies of associated variants from lipid and glycemic traits analysis in subgroups of Kuwaiti Arab population.**

## 9.4 Study V: Imputation of genome-wide variants and examination of their associations with 13 metabolic quantitative traits

Genome-wide imputation of 1298 samples genotyped in HumanOmniExpress and 1434 samples genotyped in HumanCardio-MetaboChip using MIS resulted in 48,424,667 and 46,597,973 variants from both datasets, respectively. However, MAF < 5% variants comprising a large proportion of imputed variants displayed a poor imputation quality score (Rsq), whereas MAF ≥ 5% comprising a small proportion of imputed variants showed an average Rsq > 0.75, i.e., confidently imputed in both datasets. After filtering for MAF ≥ 5% and Rsq ≥ 0.5, the number of variants that remained for association tests was 6.08 million (mean Rsq = 0.822) in OmniExpress and 1.52 million (mean Rsq = 0.759) in Cardio-MetaboChip datasets.

Subsequently, statistical association tests were performed with 13 transformed traits in individual datasets, followed by meta-analysis which revealed 978 associations involving 821 unique variants from 251 gene loci. These associations were selected in such a way that all associated variants had p-value < 0.05 and consistent effect size direction in each respective trait association test results of both the datasets. Of these 821 variants, 72 associations at genome-wide significance (p-value $<5.0 \times 10^{-08}$) comprising 70 unique variants from 9 genes were observed. Furthermore, 455 associations at borderline to genome-wide significance (p-value $<1.0 \times 10^{-06}$ and $>5.0 \times 10^{-08}$) comprising 440 unique variants from 76 genes were observed. And 451 established associations at p-values of suggestive evidence of association ($>1\times10^{-06}$ and ≤0.05) comprising 319 unique variants from 181 genes were found. These established associations were reported in the GWAS Catalog for the 313 traits related to the 13 study-specific metabolic traits.

Of the 70 unique variants at a genome-wide significant p-value, trait-wise, 63 were associated with HDL-C, 1 with TGs, 1 with LDL-C, 3 with SBP, and 2 with each of FPG and DBP. Trait-specific association of 9 indexed variants of these 70 variants is listed in **Table 18**. Of the 70 variants, 15 were established risk variants and 55 were novel risk variants resulting from seven genomic regions. Fine-mapping analysis of these seven genomic regions revealed 95% credible causal variants for nine SNP-trait association signals, corresponding to eight lead SNPs. Table 3 of study V presents a detailed list of 95% credible causal variants and the estimated regional SNP heritability variance along with SNP association statistics.

Moreover, the established variants observed were classified as direct (83 variants from 42 genes), indirect (164 variants from 77), or broad (203 variants from 131 genes) by comparing the metabolic traits associated in our cohort with those associated in the GWAS Catalog. **Figure 16** describes the count of associated variants found on mapping with 313 traits relating to the 13 study-specific metabolic traits with NHGRI-EBI GWAS catalog into three categories such as direct (association with the same trait), indirect (association with related trait class), and broad (association with any of the metabolic trait class). Overall, 349 unique variants from 187 genes known in the global population were confirmed in Arab population from direct, indirect, and

broad associations. **Figure 17** illustrates the distribution of SNP counts from "direct" relation in global populations with respect to specific study traits.

We further estimated the sample size required for replicating (at borderline and suggestive p-values) and non-replicating variants using their effect size observed from the study. The trend of effect size between these two sets of variants is shown in Figure 3 of study V. Results of this analysis suggest that replicating variants would attain genome-wide significant p-values ($5.0 \times 10^{-8}$) for most of the traits at 80% power and even more power with a sample size of 10,000. However, non-replicating variants would be less likely to replicate at genome-wide significant p-values, even at a sample size of 20,000.

Thereafter, 821 variants were prioritized based on the potential functional signatures in their vicinity. Majority of these variants were situated proximal to the transcription start site (TSS). Some interesting variants gleaned from their functional signatures were (i) the rs12740374 [CELSR2] was seen associated with non-HDL-C, LDL-C, and TC at p-values borderline to genome-wide significance carried a high score for DNase hypersensitive sites (DHSs), DNase footprint, and transcription factor binding sites (TFBS) (ii) rs3749147 [GPN1,ZNF512] were seen associated with TG at a suggestive p-value, overlapped with the promoter, located very close to the TSS and carried a high score for DHSs and TFBS (iii) rs7670 and rs4705745 of the DCP2 gene were found associated with WC at p-values borderline to genome-wide significance and carried high scores for DHSs; (iv) rs62355943 and rs72758038 [MAP3K1] were found associated with TG at p-values borderline to genome-wide significance, overlapped with the promoter & CpG shore, and carried high scores for DHSs and TFBS. The variants that were prioritized along with their functional niche are given in Supplementary Dataset S3 of research article. In addition, examination of 821 variants with GTEx expression data showed 510 variants differentially regulated the expression of 464 genes in 49 tissues, of which 385 variants upregulated 291 genes and 402 variants downregulated 294 genes. Upon considering a stringent criterion of FDR (q-value≤0.05), 62 of 821 were still found to differentially regulate 39 genes in 38 tissues, of which 34 variants upregulated 22 genes and 33 variants downregulated 19 genes. List of variants and genes that they regulate is given in Supplementary Dataset S4 of the study V research article.

EAF differences between GWAS and EAF of each subgroup were examined after LD clumping of 163 (out of 821) genotyped variants at $r^2 = 0.1$ (imputed variants were not considered for the analysis). EAF of resulting 42 variants (illustrated in **Figure 18**) evaluated by proportional test suggested that none of the variants showed significant differences between GWAS and each subgroup. Hence, association of these variants may remain significant if subgroup-specific association was performed in a larger sample dataset.

In summary, this study established the association of 66 genes with anthropometric traits, 42 genes with blood pressure traits, 25 genes with glycemic traits, and 133 genes with lipid traits at different p-value thresholds for the SNP-trait associations. Overall, 349 unique variants from 187 genes associated with metabolic traits in the global population were transferable in Arab population from direct, indirect, and broad associations. The more significant observation was

that 72 associations comprising 70 unique variants from 9 genes were seen at genome-wide significance.

| Indexed SNP_Effect allele, Chromosomal position. | Trait | Source | EAF | Effect Size | P-value | Gene, Function consequence |
|---|---|---|---|---|---|---|
| rs112861901_C, 8:19750951 | HDL-C | $OE^{imputed}$ | 0.11 | 0.2824 | 1.33E-04 | INTS10, LPL; Intergenic |
| | | $CM^{imputed}$ | 0.107 | 0.2434 | 8.54E-05 | |
| | | Meta | 0.109 | 5.4790 | 4.28E-08 | |
| rs76018028_C, 4:190161531 | HDL-C | $OE^{imputed}$ | 0.227 | -0.2082 | 3.65E-04 | LOC105377613, LOC105377614; Intergenic |
| | | $CM^{imputed}$ | 0.236 | -0.2605 | 7.69E-06 | |
| | | Meta | 0.232 | -5.7040 | 1.17E-08 | |
| rs1864163_A, 16:56997233[#] | HDL-C | $OE^{imputed}$ | 0.25 | -0.2494 | 2.63E-06 | CETP Intronic |
| | | $CM^{genotyped}$ | 0.286 | -0.2285 | 1.92E-08 | |
| | | Meta | 0.268 | -6.3650 | 1.95E-10 | |
| rs10635970_TAA, 8:19745039 | LDL-C | $OE^{imputed}$ | 0.497 | -0.1675 | 1.29E-04 | INTS10, LPL; Intergenic |
| | | $CM^{imputed}$ | 0.504 | -0.1604 | 4.31E-05 | |
| | | Meta | 0.5 | -5.6020 | 2.12E-08 | |
| rs66505542_T, 11:116623213 | TG | $OE^{imputed}$ | 0.746 | -0.1751 | 3.27E-04 | BUD13; Intronic |
| | | $CM^{imputed}$ | 0.717 | -0.1832 | 9.41E-06 | |
| | | Meta | 0.731 | -5.6920 | 1.26E-08 | |
| rs7838666_C, 8:4126701 | FPG | $OE^{imputed}$ | 0.689 | 0.236 | 3.92E-05 | CSMD1, intronic |
| | | $CM^{imputed}$ | 0.685 | 0.254 | 4.47E-06 | |
| | | Meta | 0.68 | 6.159 | 7.31E-10 | |
| rs2920844_T, 2:55341367 | DBP | $OE^{imputed}$ | 0.860 | 0.184 | 2.08E-03 | RTN4, upstream |
| | | $CM^{imputed}$ | 0.846 | 0.272 | 1.50E-06 | |
| | | Meta | 0.852 | 5.617 | 1.94E-08 | |
| rs2920844_T, 2:55341367 | SBP | $OE^{imputed}$ | 0.860 | 0.202 | 5.32E-04 | RTN4, upstream |
| | | $CM^{imputed}$ | 0.846 | 0.257 | 5.37E-06 | |
| | | Meta | 0.852 | 5.69 | 1.27E-08 | |
| rs2835788_G, 21:38906071 | SBP | $OE^{imputed}$ | 0.120 | 0.241 | 1.91E-03 | DYRK1A, LOC105372798, intergenic |
| | | $CM^{imputed}$ | 0.118 | 0.316 | 2.51E-06 | |
| | | Meta | 0.119 | 5.559 | 2.72E-08 | |

**Table 18: Summary statistics of 8 lead SNVs from 9 associations with p-value ≤ 5E-08.**

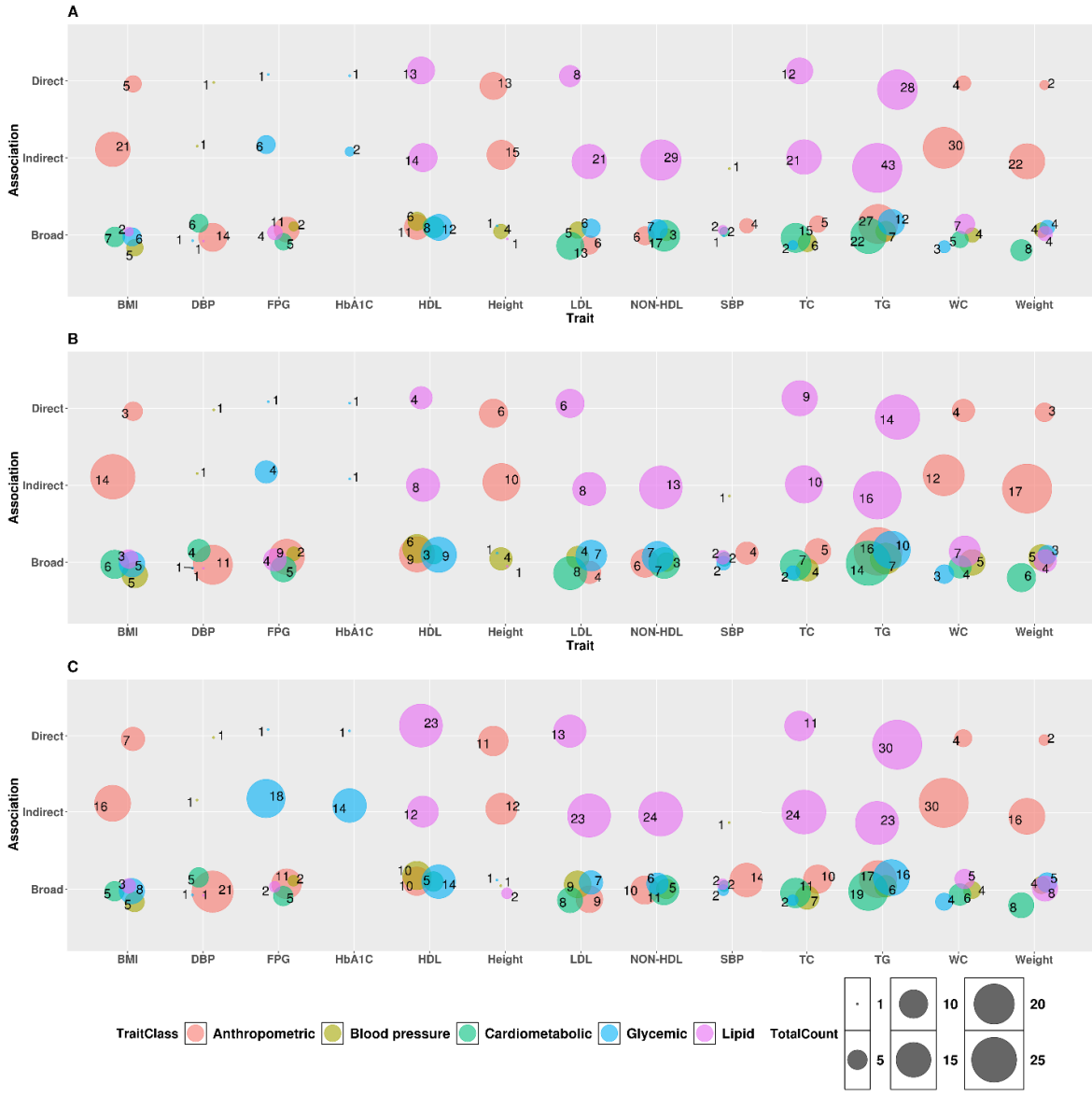


**Figure 16: Bubble plots illustrating the distributions of observed established variants (A), gene loci (B), and publications (C) reporting for the observed associations in GWAS Catalog with the study-specific metabolic traits with direct, indirect, or broad relationships.**

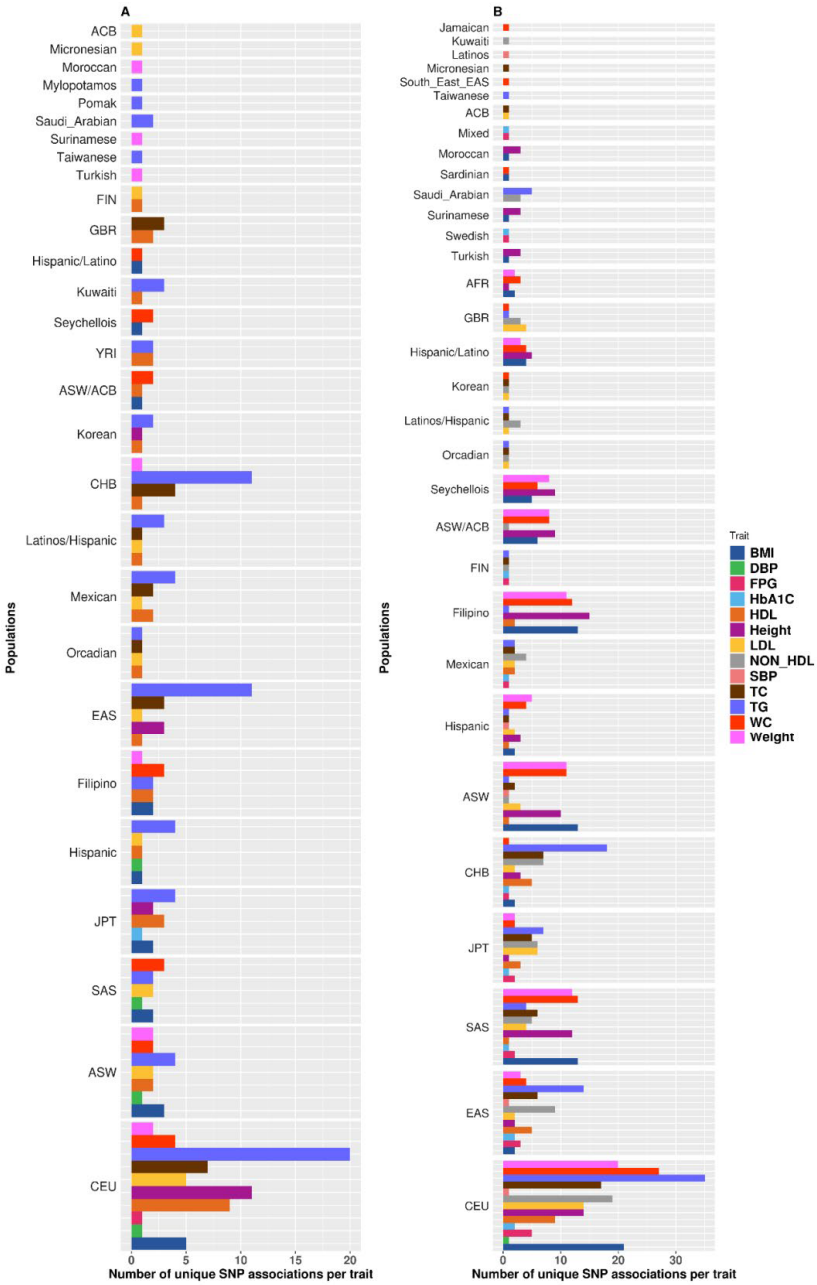


**Figure 17: Distribution of ethnic transferable SNP associations for 13 metabolic traits among global populations. A) represents SNP counts from "direct" and B) represents SNP count from "indirect" relationship.**

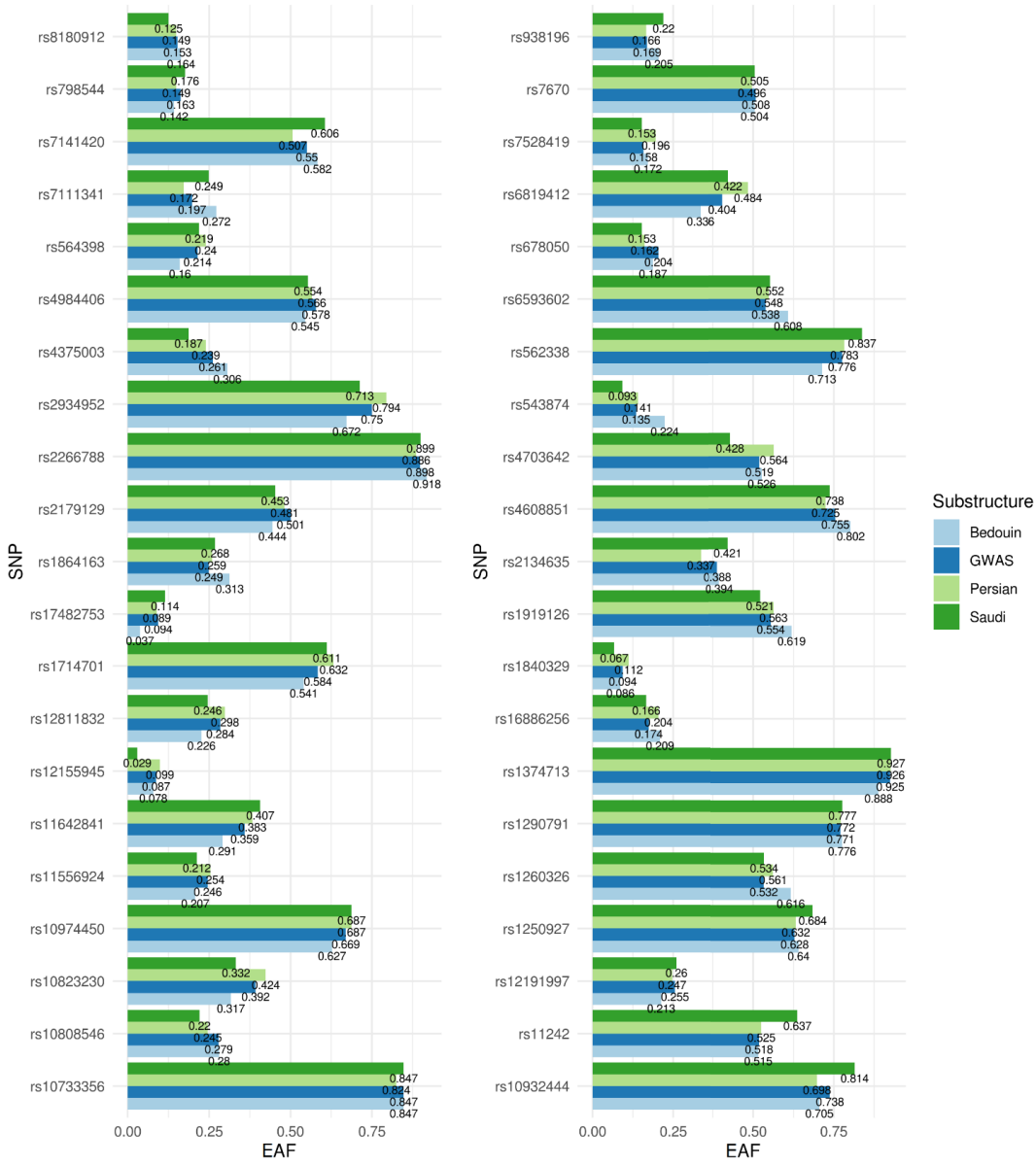


**Figure 18: Effect allele frequency distribution among subgroup population for 42 variants selected from 163 genotyped variants upon clumping at $r^2$=0.1.**

# 10. Discussion

This study is the fruit of two successful activities undertaken in DDI for studying genetic factors involved in obesity (KOGP) and T2DM (KDEP) to mitigate metabolic disorders in the Arab population. Despite high prevalence of obesity, T2DM, hypertension, dyslipidemia, CAD, and other metabolic disorders, dearth of genetic studies in such disciplines had built a void with no convincing and determined genetic risk variants in genome scale for this population. The timeliest observations from the five studies using 13 metabolic quantitative traits, variants were carefully delineated for: modulated anthropometric levels from 70 genes- 190 variants (of which 3 variants passing FDR p-value from study I, 1 variant from study II, and 186 variants from study V), lipid from 160 genes- 530 unique risk variants (of which 14 variants passed FDR p-value from study I, 16 variants from study III, and 500 variants from study V), blood pressure from 42 genes- 107 risk variants (of which 1 from study I and 106 from study V), and glycemic traits from 40 genes-59 (of which 12 from study I, 4 from study IV, and 43 from study V).

Although genome-wide variants are usually examined using additive inheritance model, due to high consanguinity in Arab population and high number of Mendelian disorders risk loci known to overlap with complex traits (Blair et al. 2013), recessive inheritance model was also used to examine variants association with 13 metabolic traits. Since rare homologous risk variants often exert severe effects on phenotype, the effect sizes of variants from association tests were validated using a power calculation (in addition to power calculation for sample size determination) to estimate the acceptable effect size (using Quanto software, which is traditionally used in candidate variant studies). Variants with effect size within the limit of the expected effect size were only considered for replication phase. Consequently, this study successfully harnessed 33 highly confident recessive signatures (either passed discovery and replication phase or BH-FDR p-value) contributing to the elevation of TG, FPG, and HbA1C levels. These variants identified at genome-wide p-value threshold using recessive model were rs1002487 [RPS6KA1] with FPG & TG (from study III and IV) and rs11805972 [LAD1], rs7761746 [OR5V1], rs39745 [CTTNBP2, LSM8], rs2934952 [PGAP3], & rs9626773 [RP11-191L9.4, CERK] with TG (from study III). Similarly at nominal p-value, rs10873925 [ST6GALNAC5], rs4663379 [SPP2, ARL4C], rs10033119 [NPY1R], rs17709449 [LINC00911, FLRT2], rs11654954 [CDK12, NEUROD2], rs9972882 [STARD3], rs7156508 [SLC10A1, SMOC1] with TG and rs487321 [CADPS] with FPG were identified. Furthermore, the variants that passed the BH-FDR test (study I variants were not replicated, instead BH-FDR test was performed) were rs12440118 [ZNF106] & rs3767494 [c1orf106] with HbA1c; rs7144734 [OTX2-AS1], rs7729384 [LOC101927421], rs3799125 [RGS17], rs747486 [ANKRD11], & rs4764409 [PIK3C2G] with FPG; and rs17501809 [PLGRKT], rs11143005 [LOC105376072], rs10860880 [IGF1], rs900543 [THSD4, NR2E3], rs12722856 [RPS3AP9, GAPDHP75], rs10014125 [LOC105377567], rs17073574 [LAMA4], rs11777524 [LY6D, GML], rs11602685 [MICAL2], chr11:45779819 [DKFZp779M0652, SLC35C1], rs17569297 [LOC105369738, LOC105369739], & rs7342999 [LOC105372082, LOC105372084] with TG.

By way of using additive model at genome-wide p-value (study-specific) rs9606756 [TCN2, I>V] with increased WC (same variant also showed association with BMI, weight, and WcHtR traits at nominal p-values), rs112861901_C [INTS10, LPL], rs76018028_C [LOC105377613, LOC105377614], & rs1864163_A [CETP] with HDL-C; rs10635970_TAA [INTS10, LPL] with LDL-C; rs66505542_T [BUD13] with TG; rs7838666_C [CSMD1] with FPG; rs2920844_T [RTN4] with DBP & SBP; and rs2835788_G [DYRK1A, LOC105372798] with SBP. Among these, variants from CETP were consistently associated with HDL-C in study I (using Cardio-MetaboChip), study III (using OmniExpress and independent replication cohort), and study V (Imputation of OmniExpress and Cardio-MetaboChip data). Similarly, variants from PGAP3 were consistently associated with TG in both recessive & additive models of study III and additive model of study V. In addition, either at borderline or suggestive p-value, 752 variants were identified from study V, of which 334 variants were trans-ethnic i.e., previously known to be associated with metabolic traits or disease in global population. Moreover, from study I, the variants that passed BH-FDR p-value were rs10005556 [GAPDHP56, LOC105377421] with weight; rs17117722 [TRA, TRD] with WcHtR; rs10005556 [GAPDHP56, LOC105377421], rs9390649 [UST] with WC; rs12722856 [RPS3AP9, GAPDHP75], rs10014125 [LOC105377567], rs17073574 [LAMA4], rs11777524 [LY6D, GML], rs11602685 [MICAL2], chr11:45779819 [DKFZp779M0652, SLC35C1], rs17569297 [LOC105369738, LOC105369739], & rs7342999 [LOC105372082, LOC105372084] with TG; rs7729384 [LOC101927421], rs3799125 [RGS17], rs747486 [ANKRD11], & rs4764409 [PIK3C2G] with FPG; and rs3767494 [C1orf106] with HbA1c.

Overall, study I (Hebbar et al. 2017b) has identified 2 risk variants from glycemic trait (rs12440118 [ZNF106, W>R], rs7144734 [OTX2-AS1, RPL3P3]) and 4 risk variants form TG (rs17501809 [PLGRKT], rs11143005 [LOC105376072], rs900543 [THSD4, NR2E3], and Chr12:101494770 [IGF1]) close to genome-wide or nominal p-value. Their relation to metabolic disease was supported by literature evidence. Interestingly, many of the identified associations were with the recessive model and of these, few associated with genetic disorders were prevalent in the Arabian Peninsula. For example, C1orf106 associated with Type 1 Diabetes (Onengut-Gumuscu et al. 2015); [OTX2-AS1,RPL3P3] with hypogonadism (Al Hayek et al. 2013; Baccetti et al. 2002; Diaczok et al. 2011); [THSD4,NR2E3] with several eye-related disorders, including enhanced S-cone syndrome (Haider et al. 2000) & Goldmann-Favre syndrome (Manayath et al. 2014); [IGF1,PAH] with diabetic retinopathy, growth retardation with deafness, & mental retardation (due to IGF1 deficiency) (Woods et al. 1996); [TRA,TRD] with immune-related disorders (Han et al. 2013; van Hagen et al. 2003); LAMA4 with diabetic nephropathy (Ewens et al. 2005); and KSR1 with diabetic vascular complications (Nemoto et al. 2012). Among them, S-cone syndrome and growth retardation with deafness & mental retardation due to IGF1 deficiency are autosomal recessive disorders. Furthermore, genes such as ZNF106, IGF1, SLC28A3, TMEM120B, and KSR1 are involved in metabolic processes related to associated traits, as shown in experimental studies (refer Table 5 of study I research article). This study also identified the larger molecular mechanisms by which gene loci operate in pathogenesis with the identification of ceramide signaling (implicated in the pathogenesis of insulin resistance and other obesity-associated

metabolic diseases), pregnenolone biosynthesis (involved in causing obesity and insulin resistance), ERK/MAPK signaling (insulin signaling pathway in controlling glucose metabolism), histidine degradation (histidine supplementation improves insulin resistance and reduces obesity), and LPS/IL-1 mediated inhibition of RXR function (involved in causing diet-induced obesity and noninsulin-dependent diabetes mellitus).

The analysis of anthropometric traits with genotyped data [of study II, (Hebbar et al. 2017a)] identified a variant rs9606756 [TCN2, I23V] at genome-wide significance its association with lower levels of HDL-C at suggestive p-value. This variant further passed the replication p-value criteria. Incidentally, interaction analysis between WC vs Apo-A1 and HDL in a randomly selected subset of replication phase showed an inverse correlation of Apo-A1 and HDL-C with obesity-related anthropometric traits (particularly with WC) at heterozygotes. Examination of eQTL data indicated the downregulation of TCN2 expression (including skin unexposed to sun) by variant rs9606756. A fascinating fact is that epidemiological studies have shown that low vitamin B12 levels were associated with obesity and being overweight (Baltaci et al. 2013; Sun et al. 2019). The lack of exposure to sunlight (due to harsh weather in the region) (Gulvady et al. 2007) and higher consumption of ultra-processed foods were significantly associated with vitamin B12 deficiency (Louzada et al. 2015). Likewise, the prevalence of vitamin B12 deficiency is common in patients with T2DM [due to standard therapy with Metformin which causes vitamin B12 deficiency (Alharbi et al. 2018; Alshammari et al. 2019; Aroda et al. 2016; de Jager et al. 2010)] and is associated with adverse lipid parameters (Adaikalakoteswari et al. 2014). However, in our study vitamin B12 levels of plasma did not show significant differences between the genotypes of rs9606756.

From study III (Hebbar et al. 2018), the analysis of genome-wide genotyped variants with lipid traits identified a set of 6 variants associated with TG at study-specific stringent genome-wide significant p-values of <6.12E-09 including rs1002487 [RPS6KA1], rs11805972 [LAD1,P>Q], rs7761746 [OR5V1], rs39745 [CTTNBP2, LSM8], rs2934952 [PGAP3], and rs9626773 [RP11-191L9.4,CERK]. Additionally, 6 variants were identified with p-values in the range of >6.12E-09 to <5E-08 with TG including rs10873925 [ST6GALNAC5], rs4663379 [SPP2, ARL4C], rs10033119 [NPY1R], rs17709449 [LINC00911, FLRT2], rs11654954 [CDK12, NEUROD2], & rs9972882 [STARD3] and 4 variants including rs7156508 [SLC10A1, SMOC1], rs3764261, rs1864163, & rs1800775 [CETP] with nominal p-value. Similarly, the analysis of FPG [from study IV (Hebbar et al. 2020a)] identified the association of 4 variants including rs1002487 [RPS6KA1] at study-specific genome-wide p-value and rs487321 [CADPS], rs707927 [VARS and VWA7] & rs12600570 [DHX58] at nominal p-value.

Implications of identified variants or genes with metabolic traits or processes or diseases were supported by previous genetic associations or literature evidence.

i. Interestingly, rs1002487 [RPS6KA1] was found to be associated with higher levels of TG, FPG, and HbA1C. Indeed, the RSK1 protein is involved in the regulation of insulin signaling and glucose metabolism in the MAPK/ERK pathway (Gao and Patel 2009). S6K1 mediates

insulin resistance and potential T2DM when there is nutrient excess (Tremblay et al. 2007; Um et al. 2006). Furthermore, rs1002487 variant from RPS6KA1 regulates DHDDS gene, which is involved in a rare genetic disorder called retinitis pigmentosa (Zuchner et al. 2011).

ii. Olfactory receptor family 5 subfamily V member 1 (OR5V1) gene is a member of a large family of G-protein-coupled receptors (GPCRs). Several studies have suggested the role of GPCRs in metabolic disorders (Recio et al. 2018; Tzameli 2016). A gene-based association study associated OR5V1 with type 1 diabetes (Qiu et al. 2014).

iii. The CERK alias ceramide kinase, an enzyme involved in the conversion of ceramide to ceramide 1-phosphate (a sphingolipid metabolite), plays an active role in glucose homeostasis, insulin signaling and ultimately, the diabetes phenotype (Holland and Summers 2008; Lipina and Hundal 2011; Straczkowski and Kowalska 2008; Summers 2010). Ceramides along with diacyl-glycerols mediate high TG and insulin resistance (Amati et al. 2011).

iv. Mutations in Ladinin 1 (LAD1) gene cause a rare form of autoimmune dermopathies [such as linear immunoglobulin A (IgA) dermatosis and dermatitis herpetiformis] with which diabetes and chronic renal failure tend to be comorbid conditions (Bhat et al. 2016; Serwin et al. 2002).

v. LSM8 is involved in retinitis pigmentosa (Schmidt-Kastner et al. 2008) and its comorbidity with diabetes has been observed in the Arab ethnic individuals (Al-Adsani and Gader 2010).

vi. PGAP3 encodes glycosylphosphatidylinositol (GPI)-specific phospholipase and is involved in the lipid remodeling of GPI. The variant rs2934952 [PGAP3] has strong LD with variants harboring from STARD3, ERBB2, GSDMA, GSDMB, ORMDL3, PNMT, PPP1R1B, RP11-690G19.3, and ZPBP2 genes. These variants were found to regulate many proximal genes including ORMDL3, a gene known to influence sphingolipids metabolism (Zhakupova et al. 2016). In addition, an established marker, rs1877031 [STARD3] associated with low levels of HDL in East Asians and Europeans, was in LD ($r^2$ =0.55) with rs9972882 [STARD3] in the study population. rs9972882 [STARD3] and rs11654954 [CDK12-NEUROD2] were in strong LD with rs2934952 [PGAP3]. STARD3 affects the regulation of macrophage cholesterol metabolism (Borthwick et al. 2009). Likewise, polymorphisms in the neuronal differentiation 2 (NEUROD2) gene affect the onset pattern of type 1 diabetes in the Japanese population (Yamada et al. 2001).

vii. SNPs close to FLRT2 were associated with age at menarche in African-American women (Demerath et al. 2013). Notably, early menarche is a risk factor for obesity and T2DM.

viii. Analysis of Iranian CAD pedigree comprising 160 individuals demonstrated that both p.Val99Met and stop-loss mutations resulted in increased sialyltransferase activity and act as risk factors for CAD (InanlooRahatloo et al. 2014). The encoded protein- Golgi type II transmembrane glycosyltransferase in ST6GALNAC5 gene catalyzes the transfer of sialic acid to cell surface proteins to modulate cell–cell interactions. Interestingly, elevated activity of sialyltransferase (from ST6GALNAC5) in blood cells and increased levels of sialic

acids are found to be associated with coronary diseases (Gopaul and Crook 2006). A microsatellite-based linkage study on Caucasian families with premature CAD and myocardial infarction showed that genes in 1p31-32 region (in which the ST6GALNAC5 is located) influence the TG level (Seidelmann et al. 2008).

ix. The members of SPP2 protein family are involved in diabetes (Jeon et al. 2011; Zhao et al. 2016) and metabolic disorders (Magnusson et al. 2013). The C-terminal of this protein binds BMP2 and TGF-β (Tian et al. 2015). Particularly, BMP2 regulates adipocyte function (Schulz and Tseng 2009) and possibly influences human insulin resistance (Zhang et al. 2010). ARL4C has been widely implicated in the cholesterol secretion pathway and intracellular vesicular transport (Engel et al. 2004; Hong et al. 2011; Wei et al. 2009) and plays a key role in lipid homeostasis (Hong and Tontonoz 2014).

x. NPY1R plays crucial roles in regulating body weight and mediating cardiovascular responses in CNS (Li et al. 2003; Pedrazzini et al. 1998). Several studies have established its involvement in metabolic disorders including obesity, pre-diabetes, and pre-hypertension (Pedrazzini 2004) by influencing food intake, psychomotor activity, and regulation of central endocrine secretion.

xi. The protein encoded by SLC10A1 gene belongs to the sodium/bile acid co-transporter family. Bile acids are the catabolic product of cholesterol metabolism and hence, this protein is important for homeostasis of cholesterol & TG levels (Watanabe et al. 2004). Bile acid is involved in the pathophysiology of obesity and T2DM (Ahmad and Haeusler 2019; Prawitt et al. 2011; Tomkin and Owens 2016).

xii. rs3764261, rs1864163, and rs1800775 [CETP] were found to be associated with lipid traits in global populations (Hiura et al. 2009; Kurano et al. 2016; Lettre et al. 2011; Sabatti et al. 2009; Spracklen et al. 2017).

xiii. CADPS regulates the recruitment of insulin granules and beta-cell function (Gandasi et al. 2017; Speidel et al. 2008). Previous global GWAS associated CADPS loci with treatment interaction of sulfonylurea (a glucose-lowering drug) and heart failure-related metabolite levels (Floyd et al. 2018; Yu et al. 2013).

xiv. VARS encodes valyl-tRNA synthetase and is associated with diabetic cataract, neurodevelopmental disorder, microcephaly, seizures, and cortical atrophy. VWA7 encodes Von Willebrand factor A domain- containing protein 7. Previous global GWAS associated the VWA7 locus with IBD, blood plasma proteome, blood protein levels, and schizophrenia. Furthermore, the risk variant and its 26 strong LD partners are from a gene-dense region, commonly referred as the HLA “class III” region (Milner and Campbell 2001) which contain several gene markers and genes from the HLA region associated with risk for type 1 diabetes (Valdes et al. 2009) and T2DM (Tuomilehto-Wolf et al. 1993).

xv. The DHX58 risk variant regulates DHX58, RAB5C, KCNH4, and HSPB9. Earlier global GWAS implicated these 4 genes in CAD (van der Harst and Verweij 2018). Variants from RAB5C were associated with fibrinogen levels, which are elevated in patients with diabetes, especially in those with foot ulcers (Li et al. 2016).

By analyzing genotype imputation data with 13 metabolic traits, study V (Hebbar et al. 2020b) harnessed 978 association signals (involving 821 variants from 251 genes). Of these, 70 variants from 9 gene loci were associated with metabolic traits at a genome-wide significance p-value. These gene loci include BUD13, CETP, CSMD1, DYRK1A, HERPUD1, INTS10, LPL, and RTN4, which are involved in metabolic processes. Specifically,

i. The variants rs10790162 (associated at borderline p-value) of BUD13 and rs4938303 (associated at suggestive p-value) of intergenic to [LOC101929011 - BUD13] associate with TG, HDL-C, and metabolic syndrome from preceding GWA studies (Kraja et al. 2011; Waterworth et al. 2010). Both variants share remarkable LD with lead variant rs66505542 of BUD13.
ii. The lead variant rs1864163 [CETP] and 17 more variants had strong LD with each other (including rs7499892, rs11508026, rs1864163, rs711752, rs1800775, rs17231506, rs821840, rs3764261, rs12149545, rs183130, rs247617, rs247616, rs173539, rs56228609, and rs56156922 at genome-wide p-value; rs1532624 and rs9939224 at borderline p-value; rs118146573 at suggestive p-value) for HDL-C, C-reactive protein levels or TG levels (pleiotropy), TC levels, WC, and CAD (Chambers et al. 2008; Ko et al. 2014; Ligthart et al. 2016; van der Harst and Verweij 2018). Importantly, study I and study III also found an association of CETP with HDL-C.
iii. The two lead variants from [INTS10, LPL], rs112861901 with HDL-C and rs10635970 with LDL-C were found to have strong LD with 17 more variants (such as rs76085257, rs7016880, rs325, rs328, rs11984636, rs13702, rs1059611, rs15285, rs10105606, rs10096633, rs17091905, rs7841189, rs17482753, rs326, rs17410962, rs17091891, and rs15285) which appeared at suggestive p-value with either HDL-C, LDL-C, TC, or non-HDL and were established for traits like hypertriglyceridemia, HDL-C, or CAD (Johansen et al. 2010; Keller et al. 2013; van der Harst and Verweij 2018; Waterworth et al. 2010).
iv. Previously, the lead variant rs76018028 [LOC105377613, LOC105377614] or any of the other 24 variants (appeared at genome-wide p-value) having strong LD with each other from the region were established for lipid traits. However, variants from LOC105377614 were reported for amyotrophic lateral sclerosis (ALS) (Keller et al. 2013). Coincidentally, abnormal lipid metabolism was a risk factor for the development of ALS (Dupuis et al. 2011).
v. Either lead variant rs7838666 [CSMD1] or other 6 variants (either at genome-wide or borderline p-values) with strong LD among them, appeared to be established for metabolic traits. However, previous studies have associated variants from CSMD1 with eating disorder (Wade et al. 2013), BMI (Zhu et al. 2020), sugar consumption (Hwang et al. 2019), T2DM (Chen et al. 2019), or HDL-C (Kulyte et al. 2020; Lettre et al. 2011).
vi. Either lead variant rs2920844 [RTN4] or other 4 variants in LD (include rs2968781, rs62138262, rs56181849, rs56377199) were found to have a previous association with metabolic traits. However, variants from RTN4 were discovered to be associated with coronary heart disease in Lithuanian families (Domarkiene et al. 2013).

vii. Either lead variant rs2835788 [DYRK1A, LOC105372798] or its LD variant rs2835799 [DYRK1A, KCNJ6] (appeared at borderline p-value) were known to have previous association with metabolic traits. Although, other variants from DYRK1A gene were associated with BMI, stroke, and metabolic syndrome (Malik et al. 2018; Zabaneh and Balding 2010; Zhu et al. 2020)

## 10.1 Limitations

The main limitation of this study was the relatively small sample size of the study cohorts; therefore, the association of variants <5% frequencies was not possible in this study. The study sample sizes had the power to establish an association of common variants only. The number of risk loci identified from a given study has a linear relationship with the size of the cohort, yet no plateau has been observed for any trait to date (Visscher et al. 2017). Hence, using small cohorts might have hindered the ability to harness more risk variants from the study population. Since large sample size is of utmost importance in GWAS to obtain true reflection of small effect size causal variants, regional collaborations with research centers to expand the sample size are in progress.

The Cardio-MetaboChip (used in study I and study V) was developed for fine mapping purposes, so it consisted only ~120,000 variants. Therefore, imputation of such spotty coverage data did not yield great SNP coverage (only 2.16 million SNPs from Cardio-MetaboChip data as supposed to 12.2 million from OmniExpress data at mean imputation quality score ≥0.5). This led to many variants not being imputed and involved in association testing.

Based on genetic heterogeneity, three genetic substructures were identified in Kuwaiti Arab population. This genetic heterogeneity warrants subgroup-specific genetic association analysis with large sample sizes, particularly for those with higher rates of inbreeding. However, our study could not perform subgroup-specific association analysis to assess risk variants for metabolic traits in each subgroup population. This was mainly because of the small number of samples in each subgroup (40% of Persian, 35% Saudi Arabian, and 25% Bedouin) which had no statistical power to assess SNP association.

A considerable number of samples (~10%) were lost due to cryptic relatedness. But, the recent advanced hybrid methods such as linear mixed models (LMM) allow the use of related samples in analysis with the construction of kinship matrix. If this alternative approach was undertaken for statistical testing of association, loss of samples due to relatedness could have been avoided.

Among the study cohorts, there were many subjects undergoing hypoglycemic and lipid lowering therapy. Although these confounding factors were considered by adjusting the association tests for medication status and performing sensitivity analysis, such individuals at risk for hyperglycemia or obesity might have also been introduced to exercise and hypocaloric diet regimes. Unfortunately, since data relating to these corrective measures were not available, such confounding factors were not considered in association analysis.

## 10.2 Future perspectives

GWAS is arguably a fishing expedition. Once variants are found with specific metabolic traits, a comprehensive analysis of various metabolic phenotypes in the context of cytokines, adipokines, insulin resistance, apo-lipoproteins, and inflammatory markers with associated variant may offer special promise for illuminating their biological basis in disease. Study design concerning diet and exercise promises the exact role of genetics in this environmental setting. The variants determined from these five studies will be further studied in a cohort of 300 (will be expanded to 1500 individuals) individuals with approximately 120 different phenotypes. Through this initiative, variants from FTO (rs1421085) (Hebbar et al. 2019a), MC4R (rs17782313) (Hammad et al. 2020), and DOCK7 (rs1748197 and rs12130333) (Alanbaei et al. 2021) have already been studied using many phenotypes and helped reveal genetic variation leading to modulation of biomolecular cascades in Arab population. For example, the studied FTO variant rs1421085, having strong LD ($R^2$> 0.90) with rs1121980, a variant found associated at suggestive p-value with WC from study V, was examined with 60 rarely studied biomarkers. This deep phenotype examination unraveled the association of rs1421085 with increased body weight, soft lean mass, and total body water through interaction with ghrelin and apolipoproteins (Hebbar et al. 2019a). Similarly, rs1748197 of DOCK7 (or ANGPTL3) having strong LD ($R^2$> 0.90) with rs1748195, a variant found associated with increased levels of TC and TG from study V, was examined in data cohort of 300 Arab individuals with 28 rarely studied biomarkers. Upon examination, rs1748197 was found to be associated with low levels of irisin and c-peptide. This illustrated ANGPTL3 as a potential link connecting lipid metabolism, insulin resistance and cardioprotection (Alanbaei et al. 2021).

The five GWA studies have enabled us to identify genetic variants associated with major metabolic diseases traits relating to obesity, T2DM, dyslipidemia, and hypertension in the study population. The general understanding is that most complex metabolic disorders (or any non-communicable disorder for that matter) are highly polygenic and manifested by hundreds or thousands of genetic variants, each having a small effect on disease risk. While genetic variants associated with a disease are precious in revealing a gene or pathway of biological relevance to the disorder, they can also be exploited to the clinical utility for the prediction of disease risk (Lewis and Vassos 2020). These identified variants can be summarized into a polygenic risk score to capture an individual's susceptibility to any metabolic diseases in question. Recently, polygenic risk scores (PRS) have been widely applied in complex metabolic research, confirming the association between PRS and disease status. Since, the generalizability of European PRS to non-European population showed reduced predictive powers due to allele frequency differences and LD patterns between populations (Duncan et al. 2019; International Schizophrenia et al. 2009; Scutari et al. 2016), studying PRS in a specific population like Arabs would benefit by population-specific early disease risk prediction.

The focus on GWAS is switching from common to rare variants, as only <10% of the genetic variance could be explained by common variants and much of the missing heritability is assumed

to be contributed from missing rare variants in most complex traits (Eichler et al. 2010; Frazer et al. 2009). In this phase of GWAS, we could assess the role of common variation with metabolic traits. However, the rare variants that are considered evolutionarily youngest forms were out of our reach due to statistically under-powered sample size. However, further efforts to discover associations driven by low-frequency (MAF 1%–5%) and rare (MAF <1%) variants are essential to identify the proportion of trait heritability explained by variants across the frequency spectrum. One of the feasible ways to assess rare variants associations given a cohort of modest sample size is variant set association test method, including the SKAT (Derkach et al. 2018; Lee et al. 2014; Wu and Pankow 2016). Unlike single variant association analysis, the variant set association analysis does not require a large sample size. Approximately, an addition of 1000 samples to the current cohort size would bolster power to our cohort to evaluate variants with MAF ≥ 0.001 (Lee et al. 2014). The recent release of trans-omics for precision medicine (TOPMED) data comprising 194,512 haplotypes has increased imputation quality and rare variant detection (Kowalski et al. 2019), which has further created opportunities for re-imputation of existing data with the intention of identifying ROH and rare variants from dense variant data. This would help identify the rare and recessive causal signatures from Arab population.

# 11. Conclusion

One of the major interests of our genome-wide genetic analysis approach was to provide breakthroughs in the understanding of molecular mechanisms involved in metabolic trait regulation in the study population. By analyzing genetic and metabolic traits data of approximately 2700 individuals with Arabian ancestry, our study has successfully pinpointed genetic determinants that modulate metabolic trait levels in Arabs. This study provided many insights into the genetic basis of 13 different quantitative traits and demonstrated the association of recessive signatures and their pleiotropic effects with the etiology of metabolic disorders, thereby attempting to deduce the molecular mechanisms through which these loci operate in complex metabolic diseases. Our findings can be used in population-specific therapeutic targets or to design risk assessment tools for diagnosing metabolic diseases in the early stage. Importantly, the present study attempted to bridge the gap in a culturally and geographically distinct population, which is under-represented in global genome survey studies.

## 12. Acknowledgements

This work would not have been possible without an excellent infrastructure facilitated by Dasman Diabetes Institute (DDI) and institutional funding of Kuwait Foundation for Advancement of Sciences (KFAS). From setting up the required state of the art high-throughput instruments to building manpower, to conduct present day research activities especially in Middle East is not an easy task; DDI has put up tremendous and commendable efforts. It would be an injustice to Dr. Kazem Behbehani (the former Director General at DDI) if I forget to thank him for his visionary roles in making DDI as a highly competent organization.

Foremost, I express my sincere gratitude to my supervisors Prof. Jaakko Tuomilehto (a former Chief Scientific Officer at DDI), Adj. Prof. Janne Pitkaniemi and Dr. Alphonse Thanaraj Thangavel, who kindly created an incredible opportunity for me to step into Doctoral Programme in Population Health of this prestigious University of Helsinki. Their patience, guidance, and motivation has helped me in the times of research and writing this thesis. I whole-heartedly appreciate their great advice for my study as it proved to be an immense help towards the success of this study. Besides my supervisors, I owe special thanks to my thesis committee members, Prof. Markus Perola, Finnish Institute for Health and Welfare, Finland and Dr. Kumaraswamy Thangaraj, CCMB, India, aside from engaging with their own batch of Ph. D students, clinic, academic and research works, they always managed to find time to evaluate annual research progress; their recommendations were always constructive and motivating.

I am immensely grateful to the reviewers of this thesis namely, Prof. Johannes Kettunen, University of Oulu, Finland and Asst. Prof. Juan Rodriguez-Flores, Weill Cornell Medical College, USA, and published articles, for their constructive criticism and time. Their review contributions have always improved the research presentation and quality of the thesis & articles, and my knowledge of the field. I would like to thank Prof. Reedik Mägi, University of Tartu, Estonia and Prof. Samuli Ripatti, Institute for Molecular Medicine Finland, University of Helsinki, Finland for accepting the Opponent's and Custos roles respectively for the public examination of this thesis.

I am indebted to Dr. Osama Alsmadi and Dr. Naser Elkum for their efforts in sample collection and genotype data generation. I thank all the participants of KOGP and KDEP project, without their willingness to participate in research, this work would not have been successful. I acknowledge all the co-authors in the original articles included in this thesis for sharing their expertise and giving valuable comments for the manuscripts; Prof. Robert Sladek, Dr. Mohamed Abu-Farha, Dr. Jehad Ahmed Abubaker, Ms. Rasheeba Nizam, Mr. Motasem Melhem, Mr. Fadi Alkayal, Ms. Daisy Thomas, Ms. Sumi Elsa John and Mr. Arshad Channanath, whose assistance was a necessary milestone in the completion of this project.

I would like to express my gratitude to Ms. Pirjo Saastamoinen, Administrative assistant at National Institute for Health and Welfare (THL) for her unconditional support during my Ph.D tenure. I pay my regards to Dr. Maisa Mahmoud for taking up discussions on creating opportunities to pursue Ph.D for DDI employees in DDI team meetings.

My sincere thanks to Dr. Qais Al Duwairi (Director General), Prof. Fahd Al Mulla (Chief Scientific Officer), Dr. Faisal H. Al-Refaei (Chief Operating Officer) and Dr. Yousuf Alkhamis (Human Resource Manager) of DDI for securing financial support to allow me participate in many scientific course works; this in turn helped me to earn required course credits for completion of Ph.D.

My special thanks to Dr. Veerendra Gadekar, University of Copenhagen, Denmark and Dr. Vishnu Modur, Cincinnati Children's Hospital Medical Center, USA for their assistance in final editing and proofreading of the thesis.

I would finally like to thank my parents and teachers who educated and supported me spiritually all throughout my life. I wish to acknowledge the support and great love of my family, my wife-Sumana; my little daughters Vidhathri and Vaidehi. They kept me going on and this work would not have been possible without their support.

# 13. References


Abecasis GR, Cardon LR, Cookson WO (2000) A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66: 279-92. doi: 10.1086/302698

Abu-Amero KK, Gonzalez AM, Larruga JM, Bosley TM, Cabrera VM (2007) Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. BMC Evol Biol 7: 32. doi: 10.1186/1471-2148-7-32

Abujbara M, Batieha A, Khader Y, Jaddou H, El-Khateeb M, Ajlouni K (2018) The Prevalence of Dyslipidemia among Jordanians. J Lipids 2018: 6298739. doi: 10.1155/2018/6298739

Adaikalakoteswari A, Jayashri R, Sukumar N, Venkataraman H, Pradeepa R, Gokulakrishnan K, Anjana RM, McTernan PG, Tripathi G, Patel V, Kumar S, Mohan V, Saravanan P (2014) Vitamin B12 deficiency is associated with adverse lipid profile in Europeans and Indians with type 2 diabetes. Cardiovasc Diabetol 13: 129. doi: 10.1186/s12933-014-0129-4

Adams TD, Hunt SC, Mason LA, Ramirez ME, Fisher AG, Williams RR (1993) Familial aggregation of morbid obesity. Obes Res 1: 261-70. doi: 10.1002/j.1550-8528.1993.tb00620.x

Afzal M (1988) Consequences of consanguinity on cognitive behavior. Behav Genet 18: 583-94. doi: 10.1007/BF01082310

Agler CS, Shungin D, Ferreira Zandona AG, Schmadeke P, Basta PV, Luo J, Cantrell J, Pahel TD, Jr., Meyer BD, Shaffer JR, Schaefer AS, North KE, Divaris K (2019) Protocols, Methods, and Tools for Genome-Wide Association Studies (GWAS) of Dental Traits. Methods Mol Biol 1922: 493-509. doi: 10.1007/978-1-4939-9012-2_38

Ahmad TR, Haeusler RA (2019) Bile acids in glucose metabolism and insulin signalling - mechanisms and research needs. Nat Rev Endocrinol 15: 701-712. doi: 10.1038/s41574-019-0266-7

Akiyama M, Okada Y, Kanai M, Takahashi A, Momozawa Y, Ikeda M, Iwata N, Ikegawa S, Hirata M, Matsuda K, Iwasaki M, Yamaji T, Sawada N, Hachiya T, Tanno K, Shimizu A, Hozawa A, Minegishi N, Tsugane S, Yamamoto M, Kubo M, Kamatani Y (2017) Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. Nat Genet 49: 1458-1467. doi: 10.1038/ng.3951

Akl C, Akik C, Ghattas H, Obermeyer CM (2017) Gender disparities in midlife hypertension: a review of the evidence on the Arab region. Womens Midlife Health 3: 1. doi: 10.1186/s40695-017-0020-z

Al-Adsani A, Gader FA (2010) Combined occurrence of diabetes mellitus and retinitis pigmentosa. Ann Saudi Med 30: 70-5. doi: 10.4103/0256-4947.59381

Al-Awadi SA, Moussa MA, Naguib KK, Farag TI, Teebi AS, el-Khalifa M, el-Dossary L (1985) Consanguinity among the Kuwaiti population. Clin Genet 27: 483-6.

Al-Daghri NM (2010) Extremely high prevalence of metabolic syndrome manifestations among Arab youth: a call for early intervention. Eur J Clin Invest 40: 1063-6. doi: 10.1111/j.1365-2362.2010.02341.x

Al-Daghri NM, Alkharfy KM, Al-Attas OS, Krishnaswamy S, Mohammed AK, Albagha OM, Alenad AM, Chrousos GP, Alokail MS (2014a) Association between type 2 diabetes mellitus-related SNP variants and obesity traits in a Saudi population. Mol Biol Rep 41: 1731-40. doi: 10.1007/s11033-014-3022-z

Al-Daghri NM, Alkharfy KM, Alokail MS, Alenad AM, Al-Attas OS, Mohammed AK, Sabico S, Albagha OM (2014b) Assessing the contribution of 38 genetic loci to the risk of type 2 diabetes in the Saudi Arabian Population. Clin Endocrinol (Oxf) 80: 532-7. doi: 10.1111/cen.12187

Al-Gazali L, Hamamy H (2014) Consanguinity and dysmorphology in Arabs. Hum Hered 77: 93-107. doi: 10.1159/000360421

Al-Gazali L, Hamamy H, Al-Arrayad S (2006) Genetic disorders in the Arab world. BMJ 333: 831-4. doi: 10.1136/bmj.38982.704931.AE

al-Gazali LI, Bener A, Abdulrazzaq YM, Micallef R, al-Khayat AI, Gaber T (1997) Consanguineous marriages in the United Arab Emirates. J Biosoc Sci 29: 491-7. doi: 10.1017/s0021932097004914
Al-Lawati JA, Mohammed AJ, Al-Hinai HQ, Jousilahti P (2003) Prevalence of the metabolic syndrome among Omani adults. Diabetes Care 26: 1781-5. doi: 10.2337/diacare.26.6.1781
Al-Lawati JA, Panduranga P, Al-Shaikh HA, Morsi M, Mohsin N, Khandekar RB, Al-Lawati HJ, Bayoumi RA (2015) Epidemiology of Diabetes Mellitus in Oman: Results from two decades of research. Sultan Qaboos Univ Med J 15: e226-33.
Al-Nozha M, Al-Khadra A, Arafah MR, Al-Maatouq MA, Khalil MZ, Khan NB, Al-Mazrou YY, Al-Marzouki K, Al-Harthi SS, Abdullah M, Al-Shahid MS, Al-Mobeireek A, Nouh MS (2005) Metabolic syndrome in Saudi Arabia. Saudi Med J 26: 1918-25.
Al-Nozha MM, Abdullah M, Arafah MR, Khalil MZ, Khan NB, Al-Mazrou YY, Al-Maatouq MA, Al-Marzouki K, Al-Khadra A, Nouh MS, Al-Harthi SS, Al-Shahid MS, Al-Mobeireek A (2007) Hypertension in Saudi Arabia. Saudi Med J 28: 77-84.
Al-Quwaidhi AJ, Pearce MS, Critchley JA, Sobngwi E, O'Flaherty M (2014) Trends and future projections of the prevalence of adult obesity in Saudi Arabia, 1992-2022. East Mediterr Health J 20: 589-95.
Al-Riyami A, Ebrahim GJ (2003) Genetic Blood Disorders Survey in the Sultanate of Oman. J Trop Pediatr 49 Suppl 1: i1-20.
Al-Safar H, Hassoun A, Almazrouei S, Kamal W, Afandi B, Rais N (2015) Association of the Genetic Polymorphisms in Transcription Factor 7-Like 2 and Peroxisome Proliferator-Activated Receptors- gamma 2 with Type 2 Diabetes Mellitus and is Interaction with Obesity Status in Emirati Population. J Diabetes Res 2015: 129695. doi: 10.1155/2015/129695
Al-Sendi AM, Shetty P, Musaiger AO (2003) Prevalence of overweight and obesity among Bahraini adolescents: a comparison between three different sets of criteria. Eur J Clin Nutr 57: 471-4. doi: 10.1038/sj.ejcn.1601560
Al-Sinani S, Al-Shafaee M, Al-Mamari A, Woodhouse N, Al-Shafie O, Hassan M, Al-Yahyaee S, Albarwani S, Jaju D, Al-Hashmi K, Al-Abri M, Rizvi S, Bayoumi R (2014a) Familial Clustering of Type 2 Diabetes among Omanis. Oman Med J 29: 51-4. doi: 10.5001/omj.2014.11
Al-Sinani S, Hassan MO, Zadjali F, Al-Yahyaee S, Albarwani S, Rizvi S, Jaju D, Comuzzie A, Voruganti VS, Bayoumi R (2014b) Utility of large consanguineous family-based model for investigating the genetics of type 2 diabetes mellitus. Gene 548: 22-8. doi: 10.1016/j.gene.2014.06.053
Al-Sinani S, Woodhouse N, Al-Mamari A, Al-Shafie O, Al-Shafaee M, Al-Yahyaee S, Hassan M, Jaju D, Al-Hashmi K, Al-Abri M, Al-Rassadi K, Rizvi S, Loic Y, Froguel P, Bayoumi R (2015) Association of gene variants with susceptibility to type 2 diabetes among Omanis. World J Diabetes 6: 358-66. doi: 10.4239/wjd.v6.i2.358
Al-Thani M, Al-Thani AA, Al-Mahdi N, Al-Kareem H, Barakat D, Al-Chetachi W, Tawfik A, Akram H (2017) An Overview of Food Patterns and Diet Quality in Qatar: Findings from the National Household Income Expenditure Survey. Cureus 9: e1249. doi: 10.7759/cureus.1249
Al Hayek AA, Khader YS, Jafal S, Khawaja N, Robert AA, Ajlouni K (2013) Prevalence of low testosterone levels in men with type 2 diabetes mellitus: a cross-sectional study. J Family Community Med 20: 179-86. doi: 10.4103/2230-8229.122006
Al Riyami AA, Afifi MM (2002) Hypertension in Oman: distribution and correlates. J Egypt Public Health Assoc 77: 383-407.
Al Safar HS, Cordell HJ, Jafer O, Anderson D, Jamieson SE, Fakiola M, Khazanehdari K, Tay GK, Blackwell JM (2013) A genome-wide search for type 2 diabetes susceptibility genes in an extended Arab family. Ann Hum Genet 77: 488-503. doi: 10.1111/ahg.12036
Al Sifri SN, Almahmeed W, Azar S, Okkeh O, Bramlage P, Junger C, Halawa I, Ambegaonkar B, Wajih S, Brudi P (2014) Results of the Dyslipidemia International Study (DYSIS)-Middle East: clinical

perspective on the prevalence and characteristics of lipid abnormalities in the setting of chronic statin treatment. PLoS One 9: e84350. doi: 10.1371/journal.pone.0084350

Al Thani M, Al Thani AA, Al-Chetachi W, Al Malki B, Khalifa SAH, Bakri AH, Hwalla N, Naja F, Nasreddine L (2018) Adherence to the Qatar dietary guidelines: a cross-sectional study of the gaps, determinants and association with cardiometabolic risk amongst adults. BMC Public Health 18: 503. doi: 10.1186/s12889-018-5400-2

Al Zenki S, Al Omirah H, Al Hooti S, Al Hamad N, Jackson RT, Rao A, Al Jahmah N, Al Obaid I, Al Ghanim J, Al Somaie M, Zaghloul S, Al Othman A (2012) High prevalence of metabolic syndrome among Kuwaiti adults--a wake-up call for public health intervention. Int J Environ Res Public Health 9: 1984-96. doi: 10.3390/ijerph9051984

Alanbaei M, Abu-Farha M, Hebbar P, Melhem M, Chandy BS, Anoop E, Cherian P, Al-Khairi I, Alkayal F, Al-Mulla F, Abubaker J, Thanaraj TA (2021) ANGPTL3 Variants Associate with Lower Levels of Irisin and C-Peptide in a Cohort of Arab Individuals. Genes (Basel) 12. doi: 10.3390/genes12050755

Albert Hourani MR (2010) A History of the Arab Peoples: With a New Afterword. Harvard University Press

Alharbi KK, Richardson TG, Khan IA, Syed R, Mohammed AK, Boustred CR, Gaunt TR, Tamimi W, Al-Daghri NM, Day IN (2014) Influence of adiposity-related genetic markers in a population of saudi arabians where other variables influencing obesity may be reduced. Dis Markers 2014: 758232. doi: 10.1155/2014/758232

Alharbi TJ, Tourkmani AM, Abdelhay O, Alkhashan HI, Al-Asmari AK, Bin Rsheed AM, Abuhaimed SN, Mohammed N, AlRasheed AN, AlHarbi NG (2018) The association of metformin use with vitamin B12 deficiency and peripheral neuropathy in Saudi individuals with type 2 diabetes mellitus. PLoS One 13: e0204420. doi: 10.1371/journal.pone.0204420

Aljohani NJ (2014) Metabolic syndrome: Risk factors among adults in Kingdom of Saudi Arabia. J Family Community Med 21: 170-5. doi: 10.4103/2230-8229.142971

Alkandari A, Longenecker JC, Barengo NC, Alkhatib A, Weiderpass E, Al-Wotayan R, Al Duwairi Q, Tuomilehto J (2018) The prevalence of pre-diabetes and diabetes in the Kuwaiti adult population in 2014. Diabetes Res Clin Pract 144: 213-223. doi: 10.1016/j.diabres.2018.08.016

Almawi WY, Nemr R, Keleshian SH, Echtay A, Saldanha FL, AlDoseri FA, Racoubian E (2013) A replication study of 19 GWAS-validated type 2 diabetes at-risk variants in the Lebanese population. Diabetes Res Clin Pract 102: 117-22. doi: 10.1016/j.diabres.2013.09.001

AlSafar HS, Al-Ali M, Elbait GD, Al-Maini MH, Ruta D, Peramo B, Henschel A, Tay GK (2019) Introducing the first whole genomes of nationals from the United Arab Emirates. Sci Rep 9: 14725. doi: 10.1038/s41598-019-50876-9

Alshammari AN, Iqbal R, Baksh IP (2019) Vitamin B12 deficiency and the knowledge and practice of physicians regarding screening for vitamin B12 deficiency among type 2 diabetic patients on metformin in selected hospitals in Riyadh, Saudi Arabia. J Family Med Prim Care 8: 2306-2311. doi: 10.4103/jfmpc.jfmpc_416_19

Alsmadi O, John SE, Thareja G, Hebbar P, Antony D, Behbehani K, Thanaraj TA (2014a) Genome at juncture of early human migration: a systematic analysis of two whole genomes and thirteen exomes from Kuwaiti population subgroup of inferred Saudi Arabian tribe ancestry. PLoS One 9: e99069. doi: 10.1371/journal.pone.0099069

Alsmadi O, Melhem M, Hebbar P, Thareja G, John SE, Alkayal F, Behbehani K, Thanaraj TA (2014b) Leptin in association with common variants of MC3R mediates hypertension. Am J Hypertens 27: 973-81. doi: 10.1093/ajh/hpt285

Alsmadi O, Thareja G, Alkayal F, Rajagopalan R, John SE, Hebbar P, Behbehani K, Thanaraj TA (2013) Genetic substructure of Kuwaiti population reveals migration history. PLoS One 8: e74913. doi: 10.1371/journal.pone.0074913

Alswat K, Aljumah AA, Sanai FM, Abaalkhail F, Alghamdi M, Al Hamoudi WK, Al Khathlan A, Al Quraishi H, Al Rifai A, Al Zaabi M, Babatin MA, Estes C, Hashim A, Razavi H (2018) Nonalcoholic fatty liver disease burden - Saudi Arabia and United Arab Emirates, 2017-2030. Saudi J Gastroenterol 24: 211-219. doi: 10.4103/sjg.SJG_122_18

Alzaabi A, Al-Kaabi J, Al-Maskari F, Farhood AF, Ahmed LA (2019) Prevalence of diabetes and cardio-metabolic risk factors in young men in the United Arab Emirates: A cross-sectional national survey. Endocrinol Diabetes Metab 2: e00081. doi: 10.1002/edm2.81

Amati F, Dube JJ, Alvarez-Carnero E, Edreira MM, Chomentowski P, Coen PM, Switzer GE, Bickel PE, Stefanovic-Racic M, Toledo FG, Goodpaster BH (2011) Skeletal muscle triglycerides, diacylglycerols, and ceramides in insulin resistance: another paradox in endurance-trained athletes? Diabetes 60: 2588-97. doi: 10.2337/db10-1221

Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N, Bagger FO, Jorgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, Baillie JK, Ishizu Y, Shimizu Y, Furuhata E, Maeda S, Negishi Y, Mungall CJ, Meehan TF, Lassmann T, Itoh M, Kawaji H, Kondo N, Kawai J, Lennartsson A, Daub CO, Heutink P, Hume DA, Jensen TH, Suzuki H, Hayashizaki Y, Muller F, Forrest ARR, Carninci P, Rehli M, Sandelin A (2014) An atlas of active enhancers across human cell types and tissues. Nature 507: 455-461. doi: 10.1038/nature12787

Arfa I, Abid A, Malouche D, Ben Alaya N, Azegue TR, Mannai I, Zorgati MM, Ben Rayana MC, Ben Ammar S, Blousa-Chabchoub S, Ben Romdhane H, Zouari B, Dellagi MK, Abdelhak S (2007) Familial aggregation and excess maternal transmission of type 2 diabetes in Tunisia. Postgrad Med J 83: 348-51. doi: 10.1136/pgmj.2006.053744

Armitage SJ, Jasim SA, Marks AE, Parker AG, Usik VI, Uerpmann HP (2011) The southern route "out of Africa": evidence for an early expansion of modern humans into Arabia. Science 331: 453-6. doi: 10.1126/science.1199113

Aroda VR, Edelstein SL, Goldberg RB, Knowler WC, Marcovina SM, Orchard TJ, Bray GA, Schade DS, Temprosa MG, White NH, Crandall JP, Diabetes Prevention Program Research G (2016) Long-term Metformin Use and Vitamin B12 Deficiency in the Diabetes Prevention Program Outcomes Study. J Clin Endocrinol Metab 101: 1754-61. doi: 10.1210/jc.2015-3754

Arthur Goldschmidt Jr. LD (2006) A concise History of the Middle East. Westview Press

Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, Mead D, Bouman H, Riveros-Mckay F, Kostadima MA, Lambourne JJ, Sivapalaratnam S, Downes K, Kundu K, Bomba L, Berentsen K, Bradley JR, Daugherty LC, Delaneau O, Freson K, Garner SF, Grassi L, Guerrero J, Haimel M, Janssen-Megens EM, Kaan A, Kamat M, Kim B, Mandoli A, Marchini J, Martens JHA, Meacham S, Megy K, O'Connell J, Petersen R, Sharifi N, Sheard SM, Staley JR, Tuna S, van der Ent M, Walter K, Wang SY, Wheeler E, Wilder SP, Iotchkova V, Moore C, Sambrook J, Stunnenberg HG, Di Angelantonio E, Kaptoge S, Kuijpers TW, Carrillo-de-Santa-Pau E, Juan D, Rico D, Valencia A, Chen L, Ge B, Vasquez L, Kwan T, Garrido-Martin D, Watt S, Yang Y, Guigo R, Beck S, Paul DS, Pastinen T, Bujold D, Bourque G, Frontini M, Danesh J, Roberts DJ, Ouwehand WH, Butterworth AS, Soranzo N (2016) The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. Cell 167: 1415-1429 e19. doi: 10.1016/j.cell.2016.10.042

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. Bioinformatics 23: 1294-6. doi: 10.1093/bioinformatics/btm108

Baccetti B, La Marca A, Piomboni P, Capitani S, Bruni E, Petraglia F, De Leo V (2002) Insulin-dependent diabetes in men is associated with hypothalamo-pituitary derangement and with impairment in semen quality. Hum Reprod 17: 2673-7. doi: 10.1093/humrep/17.10.2673

Badaruddoza, Afzal M (1993) Inbreeding depression and intelligence quotient among north Indian children. Behav Genet 23: 343-7. doi: 10.1007/BF01067435

Badaruddoza, Afzal M, Akhtaruzzaman (1994) Inbreeding and congenital heart diseases in a north Indian population. Clin Genet 45: 288-91. doi: 10.1111/j.1399-0004.1994.tb04032.x

Bahijri SM, Jambi HA, Al Raddadi RM, Ferns G, Tuomilehto J (2016) The Prevalence of Diabetes and Prediabetes in the Adult Population of Jeddah, Saudi Arabia--A Community-Based Survey. PLoS One 11: e0152559. doi: 10.1371/journal.pone.0152559

Bai H, Baik N, Kiosses WB, Krajewski S, Miles LA, Parmer RJ (2011) The novel plasminogen receptor, plasminogen receptor(KT) (Plg-R(KT)), regulates catecholamine release. J Biol Chem 286: 33125-33. doi: 10.1074/jbc.M111.218693

Bakhashab S, Filimban N, Altall RM, Nassir R, Qusti SY, Alqahtani MH, Abuzenadah AM, Dallol A (2020) The Effect Sizes of PPARgamma rs1801282, FTO rs9939609, and MC4R rs2229616 Variants on Type 2 Diabetes Mellitus Risk among the Western Saudi Population: A Cross-Sectional Prospective Study. Genes (Basel) 11. doi: 10.3390/genes11010098

Baltaci D, Kutlucan A, Turker Y, Yilmaz A, Karacam S, Deler H, Ucgun T, Kara IH (2013) Association of vitamin B12 with obesity, overweight, insulin resistance and metabolic syndrome, and body fat composition; primary care-based study. Med Glas (Zenica) 10: 203-10.

Barakat H (1993) The Arab World: Society, Culture, and State. University of California Press

Beasley TM, Erickson S, Allison DB (2009) Rank-based inverse normal transformations are increasingly used, but are they merited? Behav Genet 39: 580-95. doi: 10.1007/s10519-009-9281-0

Bener A, Al-Suwaidi J, Al-Jaber K, Al-Marri S, Dagash MH, Elbagi IE (2004) The prevalence of hypertension and its associated risk factors in a newly developed country. Saudi Med J 25: 918-22.

Bener A, Hussain R (2006) Consanguineous unions and child health in the State of Qatar. Paediatr Perinat Epidemiol 20: 372-8. doi: 10.1111/j.1365-3016.2006.00750.x

Bener A, Yousafzai MT, Al-Hamaq AO, Mohammad AG, Defronzo RA (2013) Parental transmission of type 2 diabetes mellitus in a highly endogamous population. World J Diabetes 4: 40-6. doi: 10.4239/wjd.v4.i2.40

Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M (2016) FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics 32: 1493-501. doi: 10.1093/bioinformatics/btw018

Benrahma H, Arfa I, Charif M, Bounaceur S, Eloualid A, Boulouiz R, Nahili H, Abidi O, Rouba H, Chadli A, Oudghiri M, Farouqui A, Abdelhak S, Barakat A (2011) Maternal effect and familial aggregation in a type 2 diabetic Moroccan population. J Community Health 36: 943-8. doi: 10.1007/s10900-011-9393-3

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS, Thomson JA (2010) The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol 28: 1045-8. doi: 10.1038/nbt1010-1045

Bhat ZY, Abu Minshar M, Imran N, Thompson A, Malik YO (2016) Bullous Dermatosis in an End-Stage Renal Disease Patient: A Case Report and Literature Review. Case Rep Nephrol 2016: 6713807. doi: 10.1155/2016/6713807

Bishop DT, Williamson JA (1990) The power of identity-by-state methods for linkage analysis. Am J Hum Genet 46: 254-65.

Bittles AH, Black ML (2010) Evolution in health and medicine Sackler colloquium: Consanguinity, human evolution, and complex diseases. Proc Natl Acad Sci U S A 107 Suppl 1: 1779-86. doi: 10.1073/pnas.0906079106

Blair DR, Lyttle CS, Mortensen JM, Bearden CF, Jensen AB, Khiabanian H, Melamed R, Rabadan R, Bernstam EV, Brunak S, Jensen LJ, Nicolae D, Shah NH, Grossman RL, Cox NJ, White KP, Rzhetsky A (2013) A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk. Cell 155: 70-80. doi: 10.1016/j.cell.2013.08.030

Bo-Abbas Y, Acierno JS, Jr., Shagoury JK, Crowley WF, Jr., Seminara SB (2003) Autosomal recessive idiopathic hypogonadotropic hypogonadism: genetic analysis excludes mutations in the gonadotropin-releasing hormone (GnRH) and GnRH receptor genes. J Clin Endocrinol Metab 88: 2730-7. doi: 10.1210/jc.2002-021948

Boodai SA, Cherry LM, Sattar NA, Reilly JJ (2014) Prevalence of cardiometabolic risk factors and metabolic syndrome in obese Kuwaiti adolescents. Diabetes Metab Syndr Obes 7: 505-11. doi: 10.2147/DMSO.S66156

Borthwick F, Taylor JM, Bartholomew C, Graham A (2009) Differential regulation of the STARD1 subfamily of START lipid trafficking proteins in human macrophages. FEBS Lett 583: 1147-53. doi: 10.1016/j.febslet.2009.02.042

Brown EE, Sturm AC, Cuchel M, Braun LT, Duell PB, Underberg JA, Jacobson TA, Hegele RA (2020) Genetic testing in dyslipidemia: A scientific statement from the National Lipid Association. J Clin Lipidol 14: 398-413. doi: 10.1016/j.jacl.2020.04.011

Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, ReproGen C, Psychiatric Genomics C, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control C, Duncan L, Perry JR, Patterson N, Robinson EB, Daly MJ, Price AL, Neale BM (2015) An atlas of genetic correlations across human diseases and traits. Nat Genet 47: 1236-41. doi: 10.1038/ng.3406

Cabrera VM, Marrero P, Abu-Amero KK, Larruga JM (2018) Carriers of mitochondrial DNA macrohaplogroup L3 basal lineages migrated back to Africa from Asia around 70,000 years ago. BMC Evol Biol 18: 98. doi: 10.1186/s12862-018-1211-4

Cappai G, Songini M, Doria A, Cavallerano JD, Lorenzi M (2011) Increased prevalence of proliferative retinopathy in patients with type 1 diabetes who are deficient in glucose-6-phosphate dehydrogenase. Diabetologia 54: 1539-42. doi: 10.1007/s00125-011-2099-3

Cauchi S, Ezzidi I, El Achhab Y, Mtiraoui N, Chaieb L, Salah D, Nejjari C, Labrune Y, Yengo L, Beury D, Vaxillaire M, Mahjoub T, Chikri M, Froguel P (2012) European genetic variants associated with type 2 diabetes in North African Arabs. Diabetes Metab 38: 316-23. doi: 10.1016/j.diabet.2012.02.003

Centers for Disease C, Prevention (2007) Prevalence of regular physical activity among adults--United States, 2001 and 2005. MMWR Morb Mortal Wkly Rep 56: 1209-12.

Cerf ME (2013) Beta cell dysfunction and insulin resistance. Front Endocrinol (Lausanne) 4: 37. doi: 10.3389/fendo.2013.00037

Chakravarthy MV, Semenkovich CF (2007) The ABCs of beta-cell dysfunction in type 2 diabetes. Nat Med 13: 241-2. doi: 10.1038/nm0307-241

Chaleby K (1985) Women of polygamous marriages in an inpatient psychiatric service in Kuwait. J Nerv Ment Dis 173: 56-8.

Chambers JC, Elliott P, Zabaneh D, Zhang W, Li Y, Froguel P, Balding D, Scott J, Kooner JS (2008) Common genetic variation near MC4R is associated with waist circumference and insulin resistance. Nat Genet 40: 716-8. doi: 10.1038/ng.156

Channanath AM, Farran B, Behbehani K, Thanaraj TA (2013) State of diabetes, hypertension, and comorbidity in Kuwait: showcasing the trends as seen in native versus expatriate populations. Diabetes Care 36: e75. doi: 10.2337/dc12-2451

Chen J, Sun M, Adeyemo A, Pirie F, Carstensen T, Pomilla C, Doumatey AP, Chen G, Young EH, Sandhu M, Morris AP, Barroso I, McCarthy MI, Mahajan A, Wheeler E, Rotimi CN, Motala AA (2019) Genome-wide association study of type 2 diabetes in Africa. Diabetologia 62: 1204-1211. doi: 10.1007/s00125-019-4880-7

Chen W, Larrabee BR, Ovsyannikova IG, Kennedy RB, Haralambieva IH, Poland GA, Schaid DJ (2015) Fine Mapping Causal Variants with an Approximate Bayesian Method Using Marginal Test Statistics. Genetics 200: 719-36. doi: 10.1534/genetics.115.176107

Church C, Lee S, Bagg EA, McTaggart JS, Deacon R, Gerken T, Lee A, Moir L, Mecinovic J, Quwailid MM, Schofield CJ, Ashcroft FM, Cox RD (2009) A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. PLoS Genet 5: e1000599. doi: 10.1371/journal.pgen.1000599

Clement K, Vaisse C, Lahlou N, Cabrol S, Pelloux V, Cassuto D, Gourmelen M, Dina C, Chambaz J, Lacorte JM, Basdevant A, Bougneres P, Lebouc Y, Froguel P, Guy-Grand B (1998) A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. Nature 392: 398-401. doi: 10.1038/32911

Coelho M, Oliveira T, Fernandes R (2013) Biochemistry of adipose tissue: an endocrine organ. Arch Med Sci 9: 191-200. doi: 10.5114/aoms.2013.33181

Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 305: 869-72. doi: 10.1126/science.1099870

Cohen JC, Pertsemlidis A, Fahmi S, Esmail S, Vega GL, Grundy SM, Hobbs HH (2006) Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. Proc Natl Acad Sci U S A 103: 1810-5. doi: 10.1073/pnas.0508483103

Consortium EP (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57-74. doi: 10.1038/nature11247

Consortium F, the RP, Clst, Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Haberle V, Lassmann T, Kulakovskiy IV, Lizio M, Itoh M, Andersson R, Mungall CJ, Meehan TF, Schmeier S, Bertin N, Jorgensen M, Dimont E, Arner E, Schmidl C, Schaefer U, Medvedeva YA, Plessy C, Vitezic M, Severin J, Semple C, Ishizu Y, Young RS, Francescatto M, Alam I, Albanese D, Altschuler GM, Arakawa T, Archer JA, Arner P, Babina M, Rennie S, Balwierz PJ, Beckhouse AG, Pradhan-Bhatt S, Blake JA, Blumenthal A, Bodega B, Bonetti A, Briggs J, Brombacher F, Burroughs AM, Califano A, Cannistraci CV, Carbajo D, Chen Y, Chierici M, Ciani Y, Clevers HC, Dalla E, Davis CA, Detmar M, Diehl AD, Dohi T, Drablos F, Edge AS, Edinger M, Ekwall K, Endoh M, Enomoto H, Fagiolini M, Fairbairn L, Fang H, Farach-Carson MC, Faulkner GJ, Favorov AV, Fisher ME, Frith MC, Fujita R, Fukuda S, Furlanello C, Furino M, Furusawa J, Geijtenbeek TB, Gibson AP, Gingeras T, Goldowitz D, Gough J, Guhl S, Guler R, Gustincich S, Ha TJ, Hamaguchi M, Hara M, Harbers M, Harshbarger J, Hasegawa A, Hasegawa Y, Hashimoto T, Herlyn M, Hitchens KJ, Ho Sui SJ, Hofmann OM, et al. (2014) A promoter-level mammalian expression atlas. Nature 507: 462-70. doi: 10.1038/nature13182

Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida, Biospecimen Collection Source Site N, Biospecimen Collection Source Site R, Biospecimen Core Resource V, Brain Bank Repository-University of Miami Brain Endowment B, Leidos Biomedical-Project M, Study E, Genome Browser Data I, Visualization EBI, Genome Browser Data I, Visualization-Ucsc Genomics Institute UoCSC, Lead a, Laboratory DA, Coordinating C, management NIHp, Biospecimen c, Pathology, e QTLmwg, Battle A, Brown CD, Engelhardt BE, Montgomery SB (2017) Genetic effects on gene expression across human tissues. Nature 550: 204-213. doi: 10.1038/nature24277

Cutting GR (2010) Modifier genes in Mendelian disorders: the example of cystic fibrosis. Ann N Y Acad Sci 1214: 57-69. doi: 10.1111/j.1749-6632.2010.05879.x

Cyrus C, Al-Mueilo S, Vatte C, Chathoth S, Li YR, Qutub H, Al Ali R, Al-Muhanna F, Lanktree MB, Alkharsah KR, Al-Rubaish A, Kim-Mozeleski B, Keating B, Al Ali A (2018) Assessing known chronic kidney disease associated genetic variants in Saudi Arabian populations. BMC Nephrol 19: 88. doi: 10.1186/s12882-018-0890-9

Dajani R, Li J, Wei Z, March ME, Xia Q, Khader Y, Hakooz N, Fatahallah R, El-Khateeb M, Arafat A, Saleh T, Dajani AR, Al-Abbadi Z, Abdul Qader M, Shiyab AH, Bateiha A, Ajlouni K, Hakonarson H (2017) Genome-wide association study identifies novel type II diabetes risk loci in Jordan subpopulations. PeerJ 5: e3618. doi: 10.7717/peerj.3618

Daoulah A, Lotfi A, Al-Murayeh M, Al-Kaabi S, Al-Faifi SM, Elkhateeb OE, Alama MN, Hersi AS, Dixon CM, Ahmed W, Al-Shehri M, Youssef A, Elimam AM, Abougalambou AS, Murad W, Alsheikh-Ali AA (2017) Polygamy and Risk of Coronary Artery Disease in Men Undergoing Angiography: An Observational Study. Int J Vasc Med 2017: 1925176. doi: 10.1155/2017/1925176

Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C (2016) Next-generation genotype imputation service and methods. Nat Genet 48: 1284-1287. doi: 10.1038/ng.3656

Davey Smith G, Hemani G (2014) Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 23: R89-98. doi: 10.1093/hmg/ddu328

de Jager J, Kooy A, Lehert P, Wulffele MG, van der Kolk J, Bets D, Verburg J, Donker AJ, Stehouwer CD (2010) Long term treatment with metformin in patients with type 2 diabetes and risk of vitamin B-12 deficiency: randomised placebo controlled trial. BMJ 340: c2181. doi: 10.1136/bmj.c2181

De Sanctis V, Zurlo MG, Senesi E, Boffa C, Cavallo L, Di Gregorio F (1988) Insulin dependent diabetes in thalassaemia. Arch Dis Child 63: 58-62.

Demerath EW, Liu CT, Franceschini N, Chen G, Palmer JR, Smith EN, Chen CT, Ambrosone CB, Arnold AM, Bandera EV, Berenson GS, Bernstein L, Britton A, Cappola AR, Carlson CS, Chanock SJ, Chen W, Chen Z, Deming SL, Elks CE, Evans MK, Gajdos Z, Henderson BE, Hu JJ, Ingles S, John EM, Kerr KF, Kolonel LN, Le Marchand L, Lu X, Millikan RC, Musani SK, Nock NL, North K, Nyante S, Press MF, Rodriquez-Gil JL, Ruiz-Narvaez EA, Schork NJ, Srinivasan SR, Woods NF, Zheng W, Ziegler RG, Zonderman A, Heiss G, Gwen Windham B, Wellons M, Murray SS, Nalls M, Pastinen T, Rajkovic A, Hirschhorn J, Adrienne Cupples L, Kooperberg C, Murabito JM, Haiman CA (2013) Genome-wide association study of age at menarche in African-American women. Hum Mol Genet 22: 3329-46. doi: 10.1093/hmg/ddt181

Deng HW, Chen WM, Recker RR (2000) QTL fine mapping by measuring and testing for Hardy-Weinberg and linkage disequilibrium at a series of linked marker loci in extreme samples of populations. Am J Hum Genet 66: 1027-45. doi: 10.1086/302804

Derkach A, Zhang H, Chatterjee N (2018) Power Analysis for Genetic Association Test (PAGEANT) provides insights to challenges for rare variant association studies. Bioinformatics 34: 1506-1513. doi: 10.1093/bioinformatics/btx770

DESA-UN (2018) World Urbanization Prospects 2018. United Nations

Diaczok D, DiVall S, Matsuo I, Wondisford FE, Wolfe AM, Radovick S (2011) Deletion of Otx2 in GnRH neurons results in a mouse model of hypogonadotropic hypogonadism. Mol Endocrinol 25: 833-46. doi: 10.1210/me.2010-0271

Domarkiene I, Pranculis A, Germanas S, Jakaitiene A, Vitkus D, Dzenkeviciute V, Kucinskiene Z, Kucinskas V (2013) RTN4 and FBXL17 Genes are Associated with Coronary Heart Disease in Genome-Wide Association Analysis of Lithuanian Families. Balkan J Med Genet 16: 17-22. doi: 10.2478/bjmg-2013-0026

Duggirala R, Williams JT, Williams-Blangero S, Blangero J (1997) A variance component approach to dichotomous trait linkage analysis using a threshold model. Genet Epidemiol 14: 987-92. doi: 10.1002/(SICI)1098-2272(1997)14:6<987::AID-GEPI71>3.0.CO;2-G

Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, Peterson R, Domingue B (2019) Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun 10: 3328. doi: 10.1038/s41467-019-11112-0

Dupuis L, Pradat PF, Ludolph AC, Loeffler JP (2011) Energy metabolism in amyotrophic lateral sclerosis. Lancet Neurol 10: 75-82. doi: 10.1016/S1474-4422(10)70224-6

Edwards AW (2008) G. H. Hardy (1908) and Hardy-Weinberg equilibrium. Genetics 179: 1143-50. doi: 10.1534/genetics.104.92940

Edwards SL, Beesley J, French JD, Dunning AM (2013) Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet 93: 779-97. doi: 10.1016/j.ajhg.2013.10.012

Ehret GB, Caulfield MJ (2013) Genes for blood pressure: an opportunity to understand hypertension. Eur Heart J 34: 951-61. doi: 10.1093/eurheartj/ehs455

Ehret GB, Ferreira T, Chasman DI, Jackson AU, Schmidt EM, Johnson T, Thorleifsson G, Luan J, Donnelly LA, Kanoni S, Petersen AK, Pihur V, Strawbridge RJ, Shungin D, Hughes MF, Meirelles O, Kaakinen M, Bouatia-Naji N, Kristiansson K, Shah S, Kleber ME, Guo X, Lyytikainen LP, Fava C, Eriksson N, Nolte IM, Magnusson PK, Salfati EL, Rallidis LS, Theusch E, Smith AJP, Folkersen L, Witkowska K, Pers TH, Joehanes R, Kim SK, Lataniotis L, Jansen R, Johnson AD, Warren H, Kim YJ, Zhao W, Wu Y, Tayo BO, Bochud M, consortium CH-E, consortium C-H, Wellcome Trust Case Control C, Absher D, Adair LS, Amin N, Arking DE, Axelsson T, Baldassarre D, Balkau B, Bandinelli S, Barnes MR, Barroso I, Bevan S, Bis JC, Bjornsdottir G, Boehnke M, Boerwinkle E, Bonnycastle LL, Boomsma DI, Bornstein SR, Brown MJ, Burnier M, Cabrera CP, Chambers JC, Chang IS, Cheng CY, Chines PS, Chung RH, Collins FS, Connell JM, Doring A, Dallongeville J, Danesh J, de Faire U, Delgado G, Dominiczak AF, Doney ASF, Drenos F, Edkins S, Eicher JD, Elosua R, Enroth S, Erdmann J, Eriksson P, Esko T, Evangelou E, Evans A, Fall T, Farrall M, Felix JF, Ferrieres J, Ferrucci L, Fornage M, Forrester T, et al. (2016) The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. Nat Genet 48: 1171-1184. doi: 10.1038/ng.3667

EIA (2019) OPEC Revenues Fact Sheet-2019. U.S. Energy Information Administration

Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH (2010) Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet 11: 446-50. doi: 10.1038/nrg2809

el-Hazmi MA, al-Swailem AR, Warsy AS, al-Swailem AM, Sulaimani R, al-Meshari AA (1995) Consanguinity among the Saudi Arabian population. J Med Genet 32: 623-6. doi: 10.1136/jmg.32.8.623

El Hajj Chehadeh SW, Jelinek HF, Al Mahmeed WA, Tay GK, Odama UO, Elghazali GE, Al Safar HS (2016) Relationship between MTHFR C677T and A1298C gene polymorphisms and complications of type 2 diabetes mellitus in an Emirati population. Meta Gene 9: 70-5. doi: 10.1016/j.mgene.2016.04.002

Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, Fernandez J, Prati D, Baselli G, Asselta R, Grimsrud MM, Milani C, Aziz F, Kassens J, May S, Wendorff M, Wienbrandt L, Uellendahl-Werth F, Zheng T, Yi X, de Pablo R, Chercoles AG, Palom A, Garcia-Fernandez AE, Rodriguez-Frias F, Zanella A, Bandera A, Protti A, Aghemo A, Lleo A, Biondi A, Caballero-Garralda A, Gori A, Tanck A, Carreras Nolla A, Latiano A, Fracanzani AL, Peschuck A, Julia A, Pesenti A, Voza A, Jimenez D, Mateos B, Nafria Jimenez B, Quereda C, Paccapelo C, Gassner C, Angelini C, Cea C, Solier A, Pestana D, Muniz-Diaz E, Sandoval E, Paraboschi EM, Navas E, Garcia Sanchez F, Ceriotti F, Martinelli-Boneschi F, Peyvandi F, Blasi F, Tellez L, Blanco-Grau A, Hemmrich-Stanisak

G, Grasselli G, Costantino G, Cardamone G, Foti G, Aneli S, Kurihara H, ElAbd H, My I, Galvan-Femenia I, Martin J, Erdmann J, Ferrusquia-Acosta J, Garcia-Etxebarria K, Izquierdo-Sanchez L, Bettini LR, Sumoy L, Terranova L, Moreira L, Santoro L, Scudeller L, Mesonero F, Roade L, Ruhlemann MC, Schaefer M, Carrabba M, Riveiro-Barciela M, Figuera Basso ME, Valsecchi MG, Hernandez-Tejero M, Acosta-Herrera M, D'Angio M, Baldini M, Cazzaniga M, Schulzky M, Cecconi M, Wittig M, Ciccarelli M, et al. (2020) Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N Engl J Med. doi: 10.1056/NEJMoa2020283

Engel T, Lueken A, Bode G, Hobohm U, Lorkowski S, Schlueter B, Rust S, Cullen P, Pech M, Assmann G, Seedorf U (2004) ADP-ribosylation factor (ARF)-like 7 (ARL7) is induced by cholesterol loading and participates in apolipoprotein AI-dependent cholesterol export. FEBS Lett 566: 241-6. doi: 10.1016/j.febslet.2004.04.048

Erdmann J, Kessler T, Munoz Venegas L, Schunkert H (2018) A decade of genome-wide association studies for coronary artery disease: the challenges ahead. Cardiovasc Res 114: 1241-1257. doi: 10.1093/cvr/cvy084

Ewens KG, George RA, Sharma K, Ziyadeh FN, Spielman RS (2005) Assessment of 115 candidate genes for diabetic nephropathy by transmission/disequilibrium test. Diabetes 54: 3305-18. doi: 10.2337/diabetes.54.11.3305

Ewens WJ, Spielman RS (1995) The transmission/disequilibrium test: history, subdivision, and admixture. Am J Hum Genet 57: 455-64.

Fachal L, Aschard H, Beesley J, Barnes DR, Allen J, Kar S, Pooley KA, Dennis J, Michailidou K, Turman C, Soucy P, Lemacon A, Lush M, Tyrer JP, Ghoussaini M, Moradi Marjaneh M, Jiang X, Agata S, Aittomaki K, Alonso MR, Andrulis IL, Anton-Culver H, Antonenkova NN, Arason A, Arndt V, Aronson KJ, Arun BK, Auber B, Auer PL, Azzollini J, Balmana J, Barkardottir RB, Barrowdale D, Beeghly-Fadiel A, Benitez J, Bermisheva M, Bialkowska K, Blanco AM, Blomqvist C, Blot W, Bogdanova NV, Bojesen SE, Bolla MK, Bonanni B, Borg A, Bosse K, Brauch H, Brenner H, Briceno I, Brock IW, Brooks-Wilson A, Bruning T, Burwinkel B, Buys SS, Cai Q, Caldes T, Caligo MA, Camp NJ, Campbell I, Canzian F, Carroll JS, Carter BD, Castelao JE, Chiquette J, Christiansen H, Chung WK, Claes KBM, Clarke CL, Collaborators GS, Collaborators E, Collee JM, Cornelissen S, Couch FJ, Cox A, Cross SS, Cybulski C, Czene K, Daly MB, de la Hoya M, Devilee P, Diez O, Ding YC, Dite GS, Domchek SM, Dork T, Dos-Santos-Silva I, Droit A, Dubois S, Dumont M, Duran M, Durcan L, Dwek M, Eccles DM, Engel C, Eriksson M, Evans DG, Fasching PA, Fletcher O, Floris G, Flyger H, et al. (2020) Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. Nat Genet 52: 56-73. doi: 10.1038/s41588-019-0537-1

Fakhro KA, Staudt MR, Ramstetter MD, Robay A, Malek JA, Badii R, Al-Marri AA, Abi Khalil C, Al-Shakaki A, Chidiac O, Stadler D, Zirie M, Jayyousi A, Salit J, Mezey JG, Crystal RG, Rodriguez-Flores JL (2016) The Qatar genome: a population-specific tool for precision medicine in the Middle East. Hum Genome Var 3: 16016. doi: 10.1038/hgv.2016.16

Falconer DS (1989) Introduction to quantitative genetics, 3rd edn. Longman Scientific & Technical, New York : Wiley

Fareed M, Afzal M (2014a) Estimating the inbreeding depression on cognitive behavior: a population based study of child cohort. PLoS One 9: e109585. doi: 10.1371/journal.pone.0109585

Fareed M, Afzal M (2014b) Evidence of inbreeding depression on height, weight, and body mass index: a population-based child cohort study. Am J Hum Biol 26: 784-95. doi: 10.1002/ajhb.22599

Fareed M, Afzal M (2016) Increased cardiovascular risks associated with familial inbreeding: a population-based study of adolescent cohort. Ann Epidemiol 26: 283-92. doi: 10.1016/j.annepidem.2016.03.001

Farrag NS, Cheskin LJ, Farag MK (2017) A systematic review of childhood obesity in the Middle East and North Africa (MENA) region: Prevalence and risk factors meta-analysis. Adv Pediatr Res 4. doi: 10.12715/apr.2017.4.8

Feingold KR, Anawalt, B., Boyce, A. et al (2000) Endotext. In: Feingold KR, Anawalt B, Boyce A, Chrousos G, de Herder WW, Dhatariya K, Dungan K, Grossman A, Hershman JM, Hofland J, Kalra S, Kaltsas G, Koch C, Kopp P, Korbonits M, Kovacs CS, Kuohung W, Laferrere B, McGee EA, McLachlan R, Morley JE, New M, Purnell J, Sahay R, Singer F, Stratakis CA, Trence DL, Wilson DP (eds) Endotext, South Dartmouth (MA)

Feng Y, Zang T, Xu X, Xu X (2008) Familial aggregation of metabolic syndrome and its components in a large Chinese population. Obesity (Silver Spring) 16: 125-9. doi: 10.1038/oby.2007.22

Fernandes V, Brucato N, Ferreira JC, Pedro N, Cavadas B, Ricaut FX, Alshamali F, Pereira L (2019) Genome-Wide Characterization of Arabian Peninsula Populations: Shedding Light on the History of a Fundamental Bridge between Continents. Mol Biol Evol 36: 575-586. doi: 10.1093/molbev/msz005

Fienberg Stephen E HDV (1980) R.A. Fisher: An Appreciation. Springer

Floyd JS, Sitlani CM, Avery CL, Noordam R, Li X, Smith AV, Gogarten SM, Li J, Broer L, Evans DS, Trompet S, Brody JA, Stewart JD, Eicher JD, Seyerle AA, Roach J, Lange LA, Lin HJ, Kors JA, Harris TB, Li-Gao R, Sattar N, Cummings SR, Wiggins KL, Napier MD, Sturmer T, Bis JC, Kerr KF, Uitterlinden AG, Taylor KD, Stott DJ, de Mutsert R, Launer LJ, Busch EL, Mendez-Giraldez R, Sotoodehnia N, Soliman EZ, Li Y, Duan Q, Rosendaal FR, Slagboom PE, Wilhelmsen KC, Reiner AP, Chen YD, Heckbert SR, Kaplan RC, Rice KM, Jukema JW, Johnson AD, Liu Y, Mook-Kanamori DO, Gudnason V, Wilson JG, Rotter JI, Laurie CC, Psaty BM, Whitsel EA, Cupples LA, Stricker BH (2018) Large-scale pharmacogenomic study of sulfonylureas and the QT, JT and QRS intervals: CHARGE Pharmacogenomics Working Group. Pharmacogenomics J 18: 127-135. doi: 10.1038/tpj.2016.90

Foulkes AS (2009) Applied Statistical Genetics with R. Springer

Franceschini N, Fox E, Zhang Z, Edwards TL, Nalls MA, Sung YJ, Tayo BO, Sun YV, Gottesman O, Adeyemo A, Johnson AD, Young JH, Rice K, Duan Q, Chen F, Li Y, Tang H, Fornage M, Keene KL, Andrews JS, Smith JA, Faul JD, Guangfa Z, Guo W, Liu Y, Murray SS, Musani SK, Srinivasan S, Velez Edwards DR, Wang H, Becker LC, Bovet P, Bochud M, Broeckel U, Burnier M, Carty C, Chasman DI, Ehret G, Chen WM, Chen G, Chen W, Ding J, Dreisbach AW, Evans MK, Guo X, Garcia ME, Jensen R, Keller MF, Lettre G, Lotay V, Martin LW, Moore JH, Morrison AC, Mosley TH, Ogunniyi A, Palmas W, Papanicolaou G, Penman A, Polak JF, Ridker PM, Salako B, Singleton AB, Shriner D, Taylor KD, Vasan R, Wiggins K, Williams SM, Yanek LR, Zhao W, Zonderman AB, Becker DM, Berenson G, Boerwinkle E, Bottinger E, Cushman M, Eaton C, Nyberg F, Heiss G, Hirschhron JN, Howard VJ, Karczewsk KJ, Lanktree MB, Liu K, Liu Y, Loos R, Margolis K, Snyder M, Asian Genetic Epidemiology Network C, Psaty BM, Schork NJ, Weir DR, Rotimi CN, Sale MM, Harris T, Kardia SL, Hunt SC, Arnett D, Redline S, Cooper RS, Risch NJ, et al. (2013) Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations. Am J Hum Genet 93: 545-54. doi: 10.1016/j.ajhg.2013.07.010

Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 316: 889-94. doi: 10.1126/science.1141634

Frazer KA, Murray SS, Schork NJ, Topol EJ (2009) Human genetic variation and its contribution to complex traits. Nat Rev Genet 10: 241-51. doi: 10.1038/nrg2554
Funcke JB, von Schnurbein J, Lennerz B, Lahr G, Debatin KM, Fischer-Posovszky P, Wabitsch M (2014) Monogenic forms of childhood obesity due to mutations in the leptin gene. Mol Cell Pediatr 1: 3. doi: 10.1186/s40348-014-0003-1
Gallagher MD, Chen-Plotkin AS (2018) The Post-GWAS Era: From Association to Function. Am J Hum Genet 102: 717-730. doi: 10.1016/j.ajhg.2018.04.002
Galton F (1989) Natural Inheritance. Macmillan & co., London, New York
Gandasi NR, Yin P, Riz M, Chibalina MV, Cortese G, Lund PE, Matveev V, Rorsman P, Sherman A, Pedersen MG, Barg S (2017) Ca2+ channel clustering with insulin-containing granules is disturbed in type 2 diabetes. J Clin Invest 127: 2353-2364. doi: 10.1172/JCI88491
Ganesh SK, Chasman DI, Larson MG, Guo X, Verwoert G, Bis JC, Gu X, Smith AV, Yang ML, Zhang Y, Ehret G, Rose LM, Hwang SJ, Papanicolau GJ, Sijbrands EJ, Rice K, Eiriksdottir G, Pihur V, Ridker PM, Vasan RS, Newton-Cheh C, Global Blood Pressure Genetics C, Raffel LJ, Amin N, Rotter JI, Liu K, Launer LJ, Xu M, Caulfield M, Morrison AC, Johnson AD, Vaidya D, Dehghan A, Li G, Bouchard C, Harris TB, Zhang H, Boerwinkle E, Siscovick DS, Gao W, Uitterlinden AG, Rivadeneira F, Hofman A, Willer CJ, Franco OH, Huo Y, Witteman JC, Munroe PB, Gudnason V, Palmas W, van Duijn C, Fornage M, Levy D, Psaty BM, Chakravarti A (2014) Effects of long-term averaging of quantitative blood pressure traits on the detection of genetic associations. Am J Hum Genet 95: 49-65. doi: 10.1016/j.ajhg.2014.06.002
Gao X, Patel TB (2009) Regulation of protein kinase A activity by p90 ribosomal S6 kinase 1. J Biol Chem 284: 33070-8. doi: 10.1074/jbc.M109.058453
Garcia-Giustiniani D, Stein R (2016) Genetics of Dyslipidemia. Arq Bras Cardiol 106: 434-8. doi: 10.5935/abc.20160074
Gaskin RS, Estwick D, Peddi R (2001) G6PD deficiency: its role in the high prevalence of hypertension and diabetes mellitus. Ethn Dis 11: 749-54.
Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061-73. doi: 10.1038/nature09534
Ghassibe-Sabbagh M, Haber M, Salloum AK, Al-Sarraj Y, Akle Y, Hirbli K, Romanos J, Mouzaya F, Gauguier D, Platt DE, El-Shanti H, Zalloua PA (2014) T2DM GWAS in the Lebanese population confirms the role of TCF7L2 and CDKAL1 in disease susceptibility. Sci Rep 4: 7351. doi: 10.1038/srep07351
Gogarten SM, Bhangale T, Conomos MP, Laurie CA, McHugh CP, Painter I, Zheng X, Crosslin DR, Levine D, Lumley T, Nelson SC, Rice K, Shen J, Swarnkar R, Weir BS, Laurie CC (2012) GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. Bioinformatics 28: 3329-31. doi: 10.1093/bioinformatics/bts610
Goodarzi MO (2018) Genetics of obesity: what genetic association studies have taught us about the biology of obesity and its complications. Lancet Diabetes Endocrinol 6: 223-236. doi: 10.1016/S2213-8587(17)30200-0
Gopaul KP, Crook MA (2006) Sialic acid: a novel marker of cardiovascular disease? Clin Biochem 39: 667-81. doi: 10.1016/j.clinbiochem.2006.02.010
Gosadi IM, Goyder EC, Teare MD (2014) Investigating the potential effect of consanguinity on type 2 diabetes susceptibility in a Saudi population. Hum Hered 77: 197-206. doi: 10.1159/000362447
Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, Braun LT, de Ferranti S, Faiella-Tommasino J, Forman DE, Goldberg R, Heidenreich PA, Hlatky MA, Jones DW, Lloyd-Jones D, Lopez-Pajares N, Ndumele CE, Orringer CE, Peralta CA, Saseen JJ, Smith SC, Jr., Sperling L, Virani SS, Yeboah J (2019) 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the

Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 73: e285-e350. doi: 10.1016/j.jacc.2018.11.003
Gulvady C, Pingle S, Shanbhag S (2007) Incidence of vitamin B12 / D3 deficiency among company executives. Indian J Occup Environ Med 11: 83-5. doi: 10.4103/0019-5278.34535
Gunaid AA (2002) Prevalence of known diabetes and hypertension in the Republic of Yemen. East Mediterr Health J 8: 374-85.
Gunstad J, Schofield P, Paul RH, Spitznagel MB, Cohen RA, Williams LM, Kohn M, Gordon E (2006) BDNF Val66Met polymorphism is associated with body mass index in healthy adults. Neuropsychobiology 53: 153-6. doi: 10.1159/000093341
Haber M, Nassar J, Almarri MA, Saupe T, Saag L, Griffith SJ, Doumet-Serhal C, Chanteau J, Saghieh-Beydoun M, Xue Y, Scheib CL, Tyler-Smith C (2020) A Genetic History of the Near East from an aDNA Time Course Sampling Eight Points in the Past 4,000 Years. Am J Hum Genet 107: 149-157. doi: 10.1016/j.ajhg.2020.05.008
Haberle V, Stark A (2018) Eukaryotic core promoters and the functional basis of transcription initiation. Nat Rev Mol Cell Biol 19: 621-637. doi: 10.1038/s41580-018-0028-8
Haider NB, Jacobson SG, Cideciyan AV, Swiderski R, Streb LM, Searby C, Beck G, Hockey R, Hanna DB, Gorman S, Duhl D, Carmi R, Bennett J, Weleber RG, Fishman GA, Wright AF, Stone EM, Sheffield VC (2000) Mutation of a nuclear receptor gene, NR2E3, causes enhanced S cone syndrome, a disorder of retinal cell fate. Nat Genet 24: 127-31. doi: 10.1038/72777
Halberg N, Wernstedt-Asterholm I, Scherer PE (2008) The adipocyte as an endocrine cell. Endocrinol Metab Clin North Am 37: 753-68, x-xi. doi: 10.1016/j.ecl.2008.07.002
Hammad MM, Abu-Farha M, Hebbar P, Cherian P, Al Khairi I, Melhem M, Alkayal F, Alsmadi O, Thanaraj TA, Al-Mulla F, Abubaker J (2020) MC4R Variant rs17782313 Associates With Increased Levels of DNAJC27, Ghrelin, and Visfatin and Correlates With Obesity and Hypertension in a Kuwaiti Cohort. Front Endocrinol (Lausanne) 11: 437. doi: 10.3389/fendo.2020.00437
Hammad SS, Berry DC (2017) The Child Obesity Epidemic in Saudi Arabia: A Review of the Literature. J Transcult Nurs 28: 505-515. doi: 10.1177/1043659616668398
Hamoudi R, Saheb Sharif-Askari N, Saheb Sharif-Askari F, Abusnana S, Aljaibeji H, Taneera J, Sulaiman N (2019) Prediabetes and diabetes prevalence and risk factors comparison between ethnic groups in the United Arab Emirates. Sci Rep 9: 17437. doi: 10.1038/s41598-019-53505-7
Han A, Newell EW, Glanville J, Fernandez-Becker N, Khosla C, Chien YH, Davis MM (2013) Dietary gluten triggers concomitant activation of CD4+ and CD8+ alphabeta T cells and gammadelta T cells in celiac disease. Proc Natl Acad Sci U S A 110: 13073-8. doi: 10.1073/pnas.1311861110
Han B, Pouget JG, Slowikowski K, Stahl E, Lee CH, Diogo D, Hu X, Park YR, Kim E, Gregersen PK, Dahlqvist SR, Worthington J, Martin J, Eyre S, Klareskog L, Huizinga T, Chen WM, Onengut-Gumuscu S, Rich SS, Major Depressive Disorder Working Group of the Psychiatric Genomics C, Wray NR, Raychaudhuri S (2016) A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. Nat Genet 48: 803-10. doi: 10.1038/ng.3572
Haroun D, Mechli R, Sahuri R, AlKhatib S, Obeid O, El Mallah C, Wood L, AlSuwaidi K (2018) Metabolic syndrome among adolescents in Dubai, United Arab Emirates, is attributable to the high prevalence of low HDL levels: a cross-sectional study. BMC Public Health 18: 1284. doi: 10.1186/s12889-018-6215-x
Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2: 3-19. doi: 10.1007/BF01066731
Hauner H (2005) Secretory factors from human adipose tissue and their functional role. Proc Nutr Soc 64: 163-9. doi: 10.1079/pns2005428

Hebbar P, Abu-Farha M, Alkayal F, Nizam R, Elkum N, Melhem M, John SE, Channanath A, Abubaker J, Bennakhi A, Al-Ozairi E, Tuomilehto J, Pitkaniemi J, Alsmadi O, Al-Mulla F, Thanaraj TA (2020a) Genome-wide association study identifies novel risk variants from RPS6KA1, CADPS, VARS, and DHX58 for fasting plasma glucose in Arab population. Sci Rep 10: 152. doi: 10.1038/s41598-019-57072-9

Hebbar P, Abu-Farha M, Mohammad A, Alkayal F, Melhem M, Abubaker J, Al-Mulla F, Thanaraj TA (2019a) FTO Variant rs1421085 Associates With Increased Body Weight, Soft Lean Mass, and Total Body Water Through Interaction With Ghrelin and Apolipoproteins in Arab Population. Front Genet 10: 1411. doi: 10.3389/fgene.2019.01411

Hebbar P, Abubaker JA, Abu-Farha M, Alsmadi O, Elkum N, Alkayal F, John SE, Channanath A, Iqbal R, Pitkaniemi J, Tuomilehto J, Sladek R, Al-Mulla F, Thanaraj TA (2020b) Genome-wide landscape establishes novel association signals for metabolic traits in the Arab population. Hum Genet. doi: 10.1007/s00439-020-02222-7

Hebbar P, Abubaker JA, Abu-Farha M, Tuomilehto J, Al-Mulla F, Thanaraj TA (2019b) A Perception on Genome-Wide Genetic Analysis of Metabolic Traits in Arab Populations. Front Endocrinol (Lausanne) 10: 8. doi: 10.3389/fendo.2019.00008

Hebbar P, Alkayal F, Nizam R, Melhem M, Elkum N, John SE, Abufarha M, Alsmadi O, Thanaraj TA (2017a) The TCN2 variant of rs9606756 [Ile23Val] acts as risk loci for obesity-related traits and mediates by interacting with Apo-A1. Obesity (Silver Spring) 25: 1098-1108. doi: 10.1002/oby.21826

Hebbar P, Elkum N, Alkayal F, John SE, Thanaraj TA, Alsmadi O (2017b) Genetic risk variants for metabolic traits in Arab populations. Sci Rep 7: 40988. doi: 10.1038/srep40988

Hebbar P, Nizam R, Melhem M, Alkayal F, Elkum N, John SE, Tuomilehto J, Alsmadi O, Thanaraj TA (2018) Genome-wide association study identifies novel recessive genetic variants for high TGs in an Arab population. J Lipid Res. doi: 10.1194/jlr.P080218

Hershberger RE, Siegfried JD (2011) Update 2011: clinical and genetic issues in familial dilated cardiomyopathy. J Am Coll Cardiol 57: 1641-9. doi: 10.1016/j.jacc.2011.01.015

Heymann AD, Cohen Y, Chodick G (2012) Glucose-6-phosphate dehydrogenase deficiency and type 2 diabetes. Diabetes Care 35: e58. doi: 10.2337/dc11-2527

Hilado MA, Randhawa RS (2018) A novel mutation in the proopiomelanocortin (POMC) gene of a Hispanic child: metformin treatment shows a beneficial impact on the body mass index. J Pediatr Endocrinol Metab 31: 815-819. doi: 10.1515/jpem-2017-0467

Hilal Al-Sabti KHA-R, Mirdavron M. Mukaddirov, Ali Talib Al-Hinai (2010) Prevalence and Pattern of Dyslipidemia in Omani Patients Undergoing Coronary Artery Bypass Surgery. Journal of Clinical Lipidology 4: 201-202. doi: https://doi.org/10.1016/j.jacl.2010.03.016

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 106: 9362-7. doi: 10.1073/pnas.0903103106

Hitti P (1914) The Arabs; A Short History. MacMillan & Co

Hiura Y, Shen CS, Kokubo Y, Okamura T, Morisaki T, Tomoike H, Yoshida T, Sakamoto H, Goto Y, Nonogi H, Iwai N (2009) Identification of genetic markers associated with high-density lipoprotein-cholesterol by genome-wide screening in a Japanese population: the Suita study. Circ J 73: 1119-26.

Hoffmann TJ, Ehret GB, Nandakumar P, Ranatunga D, Schaefer C, Kwok PY, Iribarren C, Chakravarti A, Risch N (2017) Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. Nat Genet 49: 54-64. doi: 10.1038/ng.3715

Holdt LM, Teupser D (2012) Recent studies of the human chromosome 9p21 locus, which is associated with atherosclerosis in human populations. Arterioscler Thromb Vasc Biol 32: 196-206. doi: 10.1161/ATVBAHA.111.232678

Holland WL, Summers SA (2008) Sphingolipids, insulin resistance, and metabolic disease: new insights from in vivo manipulation of sphingolipid metabolism. Endocr Rev 29: 381-402. doi: 10.1210/er.2007-0025

Hong C, Tontonoz P (2014) Liver X receptors in lipid metabolism: opportunities for drug discovery. Nat Rev Drug Discov 13: 433-44. doi: 10.1038/nrd4280

Hong C, Walczak R, Dhamko H, Bradley MN, Marathe C, Boyadjian R, Salazar JV, Tontonoz P (2011) Constitutive activation of LXR in macrophages regulates metabolic and inflammatory gene expression: identification of ARL7 as a direct target. J Lipid Res 52: 531-9. doi: 10.1194/jlr.M010686

Hormozdiari F, Kichaev G, Yang WY, Pasaniuc B, Eskin E (2015) Identification of causal genes for complex traits. Bioinformatics 31: i206-13. doi: 10.1093/bioinformatics/btv240

Horvath S, Xu X, Laird NM (2001) The family based association test method: strategies for studying general genotype--phenotype associations. Eur J Hum Genet 9: 301-6. doi: 10.1038/sj.ejhg.5200625

Hu Y, Liang J, Yu S (2014) High prevalence of diabetes mellitus in a five-generation Chinese family with Huntington's disease. J Alzheimers Dis 40: 863-8. doi: 10.3233/JAD-131847

Hwang LD, Lin C, Gharahkhani P, Cuellar-Partida G, Ong JS, An J, Gordon SD, Zhu G, MacGregor S, Lawlor DA, Breslin PAS, Wright MJ, Martin NG, Reed DR (2019) New insight into human sweet taste: a genome-wide association study of the perception and intake of sweet substances. Am J Clin Nutr 109: 1724-1737. doi: 10.1093/ajcn/nqz043

IDF (2019) IDF DIABETES ATLAS Ninth edition 2019. International Diabetes Federation

InanlooRahatloo K, Parsa AF, Huse K, Rasooli P, Davaran S, Platzer M, Kramer M, Fan JB, Turk C, Amini S, Steemers F, Gunderson K, Ronaghi M, Elahi E (2014) Mutation in ST6GALNAC5 identified in family with coronary artery disease. Sci Rep 4: 3595. doi: 10.1038/srep03595

International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, Hwang SJ, Pihur V, Vollenweider P, O'Reilly PF, Amin N, Bragg-Gresham JL, Teumer A, Glazer NL, Launer L, Zhao JH, Aulchenko Y, Heath S, Sober S, Parsa A, Luan J, Arora P, Dehghan A, Zhang F, Lucas G, Hicks AA, Jackson AU, Peden JF, Tanaka T, Wild SH, Rudan I, Igl W, Milaneschi Y, Parker AN, Fava C, Chambers JC, Fox ER, Kumari M, Go MJ, van der Harst P, Kao WH, Sjogren M, Vinay DG, Alexander M, Tabara Y, Shaw-Hawkins S, Whincup PH, Liu Y, Shi G, Kuusisto J, Tayo B, Seielstad M, Sim X, Nguyen KD, Lehtimaki T, Matullo G, Wu Y, Gaunt TR, Onland-Moret NC, Cooper MN, Platou CG, Org E, Hardy R, Dahgam S, Palmen J, Vitart V, Braund PS, Kuznetsova T, Uiterwaal CS, Adeyemo A, Palmas W, Campbell H, Ludwig B, Tomaszewski M, Tzoulaki I, Palmer ND, consortium CA, Consortium CK, KidneyGen C, EchoGen c, consortium C-H, Aspelund T, Garcia M, Chang YP, O'Connell JR, Steinle NI, Grobbee DE, Arking DE, Kardia SL, Morrison AC, Hernandez D, Najjar S, McArdle WL, Hadley D, Brown MJ, Connell JM, et al. (2011) Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 478: 103-9. doi: 10.1038/nature10405

International Schizophrenia C, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460: 748-52. doi: 10.1038/nature08185

Iriyama K, Teranishi T, Mori H, Nishiwaki H, Kusaka N (1984) Effects of exogenous catecholamines on glucose and fat metabolism and on triglycerides in the rat liver during total parenteral nutrition. JPEN J Parenter Enteral Nutr 8: 412-5. doi: 10.1177/0148607184008004412

Jamaluddine Z, Sibai AM, Othman S, Yazbek S (2016) Mapping genetic research in non-communicable disease publications in selected Arab countries: first step towards a guided research agenda. Health Res Policy Syst 14: 81. doi: 10.1186/s12961-016-0153-9

Jeon YK, Kim MR, Huh JE, Mok JY, Song SH, Kim SS, Kim BH, Lee SH, Kim YK, Kim IJ (2011) Cystatin C as an early biomarker of nephropathy in patients with type 2 diabetes. J Korean Med Sci 26: 258-63. doi: 10.3346/jkms.2011.26.2.258

Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, Newton-Cheh C, State MW, Levy D, Lifton RP (2008) Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nat Genet 40: 592-599. doi: 10.1038/ng.118

Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, Schwartz SM, Voight BF, Elosua R, Salomaa V, O'Donnell CJ, Dallinga-Thie GM, Anand SS, Yusuf S, Huff MW, Kathiresan S, Hegele RA (2010) Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet 42: 684-7. doi: 10.1038/ng.628

John SE, Antony D, Eaaswarkhanth M, Hebbar P, Channanath AM, Thomas D, Devarajan S, Tuomilehto J, Al-Mulla F, Alsmadi O, Thanaraj TA (2018) Assessment of coding region variants in Kuwaiti population: implications for medical genetics and population genomics. Sci Rep 8: 16583. doi: 10.1038/s41598-018-34815-8

John SE, Thareja G, Hebbar P, Behbehani K, Thanaraj TA, Alsmadi O (2015) Kuwaiti population subgroup of nomadic Bedouin ancestry-Whole genome sequence and analysis. Genom Data 3: 116-27. doi: 10.1016/j.gdata.2014.11.016

Jonathan Wallace PK (1979) The Gulf Handbook. Trade & Travel Publications

Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, Nakamura Y, Kamatani N (2010) Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat Genet 42: 210-5. doi: 10.1038/ng.531

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42: 348-54. doi: 10.1038/ng.548

Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, Kaplan L, Bennett D, Li Y, Tanaka T, Voight BF, Bonnycastle LL, Jackson AU, Crawford G, Surti A, Guiducci C, Burtt NP, Parish S, Clarke R, Zelenika D, Kubalanza KA, Morken MA, Scott LJ, Stringham HM, Galan P, Swift AJ, Kuusisto J, Bergman RN, Sundvall J, Laakso M, Ferrucci L, Scheet P, Sanna S, Uda M, Yang Q, Lunetta KL, Dupuis J, de Bakker PI, O'Donnell CJ, Chambers JC, Kooner JS, Hercberg S, Meneton P, Lakatta EG, Scuteri A, Schlessinger D, Tuomilehto J, Collins FS, Groop L, Altshuler D, Collins R, Lathrop GM, Melander O, Salomaa V, Peltonen L, Orho-Melander M, Ordovas JM, Boehnke M, Abecasis GR, Mohlke KL, Cupples LA (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet 41: 56-65. doi: 10.1038/ng.291

Kaur Y, de Souza RJ, Gibson WT, Meyre D (2017) A systematic review of genetic syndromes with obesity. Obes Rev 18: 603-634. doi: 10.1111/obr.12531

Kazma R, Bailey JN (2011) Population-based and family-based designs to analyze rare variants in complex diseases. Genet Epidemiol 35 Suppl 1: S41-7. doi: 10.1002/gepi.20648

Keller M, Schleinitz D, Forster J, Tonjes A, Bottcher Y, Fischer-Rosinsky A, Breitfeld J, Weidle K, Rayner NW, Burkhardt R, Enigk B, Muller I, Halbritter J, Koriath M, Pfeiffer A, Krohn K, Groop L, Spranger J, Stumvoll M, Kovacs P (2013) THOC5: a novel gene involved in HDL-cholesterol metabolism. J Lipid Res 54: 3170-6. doi: 10.1194/jlr.M039420

Khan SM, El Hajj Chehadeh S, Abdulrahman M, Osman W, Al Safar H (2018) Establishing a genetic link between FTO and VDR gene polymorphisms and obesity in the Emirati population. BMC Med Genet 19: 11. doi: 10.1186/s12881-018-0522-z

Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF (2010) Genomic runs of homozygosity record population history and consanguinity. PLoS One 5: e13996. doi: 10.1371/journal.pone.0013996

Klautzer L, Becker J, Mattke S (2014) The curse of wealth - Middle Eastern countries need to address the rapidly rising burden of diabetes. Int J Health Policy Manag 2: 109-14. doi: 10.15171/ijhpm.2014.33

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308: 385-9. doi: 10.1126/science.1109557

Ko A, Cantor RM, Weissglas-Volkov D, Nikkola E, Reddy PM, Sinsheimer JS, Pasaniuc B, Brown R, Alvarez M, Rodriguez A, Rodriguez-Guillen R, Bautista IC, Arellano-Campos O, Munoz-Hernandez LL, Salomaa V, Kaprio J, Jula A, Jauhiainen M, Heliovaara M, Raitakari O, Lehtimaki T, Eriksson JG, Perola M, Lohmueller KE, Matikainen N, Taskinen MR, Rodriguez-Torres M, Riba L, Tusie-Luna T, Aguilar-Salinas CA, Pajukanta P (2014) Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. Nat Commun 5: 3983. doi: 10.1038/ncomms4983

Korotayev A (1995) ANCIENT YEMEN: Some general trends of Evolution of the Sabaic Language and Sabaean culture. Journal of Semitic Studies Supplemental 5.

Korotayev A (1996) Pre-Islamic Yemen: Socio-political organization of the Sabaean cultural area in the 2nd and 3rd centuries AD. Harrassowitz Verlag, Wiesbaden

Koshy R, Ranawat A, Scaria V (2017) al mena: a comprehensive resource of human genetic variants integrating genomes and exomes from Arab, Middle Eastern and North African populations. J Hum Genet 62: 889-894. doi: 10.1038/jhg.2017.67

Kotilaine J (2010) GCC Agriculture. ECONOMIC RESEARCH. NCB Capital

Kowalski MH, Qian H, Hou Z, Rosen JD, Tapia AL, Shan Y, Jain D, Argos M, Arnett DK, Avery C, Barnes KC, Becker LC, Bien SA, Bis JC, Blangero J, Boerwinkle E, Bowden DW, Buyske S, Cai J, Cho MH, Choi SH, Choquet H, Cupples LA, Cushman M, Daya M, de Vries PS, Ellinor PT, Faraday N, Fornage M, Gabriel S, Ganesh SK, Graff M, Gupta N, He J, Heckbert SR, Hidalgo B, Hodonsky CJ, Irvin MR, Johnson AD, Jorgenson E, Kaplan R, Kardia SLR, Kelly TN, Kooperberg C, Lasky-Su JA, Loos RJF, Lubitz SA, Mathias RA, McHugh CP, Montgomery C, Moon JY, Morrison AC, Palmer ND, Pankratz N, Papanicolaou GJ, Peralta JM, Peyser PA, Rich SS, Rotter JI, Silverman EK, Smith JA, Smith NL, Taylor KD, Thornton TA, Tiwari HK, Tracy RP, Wang T, Weiss ST, Weng LC, Wiggins KL, Wilson JG, Yanek LR, Zollner S, North KE, Auer PL, Consortium NT-OfPM, Hematology TO, Hemostasis Working G, Raffield LM, Reiner AP, Li Y (2019) Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. PLoS Genet 15: e1008500. doi: 10.1371/journal.pgen.1008500

Kraja AT, Vaidya D, Pankow JS, Goodarzi MO, Assimes TL, Kullo IJ, Sovio U, Mathias RA, Sun YV, Franceschini N, Absher D, Li G, Zhang Q, Feitosa MF, Glazer NL, Haritunians T, Hartikainen AL, Knowles JW, North KE, Iribarren C, Kral B, Yanek L, O'Reilly PF, McCarthy MI, Jaquish C, Couper DJ, Chakravarti A, Psaty BM, Becker LC, Province MA, Boerwinkle E, Quertermous T, Palotie L, Jarvelin MR, Becker DM, Kardia SL, Rotter JI, Chen YD, Borecki IB (2011) A bivariate genome-wide approach to metabolic syndrome: STAMPEED consortium. Diabetes 60: 1329-39. doi: 10.2337/db10-1011

Kristjansson K, Manolescu A, Kristinsson A, Hardarson T, Knudsen H, Ingason S, Thorleifsson G, Frigge ML, Kong A, Gulcher JR, Stefansson K (2002) Linkage of essential hypertension to chromosome 18q. Hypertension 39: 1044-9. doi: 10.1161/01.hyp.0000018580.24644.18

Ku CS, Loy EY, Salim A, Pawitan Y, Chia KS (2010) The discovery of human genetic variations and their use as disease markers: past, present and future. J Hum Genet 55: 403-15. doi: 10.1038/jhg.2010.55

Kulyte A, Lundback V, Lindgren CM, Luan J, Lotta LA, Langenberg C, Arner P, Strawbridge RJ, Dahlman I (2020) Genome-wide association study of adipocyte lipolysis in the GENetics of adipocyte lipolysis (GENiAL) cohort. Mol Metab 34: 85-96. doi: 10.1016/j.molmet.2020.01.009

Kurano M, Tsukamoto K, Kamitsuji S, Kamatani N, Hara M, Ishikawa T, Kim BJ, Moon S, Jin Kim Y, Teramoto T (2016) Genome-wide association study of serum lipids confirms previously reported associations as well as new associations of common SNPs within PCSK7 gene with triglyceride. J Hum Genet 61: 427-33. doi: 10.1038/jhg.2015.170

Lee JS, Chang PY, Zhang Y, Kizer JR, Best LG, Howard BV (2017) Triglyceride and HDL-C Dyslipidemia and Risks of Coronary Heart Disease and Ischemic Stroke by Glycemic Dysregulation Status: The Strong Heart Study. Diabetes Care 40: 529-537. doi: 10.2337/dc16-1958

Lee JW, Choi AH, Ham M, Kim JW, Choe SS, Park J, Lee GY, Yoon KH, Kim JB (2011) G6PD up-regulation promotes pancreatic beta-cell dysfunction. Endocrinology 152: 793-803. doi: 10.1210/en.2010-0606

Lee S, Abecasis GR, Boehnke M, Lin X (2014) Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet 95: 5-23. doi: 10.1016/j.ajhg.2014.06.009

Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR (2012) Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics 28: 2540-2. doi: 10.1093/bioinformatics/bts474

Lettre G, Palmer CD, Young T, Ejebe KG, Allayee H, Benjamin EJ, Bennett F, Bowden DW, Chakravarti A, Dreisbach A, Farlow DN, Folsom AR, Fornage M, Forrester T, Fox E, Haiman CA, Hartiala J, Harris TB, Hazen SL, Heckbert SR, Henderson BE, Hirschhorn JN, Keating BJ, Kritchevsky SB, Larkin E, Li M, Rudock ME, McKenzie CA, Meigs JB, Meng YA, Mosley TH, Newman AB, Newton-Cheh CH, Paltoo DN, Papanicolaou GJ, Patterson N, Post WS, Psaty BM, Qasim AN, Qu L, Rader DJ, Redline S, Reilly MP, Reiner AP, Rich SS, Rotter JI, Liu Y, Shrader P, Siscovick DS, Tang WH, Taylor HA, Tracy RP, Vasan RS, Waters KM, Wilks R, Wilson JG, Fabsitz RR, Gabriel SB, Kathiresan S, Boerwinkle E (2011) Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. PLoS Genet 7: e1001300. doi: 10.1371/journal.pgen.1001300

Lewis CM, Vassos E (2020) Polygenic risk scores: from research tools to clinical instruments. Genome Med 12: 44. doi: 10.1186/s13073-020-00742-5

Li-Gao R, Wakil SM, Meyer BF, Dzimiri N, Mook-Kanamori DO (2018) Replication of Type 2 diabetes-associated variants in a Saudi Arabian population. Physiol Genomics 50: 296-297. doi: 10.1152/physiolgenomics.00100.2017

Li L, Lee EW, Ji H, Zukowska Z (2003) Neuropeptide Y-induced acceleration of postangioplasty occlusion of rat carotid artery. Arterioscler Thromb Vasc Biol 23: 1204-10. doi: 10.1161/01.ATV.0000071349.30914.25

Li XH, Guan LY, Lin HY, Wang SH, Cao YQ, Jiang XY, Wang YB (2016) Fibrinogen: A Marker in Predicting Diabetic Foot Ulcer Severity. J Diabetes Res 2016: 2358321. doi: 10.1155/2016/2358321

Ligthart S, Vaez A, Hsu YH, Inflammation Working Group of the CC, Pmi Wg XCP, LifeLines Cohort S, Stolk R, Uitterlinden AG, Hofman A, Alizadeh BZ, Franco OH, Dehghan A (2016) Bivariate genome-wide association study identifies novel pleiotropic loci for lipids and inflammation. BMC Genomics 17: 443. doi: 10.1186/s12864-016-2712-4

Lipina C, Hundal HS (2011) Sphingolipids: agents provocateurs in the pathogenesis of insulin resistance. Diabetologia 54: 1596-607. doi: 10.1007/s00125-011-2127-3

Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D (2011) FaST linear mixed models for genome-wide association studies. Nat Methods 8: 833-5. doi: 10.1038/nmeth.1681

Liu C, Kraja AT, Smith JA, Brody JA, Franceschini N, Bis JC, Rice K, Morrison AC, Lu Y, Weiss S, Guo X, Palmas W, Martin LW, Chen YD, Surendran P, Drenos F, Cook JP, Auer PL, Chu AY, Giri A, Zhao W, Jakobsdottir J, Lin LA, Stafford JM, Amin N, Mei H, Yao J, Voorman A, Consortium CHDE, Exome BPC, Go TDC, Consortium TDG, Larson MG, Grove ML, Smith AV, Hwang SJ, Chen H, Huan T, Kosova G, Stitziel NO, Kathiresan S, Samani N, Schunkert H, Deloukas P, Myocardial Infarction

G, Consortia CAE, Li M, Fuchsberger C, Pattaro C, Gorski M, Consortium CK, Kooperberg C, Papanicolaou GJ, Rossouw JE, Faul JD, Kardia SL, Bouchard C, Raffel LJ, Uitterlinden AG, Franco OH, Vasan RS, O'Donnell CJ, Taylor KD, Liu K, Bottinger EP, Gottesman O, Daw EW, Giulianini F, Ganesh S, Salfati E, Harris TB, Launer LJ, Dorr M, Felix SB, Rettig R, Volzke H, Kim E, Lee WJ, Lee IT, Sheu WH, Tsosie KS, Edwards DR, Liu Y, Correa A, Weir DR, Volker U, Ridker PM, Boerwinkle E, Gudnason V, Reiner AP, van Duijn CM, Borecki IB, Edwards TL, Chakravarti A, Rotter JI, Psaty BM, Loos RJ, Fornage M, Ehret GB, Newton-Cheh C, et al. (2016) Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. Nat Genet 48: 1162-70. doi: 10.1038/ng.3660

Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, Saleheen D, Emdin C, Alam D, Alves AC, Amouyel P, Di Angelantonio E, Arveiler D, Assimes TL, Auer PL, Baber U, Ballantyne CM, Bang LE, Benn M, Bis JC, Boehnke M, Boerwinkle E, Bork-Jensen J, Bottinger EP, Brandslund I, Brown M, Busonero F, Caulfield MJ, Chambers JC, Chasman DI, Chen YE, Chen YI, Chowdhury R, Christensen C, Chu AY, Connell JM, Cucca F, Cupples LA, Damrauer SM, Davies G, Deary IJ, Dedoussis G, Denny JC, Dominiczak A, Dube MP, Ebeling T, Eiriksdottir G, Esko T, Farmaki AE, Feitosa MF, Ferrario M, Ferrieres J, Ford I, Fornage M, Franks PW, Frayling TM, Frikke-Schmidt R, Fritsche LG, Frossard P, Fuster V, Ganesh SK, Gao W, Garcia ME, Gieger C, Giulianini F, Goodarzi MO, Grallert H, Grarup N, Groop L, Grove ML, Gudnason V, Hansen T, Harris TB, Hayward C, Hirschhorn JN, Holmen OL, Huffman J, Huo Y, Hveem K, Jabeen S, Jackson AU, Jakobsdottir J, Jarvelin MR, Jensen GB, Jorgensen ME, Jukema JW, Justesen JM, Kamstrup PR, Kanoni S, Karpe F, Kee F, Khera AV, Klarin D, Koistinen HA, Kooner JS, Kooperberg C, Kuulasmaa K, Kuusisto J, Laakso M, Lakka T, et al. (2017) Exome-wide association study of plasma lipids in >300,000 individuals. Nat Genet 49: 1758-1766. doi: 10.1038/ng.3977

Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, Investigators A, Hayward NK, Montgomery GW, Visscher PM, Martin NG, Macgregor S (2010) A versatile gene-based test for genome-wide association studies. Am J Hum Genet 87: 139-45. doi: 10.1016/j.ajhg.2010.06.009

Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, Croteau-Chonka DC, Esko T, Fall T, Ferreira T, Gustafsson S, Kutalik Z, Luan J, Magi R, Randall JC, Winkler TW, Wood AR, Workalemahu T, Faul JD, Smith JA, Zhao JH, Zhao W, Chen J, Fehrmann R, Hedman AK, Karjalainen J, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M, Bolton JL, Bragg-Gresham JL, Buyske S, Demirkan A, Deng G, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Goel A, Gong J, Jackson AU, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Mangino M, Leach IM, Medina-Gomez C, Medland SE, Nalls MA, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Shungin D, Stancakova A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Isaacs A, Albrecht E, Arnlov J, Arscott GM, Attwood AP, Bandinelli S, Barrett A, Bas IN, Bellis C, Bennett AJ, Berne C, Blagieva R, Bluher M, Bohringer S, Bonnycastle LL, Bottcher Y, Boyd HA, Bruinenberg M, Caspersen IH, Chen YI, Clarke R, Daw EW, de Craen AJM, Delgado G, Dimitriou M, et al. (2015) Genetic studies of body mass index yield new insights for obesity biology. Nature 518: 197-206. doi: 10.1038/nature14177

Loh PR, Danecek P, Palamara PF, Fuchsberger C, Y AR, H KF, Schoenherr S, Forer L, McCarthy S, Abecasis GR, Durbin R, A LP (2016) Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet 48: 1443-1448. doi: 10.1038/ng.3679

Louzada ML, Martins AP, Canella DS, Baraldi LG, Levy RB, Claro RM, Moubarac JC, Cannon G, Monteiro CA (2015) Impact of ultra-processed foods on micronutrient content in the Brazilian diet. Rev Saude Publica 49: 45. doi: 10.1590/S0034-8910.2015049006211

Lovejoy PE (2011) Transformations in slavery: a history of slavery in Africa. Cambridge University Press

Mabry RM, Reeves MM, Eakin EG, Owen N (2010a) Evidence of physical activity participation among men and women in the countries of the Gulf cooperation council: a review. Obes Rev 11: 457-64. doi: 10.1111/j.1467-789X.2009.00655.x

Mabry RM, Reeves MM, Eakin EG, Owen N (2010b) Gender differences in prevalence of the metabolic syndrome in Gulf Cooperation Council Countries: a systematic review. Diabet Med 27: 593-7. doi: 10.1111/j.1464-5491.2010.02998.x

Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z, Goodwin W, Bulbeck D, Bandelt HJ, Oppenheimer S, Torroni A, Richards M (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science 308: 1034-6. doi: 10.1126/science.1109792

MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW (2014) The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res 42: D986-92. doi: 10.1093/nar/gkt958

Magnusson M, Hedblad B, Engstrom G, Persson M, Nilsson P, Melander O (2013) High levels of cystatin C predict the metabolic syndrome: the prospective Malmo Diet and Cancer Study. J Intern Med 274: 192-9. doi: 10.1111/joim.12051

Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N, Cook JP, Schmidt EM, Wuttke M, Sarnowski C, Magi R, Nano J, Gieger C, Trompet S, Lecoeur C, Preuss MH, Prins BP, Guo X, Bielak LF, Below JE, Bowden DW, Chambers JC, Kim YJ, Ng MCY, Petty LE, Sim X, Zhang W, Bennett AJ, Bork-Jensen J, Brummett CM, Canouil M, Ec Kardt KU, Fischer K, Kardia SLR, Kronenberg F, Lall K, Liu CT, Locke AE, Luan J, Ntalla I, Nylander V, Schonherr S, Schurmann C, Yengo L, Bottinger EP, Brandslund I, Christensen C, Dedoussis G, Florez JC, Ford I, Franco OH, Frayling TM, Giedraitis V, Hackinger S, Hattersley AT, Herder C, Ikram MA, Ingelsson M, Jorgensen ME, Jorgensen T, Kriebel J, Kuusisto J, Ligthart S, Lindgren CM, Linneberg A, Lyssenko V, Mamakou V, Meitinger T, Mohlke KL, Morris AD, Nadkarni G, Pankow JS, Peters A, Sattar N, Stancakova A, Strauch K, Taylor KD, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tuomilehto J, Witte DR, Dupuis J, Peyser PA, Zeggini E, Loos RJF, Froguel P, Ingelsson E, Lind L, Groop L, Laakso M, Collins FS, Jukema JW, Palmer CNA, Grallert H, Metspalu A, et al. (2018a) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 50: 1505-1513. doi: 10.1038/s41588-018-0241-6

Mahajan A, Wessel J, Willems SM, Zhao W, Robertson NR, Chu AY, Gan W, Kitajima H, Taliun D, Rayner NW, Guo X, Lu Y, Li M, Jensen RA, Hu Y, Huo S, Lohman KK, Zhang W, Cook JP, Prins BP, Flannick J, Grarup N, Trubetskoy VV, Kravic J, Kim YJ, Rybin DV, Yaghootkar H, Muller-Nurasyid M, Meidtner K, Li-Gao R, Varga TV, Marten J, Li J, Smith AV, An P, Ligthart S, Gustafsson S, Malerba G, Demirkan A, Tajes JF, Steinthorsdottir V, Wuttke M, Lecoeur C, Preuss M, Bielak LF, Graff M, Highland HM, Justice AE, Liu DJ, Marouli E, Peloso GM, Warren HR, Exome BPC, Consortium M, Consortium G, Afaq S, Afzal S, Ahlqvist E, Almgren P, Amin N, Bang LB, Bertoni AG, Bombieri C, Bork-Jensen J, Brandslund I, Brody JA, Burtt NP, Canouil M, Chen YI, Cho YS, Christensen C, Eastwood SV, Eckardt KU, Fischer K, Gambaro G, Giedraitis V, Grove ML, de Haan HG, Hackinger S, Hai Y, Han S, Tybjaerg-Hansen A, Hivert MF, Isomaa B, Jager S, Jorgensen ME, Jorgensen T, Karajamaki A, Kim BJ, Kim SS, Koistinen HA, Kovacs P, Kriebel J, Kronenberg F, Lall K, Lange LA, Lee JJ, Lehne B, Li H, Lin KH, et al. (2018b) Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. Nat Genet 50: 559-571. doi: 10.1038/s41588-018-0084-1

Mahmoud I, Sulaiman N (2019) Dyslipidaemia prevalence and associated risk factors in the United Arab Emirates: a population-based study. BMJ Open 9: e031969. doi: 10.1136/bmjopen-2019-031969

Makansi N, Allison P, Awad M, Bedos C (2018) Fruit and vegetable intake among Emirati adolescents: a mixed methods study. East Mediterr Health J 24: 653-663. doi: 10.26719/2018.24.7.653

Malik M, Razig SA (2008) The prevalence of the metabolic syndrome among the multiethnic population of the United Arab Emirates: a report of a national survey. Metab Syndr Relat Disord 6: 177-86. doi: 10.1089/met.2008.0006

Malik R, Rannikmae K, Traylor M, Georgakis MK, Sargurupremraj M, Markus HS, Hopewell JC, Debette S, Sudlow CLM, Dichgans M, consortium M, the International Stroke Genetics C (2018) Genome-wide meta-analysis identifies 3 novel loci associated with stroke. Ann Neurol 84: 934-939. doi: 10.1002/ana.25369

Manayath GJ, Namburi P, Periasamy S, Kale JA, Narendran V, Ganesh A (2014) A novel mutation in the NR2E3 gene associated with Goldmann-Favre syndrome and vasoproliferative tumor of the retina. Mol Vis 20: 724-31.

Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, Derks EM (2018) A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res 27: e1608. doi: 10.1002/mpr.1608

Marta Mirazon Lahr RF (1994) Multiple dispersals and modern human origins. Evolutionary Anthropology 3: 48-60.

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA (2012) Systematic localization of common disease-associated variation in regulatory DNA. Science 337: 1190-5. doi: 10.1126/science.1222794

McCarthy MI, Hattersley AT (2008) Learning from molecular genetics: novel insights arising from the definition of genes for monogenic and type 2 diabetes. Diabetes 57: 2889-98. doi: 10.2337/db08-0343

McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, Timpson N, Koskinen S, Vrieze S, Scott LJ, Zhang H, Mahajan A, Veldink J, Peters U, Pato C, van Duijn CM, Gillies CE, Gandin I, Mezzavilla M, Gilly A, Cocca M, Traglia M, Angius A, Barrett JC, Boomsma D, Branham K, Breen G, Brummett CM, Busonero F, Campbell H, Chan A, Chen S, Chew E, Collins FS, Corbin LJ, Smith GD, Dedoussis G, Dorr M, Farmaki AE, Ferrucci L, Forer L, Fraser RM, Gabriel S, Levy S, Groop L, Harrison T, Hattersley A, Holmen OL, Hveem K, Kretzler M, Lee JC, McGue M, Meitinger T, Melzer D, Min JL, Mohlke KL, Vincent JB, Nauck M, Nickerson D, Palotie A, Pato M, Pirastu N, McInnis M, Richards JB, Sala C, Salomaa V, Schlessinger D, Schoenherr S, Slagboom PE, Small K, Spector T, Stambolian D, Tuke M, Tuomilehto J, Van den Berg LH, Van Rheenen W, Volker U, Wijmenga C, Toniolo D, Zeggini E, Gasparini P, Sampson MG, Wilson JF, Frayling T, de Bakker PI, Swertz MA, McCarroll S, Kooperberg C, Dekker A, Altshuler D, Willer C, Iacono W, Ripatti S, et al. (2016) A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 48: 1279-83. doi: 10.1038/ng.3643

Meo SA, Sheikh SA, Sattar K, Akram A, Hassan A, Meo AS, Usmani AM, Qalbani E, Ullah A (2019) Prevalence of Type 2 Diabetes Mellitus Among Men in the Middle East: A Retrospective Study. Am J Mens Health 13: 1557988319848577. doi: 10.1177/1557988319848577

Mersha TB, Ding L, He H, Alexander ES, Zhang X, Kurowski BG, Pilipenko V, Kottyan L, Martin LJ, Fardo DW (2015) Impact of Population Stratification on Family-Based Association in an Admixed Population. Int J Genomics 2015: 501617. doi: 10.1155/2015/501617

Michael D. Petraglia JIRe (2010) Tracking the Origin and Evolution of Human Populations in Arabia. The Evolution of Human Populations in Arabia-Paleoenvironments, Prehistory and Genetics. Springers
Miles CW, M. (2008) Quantitative trait locus (QTL) analysis. Nature Education 1: 208.
Mills MC, Rahal C (2019) A scientometric review of genome-wide association studies. Commun Biol 2: 9. doi: 10.1038/s42003-018-0261-x
Milner CM, Campbell RD (2001) Genetic organization of the human MHC class III region. Front Biosci 6: D914-26.
Minamino T, Orimo M, Shimizu I, Kunieda T, Yokoyama M, Ito T, Nojima A, Nabetani A, Oike Y, Matsubara H, Ishikawa F, Komuro I (2009) A crucial role for adipose tissue p53 in the regulation of insulin resistance. Nat Med 15: 1082-7. doi: 10.1038/nm.2014
Mohammad T, Xue Y, Evison M, Tyler-Smith C (2009) Genetic structure of nomadic Bedouin from Kuwait. Heredity (Edinb) 103: 425-33. doi: 10.1038/hdy.2009.72
Molchadsky A, Ezra O, Amendola PG, Krantz D, Kogan-Sakin I, Buganim Y, Rivlin N, Goldfinger N, Folgiero V, Falcioni R, Sarig R, Rotter V (2013) p53 is required for brown adipogenic differentiation and has a protective role against diet-induced obesity. Cell Death Differ 20: 774-83. doi: 10.1038/cdd.2013.9
Montgomery MK, De Nardo W, Watt MJ (2019) Impact of Lipotoxicity on Tissue "Cross Talk" and Metabolic Regulation. Physiology (Bethesda) 34: 134-149. doi: 10.1152/physiol.00037.2018
Mooradian AD (2009) Dyslipidemia in type 2 diabetes mellitus. Nat Clin Pract Endocrinol Metab 5: 150-9. doi: 10.1038/ncpendmet1066
Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7: 277-318.
Morton NE (1998) Significance levels in complex inheritance. Am J Hum Genet 62: 690-7. doi: 10.1086/301741
Moussa MA, Alsaeid M, Abdella N, Refai TM, Al-Sheikh N, Gomez JE (2008) Prevalence of type 2 diabetes mellitus among Kuwaiti children and adolescents. Med Princ Pract 17: 270-5. doi: 10.1159/000129604
Mtiraoui N, Turki A, Nemr R, Echtay A, Izzidi I, Al-Zaben GS, Irani-Hakime N, Keleshian SH, Mahjoub T, Almawi WY (2012) Contribution of common variants of ENPP1, IGF2BP2, KCNJ11, MLXIPL, PPARgamma, SLC30A8 and TCF7L2 to the risk of type 2 diabetes in Lebanese and Tunisian Arabs. Diabetes Metab 38: 444-9. doi: 10.1016/j.diabet.2012.05.002
Munro-Hay S (1991) Aksum: An African Civilization of Late Antiquity. Edinburgh University Press, Edinburgh
Muntner P, Shimbo D, Carey RM, Charleston JB, Gaillard T, Misra S, Myers MG, Ogedegbe G, Schwartz JE, Townsend RR, Urbina EM, Viera AJ, White WB, Wright JT, Jr. (2019) Measurement of Blood Pressure in Humans: A Scientific Statement From the American Heart Association. Hypertension 73: e35-e66. doi: 10.1161/HYP.0000000000000087
Muthukumaran S (2014) Between Archaeology and Text: The Origins of Rice Consumption and Cultivation in the Middle East and the Mediterranean. Papers from the Institute of Archaeology. Papers from the Institute of Archaeology 24. doi: http://doi.org/10.5334/pia.465
Na'amnih W, Romano-Zelekha O, Kabaha A, Rubin LP, Bilenko N, Jaber L, Honovich M, Shohat T (2014) Prevalence of consanguineous marriages and associated factors among Israeli Bedouins. J Community Genet 5: 395-8. doi: 10.1007/s12687-014-0188-y
Nalls MA, Guerreiro RJ, Simon-Sanchez J, Bras JT, Traynor BJ, Gibbs JR, Launer L, Hardy J, Singleton AB (2009) Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. Neurogenetics 10: 183-90. doi: 10.1007/s10048-009-0182-4

Nemoto S, Taguchi K, Matsumoto T, Kamata K, Kobayashi T (2012) Pravastatin normalizes ET-1-induced contraction in the aorta of type 2 diabetic OLETF rats by suppressing the KSR1/ERK complex. Am J Physiol Heart Circ Physiol 303: H893-902. doi: 10.1152/ajpheart.01128.2011

Newman B, Selby JV, King MC, Slemenda C, Fabsitz R, Friedman GD (1987) Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. Diabetologia 30: 763-8. doi: 10.1007/BF00275741

Newton-Cheh C, Larson MG, Vasan RS, Levy D, Bloch KD, Surti A, Guiducci C, Kathiresan S, Benjamin EJ, Struck J, Morgenthaler NG, Bergmann A, Blankenberg S, Kee F, Nilsson P, Yin X, Peltonen L, Vartiainen E, Salomaa V, Hirschhorn JN, Melander O, Wang TJ (2009) Association of common variants in NPPA and NPPB with circulating natriuretic peptides and blood pressure. Nat Genet 41: 348-53. doi: 10.1038/ng.328

Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, Mullany EC, Biryukov S, Abbafati C, Abera SF, Abraham JP, Abu-Rmeileh NM, Achoki T, AlBuhairan FS, Alemu ZA, Alfonso R, Ali MK, Ali R, Guzman NA, Ammar W, Anwari P, Banerjee A, Barquera S, Basu S, Bennett DA, Bhutta Z, Blore J, Cabral N, Nonato IC, Chang JC, Chowdhury R, Courville KJ, Criqui MH, Cundiff DK, Dabhadkar KC, Dandona L, Davis A, Dayama A, Dharmaratne SD, Ding EL, Durrani AM, Esteghamati A, Farzadfar F, Fay DF, Feigin VL, Flaxman A, Forouzanfar MH, Goto A, Green MA, Gupta R, Hafezi-Nejad N, Hankey GJ, Harewood HC, Havmoeller R, Hay S, Hernandez L, Husseini A, Idrisov BT, Ikeda N, Islami F, Jahangir E, Jassal SK, Jee SH, Jeffreys M, Jonas JB, Kabagambe EK, Khalifa SE, Kengne AP, Khader YS, Khang YH, Kim D, Kimokoti RW, Kinge JM, Kokubo Y, Kosen S, Kwan G, Lai T, Leinsalu M, Li Y, Liang X, Liu S, Logroscino G, Lotufo PA, Lu Y, Ma J, Mainoo NK, Mensah GA, Merriman TR, Mokdad AH, Moschandreas J, Naghavi M, Naheed A, Nand D, Narayan KM, Nelson EL, Neuhouser ML, Nisar MI, Ohkubo T, Oti SO, Pedroza A, et al. (2014) Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 384: 766-81. doi: 10.1016/S0140-6736(14)60460-8

Ng SW, Zaghloul S, Ali HI, Harrison G, Popkin BM (2011) The prevalence and trends of overweight, obesity and nutrition-related non-communicable diseases in the Arabian Gulf States. Obes Rev 12: 1-13. doi: 10.1111/j.1467-789X.2010.00750.x

O'Beirne SL, Salit J, Rodriguez-Flores JL, Staudt MR, Abi Khalil C, Fakhro KA, Robay A, Ramstetter MD, Al-Azwani IK, Malek JA, Zirie M, Jayyousi A, Badii R, Al-Nabet Al-Marri A, Chiuchiolo MJ, Al-Shakaki A, Chidiac O, Gharbiah M, Bener A, Stadler D, Hackett NR, Mezey JG, Crystal RG (2016) Type 2 Diabetes Risk Allele Loci in the Qatari Population. PLoS One 11: e0156834. doi: 10.1371/journal.pone.0156834

O'Beirne SL, Salit J, Rodriguez-Flores JL, Staudt MR, Abi Khalil C, Fakhro KA, Robay A, Ramstetter MD, Malek JA, Zirie M, Jayyousi A, Badii R, Al-Nabet Al-Marri A, Bener A, Mahmoud M, Chiuchiolo MJ, Al-Shakaki A, Chidiac O, Stadler D, Mezey JG, Crystal RG (2018) Exome sequencing-based identification of novel type 2 diabetes risk allele loci in the Qatari population. PLoS One 13: e0199837. doi: 10.1371/journal.pone.0199837

Okada Y, Sim X, Go MJ, Wu JY, Gu D, Takeuchi F, Takahashi A, Maeda S, Tsunoda T, Chen P, Lim SC, Wong TY, Liu J, Young TL, Aung T, Seielstad M, Teo YY, Kim YJ, Lee JY, Han BG, Kang D, Chen CH, Tsai FJ, Chang LC, Fann SJ, Mei H, Rao DC, Hixson JE, Chen S, Katsuya T, Isono M, Ogihara T, Chambers JC, Zhang W, Kooner JS, KidneyGen C, Consortium CK, Albrecht E, consortium G, Yamamoto K, Kubo M, Nakamura Y, Kamatani N, Kato N, He J, Chen YT, Cho YS, Tai ES, Tanaka T (2012) Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. Nat Genet 44: 904-9. doi: 10.1038/ng.2352

Onengut-Gumuscu S, Chen WM, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, Farber E, Bonnie JK, Szpak M, Schofield E, Achuthan P, Guo H, Fortune MD, Stevens H, Walker NM, Ward LD, Kundaje

A, Kellis M, Daly MJ, Barrett JC, Cooper JD, Deloukas P, Type 1 Diabetes Genetics C, Todd JA, Wallace C, Concannon P, Rich SS (2015) Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat Genet 47: 381-6. doi: 10.1038/ng.3245

Osman W, Tay GK, Alsafar H (2018a) Multiple genetic variations confer risks for obesity and type 2 diabetes mellitus in arab descendants from UAE. Int J Obes (Lond) 42: 1345-1353. doi: 10.1038/s41366-018-0057-6

Osman WM, Jelinek HF, Tay GK, Hassan MH, Almahmeed W, Khandoker AH, Khalaf K, Alsafar HS (2019) Genetics of diabetic kidney disease: A follow-up study in the Arab population of the United Arab Emirates. Mol Genet Genomic Med 7: e985. doi: 10.1002/mgg3.985

Osman WM, Jelinek HF, Tay GK, Khandoker AH, Khalaf K, Almahmeed W, Hassan MH, Alsafar HS (2018b) Clinical and genetic associations of renal function and diabetic kidney disease in the United Arab Emirates: a cross-sectional study. BMJ Open 8: e020759. doi: 10.1136/bmjopen-2017-020759

Pallotta MT, Tascini G, Crispoldi R, Orabona C, Mondanelli G, Grohmann U, Esposito S (2019) Wolfram syndrome, a rare neurodegenerative disease: from pathogenesis to future treatment perspectives. J Transl Med 17: 238. doi: 10.1186/s12967-019-1993-1

Park HS, Park JY, Cho SI (2006) Familial aggregation of the metabolic syndrome in Korean families with adolescents. Atherosclerosis 186: 215-21. doi: 10.1016/j.atherosclerosis.2005.07.019

Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, Garnaas M, Tin A, Sorice R, Li Y, Taliun D, Olden M, Foster M, Yang Q, Chen MH, Pers TH, Johnson AD, Ko YA, Fuchsberger C, Tayo B, Nalls M, Feitosa MF, Isaacs A, Dehghan A, d'Adamo P, Adeyemo A, Dieffenbach AK, Zonderman AB, Nolte IM, van der Most PJ, Wright AF, Shuldiner AR, Morrison AC, Hofman A, Smith AV, Dreisbach AW, Franke A, Uitterlinden AG, Metspalu A, Tonjes A, Lupo A, Robino A, Johansson A, Demirkan A, Kollerits B, Freedman BI, Ponte B, Oostra BA, Paulweber B, Kramer BK, Mitchell BD, Buckley BM, Peralta CA, Hayward C, Helmer C, Rotimi CN, Shaffer CM, Muller C, Sala C, van Duijn CM, Saint-Pierre A, Ackermann D, Shriner D, Ruggiero D, Toniolo D, Lu Y, Cusi D, Czamara D, Ellinghaus D, Siscovick DS, Ruderfer D, Gieger C, Grallert H, Rochtchina E, Atkinson EJ, Holliday EG, Boerwinkle E, Salvi E, Bottinger EP, Murgia F, Rivadeneira F, Ernst F, Kronenberg F, Hu FB, Navis GJ, Curhan GC, Ehret GB, Homuth G, Coassin S, Thun GA, Pistis G, Gambaro G, Malerba G, Montgomery GW, Eiriksdottir G, Jacobs G, Li G, Wichmann HE, Campbell H, Schmidt H, et al. (2016) Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat Commun 7: 10023. doi: 10.1038/ncomms10023

Pedrazzini T (2004) Importance of NPY Y1 receptor-mediated pathways: assessment using NPY Y1 receptor knockouts. Neuropeptides 38: 267-75. doi: 10.1016/j.npep.2004.05.007

Pedrazzini T, Seydoux J, Kunstner P, Aubert JF, Grouzmann E, Beermann F, Brunner HR (1998) Cardiovascular response, feeding behavior and locomotor activity in mice lacking the NPY Y1 receptor. Nat Med 4: 722-6.

Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ (2012) Genomic patterns of homozygosity in worldwide human populations. Am J Hum Genet 91: 275-92. doi: 10.1016/j.ajhg.2012.06.014

Penrose (1935) The detection of autosomal linkage in data which consist of pairs of brothers and sisters of unspecied parentage. Ann. Eugen 8: 133-138.

Peripolli E, Munari DP, Silva M, Lima ALF, Irgang R, Baldi F (2017) Runs of homozygosity: current knowledge and applications in livestock. Anim Genet 48: 255-271. doi: 10.1111/age.12526

Podolsky S, Leopold NA, Sax DS (1972) Increased frequency of diabetes mellitus in patients with Huntington's chorea. Lancet 1: 1356-8.

Popejoy AB, Fullerton SM (2016) Genomics is failing on diversity. Nature 538: 161-164. doi: 10.1038/538161a

Prawitt J, Caron S, Staels B (2011) Bile acid metabolism and the pathogenesis of type 2 diabetes. Curr Diab Rep 11: 160-6. doi: 10.1007/s11892-011-0187-x
Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904-9. doi: 10.1038/ng1847
Program NCE (2002) Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) National Heart, Lung, and Blood Institute, Bethesda
Project Team SG (2015) The Saudi Human Genome Program: An oasis in the desert of Arab medicine is providing clues to genetic disease. IEEE Pulse 6: 22-6. doi: 10.1109/MPUL.2015.2476541
Puig M, Casillas S, Villatoro S, Caceres M (2015) Human inversions and their functional consequences. Brief Funct Genomics 14: 369-79. doi: 10.1093/bfgp/elv020
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-75. doi: 10.1086/519795
Qaddoumi M, Al-Khamis Y, Channanath A, Tuomilehto J, Badawi D (2019) The Status of Metabolic Control in Patients With Type 2 Diabetes Attending Dasman Diabetes Institute, Kuwait. Front Endocrinol (Lausanne) 10: 412. doi: 10.3389/fendo.2019.00412
Qanbari S (2019) On the Extent of Linkage Disequilibrium in the Genome of Farm Animals. Front Genet 10: 1304. doi: 10.3389/fgene.2019.01304
Qiu YH, Deng FY, Li MJ, Lei SF (2014) Identification of novel risk genes associated with type 1 diabetes mellitus using a genome-wide gene-based association analysis. J Diabetes Investig 5: 649-56. doi: 10.1111/jdi.12228
Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. Nat Genet 23: 437-41. doi: 10.1038/70550
Raja'a YA, Bin Mohanna MA (2005) Overweight and obesity among schoolchildren in Sana'a City, Yemen. Ann Nutr Metab 49: 342-5. doi: 10.1159/000087723
Rankinen T, Perusse L, Weisnagel SJ, Snyder EE, Chagnon YC, Bouchard C (2002) The human obesity gene map: the 2001 update. Obes Res 10: 196-243. doi: 10.1038/oby.2002.30
Recio C, Lucy D, Iveson P, Iqbal AJ, Valaris S, Wynne G, Russell AJ, Choudhury RP, O'Callaghan C, Monaco C, Greaves DR (2018) The Role of Metabolite-Sensing G Protein-Coupled Receptors in Inflammation and Metabolic Disease. Antioxid Redox Signal 29: 237-256. doi: 10.1089/ars.2017.7168
Rees SD, Islam M, Hydrie MZ, Chaudhary B, Bellary S, Hashmi S, O'Hare JP, Kumar S, Sanghera DK, Chaturvedi N, Barnett AH, Shera AS, Weedon MN, Basit A, Frayling TM, Kelly MA, Jafar TH (2011) An FTO variant is associated with Type 2 diabetes in South Asian populations after accounting for body mass index and waist circumference. Diabet Med 28: 673-80. doi: 10.1111/j.1464-5491.2011.03257.x
Richard Tapper SZ (1996) Culinary Cultures of the Middle East. British Journal of Middle Eastern Studies 23: 239-241.
Ristow M (2004) Neurodegenerative disorders associated with diabetes mellitus. J Mol Med (Berl) 82: 510-29. doi: 10.1007/s00109-004-0552-1
Rodhan Khthir FLE (2014) The Metabolic Syndrome in Rural UAE: The Effect of Gender, Ethnicity and the Environment in its Prevalence. Journal of Metabolic Syndrome 3:4. doi: DOI: 10.4172/2167-0943.1000159
Rodriguez-Flores JL, Fakhro K, Agosto-Perez F, Ramstetter MD, Arbiza L, Vincent TL, Robay A, Malek JA, Suhre K, Chouchane L, Badii R, Al-Nabet Al-Marri A, Abi Khalil C, Zirie M, Jayyousi A, Salit J,

Keinan A, Clark AG, Crystal RG, Mezey JG (2016) Indigenous Arabs are descendants of the earliest split from ancient Eurasian populations. Genome Res 26: 151-62. doi: 10.1101/gr.191478.115
Rogan E (2015) The fall of the Ottomans : the Great War in the Middle East. Basic Books
Romeo S, Yin W, Kozlitina J, Pennacchio LA, Boerwinkle E, Hobbs HH, Cohen JC (2009) Rare loss-of-function mutations in ANGPTL family members contribute to plasma triglyceride levels in humans. J Clin Invest 119: 70-9. doi: 10.1172/JCI37118
Rongling Wu C-XM, George Casella (2007) Statistical Genetics or Quantitative Traits: Linkage, Maps and QTL. Springer
Rosa A, Brehem A (2011) African human mtDNA phylogeography at-a-glance. J Anthropol Sci 89: 25-58. doi: 10.4436/jass.89006
Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298: 2381-5. doi: 10.1126/science.1078311
Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, Barengo NC, Beaton AZ, Benjamin EJ, Benziger CP, Bonny A, Brauer M, Brodmann M, Cahill TJ, Carapetis J, Catapano AL, Chugh SS, Cooper LT, Coresh J, Criqui M, DeCleene N, Eagle KA, Emmons-Bell S, Feigin VL, Fernandez-Sola J, Fowkes G, Gakidou E, Grundy SM, He FJ, Howard G, Hu F, Inker L, Karthikeyan G, Kassebaum N, Koroshetz W, Lavie C, Lloyd-Jones D, Lu HS, Mirijello A, Temesgen AM, Mokdad A, Moran AE, Muntner P, Narula J, Neal B, Ntsekhe M, Moraes de Oliveira G, Otto C, Owolabi M, Pratt M, Rajagopalan S, Reitsma M, Ribeiro ALP, Rigotti N, Rodgers A, Sable C, Shakil S, Sliwa-Hahnle K, Stark B, Sundstrom J, Timpel P, Tleyjeh IM, Valgimigli M, Vos T, Whelton PK, Yacoub M, Zuhlke L, Murray C, Fuster V, Group G-N-JGBoCDW (2020) Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. J Am Coll Cardiol 76: 2982-3021. doi: 10.1016/j.jacc.2020.11.010
Rousset F (2002) Inbreeding and relatedness coefficients: what do they measure? Heredity (Edinb) 88: 371-80. doi: 10.1038/sj.hdy.6800065
Ryan NM, Morris SW, Porteous DJ, Taylor MS, Evans KL (2014) SuRFing the genomics wave: an R package for prioritising SNPs by functionality. Genome Med 6: 79. doi: 10.1186/s13073-014-0079-1
Saade S, Cazier JB, Ghassibe-Sabbagh M, Youhanna S, Badro DA, Kamatani Y, Hager J, Yeretzian JS, El-Khazen G, Haber M, Salloum AK, Douaihy B, Othman R, Shasha N, Kabbani S, Bayeh HE, Chammas E, Farrall M, Gauguier D, Platt DE, Zalloua PA, consortium F (2011) Large scale association analysis identifies three susceptibility loci for coronary artery disease. PLoS One 6: e29427. doi: 10.1371/journal.pone.0029427
Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J, Jones CG, Zaitlen NA, Varilo T, Kaakinen M, Sovio U, Ruokonen A, Laitinen J, Jakkula E, Coin L, Hoggart C, Collins A, Turunen H, Gabriel S, Elliot P, McCarthy MI, Daly MJ, Jarvelin MR, Freimer NB, Peltonen L (2009) Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet 41: 35-46. doi: 10.1038/ng.271
Salvatore P, Hanash CR, Kido Y, Imai Y, Accili D (1998) Identification of sirm, a novel insulin-regulated SH3 binding protein that associates with Grb-2 and FYN. J Biol Chem 273: 6989-97. doi: 10.1074/jbc.273.12.6989
Sanghera DK, Bejar C, Sharma S, Gupta R, Blackett PR (2019) Obesity genetics and cardiometabolic health: Potential for risk prediction. Diabetes Obes Metab 21: 1088-1100. doi: 10.1111/dom.13641
Saraswathi S, Al-Khawaga S, Elkum N, Hussain K (2019) A Systematic Review of Childhood Diabetes Research in the Middle East Region. Front Endocrinol (Lausanne) 10: 805. doi: 10.3389/fendo.2019.00805

Saxton SN, Clark BJ, Withers SB, Eringa EC, Heagerty AM (2019) Mechanistic Links Between Obesity, Diabetes, and Blood Pressure: Role of Perivascular Adipose Tissue. Physiol Rev 99: 1701-1763. doi: 10.1152/physrev.00034.2018

Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M (2012) Linking disease associations with regulatory information in the human genome. Genome Res 22: 1748-59. doi: 10.1101/gr.136127.111

Schmidt-Kastner R, Yamamoto H, Hamasaki D, Yamamoto H, Parel JM, Schmitz C, Dorey CK, Blanks JC, Preising MN (2008) Hypoxia-regulated components of the U4/U6.U5 tri-small nuclear riboprotein complex: possible role in autosomal dominant retinitis pigmentosa. Mol Vis 14: 125-35.

Schulz TJ, Tseng YH (2009) Emerging role of bone morphogenetic proteins in adipogenesis and energy metabolism. Cytokine Growth Factor Rev 20: 523-31. doi: 10.1016/j.cytogfr.2009.10.019

Scott EM, Halees A, Itan Y, Spencer EG, He Y, Azab MA, Gabriel SB, Belkadi A, Boisson B, Abel L, Clark AG, Greater Middle East Variome C, Alkuraya FS, Casanova JL, Gleeson JG (2016) Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. Nat Genet 48: 1071-6. doi: 10.1038/ng.3592

Scutari M, Mackay I, Balding D (2016) Using Genetic Distance to Infer the Accuracy of Genomic Prediction. PLoS Genet 12: e1006288. doi: 10.1371/journal.pgen.1006288

Segal R (2002) Islam's Black Slaves: The Other Black Diaspora. Atlantic Books, London

Seidelmann SB, Li L, Shen GQ, Topol EJ, Wang QK (2008) Identification of a novel locus for triglyceride on chromosome 1p31-32 in families with premature CAD and MI. J Lipid Res 49: 1034-8. doi: 10.1194/jlr.M700576-JLR200

Sella G, Barton NH (2019) Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. Annu Rev Genomics Hum Genet 20: 461-493. doi: 10.1146/annurev-genom-083115-022316

Serwin AB, Mysliwiec H, Laudanska H, Chodynicka B (2002) Linear IgA bullous dermatosis in a diabetic patient with chronic renal failure. Int J Dermatol 41: 778-80.

Shah SM, Aziz F, Al Meskari F, Al Kaabi J, Khan UI, Jaacks LM (2020) Metabolic syndrome among children aged 6 to 11 years, Al Ain, United Arab Emirates: Role of obesity. Pediatr Diabetes 21: 735-742. doi: 10.1111/pedi.13027

Shalata A, Ramirez MC, Desnick RJ, Priedigkeit N, Buettner C, Lindtner C, Mahroum M, Abdul-Ghani M, Dong F, Arar N, Camacho-Vanegas O, Zhang R, Camacho SC, Chen Y, Ibdah M, DeFronzo R, Gillespie V, Kelley K, Dynlacht BD, Kim S, Glucksman MJ, Borochowitz ZU, Martignetti JA (2013) Morbid obesity resulting from inactivation of the ciliary protein CEP19 in humans and mice. Am J Hum Genet 93: 1061-71. doi: 10.1016/j.ajhg.2013.10.025

Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, Manolio T, Rudan I, McKeigue P, Wilson JF, Campbell H (2011) Abundant pleiotropy in human complex diseases and traits. Am J Hum Genet 89: 607-18. doi: 10.1016/j.ajhg.2011.10.004

Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso C, Costa MD, Musilova E, Macaulay V, Richards MB, Cerny V, Pereira L (2012) The Expansion of mtDNA Haplogroup L3 within and out of Africa. Mol Biol Evol 29: 915-27. doi: 10.1093/molbev/msr245

Speakman JR (2015) The 'Fat Mass and Obesity Related' (FTO) gene: Mechanisms of Impact on Obesity and Energy Balance. Curr Obes Rep 4: 73-91. doi: 10.1007/s13679-015-0143-1

Speidel D, Salehi A, Obermueller S, Lundquist I, Brose N, Renstrom E, Rorsman P (2008) CAPS1 and CAPS2 regulate stability and recruitment of insulin granules in mouse pancreatic beta cells. Cell Metab 7: 57-67. doi: 10.1016/j.cmet.2007.11.009

Spracklen CN, Chen P, Kim YJ, Wang X, Cai H, Li S, Long J, Wu Y, Wang YX, Takeuchi F, Wu JY, Jung KJ, Hu C, Akiyama K, Zhang Y, Moon S, Johnson TA, Li H, Dorajoo R, He M, Cannon ME, Roman TS,

Salfati E, Lin KH, Guo X, Sheu WHH, Absher D, Adair LS, Assimes TL, Aung T, Cai Q, Chang LC, Chen CH, Chien LH, Chuang LM, Chuang SC, Du S, Fan Q, Fann CSJ, Feranil AB, Friedlander Y, Gordon-Larsen P, Gu D, Gui L, Guo Z, Heng CK, Hixson J, Hou X, Hsiung CA, Hu Y, Hwang MY, Hwu CM, Isono M, Juang JJ, Khor CC, Kim YK, Koh WP, Kubo M, Lee IT, Lee SJ, Lee WJ, Liang KW, Lim B, Lim SH, Liu J, Nabika T, Pan WH, Peng H, Quertermous T, Sabanayagam C, Sandow K, Shi J, Sun L, Tan PC, Tan SP, Taylor KD, Teo YY, Toh SA, Tsunoda T, van Dam RM, Wang A, Wang F, Wang J, Wei WB, Xiang YB, Yao J, Yuan JM, Zhang R, Zhao W, Chen YI, Rich SS, Rotter JI, Wang TD, Wu T, Lin X, Han BG, Tanaka T, Cho YS, Katsuya T, Jia W, et al. (2017) Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. Hum Mol Genet 26: 1770-1784. doi: 10.1093/hmg/ddx062

Straczkowski M, Kowalska I (2008) The role of skeletal muscle sphingolipids in the development of insulin resistance. Rev Diabet Stud 5: 13-24. doi: 10.1900/RDS.2008.5.13

Stutzmann F, Vatin V, Cauchi S, Morandi A, Jouret B, Landt O, Tounian P, Levy-Marchal C, Buzzetti R, Pinelli L, Balkau B, Horber F, Bougneres P, Froguel P, Meyre D (2007) Non-synonymous polymorphisms in melanocortin-4 receptor protect against obesity: the two facets of a Janus obesity gene. Hum Mol Genet 16: 1837-44. doi: 10.1093/hmg/ddm132

Summers SA (2010) Sphingolipids and insulin resistance: the five Ws. Curr Opin Lipidol 21: 128-35. doi: 10.1097/MOL.0b013e3283373b66

Sun Y, Sun M, Liu B, Du Y, Rong S, Xu G, Snetselaar LG, Bao W (2019) Inverse Association Between Serum Vitamin B12 Concentration and Obesity Among Adults in the United States. Front Endocrinol (Lausanne) 10: 414. doi: 10.3389/fendo.2019.00414

Sung J, Lee K, Song YM (2009) Heritabilities of the metabolic syndrome phenotypes and related factors in Korean twins. J Clin Endocrinol Metab 94: 4946-52. doi: 10.1210/jc.2009-1268

Surakka I, Horikoshi M, Magi R, Sarin AP, Mahajan A, Lagou V, Marullo L, Ferreira T, Miraglio B, Timonen S, Kettunen J, Pirinen M, Karjalainen J, Thorleifsson G, Hagg S, Hottenga JJ, Isaacs A, Ladenvall C, Beekman M, Esko T, Ried JS, Nelson CP, Willenborg C, Gustafsson S, Westra HJ, Blades M, de Craen AJ, de Geus EJ, Deelen J, Grallert H, Hamsten A, Havulinna AS, Hengstenberg C, Houwing-Duistermaat JJ, Hypponen E, Karssen LC, Lehtimaki T, Lyssenko V, Magnusson PK, Mihailov E, Muller-Nurasyid M, Mpindi JP, Pedersen NL, Penninx BW, Perola M, Pers TH, Peters A, Rung J, Smit JH, Steinthorsdottir V, Tobin MD, Tsernikova N, van Leeuwen EM, Viikari JS, Willems SM, Willemsen G, Schunkert H, Erdmann J, Samani NJ, Kaprio J, Lind L, Gieger C, Metspalu A, Slagboom PE, Groop L, van Duijn CM, Eriksson JG, Jula A, Salomaa V, Boomsma DI, Power C, Raitakari OT, Ingelsson E, Jarvelin MR, Thorsteinsdottir U, Franke L, Ikonen E, Kallioniemi O, Pietiainen V, Lindgren CM, Stefansson K, Palotie A, McCarthy MI, Morris AP, Prokopenko I, Ripatti S, Consortium E (2015) The impact of low-frequency and rare variants on lipid levels. Nat Genet 47: 589-97. doi: 10.1038/ng.3300

Surendran P, Drenos F, Young R, Warren H, Cook JP, Manning AK, Grarup N, Sim X, Barnes DR, Witkowska K, Staley JR, Tragante V, Tukiainen T, Yaghootkar H, Masca N, Freitag DF, Ferreira T, Giannakopoulou O, Tinker A, Harakalova M, Mihailov E, Liu C, Kraja AT, Fallgaard Nielsen S, Rasheed A, Samuel M, Zhao W, Bonnycastle LL, Jackson AU, Narisu N, Swift AJ, Southam L, Marten J, Huyghe JR, Stancakova A, Fava C, Ohlsson T, Matchan A, Stirrups KE, Bork-Jensen J, Gjesing AP, Kontto J, Perola M, Shaw-Hawkins S, Havulinna AS, Zhang H, Donnelly LA, Groves CJ, Rayner NW, Neville MJ, Robertson NR, Yiorkas AM, Herzig KH, Kajantie E, Zhang W, Willems SM, Lannfelt L, Malerba G, Soranzo N, Trabetti E, Verweij N, Evangelou E, Moayyeri A, Vergnaud AC, Nelson CP, Poveda A, Varga TV, Caslake M, de Craen AJ, Trompet S, Luan J, Scott RA, Harris SE, Liewald DC, Marioni R, Menni C, Farmaki AE, Hallmans G, Renstrom F, Huffman JE, Hassinen M, Burgess S, Vasan RS, Felix JF, Consortium CH-HF, Uria-Nickelsen M, Malarstig A, Reily DF, Hoek

M, Vogt T, Lin H, Lieb W, EchoGen C, Traylor M, Markus HF, Consortium M, Highland HM, Justice AE, Marouli E, Consortium G, et al. (2016) Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. Nat Genet 48: 1151-1161. doi: 10.1038/ng.3654

Tadmouri GO, Al Ali MT, Al-Haj Ali S, Al Khaja N (2006) CTGA: the database for genetic disorders in Arab populations. Nucleic Acids Res 34: D602-6. doi: 10.1093/nar/gkj015

Tadmouri GO, Nair P, Obeid T, Al Ali MT, Al Khaja N, Hamamy HA (2009) Consanguinity and reproductive health among Arabs. Reprod Health 6: 17. doi: 10.1186/1742-4755-6-17

Tadmouri GO, Sastry KS, Chouchane L (2014) Arab gene geography: From population diversities to personalized medical genomics. Glob Cardiol Sci Pract 2014: 394-408. doi: 10.5339/gcsp.2014.54

Teeuw ME, Henneman L, Bochdanovits Z, Heutink P, Kuik DJ, Cornel MC, Ten Kate LP (2010) Do consanguineous parents of a child affected by an autosomal recessive disease have more DNA identical-by-descent than similarly-related parents with healthy offspring? Design of a case-control study. BMC Med Genet 11: 113. doi: 10.1186/1471-2350-11-113

Thareja G, John SE, Hebbar P, Behbehani K, Thanaraj TA, Alsmadi O (2015) Sequence and analysis of a whole genome from Kuwaiti population subgroup of Persian ancestry. BMC Genomics 16: 92. doi: 10.1186/s12864-015-1233-x

Theyab JB, Al-Bustan S, Crawford MH (2012) The genetic structure of the Kuwaiti population: mtDNA Inter- and intra-population variation. Hum Biol 84: 379-403. doi: 10.3378/027.084.0403

Thurner M, van de Bunt M, Torres JM, Mahajan A, Nylander V, Bennett AJ, Gaulton KJ, Barrett A, Burrows C, Bell CG, Lowe R, Beck S, Rakyan VK, Gloyn AL, McCarthy MI (2018) Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. Elife 7. doi: 10.7554/eLife.31977

Tian H, Li CS, Zhao KW, Wang JC, Duarte ME, David CL, Phan K, Atti E, Brochmann EJ, Murray SS (2015) A carboxy terminal BMP/TGF-beta binding site in secreted phosphoprotein 24 kD independently affects BMP-2 activity. J Cell Biochem 116: 667-76. doi: 10.1002/jcb.25023

Tim Armstrong AB, Joanne Davies (2000) Physical activity patterns of Australian adults. Canberra: Australian Institute of Healthand Welfare

Tischfield JA (1997) Loss of heterozygosity or: how I learned to stop worrying and love mitotic recombination. Am J Hum Genet 61: 995-9. doi: 10.1086/301617

Tomei S, Mamtani R, Al Ali R, Elkum N, Abdulmalik M, Ismail A, Cheema S, Rouh HA, Aigha, II, Hani F, Al-Samraye S, Taher Aseel M, El Emadi N, Al Mujalli A, Abdelkerim A, Youssif S, Worschech A, El Sebakhy E, Temanni R, Khanna V, Wang E, Kizhakayil D, Al-Thani AA, Al-Thani M, Lowenfels A, Marincola FM, Sheikh J, Chouchane L (2015) Obesity susceptibility loci in Qataris, a highly consanguineous Arabian population. J Transl Med 13: 119. doi: 10.1186/s12967-015-0459-3

Tomkin GH, Owens D (2016) Obesity diabetes and the role of bile acids in metabolism. J Transl Int Med 4: 73-80. doi: 10.1515/jtim-2016-0018

Tremblay F, Brule S, Hee Um S, Li Y, Masuda K, Roden M, Sun XJ, Krebs M, Polakiewicz RD, Thomas G, Marette A (2007) Identification of IRS-1 Ser-1101 as a target of S6K1 in nutrient- and obesity-induced insulin resistance. Proc Natl Acad Sci U S A 104: 14056-61. doi: 10.1073/pnas.0706517104

Triki-Fendri S, Alfadhli S, Ayadi I, Kharrat N, Ayadi H, Rebai A (2010) Genetic structure of Kuwaiti population revealed by Y-STR diversity. Ann Hum Biol 37: 827-35. doi: 10.3109/03014461003720296

Tuomilehto-Wolf E, Tuomilehto J, Hitman GA, Nissinen A, Stengard J, Pekkanen J, Kivinen P, Kaarsalo E, Karvonen MJ (1993) Genetic susceptibility to non-insulin dependent diabetes mellitus and glucose intolerance are located in HLA region. BMJ 307: 155-9.

Turchin MC, Chiang CW, Palmer CD, Sankararaman S, Reich D, Genetic Investigation of ATC, Hirschhorn JN (2012) Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nat Genet 44: 1015-9. doi: 10.1038/ng.2368

Tzameli I (2016) GPCRs - Pivotal Players in Metabolism. Trends Endocrinol Metab 27: 597-599. doi: 10.1016/j.tem.2016.07.003

Um SH, D'Alessio D, Thomas G (2006) Nutrient overload, insulin resistance, and ribosomal protein S6 kinase 1, S6K1. Cell Metab 3: 393-402. doi: 10.1016/j.cmet.2006.05.003

Usanga EA, Ameen R (2000) Glucose-6-phosphate dehydrogenase deficiency in Kuwait, Syria, Egypt, Iran, Jordan and Lebanon. Hum Hered 50: 158-61. doi: 10.1159/000022906

Valdes AM, Thomson G, Type 1 Diabetes Genetics C (2009) Several loci in the HLA class III region are associated with T1D risk after adjusting for DRB1-DQB1. Diabetes Obes Metab 11 Suppl 1: 46-52. doi: 10.1111/j.1463-1326.2008.01002.x

van den Berg L, van Beekum O, Heutink P, Felius BA, van de Heijning MP, Strijbis S, van Spaendonk R, Piancatelli D, Garner KM, El Aouad R, Sistermans E, Adan RA, Delemarre-van de Waal HA (2011) Melanocortin-4 receptor gene mutations in a Dutch cohort of obese children. Obesity (Silver Spring) 19: 604-11. doi: 10.1038/oby.2010.259

van der Harst P, Verweij N (2018) Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. Circ Res 122: 433-443. doi: 10.1161/CIRCRESAHA.117.312086

van Hagen PM, Hooijkaas H, Vd Beemd MW, Verjans G, Baarsma GS (2003) T-gammadelta receptor restriction in peripheral lymphocytes of patients with Behcet's disease. Adv Exp Med Biol 528: 267-8. doi: 10.1007/0-306-48382-3_53

Verge CF, Vardi P, Babu S, Bao F, Erlich HA, Bugawan T, Tiosano D, Yu L, Eisenbarth GS, Fain PR (1998) Evidence for oligogenic inheritance of type 1 diabetes in a large Bedouin Arab family. J Clin Invest 102: 1569-75. doi: 10.1172/JCI3379

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 101: 5-22. doi: 10.1016/j.ajhg.2017.06.005

Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, Burtt NP, Fuchsberger C, Li Y, Erdmann J, Frayling TM, Heid IM, Jackson AU, Johnson T, Kilpelainen TO, Lindgren CM, Morris AP, Prokopenko I, Randall JC, Saxena R, Soranzo N, Speliotes EK, Teslovich TM, Wheeler E, Maguire J, Parkin M, Potter S, Rayner NW, Robertson N, Stirrups K, Winckler W, Sanna S, Mulas A, Nagaraja R, Cucca F, Barroso I, Deloukas P, Loos RJ, Kathiresan S, Munroe PB, Newton-Cheh C, Pfeufer A, Samani NJ, Schunkert H, Hirschhorn JN, Altshuler D, McCarthy MI, Abecasis GR, Boehnke M (2012) The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genet 8: e1002793. doi: 10.1371/journal.pgen.1002793

Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, Huffman JE, Assimes TL, Lorenz K, Zhu X, Hilliard AT, Judy RL, Huang J, Lee KM, Klarin D, Pyarajan S, Danesh J, Melander O, Rasheed A, Mallick NH, Hameed S, Qureshi IH, Afzal MN, Malik U, Jalal A, Abbas S, Sheng X, Gao L, Kaestner KH, Susztak K, Sun YV, DuVall SL, Cho K, Lee JS, Gaziano JM, Phillips LS, Meigs JB, Reaven PD, Wilson PW, Edwards TL, Rader DJ, Damrauer SM, O'Donnell CJ, Tsao PS, Consortium H, Regeneron Genetics C, Program VAMV, Chang KM, Voight BF, Saleheen D (2020) Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. Nat Genet 52: 680-691. doi: 10.1038/s41588-020-0637-y

Vyas DN, Mulligan CJ (2019) Analyses of Neanderthal introgression suggest that Levantine and southern Arabian populations have a shared population history. Am J Phys Anthropol 169: 227-239. doi: 10.1002/ajpa.23818

Wade TD, Gordon S, Medland S, Bulik CM, Heath AC, Montgomery GW, Martin NG (2013) Genetic variants associated with disordered eating. Int J Eat Disord 46: 594-608. doi: 10.1002/eat.22133

Wallace C, Xue MZ, Newhouse SJ, Marcano AC, Onipinla AK, Burke B, Gungadoo J, Dobson RJ, Brown M, Connell JM, Dominiczak A, Lathrop GM, Webster J, Farrall M, Mein C, Samani NJ, Caulfield MJ, Clayton DG, Munroe PB (2006) Linkage analysis using co-phenotypes in the BRIGHT study reveals novel potential susceptibility loci for hypertension. Am J Hum Genet 79: 323-31. doi: 10.1086/506370

Wang X, Zhao X, Gao X, Mei Y, Wu M (2013) A new role of p53 in regulating lipid metabolism. J Mol Cell Biol 5: 147-50. doi: 10.1093/jmcb/mjs064

Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, Ntalla I, Surendran P, Liu C, Cook JP, Kraja AT, Drenos F, Loh M, Verweij N, Marten J, Karaman I, Lepe MP, O'Reilly PF, Knight J, Snieder H, Kato N, He J, Tai ES, Said MA, Porteous D, Alver M, Poulter N, Farrall M, Gansevoort RT, Padmanabhan S, Magi R, Stanton A, Connell J, Bakker SJ, Metspalu A, Shields DC, Thom S, Brown M, Sever P, Esko T, Hayward C, van der Harst P, Saleheen D, Chowdhury R, Chambers JC, Chasman DI, Chakravarti A, Newton-Cheh C, Lindgren CM, Levy D, Kooner JS, Keavney B, Tomaszewski M, Samani NJ, Howson JM, Tobin MD, Munroe PB, Ehret GB, Wain LV, International Consortium of Blood Pressure GA, Consortium B, Lifelines Cohort S, Understanding Society Scientific g, Consortium CHDE, Exome BPC, Consortium TDG, Go TDC, Cohorts for H, Ageing Research in Genome Epidemiology BPEC, International Genomics of Blood Pressure C, group UKBCCBw (2017) Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat Genet 49: 403-415. doi: 10.1038/ng.3768

Warsy AS, Al-Jaser MH, Albdass A, Al-Daihan S, Alanazi M (2014) Is consanguinity prevalence decreasing in Saudis?: A study in two generations. Afr Health Sci 14: 314-21. doi: 10.4314/ahs.v14i2.5

Warsy AS, El-Hazmi MA (2001) G6PD deficiency, distribution and variants in Saudi Arabia: an overview. Ann Saudi Med 21: 174-7. doi: 10.5144/0256-4947.2001.174

Watanabe M, Houten SM, Wang L, Moschetta A, Mangelsdorf DJ, Heyman RA, Moore DD, Auwerx J (2004) Bile acids lower triglyceride levels via a pathway involving FXR, SHP, and SREBP-1c. J Clin Invest 113: 1408-18. doi: 10.1172/JCI21025

Waterworth DM, Ricketts SL, Song K, Chen L, Zhao JH, Ripatti S, Aulchenko YS, Zhang W, Yuan X, Lim N, Luan J, Ashford S, Wheeler E, Young EH, Hadley D, Thompson JR, Braund PS, Johnson T, Struchalin M, Surakka I, Luben R, Khaw KT, Rodwell SA, Loos RJ, Boekholdt SM, Inouye M, Deloukas P, Elliott P, Schlessinger D, Sanna S, Scuteri A, Jackson A, Mohlke KL, Tuomilehto J, Roberts R, Stewart A, Kesaniemi YA, Mahley RW, Grundy SM, Wellcome Trust Case Control C, McArdle W, Cardon L, Waeber G, Vollenweider P, Chambers JC, Boehnke M, Abecasis GR, Salomaa V, Jarvelin MR, Ruokonen A, Barroso I, Epstein SE, Hakonarson HH, Rader DJ, Reilly MP, Witteman JC, Hall AS, Samani NJ, Strachan DP, Barter P, van Duijn CM, Kooner JS, Peltonen L, Wareham NJ, McPherson R, Mooser V, Sandhu MS (2010) Genetic variants influencing circulating lipid levels and risk of coronary artery disease. Arterioscler Thromb Vasc Biol 30: 2264-76. doi: 10.1161/ATVBAHA.109.201020

Wei SM, Xie CG, Abe Y, Cai JT (2009) ADP-ribosylation factor like 7 (ARL7) interacts with alpha-tubulin and modulates intracellular vesicular transport. Biochem Biophys Res Commun 384: 352-6. doi: 10.1016/j.bbrc.2009.04.125

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M (1992) A second-generation linkage map of the human genome. Nature 359: 794-801. doi: 10.1038/359794a0

Wells QS, Becker JR, Su YR, Mosley JD, Weeke P, D'Aoust L, Ausborn NL, Ramirez AH, Pfotenhauer JP, Naftilan AJ, Markham L, Exil V, Roden DM, Hong CC (2013) Whole exome sequencing identifies a causal RBM20 mutation in a large pedigree with familial dilated cardiomyopathy. Circ Cardiovasc Genet 6: 317-26. doi: 10.1161/CIRCGENETICS.113.000011

Wheeler E, Marenne G, Barroso I (2017) Genetic aetiology of glycaemic traits: approaches and insights. Hum Mol Genet 26: R172-R184. doi: 10.1093/hmg/ddx293

Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26: 2190-1. doi: 10.1093/bioinformatics/btq340

Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, Beckmann JS, Bragg-Gresham JL, Chang HY, Demirkan A, Den Hertog HM, Do R, Donnelly LA, Ehret GB, Esko T, Feitosa MF, Ferreira T, Fischer K, Fontanillas P, Fraser RM, Freitag DF, Gurdasani D, Heikkila K, Hypponen E, Isaacs A, Jackson AU, Johansson A, Johnson T, Kaakinen M, Kettunen J, Kleber ME, Li X, Luan J, Lyytikainen LP, Magnusson PKE, Mangino M, Mihailov E, Montasser ME, Muller-Nurasyid M, Nolte IM, O'Connell JR, Palmer CD, Perola M, Petersen AK, Sanna S, Saxena R, Service SK, Shah S, Shungin D, Sidore C, Song C, Strawbridge RJ, Surakka I, Tanaka T, Teslovich TM, Thorleifsson G, Van den Herik EG, Voight BF, Volcik KA, Waite LL, Wong A, Wu Y, Zhang W, Absher D, Asiki G, Barroso I, Been LF, Bolton JL, Bonnycastle LL, Brambilla P, Burnett MS, Cesana G, Dimitriou M, Doney ASF, Doring A, Elliott P, Epstein SE, Ingi Eyjolfsson G, Gigante B, Goodarzi MO, Grallert H, Gravito ML, Groves CJ, Hallmans G, Hartikainen AL, Hayward C, Hernandez D, Hicks AA, Holm H, Hung YJ, Illig T, Jones MR, Kaleebu P, Kastelein JJP, Khaw KT, Kim E, et al. (2013) Discovery and refinement of loci associated with lipid levels. Nat Genet 45: 1274-1283. doi: 10.1038/ng.2797

William S Klug MC, Charlotte A. Spencer, Michael A Palladino, Darrell Killian (2018) Concepts of Genetics, 12 edn. Pearson

Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Magi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AA, Westra HJ, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Mateo Leach I, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancakova A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Arnlov J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Bluher M, Bolton JL, Bottcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, Deelen J, Delgado G, et al. (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 46: 1173-86. doi: 10.1038/ng.3097

Woods KA, Camacho-Hubner C, Savage MO, Clark AJ (1996) Intrauterine growth retardation and postnatal growth failure associated with deletion of the insulin-like growth factor I gene. N Engl J Med 335: 1363-7. doi: 10.1056/NEJM199610313351805

Wu B, Pankow JS (2016) On Sample Size and Power Calculation for Variant Set-Based Association Tests. Ann Hum Genet 80: 136-43. doi: 10.1111/ahg.12147

Yamada S, Motohashi Y, Yanagawa T, Maruyama T, Kasuga A, Hirose H, Matsubara K, Shimada A, Saruta T (2001) NeuroD/beta2 gene G-->A polymorphism may affect onset pattern of type 1 diabetes in Japanese. Diabetes Care 24: 1438-41.

Yasser Taher Al-Hassan ELF, Edric Estrella, Mohammad Aatif (2018) Prevalence and Determinants of Dyslipidemia: Data from a Saudi University Clinic. The Open Public Health Journal 11: 416-424. doi: DOI: 10.2174/1874944501811010416

Yousri NA, Fakhro KA, Robay A, Rodriguez-Flores JL, Mohney RP, Zeriri H, Odeh T, Kader SA, Aldous EK, Thareja G, Kumar M, Al-Shakaki A, Chidiac OM, Mohamoud YA, Mezey JG, Malek JA, Crystal RG, Suhre K (2018) Whole-exome sequencing identifies common and rare variant metabolic QTLs in a Middle Eastern population. Nat Commun 9: 333. doi: 10.1038/s41467-017-01972-9

Yu B, Zheng Y, Alexander D, Manolio TA, Alonso A, Nettleton JA, Boerwinkle E (2013) Genome-wide association study of a heart failure related metabolomic profile among African Americans in the Atherosclerosis Risk in Communities (ARIC) study. Genet Epidemiol 37: 840-5. doi: 10.1002/gepi.21752

Zabaneh D, Balding DJ (2010) A genome-wide association study of the metabolic syndrome in Indian Asian men. PLoS One 5: e11961. doi: 10.1371/journal.pone.0011961

Zadjali F, Al-Yahyaee S, Hassan MO, Albarwani S, Bayoumi RA (2013) Association of adiponectin promoter variants with traits and clusters of metabolic syndrome in Arabs: family-based study. Gene 527: 663-9. doi: 10.1016/j.gene.2013.06.057

Zarrei M, MacDonald JR, Merico D, Scherer SW (2015) A copy number variation map of the human genome. Nat Rev Genet 16: 172-83. doi: 10.1038/nrg3871

Zayed AA, Beano AM, Haddadin FI, Radwan SS, Allauzy SA, Alkhayyat MM, Al-Dahabrah ZA, Al-Hasan YG, Yousef AF (2016) Prevalence of short stature, underweight, overweight, and obesity among school children in Jordan. BMC Public Health 16: 1040. doi: 10.1186/s12889-016-3687-4

Zhakupova A, Debeuf N, Krols M, Toussaint W, Vanhoutte L, Alecu I, Kutalik Z, Vollenweider P, Ernst D, von Eckardstein A, Lambrecht BN, Janssens S, Hornemann T (2016) ORMDL3 expression levels have no influence on the activity of serine palmitoyltransferase. FASEB J 30: 4289-4300. doi: 10.1096/fj.201600639R

Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ (2016) RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics 32: 1423-6. doi: 10.1093/bioinformatics/btw079

Zhang H, Schulz TJ, Espinoza DO, Huang TL, Emanuelli B, Kristiansen K, Tseng YH (2010) Cross talk between insulin and bone morphogenetic protein signaling systems in brown adipogenesis. Mol Cell Biol 30: 4224-33. doi: 10.1128/MCB.00363-10

Zhao W, Pan J, Li H, Huang Y, Liu F, Tao M, Jia W (2016) Relationship between High Serum Cystatin C Levels and the Risk of Gestational Diabetes Mellitus. PLoS One 11: e0147277. doi: 10.1371/journal.pone.0147277

Zhu Z, Guo Y, Shi H, Liu CL, Panganiban RA, Chung W, O'Connor LJ, Himes BE, Gazal S, Hasegawa K, Camargo CA, Jr., Qi L, Moffatt MF, Hu FB, Lu Q, Cookson WOC, Liang L (2020) Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. J Allergy Clin Immunol 145: 537-549. doi: 10.1016/j.jaci.2019.09.035

Zimmerman M SB (2020) An Introduction to Nutrition. Unnamed Publisher

Zuchner S, Dallman J, Wen R, Beecham G, Naj A, Farooq A, Kohli MA, Whitehead PL, Hulme W, Konidari I, Edwards YJ, Cai G, Peter I, Seo D, Buxbaum JD, Haines JL, Blanton S, Young J, Alfonso E, Vance JM, Lam BL, Pericak-Vance MA (2011) Whole-exome sequencing links a variant in DHDDS to retinitis pigmentosa. Am J Hum Genet 88: 201-6. doi: 10.1016/j.ajhg.2011.01.001

Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, Daly MJ, Neale BM, Sunyaev SR, Lander ES (2014) Searching for missing heritability: designing rare variant association studies. Proc Natl Acad Sci U S A 111: E455-64. doi: 10.1073/pnas.1322563111

# Original Publications