# Genetic Determinants of Type 2 Diabetes

9

Miguel Cruz, Adán Valladares-Salgado,
Eugenia Flores-Alfaro, and José de Jesús Peralta Romero

## Definition of Genetic Polymorphisms

The genetic information of modern man, or *Homo sapiens*, is kept along 23 pairs of chromosomes located in the nucleus of every diploid cell. Diploid is understood as the cells that have in their nucleus a double number of chromosomes, that is, two complete copies of the genomes inherited from the parents, which correspond to maternal and paternal alleles.

Sequencing studies have described that the human haploid genome is made up of approximately 3300 million pairs of bases (3300 Mb), of which approximately 25,000 genes have a coding function and, of these, only 8% (8000) have a known function and/or action mechanism [1].

Genes are considered the unit of genetic information that codes a functional product and as a unit of inheritance, which are distributed throughout the chromatids of the chromosomes, is a specific position of DNA, known as the locus. During the process of transcription, a copy is made of the DNA, which is known as the heterogeneous nuclear RNA, which proceeds to form mRNA, which codes for structural and functional proteins.

Approximately a 99.9% of the DNA sequence is identical in humans, and the remaining 0.01% represents genetic or allelic variations, also called SNPs. The presence of SNPs varies in different populations and can explain evolution theories, migrations, and even the ethnic origin of different populations. In addition, it offers information about the phenotypical diversity within the same species, which describes a proportion of relative susceptibility to certain diseases among individuals.

## The Importance of Studying Genetic-Environmental Variant Interaction and Its Perspective in Clinical Application

There are various kinds of polymorphisms which are characterized by their presence or absence and their shape or size, the largest being insertions and deletions. In addition there are other genetic variants known as repetitions of the copy number (CNV, copy number variation) and SNPs. Unlike mutations, SNPs are changes with a frequency greater than 1% of the population. If SNPs are characterized by a simple exchange of nucleotides of adenine, cytosine, thymine, or guanine in the alleles, they are extremely important due to the fact that they are responsible for almost 90% of human phenotypical diversity. In 2008, for the first time, "1000 Genomes Project" initiative was proposed, to analyze the genetic material of 1000 people around the world and to study genetic variability. Finally, in 2015, the number of subjects analyzed reached 2500; the data suggested that in every healthy individual, there are around 150 variants that cause premature ending of proteins and another 30 implicated in the appearance of rare diseases. In addition to the presence of more than 84 million SNPs in the human genome, they located between 100 and 300 pairs of bases throughout the genome [2, 3], generated by the genetic recombination or missense. (http://www.internationalgenome.org/data#download).

It has been reported that approximately 88% of the SNPs associated with disease are located in intronic and intergenic noncoding regions, which are found in areas not related with sequences that contain essential information for the expression of a gene [4]. The remaining 12% of the SNPs are called "coding," integrated in exonic areas, giving way, in the majority of cases, to proteins that can differ in their composition and biological functions. Also, exonic SNPs may be synonymous and not produce a change in amino acid or not synonymous and change the sequence of amino acids that would alter the

M. Cruz
Unidad de Investigación Médica en Bioquímica, Hospital de Especialidades, Nuevo Laredo, Mexico

Centro Médico Nacional Siglo XXI IMSS, Mexico City, Mexico

A. Valladares-Salgado (✉) · J. de Jesús Peralta Romero
Unidad de Investigación Médica en Bioquímica, Centro Medico Nacional Siglo XXI, IMSS, Mexico City, Mexico

E. Flores-Alfaro
Laboratorio de Investigación en Epidemiología Clínica y Molecular, Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Guerrero, Chilpancingo, Guerrero, Mexico



J. Rodriguez-Saldana (ed.), *The Diabetes Textbook*, https://doi.org/10.1007/978-3-030-11815-0_9

structure, conformation, and shape of the protein. Due to the fact that SNPs are genetically stable, they are maintained for various generations and can act as true biological signals. Currently SNPs are considered ancestral risk or protector markers for diseases, from the clinical point of view. Nevertheless, their analysis is complex due to the fact that clinical phenotypes are the result of the interaction between the genotype and the exome that involves personal pathological background and unhealthy lifestyle, which contribute to the metabolic alteration present in T2D. Therefore, the evaluation of the gene-environment correlation in cohorts will allow a better understanding and interpretation of the physiopathology of the genic behavior of complex metabolic diseases, which hold first place in global morbi-mortality. In addition they will create useful tools in early detection, prevention, and more effective treatment in order to reach adequate therapeutic goals [5].

## Genome-Wide Association Study (GWAS)

Identifying the genetic determinants associated with T2D has been a complex task, due to the role that is also played by the environment in the development of the disease. Nevertheless, currently there are various genetic markers, distributed throughout the genome. Analysis of previously reported candidate genes has allowed confirmation of the association of the genes with the disease in various populations; however, replication is not always successful due to phenotypical variation and ancestry. GWAS is a method that bases its analysis on statistical and biological associations among various SNPs and phenotypes of the diseases.

The rapid development of genotyping techniques and the reduction in costs has allowed a greater number of GWAS. These studies use microarrays with more than 1,000,000 SNPs and have transformed research into the genetics of complex diseases, diabetes being outstanding. GWAS are characterized by possession of a greater power to discover variants with a modest effect, whose association is not previously known. The first studies confirmed the associations between T2D and various genetic variants located on *PPARG* genes, adding six new loci (*CDKAL1*, *HHEX*, *SLC30A8*, *IGF2BP2*, *CDKNA2A*, and *FTO*). Typically, each copy of these susceptibility alleles increases the risk of suffering diabetes by 10–15% [6].

Initially, GWAS was performed on European population and later in other populations from the African and American continents with different ethnic groups, which has contributed to identification of a greater number of genes associated with T2D. Table 9.1 shows the association of various SNPs with susceptibility to developing T2D in a trans-ethnic meta-analysis that included 1000 of cases and controls with European, East Asian, South Asian, Mexican, and Mexican American ancestry groups [7]. To date, there are more than 80 SNPs, among which variants in genes *WFS1*, *HNF1A*, *HNF1B*, *IRS1*, and *MTNR1B*. The importance of the genetic

**Table 9.1** Association of SNPs with susceptibility to developing T2D in a trans-ethnic meta-analysis that included 1000 of cases and controls with ancestry group European, East Asian, South Asian, Mexican, and Mexican American

| | | | Alleles | | Trans-ethnic meta-analysis | |
|---|---|---|---|---|---|---|
| Locus | Lead SNP | Chr | Risk | Other | *p*-value | Cochran's *Q p*-value |
| *TCF7L2* | rs7903146 | 10 | T | C | 7.8E-75 | 5.5E-04 |
| *PEPD* | rs3786897 | 19 | A | G | 3.3E-04 | 5.5E-04 |
| *KLF14* | rs13233731 | 7 | G | A | 7.0E-04 | 6.4E-04 |
| *CDKAL1* | rs7756992 | 6 | G | A | 1.6E-26 | 2.6E-03 |
| *VPS26A* | rs1802295 | 10 | T | C | 1.4E-03 | 4.4E-03 |
| *GCC1* | rs6467136 | 7 | G | A | 2.0E-01 | 5.6E-03 |
| *TSPAN8* | rs7955901 | 12 | C | T | 1.6E-03 | 6.1E-03 |
| *GCKR* | rs780094 | 2 | C | T | 1.0E-05 | 8.7E-03 |
| *GRB14* | rs3923113 | 2 | A | C | 1.5E-06 | 1.3E-02 |
| *BCAR1* | rs7202877 | 16 | T | G | 5.7E-04 | 1.3E-02 |
| *ZFAND3* | rs9470794 | 6 | C | T | 3.6E-03 | 1.4E-02 |
| *PSMD6* | rs831571 | 3 | C | T | 3.7E-04 | 1.5E-02 |
| *CILP2* | rs10401969 | 19 | C | T | 9.7E-03 | 2.0E-02 |
| *RASGRP1* | rs7403531 | 15 | T | C | 1.5E-01 | 2.1E-02 |
| *RBMS1* | rs7593730 | 2 | C | T | 4.7E-04 | 2.7E-02 |
| *TLE4* | rs17791513 | 9 | A | G | 3.2E-08 | 3.0E-02 |
| *ZBED3* | rs6878122 | 5 | G | A | 6.3E-05 | 3.1E-02 |
| *HHEX/IDE* | rs1111875 | 10 | C | T | 3.2E-19 | 3.4E-02 |
| *CDC123* | rs11257655 | 10 | T | C | 2.6E-09 | 4.3E-02 |
| *ARAP1 (CENTD2)* | rs1552224 | 11 | A | C | 1.2E-07 | 5.5E-02 |
| *KCNQ1* | rs163184 | 11 | G | T | 1.7E-14 | 5.8E-02 |
| *NOTCH2* | rs10923931 | 1 | T | G | 1.7E-02 | 6.8E-02 |
| *JAZF1* | rs849135 | 7 | G | A | 1.7E-09 | 6.9E-02 |
| *KCNJ11* | rs5215 | 11 | C | T | 3.2E-11 | 7.2E-02 |
| *DGKB* | rs17168486 | 7 | T | C | 3.4E-07 | 7.6E-02 |
| *THADA* | rs10203174 | 2 | C | T | 4.8E-05 | 8.3E-02 |
| *KCNK16* | rs1535500 | 6 | T | G | 7.5E-06 | 9.2E-02 |
| *ST64GAL1* | rs16861329 | 3 | C | T | 8.5E-06 | 1.1E-01 |
| *MTNR1B* | rs10830963 | 11 | G | C | 2.0E-07 | 1.2E-01 |
| *PTPRD* | rs17584499 | 9 | T | C | 6.0E-01 | 1.2E-01 |
| *PROX1* | rs2075423 | 1 | G | T | 2.2E-06 | 1.4E-01 |
| *HNF4A* | rs4812829 | 20 | A | G | 4.6E-08 | 1.5E-01 |
| *GIPR* | rs8108269 | 19 | G | T | 4.9E-06 | 1.5E-01 |
| *HMGA2* | rs2261181 | 12 | T | C | 3.6E-08 | 1.8E-01 |
| *SPRY2* | rs1359790 | 13 | G | A | 5.8E-06 | 2.2E-01 |
| *AP3S2* | rs2028299 | 15 | C | A | 5.2E-07 | 2.4E-01 |
| *ADAMTS9* | rs6795735 | 3 | C | T | 2.1E-04 | 2.5E-01 |
| *GCK* | rs10278336 | 7 | A | G | 1.3E-01 | 2.6E-01 |
| *ZFAND6* | rs11634397 | 15 | G | A | 1.4E-05 | 2.8E-01 |
| *FTO FTO* | rs9936385 | 16 | C | T | 1.2E-12 | 3.0E-01 |
| *GLIS3* | rs7041847 | 9 | A | G | 5.4E-06 | 3.1E-01 |
| *CCND2* | rs11063069 | 12 | G | A | 7.5E-04 | 3.2E-01 |
| *IGF2BP2* | rs4402960 | 3 | T | G | 9.5E-18 | 3.3E-01 |
| *TMEM163* | rs6723108 | 2 | T | G | 4.0E-01 | 3.3E-01 |
| *PPARG* | rs1801282 | 3 | C | G | 5.7E-10 | 3.5E-01 |
| *HNF1B* | rs4430796 | 17 | G | A | 8.9E-10 | 3.6E-01 |
| *PRC1* | rs12899811 | 15 | G | A | 5.7E-07 | 3.9E-01 |
| *CDKN2A/B* | rs10811661 | 9 | T | C | 1.1E-27 | 3.9E-01 |
| *HNF1A* | rs12427353 | 12 | G | C | 3.9E-06 | 3.9E-01 |
| *GRK5* | rs10886471 | 10 | C | T | 6.1E-01 | 4.3E-01 |

**Table 9.1** (continued)

| Locus | Lead SNP | Chr | Alleles | | Trans-ethnic meta-analysis | |
|---|---|---|---|---|---|---|
| | | | Risk | Other | *p*-value | Cochran's *Q p*-value |
| *ANK1* | rs516946 | 8 | C | T | 1.5E-07 | 4.4E-01 |
| *SRR* | rs391300 | 17 | C | T | 6.8E-01 | 5.1E-01 |
| *KLHDC5* | rs10842994 | 12 | C | T | 7.9E-06 | 5.3E-01 |
| *TP53INP1* | rs7845219 | 8 | T | C | 6.4E-08 | 5.4E-01 |
| *C2CD4A* | rs7163757 | 15 | C | T | 3.6E-06 | 5.5E-01 |
| *BCL11A* | rs243088 | 2 | T | A | 3.2E-06 | 5.5E-01 |
| *DUSP8* | rs2334499 | 11 | T | C | 1.0E-03 | 5.6E-01 |
| *SLC30A8* | rs3802177 | 8 | G | A | 1.8E-18 | 6.2E-01 |
| *WFS1* | rs4458523 | 4 | G | T | 2.1E-09 | 6.2E-01 |
| *ANKRD55* | rs459193 | 5 | G | A | 8.9E-04 | 6.7E-01 |
| *TLE1* | rs2796441 | 9 | G | A | 1.6E-06 | 7.7E-01 |
| *IRS1* | rs2943640 | 2 | C | A | 7.2E-09 | 7.9E-01 |
| *UBE2E2* | rs7612463 | 3 | C | A | 6.7E-09 | 8.3E-01 |
| *HMG20A* | rs7178572 | 15 | G | A | 1.5E-11 | 8.4E-01 |
| *ZMIZ1* | rs12571751 | 10 | A | G | 2.4E-10 | 9.3E-01 |
| *ADCY5* | rs11717195 | 3 | T | C | 2.2E-08 | 9.4E-01 |
| *MC4R* | rs12970134 | 18 | A | G | 2.6E-08 | 9.5E-01 |
| *RND3* | rs7560163 | 2 | C | G | 4.7E-01 | 9.9E-01 |
| *MAEA* | rs6815464 | 4 | C | G | 4.4E-04 | N/A |

Taken from Mahajan et al. [7]

component of T2D is clear when a concordance of 70–90% of the disease is observed between identical twins.

GWAS has allowed us to understand with greater precision the physiopathology of T2D, in order to establish better opportunities for treatment, diagnosis, and patient monitoring. From the genetic viewpoint, T2D is a multifactorial disease where the phenotype of a group of genes is modulated by environmental factors. The action mechanisms involved in the majority of signs associated with T2D offered by GWAS are involved in reduction in the secretion of insulin (be it due to dysfunction of the pancreatic beta cells or through reduction of cellular mass) or insulin resistance (associated with obesity). In conclusion, GWAS has offered important knowledge of the genetic variants most associated with T2D in the world [8]. Another focus for complex diseases is whole-exome sequencing. This has been successful in the study of low frequency variants.

## The Importance of Ancestry in Association Studies

It is known that in populations native to the American continent, there was a process of miscegenation that took place when the Amerindians and Europeans met in the New World five centuries ago. The latest studies show that the genetic composition is different in each country, and within the same country, there are regional differences. For example, in Mexico it has been shown that Mexico City has the following percentages: 65% native American, 30% European, and 5% African, while Monterrey City, N.L. the percentage was 56% native American, 38% European, and 6% African [9, 10]. Even, in Mexico City there are variations in the proportions of ancestry when we compare the IMSS vs INMEGEN studies [11]. Recently, a high prevalence of Amerindian ancestry was reported in the Montaña region of the State of Guerrero, reaching 80% of Amerindian [12].

In Mexico there is a very high degree of stratification, where the differences in allele frequencies between groups and controls can lead to false associations [9]. The admixture mapping method avoids these false associations and requires markers that may be informative concerning ancestry, that is, those for which allele frequencies differ between mixed populations. With this method data is combined from all the markers to obtain information about ancestral alleles of each marker locus and then for the association of the disease with ancestral background. We can combine the information from multiple markers in a multivariate analysis to obtain information about the ancestral alleles of each locus of each individual in the admixture.

The importance of this chapter is to describe the most important genetic variants associated with T2D (Table 9.1), for more information about frequencies, haplotypes, etc. can be consulted in http://www.internationalgenome.org/ and variants associated with diseases like T2D in the link with the ENCODE Project, https://www.genome.gov/10005107/.

## Genes Associated with T2D

### Variants of the *TCF7L2* gene

The *TCF7L2* gene has a clinical relevance because it is implicated in a wide variety of signals, insulin resistance, and T2D, specifically the variant rs7903146 in European populations, later also in Latin peoples. In 2006, the first gene implicated in susceptibility to T2D was identified through microsatellite markers, being identified without previous biological knowledge and with an important power of association, which was named *transcription factor 7-like 2 gene* (*TCF7L2*; *TCF4*). It is known that *TCF7L2* is a transcription factor that influences the transcription of various genes, thus exercising a great variety of functions within the cell. This transcription factor is a member of the signaling pathways of Wingless Int (WNT), located on chromosome 10q25. Stimulation of the WNT pathway goes along with the association of β-catenine with BCL9 and its translocation to the nucleus associated with *TCF7L2*, which results in the activation of WNT target genes, specifically in the repression of synthesis of proglucagon in enteroendocrine cells. The noncoding area contains cis-regulatory elements that lead to expression of *TCF7L2* in various tissues involved in the

homeostasis of glucose, which suggests that the variants are probably regulating the expression of this gene. The T risk allele of *rs7903146* presents greater expression in the pancreas than the C protector allele [13]. Markers located on intron 3, *DG10S478*, and SNPs *rs12255372* (allele G > T) and *rs7903146* (allele C > T) were the first markers associated with T2D in individuals in Iceland [14]. Later, this association was replicated in various populations of the world, so that this gene susceptible to T2D has become the most important worldwide. In European population, each copy of the susceptibility allele increases the risk of developing T2D 1.4–1.5 times. In Mexican population, the risk is 1.78 for each copy of the T allele for rs12255372, after adjusting with ancestral markers [15].

Lyssenko et al. showed that the risk given by the T allele of *rs7903146* associates with a lack of insulin secretion, with the effect of incretin and increase in the production of hepatic glucose. In addition, a cohort in Bosnia and another in Malmö showed how diabetes-free survival is greater in individuals with genotype CC than in individuals with CT/TT del *rs7903146* [12].

## Variants of Genes *ABCC8* and *KCNJ11*

Genes of the family *ABCC8* (union cassette ATP, subfamily C, member 8; SUR1) and *KCNJ11* (inwardly rectifying potassium channel, subfamily J, member 11; KIR 6.2) are located on chromosome 11p15.1; it has been observed that both are expressed in beta cells, and it has been reported that various polymorphism versions on these genes associate with insulin secretion disorders [16].

It has been noted that carriers of the variant *p.Arg1420His* of gene *ABCC8* have twice the risk of developing T2D, mainly among Pima Indians, although this also applies to subjects with mostly Native American ancestry [17].

In Europeans the association has been reported with variant *KCNJ11 E23K* (OR 1.23), but not with *ABCC8* (15). Nevertheless, between these two genes, there is a high degree of linkage disequilibrium (LD), which makes it harder to identify the variant causing the risk of the disease [18].

## Variants of the *CAPN10* Gene

Calpain is a cysteine protease, which participates in various functions such as apoptosis, exocytosis, mitochondrial metabolism, and remodeling of the cytoskeleton and insulin secretion. Its expression is very high in metabolically important organs such as the heart, liver, pancreas islets, and muscle. Known as the common gene in diabetes, it is located on chromosome 2q37.3, formed by 15 exons and showing 8 isoforms [19]. The most recent meta-analysis showed that the C allele of *rs2975760* of *CAPN10* was the best associated with increased risk of T2D [20]. However, an analysis by haplotypes showed that individuals with haplotype 1121/1121 for SNP-44, SNP-43, SNP-19, or SNP-63 presented twice the risk of T2D than only SNP-*43* [21]. This haplotype is not associated in other populations, which means that the genetic structure of each population is important and should be considered, as in other SNPs.

## Variants of the *PPARγ* Gene (*Peroxisome Proliferator-Activated Receptor Gamma*)

PPAR is a protein, member of a superfamily of nuclear receptors, which has a weight of approximately 56 kD. PPAR affects mechanisms present in the control of steroid hormones, of glucocorticoids, or thyroxine, of retinoic acid and of vitamin D, but mainly acts in the regulation of the expression of specific genes through a mechanism that is common to members of the nuclear receptor superfamily. It has been reported that the PPAR family is comprised of various subtypes, known as PPARα, PPARβ/δ, and PPARγ. This latter is coded by three different genes: *PPARγ1*, *PPARγ2*, and *PPARγ3*. The main function is the regulation of genes that participate in lipid and glucose metabolism. Variant of *PPARγ2* in 3p25 only is expressed in adipose tissue and regulates the differentiation, storage of lipids, and control of the transcription of various genes implicated in the metabolism, and it also participates in insulin sensitivity [22]. Various studies have shown that PPAR antagonists improve hyperlipidemia and glucose levels.

*Pro12Ala* (*rs1801282*) has been associated with T2D in different populations. *Pro12Ala* has a prevalence of 12% in Caucasian population, 10% in Native Americans, and 1% in Chinese. This change in amino acid near the extreme amino terminal (NH2-terminus) modulates the transcriptional activity. Alanine favors the formation of alpha-helix, which does not occur with proline, which forms alanine isoforms and stimulates deficiency in the target genes of the gene, carrying to the individual carriers a lesser accumulation of adipose tissue. In the latest meta-analysis, an OR of 0.86 was calculated, but unfortunately, the majority of the population at the global level carries the allele Pro12, which generates a high risk of T2D [23].

On the other hand, it has been noted that *PPARγ* has been highly studied, due to the fact that its ligands interact with thiazolidinediones, drugs used in the treatment of T2D. The effects of ligands of *PPARγ* are diverse, but the total effect is improvement in insulin sensitivity, in addition to regulation of other genes that have functions in glucose homeostasis and adipocyte differentiation.

## Variants of the *CDKN2A/B* Gene

*CDKN2A/B* gene is located in region 9p21 and codifies for a protein p16, which has the function of inhibiting cyclin-dependent kinase p16 (INK4A) and p15 (INK4B), coded by the gene *CDKN2A* and a long noncoding RNA known as *ANRIL* (*CDKN2B-AS*) [24]. It participates in the cellular cycle and helps maintain pancreas beta cell mass, but the mechanism by which *CDKN2A/B* influences diabetes risk is not yet clear. The risk allele of marker *rs10811661* has been associated with reduced insulin secretion in European population [25], while genes *MTNR1B*, *TCF7L2*, and *KCNJ11* associate with the dysfunction of β cells; both pathways are related with reduction of insulin secretion [16].

## Variants of the *FTO* Gene

Association of the fat mass and obesity-associated (*FTO)* gene with obesity was first reported in a European GWAS performed in individuals with T2D [26]. The power of association of the variant of the *FTO* gene with T2D was lost when correcting for body mass index (BMI), which suggested that susceptibility was being measured through obesity. Other studies have reported that the association between the variant and risk of T2D is maintained after adjusting for BMI. It appears that the main cause for the variability of results is related to the time when BMI was measured. The association has been demonstrated, before the development of T2D, when BMI is more elevated, and is reduced or lost with greater time of evolution of the disease.

Studies confirm the association between the variant *rs9939609 (T/A)* of *FTO* and obesity as the main risk factor for developing T2D. In other populations, such as the Mexican, the association is not as evident, particularly in children [27]. European homozygote populations for the risk allele (AA) of *rs9939609* have 1.7 times the risk of developing obesity and on average have 3 Kg more weight than the average population. Some studies have tried to identify the mechanism by which this association exists. In a metabolomic focus, metabolites have been identified, such as valine amino acid, a hexose, and other metabolites relevant to the phosphatidylcholine pathway. The alteration of valine metabolism leads to the accumulation of branched-chain amino acid in relation with the risk allele of *FTO*. The branched-chain amino acids and their derivatives seem to be an early manifestation of insulin resistance, probably via *mTOR/S6K1* kinase, which results in the phosphorylation of various residues of serine in the substrate of the insulin receptor (IRS-1). Metabolites of phosphatidylcholine are associated with apolipoprotein B, and it has been demonstrated that the risk allele of *FTO* is associated with the particles that form part apolipoprotein B.

## Variants of the *IRS-1* Gene

The molecules of IRS are important mediators in the signaling of insulin, in addition to playing an important role in metabolism, growth, and survival of the cell. The IRS family is formed by four members, IRS-1 to IRS-4, presenting a different tissue distribution and therefore different expression. IRS-1 and IRS-2 are key for insulin action and glucose homeostasis. IRS-1 is coded in chromosome 2q36.3. Polymorphism *Gly972Arg* of *IRS-1* has been the most associated with the development of T2D. The union of insulin to its active phosphorylated receptor to *IRS-1*, phosphorylating tyrosine residue, serine threonine) (Ser/Thr), which join and activate PI3K, which contains the subunit p85, and p110 phosphorylates PI, and this allows it to join with akt and PDK1. The phosphorylation of tyrosine residue accompanies the mobilization of glucose transporters (GLUT 4) that mediate the internalization of the same. However, when the serines or threonines are phosphorylated, it leads to an accelerated degradation of the IRS protein, which generates an alteration in insulin signaling and insulin resistance and a decrease in the translocation of GLUT4.

As mentioned earlier, the polymorphism *Gly972Arg* has been the most reported in studies of association with T2D, in combination with environmental factors such as diet, age, and physical activity. Like other genes and depending on the population, important associations have also been reported (such as in Europeans), weak ones as with the Japanese, or absence of, as with the Pimas [28]. In Mexican population, variant *Arg* has been observed in 2.6% in controls and 7.9% in cases [29].

## Variants of the Hepatocyte Nuclear Factor 1-Alpha (*HNF1A)* Gene

HNF1A is coded on chromosome 12q24.31. The protein joins inverted palindrome 5′-GTTAATNATTAAC-3′ for the activation and regulation of gene expression, mainly in the cells of the pancreatic islets and the liver. Some variants of the gene have been found to be associated with maturity-onset diabetes of the young 3 (MODY3). Through the study of exome sequencing, the variant pE508K has been identified and associated with T2D. This variant generates a reduction in the function of the protein, unlike MODY3 diabetes, where function is almost lost. The mechanism related with the affinity of the protein for joining DNA sequence does not appear to be altered. It seems the reduction in activity occurs mainly through a reduction in expression and the protein shows altered localization in the nucleus.

The effect of the variant on European populations is very high, with results similar to two studies in Latin population.

Carriers of the variant have up to fivefold increased prevalence of T2D. Interesting from a clinical viewpoint, carriers of the variant respond better to treatment with sulfonylureas than with metformin, a drug of choice in the treatment of T2D [30].

### Variants of the *Solute Carrier Family 30 Member 8 (SLC30A8)* Gene

This transporter, coded on chromosome 8q24.11 and expressed importantly in the islets of Langerhans in the pancreas, participates in the packaging of proinsulin in secretory granules and liberation. These processes require the presence of ions $Zn^{2+}$ and $Ca^{2+}$, which form complexes with proinsulin. The ions of $Zn^{2+}$ are transported by transporter 8, which is found in abundance in the pancreas beta cells, also located in alpha cells, and participates in the liberation of glucagon. GWAS have associated the gene with susceptibility to developing T2D. A recent study showed that the marker associated with greatest frequency in European, Asian, and African population is rs13266634 [31]. However, other authors have not found this gene to be associated with T2D [32].

### Other Variants Associated with Insulin Resistance and Dyslipidemias

Variant *R230C* of gene *ABCA1* of the HDL receptor participates in the reverse transport of cholesterol, associated with early-onset diabetes and obesity, particularly in Mexican population, with values of $p = 10^{-6}$ [11]. Also, in Japanese population, the presence of a haplotype with an OR of 2.59 has been reported associated with T2D [33].

In a meta-analysis of Mexican and Mexican-American samples to characterize genes associated with T2D in Hispanics, the following genes were identified, with values of $<10^{-5}$: gene *ATP2B2*, located on chromosome 3, *UNC5C* on chromosome 4, and *PIWIL4* on chromosome 11, in addition to three independent intergenic regions located on chromosome 10 and an *EST* (expressed sequence tag) located near the area of gene *RXRA* on chromosome 9. Upon adjusting for BMI, two additional groups of markers were observed, one in the intergenic area of chromosome 20 and the other within genes *C22orf30/DEPDC5*, located on chromosome 22. This meta-analysis showed SNPs with a high level of significance in ten genomic areas. In addition, two additional regions were identified when BMI was incorporated, in particular an intronic variant of *ANK2* gene and two intronic variants of *MCPH* gene [34, 35]. Other population studies have identified genes such as *HNF1A*, *KCNQ1*, and *PTPRD*. Also, two other genes identified, *CSMD1* and *ANK2*, were relevant due to their functionality in metabolic regulation. Other regions associated with T2D showed statistical significance, with *CDKN2A/CDKN2B* and *IGF2BP2* genes.

## Biological Validation Studies

After the identification of the genes associated with a disease, what is sought is to know their biological function, so that the genes mentioned above have been studied for their expression in adipose tissue, skeletal muscle, and lymphoblast cell lines. One of the most significant signals of SNP *rs202983*, located within *CIT* gene (chromosome 12), showed an important effect on the regulation of gene *WFS1*. It has been documented that mutations of *WFS1* gene cause monogenic diabetes and common variants of this gene have associated with T2D. Lineal regression analysis of these genetic markers with five parameters (BMI, total cholesterol, HDL-C, LDL-C, and triglycerides) showed values of association at the genomic level in polymorphisms near *APOA5* gene, which is located on chromosome 11. The variant *rs964184* showed the lowest value at $p = 2.3 \times 10^{-9}$. Other variants of interest are those of *SYNE1* gene, which is found on chromosome 6, for triglycerides (*rs998147*, $p = 5.3 \times 10^{-7}$) and an area near *MAD2L1* gene on chromosome 4 for HDL-C (*rs4568220*, $p = 7.1 \times 10^{-7}$) [34, 35].

## Conclusions

T2D is a complex disease that presents differences in prevalence between populations. Epidemiological data indicate that the risk of suffering the disease is higher in Amerindian populations than those of European origin. There is evidence on the influence of genetic factors in populations; to date, over 80 loci associated with T2D have been identified, which do not always replicate among populations. Analysis by admixture mapping has been specifically designed to identify genes involved in complex diseases that show differences in prevalence among populations. Given the history of miscegenation in Mexican population, admixture mapping is an ideal method for identifying the genetic factors that increase the risk of suffering T2D. The first GWAS performed in patients with T2D in Mexico showed that less than 10% of the 46 candidate genes reported in 2011 in European population were found associated in our population. These populations are characterized mainly by low levels of HDL-C, high levels of LDL-C, and elevated triglycerides. The genetic factors most associated with these alterations have been variants of *ZNF259/APOA5* genes, such as rs964184, associated with triglycerides, rs2367970 of the same gene, and rs2472386 of *ABCA1* gene associated with HDL.

It is a priority to establish the genetic history of the Mexican, in order to have risk markers for developing T2D, markers associated with complications and metabolic disorders, a condition very evident in our population thanks to current lifestyles.

## Multiple Choice Questions

1. A gene is considered as:
   (a) A sequence of nitrogenated bases
   (b) The unit of genetic and inherited information
   (c) The chromatid unit that forms chromosomes
   (d) A sequence of nucleosides
   (e) Triplets of bases
2. What percentage of DNA sequence is identical among humans?
   (a) 99.9
   (b) 98.0
   (c) 95.0
   (d) 98.5
   (e) 99.0
3. The main difference between mutation and a SNP is:
   (a) A mutation is lethal and a SNP no
   (b) In mutation there is a change of various bases
   (c) SNPs occur only in introns
   (d) The frequency of a SNP is greater than 1%
   (e) A SNP is presented at any stage of life
4. All are characteristics of SNPs except:
   (a) They are generally biallelic
   (b) They are presented throughout the structure of the gene
   (c) They are only present in exons and introns
   (d) They are inherited
   (e) They allow identification of an individual
5. The gene most frequently associated with T2D worldwide is:
   (a) IRS-1
   (b) CAPN10
   (c) TCF7L2
   (d) PPARg
   (e) FTO
6. Which is the action mechanism of variant rs1801282 of the gene *PPARγ*?
   (a) Transcriptional modulation in the change of alanine
   (b) Oxidation of free fatty acids
   (c) Transcriptional modulation of the signaling pathways of TZD
   (d) All of the above
   (e) None of the above
7. What is the main problem of low replication of the association of obesity with T2D of the various genetic variants of the gene *FTO* upon analyzing it in different populations?
   (a) The loss of statistical power in meta-analysis
   (b) The ancestry of various populations
   (c) The time of evolution of the disease and the difficulty in performing metabolomics studies
   (d) All of the above
   (e) None of the above
8. Why is it important to determine the genetic component in metabolic diseases?
   (a) To identify risk or protector markers associated with the disease
   (b) To perform studies in metabolomics
   (c) All of the above
   (d) None of the above
9. What is the function of the gene of *CAPN10*?
   (a) Participate in apoptosis, exocytosis, mitochondrial metabolism, and remodeling of the cytoskeleton
   (b) The gene codes for calpain-10, an atypical cysteine-protease that participates in the mechanism of insulin secretion
   (c) Participate in the oxidative use of glucose for skeletal muscle
   (d) All of the above
   (e) None of the above
10. Characteristics of gene *SLC30A8* include:
    (a) The transporter is coded on chromosome 8q24.11. It is expressed at high level in the pancreas, particularly in the Islets of Langerhans
    (b) It participates mainly in the packaging of proinsulin in secretory granules, the hepatic liberation and elimination of insulin
    (c) It processes and requires the presence of ions Zn2+ and Ca2+ which form complexes with proinsulin
    (d) All of the above
    (e) None of the above

## Correct Answers

1. (b) The unit of genetic and inherited information
2. (a) 99.9
3. (d) The frequency of a SNP is greater than 1%
4. (c) They are only present in exons and introns
5. (c) *TCF7L2*
6. (a) Transcriptional modulation in the change of alanine
7. (d) All of the above
8. (d) All of the above
9. (d) All of the above
10. (d) All of the above

## Glossary

Some definitions are found on the page: https://ghr.nlm.nih.gov/.

**Ancestry** refers to the geographical origin of populations, for example, "individuals of European ancestry" or the line of heritage or descent of a group.

**Diabetes** is a group of metabolic diseases characterized by hyperglycemia resulting from defects in insulin secretion, insulin action, or both. The chronic hyperglycemia of diabetes is associated with long-term damage, dysfunction, and failure of different organs, especially the eyes, kidneys, nerves, heart, and blood vessels (American Diabetes Association).

**Genetic marker** is a gene or (a fragment of) DNA sequence having a known location on a chromosome. It has an easily identifiable phenotype and whose inheritance pattern can be followed. Genetic markers act as chromosomal landmarks. They are used to trace or identify specific region of a gene (especially one that is associated with an inherited disease) on a chromosome. They are also used to determine a linkage group or a recombination event.

**Genome-wide association study (GWAS)** is a relatively new way to identify genes involved in human disease. This method searches the genome for small variations, called single nucleotide polymorphisms or SNPs (pronounced "snips") that occur more frequently in people with a particular disease than in people without the disease. Each study can look at 100 or 1000 of SNPs at the same time. Researchers use data from this type of study to pinpoint genes that may contribute to a person's risk of developing a certain disease.

**Microarrays** is a hybridization of a nucleic acid sample (target) to a very large set of oligonucleotide probes, which are attached to a solid support, to determine sequence or to detect variations in a gene sequence or expression or for gene mapping.

**Single nucleotide polymorphisms (SNPs)** are the most common type of genetic variation among people. Each SNP represents a difference in a single-DNA building block, called a nucleotide. For example, a SNP may replace the nucleotide cytosine (C) with the nucleotide thymine (T) in a certain stretch of DNA.

## References


1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931–45.
2. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68–74.
3. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. Nature. 2015;526(7571):75–81.
4. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74.
5. Franks PW, Pare G. Putting the genome in context: gene-environment interactions in type 2 diabetes. Curr Diab Rep. 2016;16(7):57.
6. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science. 2007;316(5829):1341–5.
7. Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, Ferreira T, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet. 2014;46(3):234–44.
8. Parra EJ, Below JE, Krithika S, Valladares A, Barta JL, Cox NJ, et al. Genome-wide association study of type 2 diabetes in a sample from Mexico City and a meta-analysis of a Mexican-Amerian sample from Starr County, Texas. Diabetologia. 2011;54:2038–46.
9. Martinez-Marignac VL, Valladares A, Cameron E, Chan A, Perera A, Globus-Goldberg R, et al. Admixture in Mexico City: implications for admixture mapping of type 2 diabetes genetic risk factors. Hum Genet. 2007;120(6):807–19.
10. Martinez-Fierro ML, Beuten J, Leach RJ, Parra EJ, Cruz M, Rangel-Villalobos H, et al. Ancestry informative markers and admixture proportions in northeastern Mexico. J Hum Genet. 2009;54:504–9.
11. Villarreal-Molina MT, Flores-Dorantes MT, Arellano-Campos O, Villalobos-Comparan M, Rodriguez-Cruz M, Miliar-Garcia A, et al. Association of the ATP-binding cassette transporter A1 R230C variant with early-onset type 2 diabetes in a Mexican population. Diabetes. 2008;57(2):509–13.
12. Cahua-Pablo JA, Cruz M, Tello-Almaguer PV, del Alarcón-Romero LC, Parra EJ, Villerías-Salinas S, Valladares-Salgado A, Tello-Flores VA, Méndez-Palacios A, Pérez-Macedonio CP, Flores-Alfaro E. Analysis of admixture proportions in seven geographical regions of the State of Guerrero, Mexico. Am J Hum Biol. 2017;29. (Submitted).
13. Lyssenko V, Lupi R, Marchetti P, Del Guerra S, Orho-Melander M, Almgren P, et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. J Clin Invest. 2007;117(8):2155–63.
14. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet. 2006;38(3):320–3.
15. Parra EJ, Cameron E, Simmonds L, Valladares A, McKeigue P, Shriver M, et al. Association of TCF7L2 polymorphisms with type 2 diabetes in Mexico City. Clin Genet. 2007;71(4):359–66.
16. Gloyn AL, Weedon MN, Owen KR, Turner MJ, Knight BA, Hitman G, et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. Diabetes. 2003;52(2):568–72.
17. Baier LJ, Muller YL, Remedi MS, Traurig M, Piaggi P, Wiessner G, et al. ABCC8 R1420H loss-of-function variant in a southwest American Indian community: association with increased birth weight and doubled risk of type 2 diabetes. Diabetes. 2015;64(12):4322–32.
18. Florez JC, Burtt N, de Bakker PI, Almgren P, Tuomi T, Holmkvist J, et al. Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region. Diabetes. 2004;53(5):1360–8.
19. Horikawa Y. Calpain-10 (NIDDM1) as a susceptibility gene for common type 2 diabetes. Endocr J. 2006;53(5):567–76.

20. Yan ST, Li CL, Tian H, Li J, Pei Y, Liu Y, et al. Association of calpain-10 rs2975760 polymorphism with type 2 diabetes mellitus: a meta-analysis. Int J Clin Exp Med. 2014;7(10):3800–7.
21. Orho-Melander M, Klannemark M, Svensson MK, Ridderstrale M, Lindgren CM, Groop L. Variants in the calpain-10 gene predispose to insulin resistance and elevated free fatty acid levels. Diabetes. 2002;51(8):2658–64.
22. Estivalet AA, Leiria LB, Dora JM, Rheinheimer J, Bouças AP, Maia AL, et al. Thr92Ala and PPARγ2 Pro12Ala polymorphisms interact in the modulation of insulin resistance in type 2 diabetic patients. Obesity. 2011;19:825–32.
23. Gouda HN, Sagoo GS, Harding AH, Yates J, Sandhu MS, Higgins JP. The association between the peroxisome proliferator-activated receptor-gamma2 (PPARG2) Pro12Ala gene variant and type 2 diabetes mellitus: a HuGE review and meta-analysis. Am J Epidemiol. 2010;171(6):645–55.
24. Kong Y, Sharma RB, Nwosu BU, Alonso LC. Islet biology, the CDKN2A/B locus and type 2 diabetes risk. Diabetologia. 2016;59(8):1579–93.
25. Hribal ML, Presta I, Procopio T, Marini MA, Stancakova A, Kuusisto J, et al. Glucose tolerance, insulin sensitivity and insulin release in European non-diabetic carriers of a polymorphism upstream of CDKN2A and CDKN2B. Diabetologia. 2011;54(4):795–802.
26. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007;316(5826):889–94.
27. Mejía-Benítez A, Klünder-Klünder M, Yengo L, Meyre D, Aradillas C, Cruz E, et al. Analysis of the contribution of FTO, NPC1, ENPP1, NEGR1, GNPDA2 and MC4R genes to obesity in Mexican children. BMC Med Genet. 2013;14:21.
28. Sesti G. Insulin receptor substrate polymorphisms and type 2 diabetes mellitus. Pharmacogenomics. 2000;1(3):343–57.
29. Burguete-Garcia AI, Cruz M, Madrid-Marina V, Lopez-Ridaura R, Hernández-Avila M, Cortina B, et al. Association of Gly972Arg polymorphism of IRS1 gene with type 2 diabetes mellitus in lean participants of a national health survey in Mexico: a candidate gene study. Metabolism. 2010;59:38–45.
30. SIGMA Type 2 Diabetes Consortium, Estrada K, Aukrust I, Bjørkhaug L, Burtt NP, Mercader JM, et al. Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. JAMA. 2014;311(22):2305–14.
31. Fan M, Li W, Wang L, Gu S, Dong S, Chen M, Yin H, Zheng J, Wu X, Jin J, Jiang X, Cai J, Liu P, Zheng C. Association of SLC30A8 gene polymorphism with type 2 diabetes, evidence from 46 studies: a meta-analysis. Endocrine. 2016;53(2):381–94.
32. Kulkarni H, Mamtani M, Peralta JM, Diego V, Dyer TD, Goring H, Almasy L, Mahaney MC, Williams-Blangero S, Duggirala R, Curran JE, Blangero J. Lack of association between SLC30A8 variants and type 2 diabetes in Mexican American families. J Diabetes Res. 2016;2016:6463214.
33. Daimon M, Kido T, Baba M, Oizumi T, Jimbu Y, Kameda W, et al. Association of the ABCA1 gene polymorphisms with type 2 DM in a Japanese population. Biochem Biophys Res Commun. 2005;329(1):205–10.
34. Parra EJ, Below JE, Krithika S, Valladares A, Barta JL, Cox NJ, et al. Genome-wide association study of type 2 diabetes in a sample from Mexico City and a meta-analysis of a Mexican-American sample from Starr County, Texas. Diabetologia. 2011;54(8):2038–46.
35. Below JE, Gamazon ER, Morrison JV, Konkashbaev A, Pluzhnikov A, McKeigue PM, et al. Genome-wide association and meta-analysis in populations from Starr County, Texas, and Mexico City identify type 2 diabetes susceptibility loci and enrichment for expression quantitative trait loci in top signals. Diabetologia. 2011;54(8):2047–55.