# SNPs and Type 2 Diabetes.

Maria Poulasouchidou[1], Alexandros Lambropoulos[2]

[2]*Laboratory of Molecular Biology, 1st Department of Obstetrics and Gynecology, General Regional Hospital "Papageorgiou"*

[1,2]*School of Medicine, Aristotle University of Thessaloniki, Greece*

**ABSTRACT:** Type 2 diabetes is considered a multifactorial trait in which multiple genetic and environmental factors interact in complex, non-linear ways to produce the common phenotype of hyperglycemia. Until recently, research efforts to identify the genetic variants that contribute to individual differences in predisposition to T2D were met with slow progress and limited success. Over the past three years, the advent of genome-wide association (GWA) scan has ushered in a new era regarding the capacity of identifying common genetic variants that contribute to predisposition to complex multifactorial phenotypes such as type 2 diabetes. The identification of the variants, genes and pathways implicated in T2D pathogenesis might facilitate its diagnosis and prevention and offer a route to new therapies. The aim of this paper is to review the literature regarding SNPs that have been associated with T2D predisposition.

*Key Words: Single nucleotide polymorphism, Type 2 Diabetes, Genome-wide association scan.*

## INTRODUCTION

Type 2 diabetes (T2D) encompasses a heterogeneous group of metabolic disorders characterized by varying degrees of insulin resistance, impaired insulin secretion and increased glucose production[1].

According to the World Confederation for combating diabetes, there has been a dramatic increase in the prevalence of the disease over the last two decades, while the number of patients is expected to rise worldwide from 171 million in 2000 to 336 million in 2030[2]. The increasing prevalence of type 2 diabetes worldwide is attributed to aging, population growth, urbanization and the increasing prevalence of obesity and physical inactivity[2]. In addition, there is compelling evidence that genetic factors make a major contribution to the development of T2D. Indeed, nowadays, type 2 diabetes is best described as a multifactorial trait in which multiple genetic and environmental factors interplay in complex and non-linear ways to produce the common phenotype of hyperglycemia. Understanding the genetic component of T2D pathogenesis will facilitate its treatment, diagnosis and prevention[3,4].

Over the past two decades, extensive research efforts have taken place in order to identify the genetic variants that contribute to individual differences in predisposition to T2D. However, until recently, these efforts were characterized by slow progress and limited success. Until 2006, the main approaches used to identify common genetic variants influencing common dichotomous traits, such as T2D, were linkage analysis and candidate gene studies. Both approaches suffered from certain limitations and neither of them proved particularly successful in detecting robustly replicating T2D-associated loci. The first approach suffered from being underpowered because linkage analysis is best placed to detect variants with high penetrance. Thus far, there is no evidence that common variants with high penetrance make a substantial contribution to the risk of common forms of T2D. The second approach had difficulties mainly associated with choosing credible gene candidates. For the purposes of a

*Corresponding author: Maria Poulasouchidou, MD, First Department of Obstetrics and Gynecology, "Papageorgiou" General Hospital, Ring Road, Nea Efkarpia, 546 23 Thessaloniki, Greece, Tel. +30-24630-28422, email: maria.poulasouchidou@gmail.com*

**Table 1.** GWA scans for type 2 diabetes included in this review[a].

| Study | Year published | Initial sample source | Number of cases/controls | Genotyping platform | Refs |
|---|---|---|---|---|---|
| Diabetes Gene Discovery Group 1st stage by Sladek et al. | 2007 | France | 694 cases/669 controls | Illumina Human-1 Illumina HumanHap300 | (20) |
| deCODE Genetics | 2007 | Iceland | 1399 cases/5275 controls | Illumina HumanHap300 | 21) |
| Wellcome Trust Case Control Consortium (WTCCC) | 2007 | UK | 1924 cases/2938 controls | Affymetrix SNP Array 5.0 | (22) |
| Diabetes Genetics Initiative (DGI) | 2007 | Finland Sweden | 1464 cases/1467 controls | Affymetrix SNP Array 5.0 | (23) |
| Finland-US investigation of NIDDM Genetics (FUSION) | 2007 | Finland | 1161 cases/1174 controls | Illumina HumanHap300 | (24) |
| DiaGen | 2007 | East Finland, Germany, UK, Ashkenazi | 500 cases/497 controls | Illumina HumanHap300 | (25) |
| BioBank Japan | 2008 | Japanese | 194 cases/1556 controls | Custom set of 268k SNPs | (26) |
| Takeuchi et al. | 2009 | Japanese | 519 cases/503 controls | Illumina HumanHap550 | (27) |
| Diabetes Gene Discovery Group 2nd stage by Rung et al. | 2009 | France | 679 cases/697 controls | Illumina Human-1 Illumina HumanHap300 | (28) |
| HanChinese | 2010 | Chinese | 995 cases/894 controls | Illumina HumanHap550 | (29) |

[a]*Studies are listed in ascending order of publication.*

candidate gene study, the selection of genes was typically based on hypotheses about probable biological mechanisms involved in T2D pathogenesis. However, the poor characterisation of the function of much of the genome rendered the selection of candidate genes difficult. In addition, poor understanding of the architecture of genetic variation, low-throughput genotyping platforms available at the time, the small sample sizes deployed and the use of liberal thresholds for declaring significance were some of the factors that hindered candidate-gene approaches from identifying reproducibly associated T2D variants[5-7].

Hence, until 2006, only two of the many T2D-associated variants reported by candidate gene studies had been convincingly replicated: the Pro12Ala and Glu23Lys variants in *PPARG* and *KCNJ11* genes respectively[8,9]. The protein encoded by the *PPARG* (peroxisome proliferator-activated receptor gamma) gene is a regulator of adipocyte differentiation and represents the target for the thiazolidinedione class of drugs used to treat T2D[10,11]. The *KCNJ11* (potassium inwardly-rectifying channel, subfamily J, member 11) gene encodes a subunit of an inwardly rectifying ATP-sensitive potassium channel. In pancreatic beta

**Table 2.** GWA scans for type 2 diabetes not included in this review[a].

| Study | Year published | Initial sample source | Number of cases/controls | Genotyping platform | Refs |
|---|---|---|---|---|---|
| Hanson et al. | 2007 | Pima Indians | 300 cases/334 controls | Affymetrix 100k | (31) |
| Hayes et al. | 2007 | Mexican Americans | 281 cases/280 controls | Affymetrix 100k | (32) |
| Rampersaud et al | 2007 | Amish | 124 cases/295 controls | Affymetrix 100k | (33) |
| Florez et al. | 2007 | Massachusetts | 91 cases/1087 controls | Affymetrix 100k | (34) |
| Yasuda et al. | 2008 | Japanese | 187 cases/1504 controls | JSNP 100k SNPs | (35) |

[a]*Studies are listed in ascending order of publication.*

cells, these channels are crucial for the regulation of glucose-induced insulin secretion and are the target for the sulfonylureas[10,11]. The strongest, thus far, known association with T2D was identified in 2006 through linkage fine-mapping study on chromosome 10[12]. It resides within *TCF7L2* (transcription factor 7 like 2) gene and has been replicated in multiple populations[13-18].

Over the past three years, the advent of genome-wide association (GWA) scan has ushered in a new era regarding the capacity of identifying common genetic variants that contribute to predisposition to complex multifactorial phenotypes such as type 2 diabetes. The implementation of the GWA approach was the result of three components: 1) the development of high-throughput genotyping platforms that enabled the conduct of massive SNP typing at high accuracy and low cost, 2) the availability of large sample collections which enhanced power in association studies and 3) a better understanding of patterns of sequence variation resulting from international efforts such as the Human Genome Sequencing Project and the HapMap[19].

*GWA scans for type 2 diabetes*

For the purposes of this review, we examined GWA scans that have genotyped over 150,000 SNPs. We found ten such GWA studies (20-29): seven conducted in European populations, two conducted in Japanese population and one carried out in a Han Chinese population (Table 1). Moreover, a meta-analysis of three such GWA scans, which is in fact the first of its kind, is reported in this paper (30). Five GWA scans[31-35] that have genotyped less than 150,000 SNPs were not included (Table 2).

Common features of most of the studies reported in this paper are the case-control study design and the employment of commercially available, fixed content genotyping platforms. Other aspects such as the specific populations examined, sample size and ascertainment, the extent to which cases and controls were matched and follow-up strategies differ among studies.

The first published GWA study for T2D was conducted by Sladek et al.[20]. In particular, the Diabetes Gene Discovery Group undertook a two-stage GWA scan, the fist one performed by Sladek et al. and the second by Rung et al. In the first stage of the study, 392,935 SNPs were genotyped in 694 T2D cases and 669 controls from France. Genotypes for each study subject were obtained using two platforms: Illumina Human-1 to assay SNPs chosen using a gene-centred design and Illumina HumanHap300 to assay SNPs chosen to tag haplotype blocks identified by the Phase 1 Hap Map. Selection criteria such as family history of at least one affected first degree relative, age at onset under 45 years and body mass index (BMI) $< 30$ Kg/m$^2$ were chosen for the diabetic subjects in order to decrease phenotypic heterogeneity and enrich for variants that determine insulin resistance and β-cell dysfunction through mechanisms other than severe obesity. Because of unequal male/female ratios in cases and controls, 12,666 X-linked SNPs were separately analyzed for each gender, however none of them attained significance. In total, 59 autosomal

SNPs showing significant association at stage 1 were genotyped in a bigger sample of 2,617 T2D cases and 2,894 controls. The study confirmed the previously known association of *TCF7L2* (rs7903146; OR = 1.65 ± 0.19) and identified significant associations for 7 SNPs representing four novel T2D susceptibility loci: *SLC30A8* (rs13266634; OR = 1.18 ± 0.25), *HHEX/IDE* (rs7923837; OR = 1.22 ± 0.21), *LOC387761* (rs7480010; OR = 1.14 ± 0.13) and *EXT2* (rs11037909; OR = 1.27 ± 0.30).

Four further GWA studies were published shortly afterwards[21-24]. Steinthorsdottir et al.[21] using Illumina Human-1 genotyped 313,179 SNPs in a sample of 1,399 T2D cases and 5,275 controls from Iceland. In addition, 339,846 two-marker haplotypes were tested. The previously identified SNP, rs7903146, in TCF7L2 gave the most significant results with OR = 1.38 and P = $1.82x10^{-10}$ in all individuals with type 2 diabetes, whereas no other SNP or haplotype was significant after adjustment for the number of tests performed. In total, 51 SNPs (single SNPs and two-marker haplotypes with P < 0.00005) were selected for replication in a sample of 1,110 T2D cases and 2,272 controls from Denmark. In the Danish group of T2D cases, two SNPs were significantly associated: rs7756992 and rs13266634 (OR = 1.24, P = 0.00013 and OR = 1.20, P = 0.0012 respectively). These SNPs were additionally genotyped in three other T2D case-control groups of European ancestry from Denmark, the Netherlands and Philadelphia (of total sample size 3,836 cases and 12,562 controls) as well as in case-control groups from Hong Kong (1,457 cases/986 controls) and West Africa (865 cases/1.106 controls). The SNP rs7756992 in the CDKAL1 gene was associated with T2D in individuals of European ancestry (OR = 1.20, P = $7.7x10^{-9}$) and individuals from Hong Kong of Han Chinese ancestry (OR = 1.25, P = 0.00018). In conclusion, the deCODE study confirmed the previously known association of TCF7L2 gene with T2D and identified CDKAL1 as new T2D-susceptibility locus.

The CDKAL1 locus was independently identified by three further contemporaneous studies performed by the Wellcome Trust Case Control Consortium (WTCCC), Diabetes Genetics Initiative (DGI) and Finland-United States Investigation of NIDDM genetics (FUSION)[22-24]. These three studies collaborated by sharing data with respect to replication stage and therefore in this paper we present the findings of their shared efforts. WTCCC[22] is a joint GWA scan of 7 common diseases: bipolar disorder, coronary artery disease, hypertension, Crohn's disease, rheumatoid arthritis, type 1 diabetes and type 2 diabetes. The study was conducted in the British population using groups of ~2,000 cases for each of the 7 examined diseases and a shared group of ~3,000 controls. The genotyping platform, Affymetrix SNP array 5.0, was used to genotype a total number of 469,557 SNPs. The group of T2D cases was comprised of 1,924 individuals while the control group was comprised of 2,938 individuals from the UK. In order to decrease the phenotypic heterogeneity of T2D the researchers employed selection criteria such as: British/Irish ancestry, family history of type 2 diabetes and age at onset under 65 years. The study identified 24 independent association signals for the seven diseases examined, of which three were association signals for T2D: TCF7L2, FTO and CDKAL1. The DGI[23] study employed the same genotyping platform, Affymetrix SNP array 5.0, and successfully genotyped 386,731 SNPs in 1,464 T2D cases and 1,467 controls from Finland and Sweden. In addition, 284,968 haplotypes were examined. Case and control groups were matched for sex, age, BMI and place of origin. 107 SNPs were selected for replication in a sample of 10,850 individuals (cases and controls) from Sweden, Poland and U.S.A. Scott et al[24], using the genotyping platform Illumina HumanHap 300, assayed 317,503 SNPs in 1.161 T2D cases and 1,174 controls from Sweden. From the first stage of the study, 82 SNPs were chosen (after comparison with WTCCC and DGI studies) for replication in 1,215 T2D cases and 1,258 controls. The shared efforts of these three GWA scans resulted in the identification of three new T2D-susceptibility loci: CDKN2A/2B (rs10811661; OR = 1.20, CI = 1.14-1.25, P = 7.8 x $10^{-10}$), FTO (rs8050136; OR = 1.17, CI = 1.12-1.22, P = 1.3 x $10^{-12}$) and IGF2BP2 (rs4402960; OR = 1.14, CI = 1.11-1.18, P = $8.9x10^{-10}$). Furthermore, the association of previously reported loci with T2D was confirmed: PPARG (rs1801282), KCNJ11 (rs5219), TCF7L2 (rs7903146) and CDKAL1 (rs7754840).

Salonen et al[25], using the genotyping platform Illumina HumanHap300, genotyped almost 300,000

SNPs in 500 cases of T2D and 497 controls originated from four different populations. In particular, the study was conducted in 200 cases and 200 controls from Eastern Finland, 200 cases and 197 controls from Israel and finally 99 cases and 100 controls from Germany and England. The phenotypic heterogeneity of T2D cases was restricted using selection criteria such as: family history of T2D and age at onset under 60 years. Initially, 315,917 SNPs were genotyped in the sample mentioned above, whereas in the replication study, 10 SNPs showing the strongest statistically significant association where chosen to be genotyped in a sample of 2,573 T2D cases and 2,776 controls. It should be noted that this replication study sample was the same one used for replication by Sladek et al. The study confirmed the previously known association of the rs7903146 SNP in *TCF7L2* gene but, mainly due to restricted power failed to identify any novel T2D association signals.

Unoki et al.[26] performed the first GWA scan in a population of non European descent. In particular, 207,097 SNPs that cover more than 50% of common SNPs in the Japanese were successfully genotyped in 194 cases with T2D and diabetic retinopathy and in 1,558 controls. Subsequently, 8,323 SNPs with the highest levels of significance were genotyped in 1,367 individuals with T2D and diabetic retinopathy and in 1,266 controls. Of 6,731 SNPs for which data were obtained successfully, the researchers selected 9 SNPs ($p < 0.0001$) and genotyped them in a third set of cases and controls comprised of 3,557 Japanese individuals with T2D and 1,352 controls. The study confirmed the association of the *CDKAL1* and *IGF2BP2* loci and additionally identified *KCNQ1* (rs2283228, OR = 1.26, 95% CI = 1.18-1.34) as a novel T2D susceptibility locus. A smaller GWA study performed in the Japanese by Yasuda et al.[35], also identified *KCNQ1* as T2D susceptibility locus.

Takeuchi et al[27] performed a three-staged genome-wide association study in the Japanese population. Using Illumina HumanHap 550, a total number of 482,625 SNPs was successfully genotyped in 519 T2D cases and 503 controls. Subsequently, 1,456 SNPs were genotyped in the second stage sample comprised of 1,110 T2D cases and 1,014 controls. At this stage, 30 SNPs representing 17 unique loci attained significance (defined as $p < 7 \times 10^{-9}$). One SNP of each of these 17 loci was selected for typing in the third stage sample of the study comprised of 4,000 T2D cases and 4,889 controls. Genome-wide significant association was revealed for three previously reported loci: *CDKAL1* (rs4712523; OR = 1.27, CI = 1.21-1.33, $P = 7.2 \times 10^{-20}$), *CDKN2A/CDKN2B* (rs2383208; OR = 1.34, CI = 1.27-1.41, $P = 2.1 \times 10^{-20}$) and *KCNQ1* (rs2237892; OR = 1.33, CI = 1.27-1.41, $P = 1.1 \times 10^{-20}$).

The second stage of Diabetes Genes Discovery study, performed by Rung et al[28] (the first one was performed by Sladek et al), was recently published. Of 392,365 SNPs that were initially genotyped in a total sample of 1,376 individuals from France, the researchers selected 16,273 SNPs and genotyped them in 2,245 T2D cases and in 2,732 controls from France. Subsequently, 28 SNPs were chosen for replication study in 7,698 Danish individuals comprised of 3,334 T2D cases and 4,364 controls. The study confirmed the association of the previously reported loci *TCF7L2* (rs7903146), *CDKAL1* (rs4712523) and *WFS1* (rs4689388), and identified one SNP (rs2943641; OR = 1.19, CI = 1.13-1.25) located adjacent to *IRS1* gene as a novel T2D-susceptibility loci.

The last to date published GWA scan was conducted in a Han Chinese population residing in Taiwan. Tsai et al.[29] undertook a two-stage genome-wide association study, in which 995 T2D cases and 894 controls were genotyped using the Illumina HumanHap550 chip for the first genome scan stage. In this stage, high-quality genotypes were obtained for 517,737 SNPs. Subsequently, 39 SNPs were selected for replication in 1,803 T2D cases and 1,473 controls in stage 2. The study confirmed the association of *KCNQ1* gene with T2D risk (rs2237895; OR = 1.29, CI = 1.19-1.40, $P = 9.65 \times 10^{-10}$) and identified two novel T2D susceptibility loci: *PTPRD* (rs17584499; OR = 1.57, CI = 1.36-1.82, $P = 8.54 \times 10^{-10}$) and *SRR* (rs391300; OR = 1.28, CI = 1.18-1.39, $P = 3.06 \times 10^{-9}$). Both *PTPRD* and *SRR* genes have not been associated to T2D in previous GWA scans and need further replication in additional populations.

*Meta-Analysis of three GWA scans for type 2 diabetes*

Although GWA scans constitute the present state-of-

the-art in efforts of elucidating the genetic component of complex diseases, they still face limitations regarding the power of individual studies to detect common SNPs with small or modest effects. Statistical power can be enhanced through the combination of datasets and through the concomitant examination of both directly genotyped and imputed SNPs[7,19]. SNPs that are directly genotyped are typically analysed by single-point methods[19]. Imputation approaches, on the other hand, are multipoint methods that take into account information from multiple surrounding markers and can detect SNPs that have not been directly typed[36]. Hence, imputation allows the examination of more than 2 million genetic markers genome-wide in parallel and moreover enables the combination of data from GWA scans carried out on different genotyping platforms[19,36].

The DIAGRAM Consortium (Diabetes Genetics, Replication And Meta-Analysis) is the first complex disease GWA scan meta-analysis conducted by Zeggini et al (30). DIAGRAM combined data across 4,549 T2D cases and 5,579 controls, all of European descent, from the studies WTCCC, DGI and FUSION. The genotyped autosomal SNPs that passed quality control filters in each study were: a) 393,143 SNPs from the Affymetrix SNP Array 5.0 in WTCCC (MAF > 0.01, 1,924 T2D cases and 2,938 controls), b) 378,860 SNPs from the Affymetrix SNP Array 5.0 in DGI (MAF > 0.01, 1,464 T2D cases and 1,476 controls) and c) 306,222 SNPs from the Illumina HumanHap300 in FUSION (MAF > 0.01, 1,161 T2D cases and 1.174 controls). There were 44,700 SNPs directly genotyped in all three studies across the two platforms. In addition, 1,570,311 SNPs were imputed in each sample resulting in the final examination of 2,202,892 variants across a total sample of 10,128 individuals. Sixty nine SNPs showing the strongest associations (meta-analysis p value $< 10^{-4}$ and meta-analysis heterogeneity p value $> 10^{-4}$), were selected for replication in 22,426 individuals of European descent. The top 11 signals ($p < 10^{-5}$) emerging from this second stage were further genotyped in 57,366 individuals of European descent in stage 3.

After integrating data from all three stages of the meta-analysis, six signals reached genome-wide significance levels ($p \leq 5 \times 10^{-8}$) for T2D association: *JAZF1* (OR = 1.10, CI = 1.07-1.13), *CDC123*/*CAMK1D* (OR = 1.11, CI = 1.07-1.14), *TSPAN8*/*LGR5* (OR = 1.09, CI = 1.06-1.12), *THADA* (OR = 1.15, CI = 1.10-1.20), *ADAMTS9* (OR = 1.09, CI = 1.06-1.12), and *NOTCH2* (OR = 1.13, CI = 1.08-1.17).

*The genetic component of T2D pathogenesis: knowledge gained*

Thus far, the literature counts 18 robustly replicating T2D-susceptibility loci (Table 3). Susceptibility loci are selected on the basis of their colocalization to the same interval as an association signal. However, colocalization does not prove causation because genetic variants may be exerting their influence on T2D pathogenesis through regulatory effects on genes whose coding sequences lie some distance away. That is the case with several of the T2D-associated signals that map to 'gene deserts', whereas others map to regions containing good candidates, such as *SLC30A8* and *HHEX/IDE* genes[37]. In addition, the task of moving from an association signal to mechanistic insights is difficult and requires identification of the causative variants through resequencing, fine-mapping and functional studies. Hence, early inferences regarding etiopathological mechanisms should be cautious and it should be expected that occasionally, genes other than the regional candidates might prove out to be responsible for the susceptibility effect[37, 5].

Most of the T2D-associated signals identified to date (involving SNPs within or adjacent to *KCNJ11, HNF1B, SLC30A8, HHEX/IDE, CDKAL1, CDKN2A/B, IGFBP2, JAZF1 and TCF7L2*) appear to exert their primary effect on diabetes pathogenesis through disruption of insulin secretion[7]. This is consistent with growing evidence supporting the substantial role of defects in β-cell function and mass in the development of diabetes[38] and it could be stated that the inherited component of T2D predisposition is predominantly characterized by alterations of β-cell function and mass[37]. However, we should also consider that the genetic variants identified thus far account for only a small proportion of the overall genetic risk of the disease. Moreover, selection criteria employed by several GWA studies, such as lean T2D cases, may have generated bias against the detection of variants affecting insulin action[37].

**Table 3.** Summary of confirmed T2D-associated loci and SNPs[a].

| Locus/gene | Chr | Index SNP* | Position | Effect size ** | Risk-allele frequency ** | Year association identified | Study type | Probable mechanism |
|---|---|---|---|---|---|---|---|---|
| PPARG | 3 | rs1801282 | 12368125 | 1.14 | 0.87 | 2000 | Candidate gene | Insulin action |
| KCNJ11 | 11 | rs5219 | 17366148 | 1.15 | 0.40 | 2003 | Candidate gene | β-cell dysfunction |
| TCF7L2 | 10 | rs7901695 | 114744078 | 1.37 | 0.31 | 2006 | Linkage peak fine-mapping | β-cell dysfunction |
| HHEX/IDE | 10 | rs5015480 | 94455539 | 1.13 | 0.63 | 2007 | GWA | β-cell dysfunction |
| SLC30A8 | 8 | rs13266634 | 118253964 | 1.15 | 0.69 | 2007 | GWA | β-cell dysfunction |
| CDKAL1 | 6 | rs10946398 | 20769013 | 1.14 | 0.32 | 2007 | GWA | β-cell dysfunction |
| FTO | 16 | rs8050136 | 52373776 | 1.17 | 0.40 | 2007 | GWA | Altered BMI |
| CDKN2A/B | 9 | rs10811661 | 22124094 | 1.20 | 0.83 | 2007 | GWA | β-cell dysfunction |
| IGF2BP2 | 3 | rs4402960 | 186994381 | 1.14 | 0.32 | 2007 | GWA | β-cell dysfunction |
| HNF1B (TCF2) | 17 | rs757210 | 33170628 | 1.08 | 0.43 | 2007 | Large-scale association | β-cell dysfunction |
| WFS1 | 4 | rs10010131 | 6343816 | 1.12 | 0.60 | 2007 | Large-scale association | Unknown |
| KCNQ1 | 11 | rs2237892 | 2796327 | 1.29 | 0.93 | 2008 | GWA | β-cell dysfunction |
| JAZF1 | 7 | rs864745 | 28147081 | 1.10 | 0.50 | 2008 | GWA meta-analysis | β-cell dysfunction |
| CDC123/ CAMK1D | 10 | rs12779790 | 12368016 | 1.11 | 0.18 | 2008 | GWA meta-analysis | Unknown |
| TSPAN8/ LGR5 | 12 | rs7961581 | 69949369 | 1.09 | 0.27 | 2008 | GWA meta-analysis | Unknown |
| **THADA** | 2 | rs7578597 | 43586327 | 1.15 | 0.90 | 2008 | GWA meta-analysis | Unknown |
| **ADAMTS9** | 3 | rs4607103 | 64686944 | 1.09 | 0.76 | 2008 | GWA meta-analysis | Unknown |
| **NOTCH2** | 1 | rs10923931 | 120319482 | 1.13 | 0.10 | 2008 | GWA meta-analysis | Unknown |

[a]*Data populating this table are derived through refs 5, 7 and 37.*

**A representative SNP with strong statistical significance is presented for each of these loci, because the true causal variants have not yet been identified.*

***Effect size (estimated as per allele odds ratios) and risk allele frequencies are calculated in European populations.*

A compelling candidate for type 2 diabetes pathogenesis is *SLC30A8* (solute carrier family 30, member 8) gene. This gene encodes a zinc transporter (ZnT8) which is expressed solely in pancreatic islets of Langerhans and colocalizes with insulin in the secretory pathway granules of the insulin-secreting INS-1 cells. ZnT8 knockout mice showed normal glucose tolerance and insulin sensitivity when fed on a control diet. However, after high-fat diet feeding the mice became glucose intolerant or diabetic. These data show that ZnT8 transporter is essential for the formation of insulin crystals in beta cells, contributing to the packaging efficiency of stored insulin[39]. Overexpression of *SLC30A8* in insullinoma cells increases glucose-stimulated insulin secretion[40]. On chromosome 10, the association signal maps to a region that contains the coding sequences of two strong candidates: *HHEX* and *IDE* genes[5]. The protein encoded by the *HHEX* (hematopoietically expressed homeobox) gene is a member of the homeobox family of transcription factors, many of which are involved in developmental processes[10,11]. *HHEX* has been implicated in pancreas development and the regulation of insulin secretion[41]. This gene, as well as TCF7L2, is involved in the regulation of Wnt-signalling, indicating that dysfunction of this pathway may contribute to the development of diabetes[7]. *IDE* (insulin-degrading enzyme) gene codes for a zinc metallopeptidase that degrades intracellular insulin, and thereby terminates insulin's activity[10,11].

At least four of the T2D-associated signals map to regions (*CDKAL1, CDKN2A/B, CDC23 and KCNQ1*) that are thought to be involved in cell cycle regulation[7]. The function of the protein encoded by *CDKAL1* [cyclin-dependent kinase 5 (CDK5) regulatory subunit associated protein 1-like 1] gene has not yet been identified. However, this protein product is similar to CDK5 regulatory subunit associated protein 1, which has been implicated in the loss of β-cell function under conditions of stimulation by high glucose[42]. The *CDKN2A* and *CDKN2B* genes encode cyclin-dependent kinase inhibitors that are involved in cell-cycle regulation in the β-cell. Transgenic mice with overexpression of the CDKN2A demonstrated decreased islet proliferation after exposure to a specific β-cell toxin. These data demonstrate that altered expression of the gene recapitulates the T2D phenotype[43].

Two of the reported T2D-susceptibility loci, *WFS1* and *HNF1B*, have emerged from large scale association studies. *WFS1* (Wolfram syndrome 1) gene is highly expressed in brain, pancreas, heart, and insullinoma beta-cell lines and encodes for a transmembrane protein, which is located primarily in the endoplasmic reticulum[10,11]. Mutations in this gene are associated with Wolfram syndrome, an autosomal recessive disorder, featured by diabetes insipidus, diabetes mellitus, optic atrophy and deafness[44]. Hepatocyte nuclear factor 1 homeobox B (*HNF1B*) gene, also known as *TCF2*, encodes a transcription factor which is involved in the development of embryonic pancreas. Mutations in this gene cause monogenic forms of diabetes[45]. Apart from these genes, two other T2D-associated loci are also known to harbour rare mutations that are causal for monogenic forms of diabetes: *PPARG*[46] and *KCNJ11*[47].

Finally, several of the genes and variants associated with T2D susceptibility appear to influence predisposition to other common disease traits as well. Thus, variation in *CDKAL1* has been revealed to have effect on susceptibility to Crohn's disease[48], variants at *HNF1B* and *JAZF1*[49,50] have been associated to prostate cancer, whereas the CDKN2A and CDKN2B genes are also known to play a role in the development of cancer[51]. These pleiotropic effects might indicate overlap in pathogenetic pathways, though in some cases, the pleiotropic relationships cross the accepted boundaries of comorbidity[7].

## DISCUSSION

The discipline of molecular genetics has always been driven by the force of gaining novel biological insights. A crucial step towards this direction is the identification of genes implicated in disease pathogenesis. From this perspective, GWA studies turned out to be the most fruitful approach in efforts of elucidating variants and genes that predispose to complex disease traits such as T2D. However the road ahead in T2D genetics is still long.

The 18 T2D-susceptibility loci identified to date, harbor common genetic variants with modest effect sizes[5-7,37]. As table 3 attests, the risk allele frequency of the representative genetic variants ranges from 0.10 to 0.93. Per-allele effect sizes range from 1.08 to 1.29, with the exception of the *TCF7L2* gene where

the strongest association signal has been reported with per-allele effect size of 1.37. These spectrums of risk allele frequencies and effect sizes are typical of what is observed in GWA scans for other complex multifactorial diseases[52]. Less than 12 genetic variants associated with complex diseases, have been detected thus far having high effect size ranging from 2 to 10[52]. Hence, it is indicated that only a limited amount of the genetic variance predisposing to complex traits has been identified so far. In the case of T2D, the 18 known variants explain less than 4% of the inherited predisposition, which implies that many more genes contributing to T2D predisposition remain to be discovered[53]. Assuming that the undiscovered variants are of the same genetic architecture with those identified to date, then more than 800 of such genetic variants will be needed in order to explain 40% of T2D heritability[53].

The missing heritability of T2D is assumed to lie in rare variants with intermediate penetrance, as well as structural variants[54,7]. Rare variants are likely to have larger effect sizes than those of common variants and might be population specific[52]. The GWA scans performed thus far have focused on the detection of susceptibility effects attributable to common SNPs and even with high density platforms, a proportion of such variants remains undetected. Currently employed genotyping arrays offer limited power to detect rare and structural variants and next-generation technologies are expected to address these challenges[54].

For the majority of the genetic variants associated with T2D thus far, the identified association remains probabilistic and not causal. Extensive resequencing, fine-mapping and functional studies are required in order to identify the causative variants and define the molecular mechanisms through which they function[52,5]. Gene redundancy, functional overlap of genes and gene-environment interactions render the identification of the truly causative variants difficult. In addition, we have limited knowledge of the extent to which epigenetic effects might mimic genetic predisposition[7].

In summary, future steps in type 2 diabetes research, include: a) the identification of new SNP associations with the conduct of GWA scans in different populations and enlarged meta-analyses, b) the detection of rare and structural variants contributing to T2D risk, c) the definition of causative variants and d) the conduct of animal or cellular studies in order to understand the function of causative variants[54, 5].

The knowledge to be gained through these efforts is expected to affect the management of T2D. In the short term, new insights in molecular pathways implicated in T2D pathogenesis are expected to reveal new therapeutic targets. The second clinical translation of this knowledge lies in the ability of using genetic variants as tools for predicting individual disease risk and response to drugs. The major challenge in T2D genetics is to establish how the knowledge derived from T2D predisposing variants can be used to improve the care of the diabetic patient[37,52].

# SNPs και Σακχαρώδης Διαβήτης τύπου ΙΙ.

Μαρία Πουλασουχίδου[1], Αλέξανδρος Λαμπρόπουλος[2]

[1]*Εργαστήριο Μοριακής Βιολογίας, Α΄ Μαευτική - Γυναικολογική Κλινική ΑΠΘ, «Παπαγεωργίου» Γενικό Νοσοκομείο Θεσσαλονίκης*
[2]*Ιατρική Σχολή, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Ελλάδα*

**ΠΕΡΙΛΗΨΗ:** Ο Σακχαρώδης Διαβήτης τύπου 2, θεωρείται πολυπαραγοντικό νόσημα στο οποίο ποικίλοι γενετικοί και περιβαλλοντικοί παράγοντες αλληλεπιδρούν με σύνθετους, μη γραμμικούς τρόπους για να προκαλέσουν τον κοινό φαινότυπο της υπεργλυκαιμίας. Μέχρι πρόσφατα, οι προσπάθειες να προσδιοριστούν οι γενετικές ποικιλομορφίες που συμβάλλουν στις ατομικές διαφορές προδιάθεσης για ΣΔ τύπου 2, χαρακτηριζόταν από αργή πρόοδο και περιορισμένη επιτυχία. Κατά τα τελευταία τρία χρόνια, με την έλευση της ευρείας συσχέτισης γονιδιωματικής ανάλυσης: genome-wide association (GWA) scan, υπήρξε μια θεαματική αλλαγή στην ικανότητα προσδιορισμού κοινών γενετικών παραλλαγών που συμβάλλουν στην προδιάθεση για πολύπλοκους πολυπαραγοντικούς φαινότυπους όπως ο ΣΔ τύπου 2. Η αναγνώριση των γενετικών παραλλαγών, γονιδίων και παθοφυσιολογικών δρόμων που εμπλέκονται στην παθογένεια του ΣΔ τύπου 2 μπορεί να συμβάλλει στην ανάπτυξη νέων θεραπευτικών προσεγγίσεων καθώς επίσης και στην καλύτερη διάγνωση και πρόληψη της νόσου. Σκοπός της παρούσας εργασίας είναι η ανασκόπηση της βιβλιογραφίας αναφορικά με SNPs που έχουν συσχετιστεί με προδιάθεση για ΣΔ τύπου 2.

*Λέξεις Κλειδιά: Πολυμορφισμός μονών νουκλεοτιδίων, Σακχαρώδης Διαβήτης τύπου 2, Ευρείας συσχέτισης γονιδιωματική ανάλυση.*

## REFERENCES

1. Kasper DL, Fauci AS, Longo DL, Braunwald E, Hauser SL, Jameson JL, et al. ,εκδότες. Harrison: εσωτερική παθολογία. 16η εκ. Επιστημονικές Εκδόσεις Παρισιάνου; 2005. vol 3 p.2319.
2. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 2004;27:1047-53.
3. Gloyn AL, McCarthy MI. The genetics of type 2 diabetes. Best Pract Res Clin Endocrinol Metabol 2001; 15(3):293-308.
4. Owen KR, McCarthy MI. Genetics of type 2 diabetes. Curr Opin Genet Dev 2007; 17:239-244.
5. Prokopenko I, McCarthy MI, Lindgren CM. Type 2 diabetes: new genes, new understanding. Trends Genet 2008; 24(12):613-621.
6. McCarthy MI, Zeggini E. Genome-wide association scans for type 2 diabetes: new insights into biology and therapy. Trends Pharmacol Sci 2007; 28(12):598-601.
7. McCarthy MI, Zeggini E. Genome-wide association studies in type 2 diabetes. Curr Diab Rep 2009; 9(2):164-171.
8. Stumvoll M, Haring H. The peroxisome proliferator-activated receptor-gamma2 Pro12Ala polymorphism. Diabetes 2002; 51(8):22341-7.
9. Nielsen EM, Hansen L, Carstensen B, Echwald SM, Drivsholm T, Glumer C, et al. The E23K variant of Kir6.2 associates with impaired post-OGTT serum insulin response and increased risk of type 2 diabetes. Diabetes 2003; 52(2):573-7.
10. Entrez Gene: URL: http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd
11. OMIM: Online mendelian inheritance in man: URL: http://www.ncbi.nlm.nih.gov/omim
12. 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.
13. Zhang C, Qi L, Hunter DG, Meigs JB, Manson JE, van Dam RM, et al. Variation of transcription factor 7-like 2 (TCF7L2) gene and the risk of type 2 diabetes in large cohorts of U.S. women and men. Diabetes 2006; 55(9):2645-8.
14. Scott LJ, Bonnycastle LL, Willer CJ, Sprau AG, Jackson AU, Narisu N, et al. Association of transcription

factor 7-like 2 (TCF7L2) variants with type 2 diabetes in a Finnish sample. Diabetes 2006; 55(9):2649-53.
15. Chandak GR, Janipalli CS, Bhaskar S, Kulkarni SR, Mohankrishna P, Hattersley AT, et al. Common variants in the TcF7L2 gene are strongly associated with type 2 diabetes mellitus in the Indian population. Diabetologia 2007; 50(1):63-7.
16. Ng MC, Tam CH, Lam VK, So WY, Ma RC, et al. Replication and identification of novel variants at TCF7L2 associated with type 2 diabetes in Hong Kong Chinese. J Clin Endocrinol Metab 2007; 92:3733-37.
17. Groves CJ, Zeggini E, Minton J, Frayling TM, Weedon MN, Rayner NW, et al. Association analysis of 6,736 U.K. subjects provides replication and confirms TCF7L2 as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. Diabetes 2006; 55:2640-44.
18. Tong Y, Lin Y, Zhang Y, Yang J, ZhangY, Liu H, et al. Association between TCF7L2 gene polymorphisms and susceptibility to type 2 diabetes mellitus: a large Human Genome Epidemiology (HuGE) review and meta-analysis. BMC Med Genet 2009;10:15.
19. Panoutsopoulou K, Zeggini E. Finding common susceptibility variants for complex disease: past, present, future. Brief Funct Genomic Proteomic 2009; 8(5):345-352.
20. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007; 445:881-885.
21. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, et al. A variant in CDKAL 1 influences insulin response and risk of type 2 diabetes. Nat Genet 2007; 39:770-775.
22. Welcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007; 447:661-678.
23. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007; 316(5829):1331-36.
24. 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; 316:1341-45.
25. Salonen JT, Uimari P, Aalto JM, Pirskanen M, Kaikkonen J, Todorova B, et al. Type 2 diabetes whole-genome association study in four populations: The DiaGen Consortium. Am J Hum Genet 2007;81:338-345.
26. Unoki H, Takahashi A, Kawaguchi T, Hara K, Horikoshi M, Andersen G, et al. SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. Nat Genet 2008; 40:1098-1102.
27. Takeuchi F, Serizawa M, Yamamoto K, Fujisawa T, Nakashima E, Ohnaka K, et al. Confirmation of multiple risk loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. Diabetes 2009; 58(7):1690-9.
28. Rung J, Cauchi S, Albrechen A, Shen L, Rocheleau G, Cavalcanti-Proenca C, et al. Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. Nat Genet 2009; 41:1110-17.
29. Tsai FJ, Yang CF, Chen CC, Chuang LM, Lu CH, Chang CT, et al. A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. Plos Genet 2010; 6(2):e1000847.
30. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 2008; 40(5):638-645.
31. Hanson RL, Bogardus C, Duggan D, Kobes S, Knowlton M, Infante AM, et al. A search for variants associated with young-onset type 2 diabetes in American Indians in a 100k genotyping array. Diabetes 2007; 56(12):3045-52.
32. Hayes MG, Pluzhnikov A, Miyake K, Sun Y, Ng MC, Roe CA, et al. Identification of type 2 diabetes genes in Mexican Americans through genome-wide association studies. Diabetes 2007; 56(12):3033-44.
33. Rampersaud E, Damcott CM, Fu M, Shen H, McArdle P, Shi X, et al. Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: evidence for replication from diabetes related quantitative traits and from independent populations. Diabetes 2007; 56(12):3053-62.
34. Florez JC, Manning AK, Dupuis J, McAteer J, Irenze K, Gianniny L, et al. A 100k genome-wide association scan for diabetes and related traits in the Framingham Heart Study: replication and integration with other genome-wide datasets. Diabetes 2007; 56(12):3063-74.
35. Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, et al. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat Genet 2008; 40(9):1092-7.
36. Marchini J, Howie B, Myers S, McVean G, Donnely P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 2007; 39(7):906-13.
37. McCarthy MI, Hattersley AT. Learning from molecular genetics: novel insights arising from the definition of genes for monogenic and type 2 diabetes. Diabetes 2008; 57:2889-98.
38. Leahy JL. Mary, Mary, quite contrary, how do your

beta-cells fail? Diabetes 2008; 57(10):2563-64.
39. Lemaire K, Ravier MA, Schraenen A, Creemers JW, Van de Plas R, Granvik M, et al. Insulin crystallization depends on zinc transporter ZnT8 expression, but is not required for normal glucose homeostasis in mice. Proc Natl Acad Sci U S A 2009; 106(35):14872-7.
40. Chimienti F, Devergnas S, Pattou F, Schuit F, Garcia-Cuenca R, Vandewalle B, et al. In vivo expression and functional characterization of the zinc transporter ZnT8 in glucose induced insulin secretion. J Cell Sci 2006; 119:4199-206.
41. Bort R, Martinez-Barbera JP, Beddington RS, Zaret KS. Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. Development 2004; 131(4):797-806.
42. Wei FY, Nagashima K, Ohshima T, Saheki Y, Lu YF, Matsusita M, et al. Cdk5-dependent regulation of glucose-stimulated insulin secretion. Nat Med 2005; 11(10):1104-8.
43. Krishnamurthy J, Ramsey MA, Ligon KL, Torrice C, Koh A, Bonner-Weir S, et al. p16INK4a induces an age-dependent decline in islet regenerative potential. Nature 2006; 443:453-7.
44. Domenech E, Gomez-Zaera M, Nunes V. Wolfram/DIDMOAD syndrome, a heterogenic and molecularly complex neurodegenerative disease. Pediatr Endocrinol Rev 2006; 3(3):249-57.
45. Maestro MA, Cardalda C, Boj SF, Luco RF, Servitja JM, Ferrer J. Distinct roles of HNF1β, HNF1α and HNF4α in regulating pancreas development, beta-cell function and growth. Endocr Dev 2007; 12:33-45.
46. Barroso I, Gurnell M, Crowley VE, Agostini M, Schwabe JW, Soos MA, et al. Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. Nature 1999; 402:880-3.
47. Gloyn Al, Pearson ER, Antcliff JF, Proks P, Bruining GJ, Slingerland AS, et al. Activating mutations in the gene encoding the ATP-sensitive potassium-channel kir6.2 and permanent neonatal diabetes. N Engl J Med 2004; 350:1838-49.
48. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, et al. Genome-wide association defines more than thirty distinct susceptibility loci for Crohn's disease. Nat Genet 2008; 40(8):955-62
49. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet 2007; 39(8):977-83.
50. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, et al. Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 2008; 40(3):310-5.
51. Finkel T, Serrano M, Blasco MA. The common biology of cancer and ageing. Nature 2007; 448(7155):767-74.
52. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. Nat Rev Genet 2009; 10(4):241-51.
53. Pawitan Y, Seng KC, Magnusson PK. How many genetic variants remain to be discovered? PLoS One 2009; 4(12):e7969.
54. McCarthy MI, Abecasis CR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 2008; 9(5):356-369.