Genetics of Type 2 Diabetes- A Review Article

Ensieh Shahvazian1, Ehsan Farashahi Yazd2*, Mohammad Hasan Sheikhha3, Masoud Rahmanian4

Abstract
Objective: Type 2 diabetes (T2D) as a complex disease is the result of genetically heterogeneous factors and environmental issues interaction. Linkage and small-scale candidate gene studies were successful in identification of genetic susceptibilities of monogenic form of diseases. However, they were largely unsuccessful while applying to the more common forms of disease. By designing Genome Wide Association studies (GWAs), the new windows open to genetic approaches. So far, around 153 variant were discovered for T2D and missing rare variants are waiting to be discovered. The new findings are beneficial to explain molecular signaling and pathways responsible for pathophysiology of T2D, which offered opportunities for the development of novel therapeutic and preventive tactics. The GWAs findings need to be confirmed in on going researches. In this review, we address, genetic susceptibilities related to T2D since 2007. Also challenge advantages and disadvantages of GWAs and discuss about the next confirmatory approaches need to be done.

Keywords: Genome Wide Association Studies, T2D, Insulin resistance, High blood glucose, Re-sequencing.

Introduction

Type 2 diabetes (T2D) is a worldwide concern. Based on International Diabetes Federation, 285 million adults aged 20–79 years suffered from diabetes in 2010. About 60% of them located in Asia. Almost 80% of people with diabetes live in developing countries. Around 95% of diabetic patients suffer from T2D (1-3). Worryingly, new findings show decrease in the onset age of T2D. The disease is more frequently reported among young Africans and Pima Indians (4). T2D in Asia differs from the other world regions; since it has developed in younger age group, and in people with much lower body-
mass index (BMI) (5). This review will highlight the genetic discoveries of T2D.

**T2D and Metabolic traits**

T2D is a complex metabolic disease (6). Complex disease like T2D and cancers are the results of both environmental and genetic risk factors and also intrauterine environment. However they do not follow a Mendelian pattern of inheritance and cannot be described by a single gene disorder, so this hypothesis “common illnesses common variants” help to understanding mechanisms of these kind of diseases (7,8).

Environmental factors are largely responsible for T2D. Decreased physical activity and increased fat consumption lead to augmentation of nutrient storage. This kind of Long-term fatty diet is associated not only with the progress of obesity but also with reductions in insulin release (9).

Environmental factors will be bolded while the genetic abnormalities already exist. Moreover, changes in the amounts of dietary carbohydrate and fat can affect both insulin sensitivity and insulin release in three days, while obesity does not exist (10). Other environmental condition that should be considered is in utero environment. Poor diet can modify metabolisms, results in a tissue adaptation that favors the storage of nutrients (11). Entirely, these environmental changes and genetic risk factors are responsible for developing T2D.

Studies of first-degree relatives highlight that Insulin resistance and impaired beta cell function are the main complication of diabetes, which leads to high blood glucose (12). Normal glucose homeostasis is conserved by a fine balance between insulin secretion and insulin sensitivity of the peripheral tissues. Insulin resistance is a key feature of the metabolic syndrome and often progresses to T2D. Also Insulin resistance is the main connection between obesity and T2D. Insulin resistance in both of these conditions is manifested by decreased insulin-stimulated glucose transport and metabolism in adipocytes and skeletal muscle and by impaired suppression of hepatic glucose output (13).

**Genetics of T2D before GWAs**

Uncovering genetic risk factor provided better vision of underlying mechanisms. Genetic studies of T2D before the GWA revolution could be divided in 4 groups; studying the pattern of inheritance, linkage studies, association studies on candidate genes and microarray studies.

The genes involved in monogenic form of diabetes mellitus (MODY, Neonatal diabetes, mitochondrial diabetes and syndromes of severe insulin resistance) were discovered by studying pattern of inheritance. The discovered genes related to beta cell dysfunction in MODY diabetes are; HNF4A (14-16), GCK (15), PDX1 (16), TCF2 (17), NEUROD1 (18) and KLF11 (19). The genes; KCNJ11 (20), ABCC8 (21), EIF2AK3 (20), PLAGL1 (22), HYMAI (23), PTF1A (24) and INS (25) are reported in monogenic form of Neonatal diabetes which cause beta cell dysfunction. INSR(26) and ACT2 are monogenic cause of insulin resistance in severe insulin resistance syndrome.

Candidate genes, based on their indirect physiologic role, seem to contribute to the disease. Studying variations in candidate genes instead of discovering new pathways leading toT2D, verifies the defined role of genes. These surveys usually have been based on linkage analysis. Since 2006, three studies reported by linkage analysis; amino acid substitution in the nuclear receptor and adipogenic transcription factors, PPARγ (27), Glu23Lys variant in the KCNJ11 gene associated with insulin signaling(28), polymorphism (K121Q) in the insulin action inhibitor ENPP1 and IRS1, IRS2 and phosphatidylinositol 3- kinase (29-31). One of the outstanding exception gene discovery by linkage studies is TCF7L2 that confirmed in later studies (32).
Microarray approach is another way of defining genetic variation in T2D. Two discoveries since 2006 are covering here. One of them is defects in skeletal muscle of T2D patients related to decrease in the expression of nuclear-encoded genes (PGC-1α and PGC-1β) involved in mitochondrial oxidative phosphorylation (33). The other one is transcription factor ARNT/Hif1b that influenced on the expression of many other genes related to insulin pathway and glucose sensing (34).

Table 1. Genetic susceptibilities related to type 2 diabetes since 2007

<table>
<thead>
<tr>
<th>Year</th>
<th>Locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>IGF2BP2 (47,48), CDKAL1 (47,49,50), SLC30A8 (46-48), CDKN2A/2B (47,49,50), HHEX (46,47), FTO (48,50-52), HNF1B (53), PPARG (47,48), WSF1 (54), TCF7L2 (46-51,55), KCN11 (47,48,50)</td>
</tr>
<tr>
<td>2008</td>
<td>NOTCH2 (56,57), THADA (56,57), ADAMST9 (56,57), JAZF1 (56,57), CDC123/CAMK1D (56,57), TSPAN8/LGR5 (56,57), GSPIF1 (56), TCF7L2 (57), KCNQ1 (58), DCDC2 (57), MC4R (59,60), FTO (57)</td>
</tr>
<tr>
<td>2009</td>
<td>IRS1 (61), HCCAC2 (62), SERS10 (63), WSF1 (61), CDKAL1 (64), SLC30A8 (64), CDKN2A/2B (64), HHEX (64), TCF7L2 (64,65), KCN11 (65), KCNQ1 (64), MTTR1B (66), MC4R (63), FTO (65)</td>
</tr>
<tr>
<td>2010</td>
<td>DUSP9 (67), GCKR (68), BCL11A (67), G6PC2 (68), ADCY5 (68,69), WFS1 (67), ZBED3 (67), DGKB/TMEM195 (68), GCK (68), KLF14 (68), TNP3INP1 (67), GLIS3 (68), TLE4 (CHIC9) (67), ADRA2A (68,70), CENTD2 (67), CRY2 (68), FADS1 (68), MTTR1B (67), MC4R (63), FTO (65)</td>
</tr>
<tr>
<td>2011</td>
<td>IRS1 (61), HCCAC2 (62), SERS10 (63), WSF1 (61), CDKAL1 (64), SLC30A8 (64), CDKN2A/2B (64), HHEX (64), TCF7L2 (64,65), KCN11 (65), KCNQ1 (67), SPRY2 (74), FTO (67)</td>
</tr>
<tr>
<td>2012</td>
<td>ST6GAL1 (75), C6orf57 (76), VPS26A (75), ARAP1 (77), MADD (77), HNF1A (78), AP3S2 (75), HMG20A (75), VPS13C/C2CD4A/B (76), HNF4A (75)</td>
</tr>
<tr>
<td>2013</td>
<td>MACF1 (79), BCL11A (6), GRB14 (6,75), RND3 (80), THADA (81), TME163 (82), IGF2BP2 (83), PPARG (84), PSMID6 (84), MAEA (84,85), ANKR355 (6), CDKAL1 (86), KCN16 (84), ZFAND3 (84), DGKB (6), GCC1-PAVX (84), JAZF1 (81), ANK1 (6,85), SLC30A8 (86), CDKN2A/2B (86), GLIS3 (84), TLE1 (6), HHEX (86), TCF7L2 (86), ZMIZ1 (6), KCNQ1 (6), CCND2 (6), HMG2A (81), KLHC6 (6), HMG20A (6,86), BCR1 (6), LAMA1 (86), MC4R (6), CATAD2A/CILP2 (6,81), GPCR (6), PEPD (84), SUGP1/CILP2 (6,81), FITM2/RHDML-HNF4A (84), FTO (86)</td>
</tr>
<tr>
<td>2014</td>
<td>COBLL1 (79), CDKAL1 (87), CDKN2A/B (87), KCNQ1 (87), CDC123/CAMK1D (87), GLIS3 (87), HNF1B (87), DUSP9 (87), GRK5 (87), RASGR1 (87), INSIG2 (70), MPHOSPH9 (89)</td>
</tr>
</tbody>
</table>

Table 2. Related genetic susceptibility associated with glycemic traits

<table>
<thead>
<tr>
<th>Locus</th>
<th>Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNX7</td>
<td>Fasting proinsulin levels adjusted for fasting glucose (77)</td>
</tr>
<tr>
<td>LYPFLA1</td>
<td>Fasting insulin (91)</td>
</tr>
<tr>
<td>DPYSL5</td>
<td>Fasting glucose (91)</td>
</tr>
<tr>
<td>GIPR</td>
<td>2h-glucose/insulinogenic index/AUCins/gluc/2-h insulin, adjusted for 2h glucose/T2D (69)</td>
</tr>
<tr>
<td>IRS1</td>
<td>Fasting glucose/HOMA-IR, Fasting insulin, CAD (68,91)</td>
</tr>
<tr>
<td>SLC2A2</td>
<td>Fasting glucose/HOMA B/HBA1C (68)</td>
</tr>
<tr>
<td>TAF11</td>
<td>Fasting insulin (91)</td>
</tr>
<tr>
<td>GRB10</td>
<td>FG, F1 (83)</td>
</tr>
<tr>
<td>LARP6</td>
<td>Fasting proinsulin levels (77)</td>
</tr>
<tr>
<td>SGSM2</td>
<td>Fasting proinsulin levels (77)</td>
</tr>
<tr>
<td>VPS13C</td>
<td>2h-glucose/2h-insulin, adjust for 2h-glocue (69)</td>
</tr>
<tr>
<td>PDX1-AS1</td>
<td>Fasting glucose (91)</td>
</tr>
<tr>
<td>PCSK1</td>
<td>Fasting proinsulin levels/ fasting glucose (77,91)</td>
</tr>
<tr>
<td>PDGFC</td>
<td>Fasting glucose (91)</td>
</tr>
<tr>
<td>MSM01</td>
<td>Fasting insulin, insulin resistance (92)</td>
</tr>
<tr>
<td>HECTD4/C12orf51</td>
<td>1-h plasma glucose (93)</td>
</tr>
<tr>
<td>OAS1</td>
<td>Fasting glucose (91)</td>
</tr>
<tr>
<td>MADD</td>
<td>Fasting glucose/ Fasting proinsulin/HOMA B (68,77)</td>
</tr>
<tr>
<td>TCERG1L</td>
<td>Fasting insulin, insulin resistance (92)</td>
</tr>
<tr>
<td>PPP1R3B</td>
<td>Fasting glucose (91)</td>
</tr>
<tr>
<td>FOXA2/LINC00261</td>
<td>Fasting glucose (91)</td>
</tr>
</tbody>
</table>

Microarray approach is another way of defining genetic variation in T2D. Two discoveries since 2006 are covering here. One of them is defects in skeletal muscle of T2D patients related to decrease in the expression of nuclear-encoded genes (PGC-1α and PGC-1β) involved in mitochondrial oxidative phosphorylation (33). The other one is transcription factor ARNT/Hif1b that influenced on the expression of many other genes related to insulin pathway and glucose sensing (34).

Overview of GWA studies

Before 2005, family linkage and SNPs-based studies were available for discovering genomic loci having influence on complex disease and traits; however on 2005, a new generation of
association survey on the entire human genome is called Genome Wide Association study or GWAs founded by National Institutes of Health to cover hundreds to thousands of genotyping (35,36). By 2013, the catalog of published GWAs include near two thousand papers (37).

GWAs, usually are held as case control study and assaying fewer than 100,000 SNPs to more than one million, commonly by the Affymetrix 500k and Illumina HumanHap300 arrays. The accepted threshold for statistical significance in GWAs is <5*10^{-8} (38-40). A precise maps of usual single nucleotide polymorphisms (SNPs) all over the genome and also patterns of linkage disequilibrium, which accomplished by International Hap Map project and affordable High-throughput genotyping technologies and map of copy number variants are the primary requirements for GWA studies (41,42).

GWA studies not only identify genetic associations with observable traits, but also demonstrate gene-gene interactions. It also could be carried out for quantitative traits, which explain underlying physiology of disease (43,44).

Genetic risk factor of T2D:
GWAs have both confirmed known genes and discovered many new susceptible genes for T2D since 2007 (45). The first GWAs for T2D was conducted in French population composed of 661 cases of T2D and 614 non-diabetic controls were genotyped on two genotyping platforms. In total, 392,935 SNPs were analyzed for association with T2D. This study revealed novel and reproducible association signals at SLC30A8 and HHEX and confirmed the famous association at TCF7L2. However, LOC387761, EXT2 associations were not reproducible in follow-up studies (46). Till now around 153 variants, more than 120 loci were identified to be associated with T2D or related traits and many new hints are waiting to be discovered. The summery of GWAs and meta-analysis related to T2D since 2007 are listed in Table 1. There are many other loci, reported to be associated with T2D complication that were listed in Table 2.

Confirmatory of GWAs
However, the loci identified in GWAs range from 10 to 100 kb, did not cover causal variants and genes. Further techniques, including re-sequencing of associated regions can be essential to entirely understand associations (94). In addition, data gathering from different GWAs should be confirmed in re-sequencing, genotyping and bioinformatics studies to elucidate that the results are reproducible in diverse population group and also the function of susceptible new candidate gene become clear (95-97).

Pros and cons of GWAs
Although several predominant limitations of using GWAs exist, GWAs methodologies are used for both monogenic and complex genetic disease. The benefits of GWAs are listed.
- No need to know the biological pathway of disease
- Documents specific genetic associations
- It is likely to discover new candidate genes that is difficult to reach by other approaches
- Provides strong data on sequence and copy-number variations

However, there are some limitations in GWAs
- Detects association related to disease not their causation
- Identifies a specific location in genome not a specific gene. So it could be the coding sequence of protein or sequence between genes that yet dose not assumed to be related to disease. These variants are hard to check by biochemically experimental works (98).
- Defines common variations in society (more that 5% frequency) not rare ones
- Unable to release all interactions between SNPs and SNPs-environment.
• Requires large sample size to studied and the results need to be reproducible in other populations (98).

**Post-GWAs prioritization**

Many GWAs have been directed, with the consumption of high-throughput and affordable genotyping and next generation sequencing platforms. Therefore, many DNA variants associated with complex traits have been discovered but these are just small proportion of heritability related to these diseases. Variants with frequencies less than 5% and variants with small effect size that sharply influence the mechanisms of disease are undiscovered. So many chips are design to fine map certain genomic region. Also a number of whole exome sequencing created to find rare variant. To raise the power, meta-analysis studies are designed. In this way so many other gaps are discovered (99,100).

**Comment and conclusion**

Since 2007, GWAs have critical role to uncover genetic risk factors associated to T2D. Around 120 loci related to the disease are discovered. However, much of the risk factors still remain unexplained. This could be due to, not covering every variant in GWA studies. Rare variant may have relatively large effects on T2D which is missing in this kind of studies. And also some risk factors are due to epigenetic mechanisms that do not make any changes on genome sequences. Therefore, better understanding T2D not only needs applying many GWAs on different populations and doing meta-analysis on them but also needs knowledge of heritability and studying rare variants in different populations. Identification of such rare variants, will require re-sequencing of the entire genome of type 2 diabetes cases and controls. However, the much more important part is to validate GWAs results by re-sequencing on different groups of same population and also diverse population. It seems that discovering mechanisms of complex diseases need a wide-range of further studies.

**References**

Genetics of type 2 diabetes


