

Identification and Validation Novel Risk Genes for Type 1 Diabetes – A Meta-Analysis

Puqiang Yang¹, Ashley Chorath², Wenjuan Jiang^{3*}

¹ General Internal Medicine, Tongling Hospital of Traditional Chinese Medicine, Tongling, Anhui 244000, P.R. China;

² Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bethesda, MD 20892, USA;

³ Department of Pharmaceutic Science, School of Pharmacy, University of Maryland, Baltimore, MD 21201, USA.

***Corresponding to:** Dr. Wenjuan Jiang, Department of Pharmaceutic Science, School of Pharmacy, University of Maryland, 20 Penn Street, MD 21201, Baltimore, USA. Email: wjiang@rx.umaryland.edu; Tel: 3157085712.

ABSTRACT

Background: Type 1 diabetes (T1D) cannot be explained by environmental factors alone and gene mutations in T1D should be also taken seriously. The aim of this study was to investigate the novel genetic risk of T1D by systematically reviewing the published literature and performing a meta-analysis.

Methods: A comprehensive search of electronic databases was completed using Illumina BioEngine. Seven T1D case/control bio-sets from four different studies were selected, including 286 T1D cases and 162 controls. The selected top T1D risk genes were further analyzed integrating an online open source T1D genetic database. Pathway enrichment analysis (PEA) and network connectivity analysis (NCA) were conducted to identify potential functional association between target genes and T1D.

Results: Six genes were identified through the meta-analysis as top target genes for T1D: HLA-DQB1, OAS1, CMKLR1, NFE2, MND4 and GPR56. These genes play roles within multiple T1D genetic pathways and demonstrate solid connections with known T1D target genes. NCA results also revealed strong functional association between these genes and T1D.

Conclusion: This study identified known as well as novel T1D target genes and their functional pathways that influence the T1D pathogenesis. Our results may provide new insights for the understanding of the genetic mechanisms of T1D.

Key words: Type 1 diabetes, ResNet Database, pathway enrichment analysis, Network connectivity analysis



<http://mo.qingres.com>

OPEN ACCESS

DOI: 10.20900/mo.20170003

Received: November 29, 2016

Accepted: January 18, 2017

Published: February 25, 2017

Copyright: ©2017 Cain *et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1 INTRODUCTION

Type 1 diabetes (T1D) is perceived as a progressive immune-mediated disease that accounts for approximately 5 % of all existing cases of diabetes in the USA^[10]. There has been a considerable worldwide increase in incidence over the past few decades^[2]. Environmental factors have been implicated in the pathogenesis of T1D as both triggers and potentiators of beta-cell destruction^[3, 4]. However, gene mutations should also be considered in response to the heterogeneity of T1D, which cannot be explained purely by environmental factors.

Numerous genetic studies of T1D have been conducted to explore candidate genes for the disease with both case-control studies and family-based studies^[5, 6]. In the 1970s, association and affected-sib pair linkage studies established the role of HLA genes in T1D predisposition^[7]. Recently, multiple modality genetic data from peripheral blood monocyte gene were employed for the continued efforts to identify T1D genetic determinants^[8-10]. These previously studies built a solid background for T1D genetic research, which could be leveraged for the discovery and evaluation of novel risk genes.

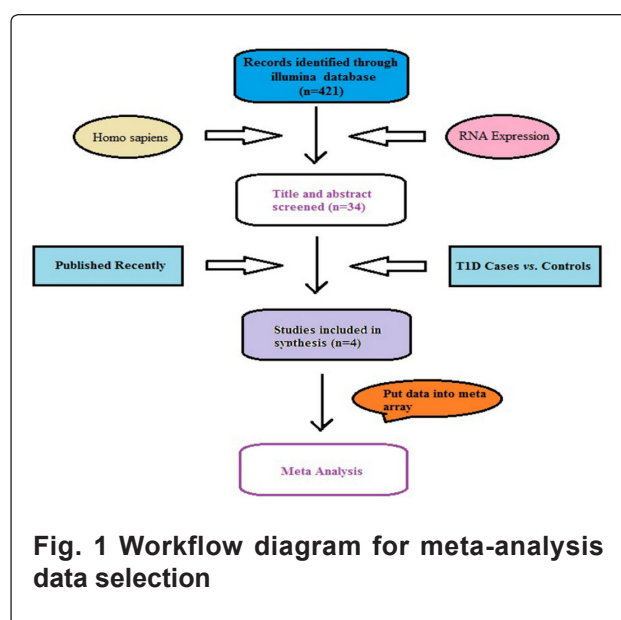
However, the risk estimates from individual studies often lack statistical power due to limited sample sizes and sample specificities in terms of phenotype characteristics. It is also difficult to come to a consistent conclusion as results are spread over a large number of independent studies. Therefore, a meta-analysis of multiple studies could provide a higher power assessment of the genetic risk factors of T1D.

In this study, a meta-analysis was performed based on four recent studies (2009-2014). The top genes from the study were further analyzed integrating a curated T1D genetic database (T1D_GD). The T1D_GD database was constructed using a large scale literature knowledge database, Pathway Studio (PS) ResNet database. In recent years, PS ResNet database has been widely used to study modeled relationships between proteins, genes, complexes, cells, tissues and diseases (<http://pathwaystudio.gousinfo.com/Mendeley.html>). Our study identified novel T1D genes and evaluated the effectiveness of integrating meta-analysis and PS ResNet database to identify and evaluate novel T1D risk genes.

2 METHODS AND MATERIALS

2.1 Genetic data selection

A systematic search of electronic databases was conducted using Illumine BaseSpace Correlation Engine (<http://www.illumina.com>). Fig. 1 presents the diagram for the data selection. The use of the keyword 'Type 1 Diabetes' searched and identified 421 T1D studies. Further filter criteria included: (1) the data type is RNA expression, (2) the data organism is Homo sapiens (3) The study is recent (within the last 10 years) (4) The study is T1D case vs. healthy control study (or include case/control bio-sets). In total, seven bio-sets (T1D case/control comparisons) from four studies satisfied the study selection criteria and were included in this systematic review and meta-analysis.



2.2 Genetic database T1D_GD

The T1D_GD is T1D targeted knowledge database online available at 'Bioinformatics Database' (<http://database.gousinfo.com>). The database is updated monthly or upon request. The current version of T1D_GD is composed of 1165 T1D target genes (T1D_GD → Related Genes), 646 small molecular/drugs (T1D_GD → Related Drugs), 120 pathways (T1D_GD → Related Pathways), and 96 related diseases (T1D_GD → Related Diseases). The database also provides supporting references for each T1D-Gene and T1D-Drug relation, including the titles and the sentences where the relation has been identified (see T1D_GD → Ref for Related Genes and T1D_GD → Ref for Related Drugs, respectively). This information could be used to locate a detailed description of how a candidate gene/drug is related to T1D.

Using T1D_GD, further analyses of the T1D target genes from the meta-analysis were conducted, including identifying their related T1D pathways (T1D_GD → Related Pathways), genes (T1D_GD → Related Genes), and drugs (T1D_GD → Related Drugs). Here we defined two genes as functionally related if they play roles within same genetic pathway. Pathway enrichment analysis (PEA) was conducted using Pathway Studio to identify genetic pathways potentially linked to T1D^[11]. The gene-drug relationships were identified using the network building module of Pathway Studio.

3 RESULTS

3.1 Selected Datasets

After screening against the selection criteria, seven T1D case/control comparison bio-sets from four independent studies were retrieved and assessed for eligibility. One of the identified datasets contained four separate case/control studies^[10], including: (1) children genetically predisposed to T1D vs. normal adults, (2) long standing T1D patients vs. normal adults (3) newly diagnosed T1D patients vs. normal adults (4) T1D vs. normal adults. The data is available at www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1724/. Each of other three datasets contained one case/control study^[8,9,12]. These three datasets were available at NCBI GEO (ID: GSE10586, GSE33440 and GSE55098, respectively). Statistics of the included bio-sets are presented in Table 1.

Table 1. Characteristics of the selected studies ordered by publication date

First Author	Design	Cases			Controls			Country
		n	Age	Sources	n	Age	Sources	
Jailwala <i>et al.</i> , 2009	c-c	12,	8-18	Children's Hospital of Wisconsin	15,	21-49	Children's Hospital of Wisconsin; Medical College of Wisconsin	USA
Irvine <i>et al.</i> , 2012	c-c	16	8.6-14.4	Mater Children's Hospital	6	23.8-40.2	Mater Children's Hospital	Australia
Ferreira <i>et al.</i> , 2014	c-c	49, 15, 25, 109	6-34, 22-35, 19-61, 0.2-9.1	Royal Free Hospital & Medical School research ethics committee; Cambridge BioResource	93	22-52	Cambridge BioResource	England
Yang <i>et al.</i> , 2014	c-c	60	13.82-24.4	Shanghai Ruijin Hospital	48	17.54-27.3	Shanghai Ruijin Hospital	China

Note: c-c refers to case vs. control.

3.2 Meta-analysis results

The Meta-analysis results were deposited into the 'Bioinformatics Database' (<http://database.gousinfo.com>), named as T1D_Meta. The top 6 genes (Score ≥ 85) from the meta-analysis appear in Table 2, with more detailed statistics presented in T1D_Meta \rightarrow Top 6 Genes. The full Meta-analysis

results are presented in T1D_Meta \rightarrow Full Gene List. A gene's score is defined by the meta-analysis Illumina BaseSpace Correlation Engine (<http://www.illumina.com>), which is based on the statistical significance and consistency of the gene across the queried bio-sets. The higher the score, the greater the importance of the gene for the case/control comparison.

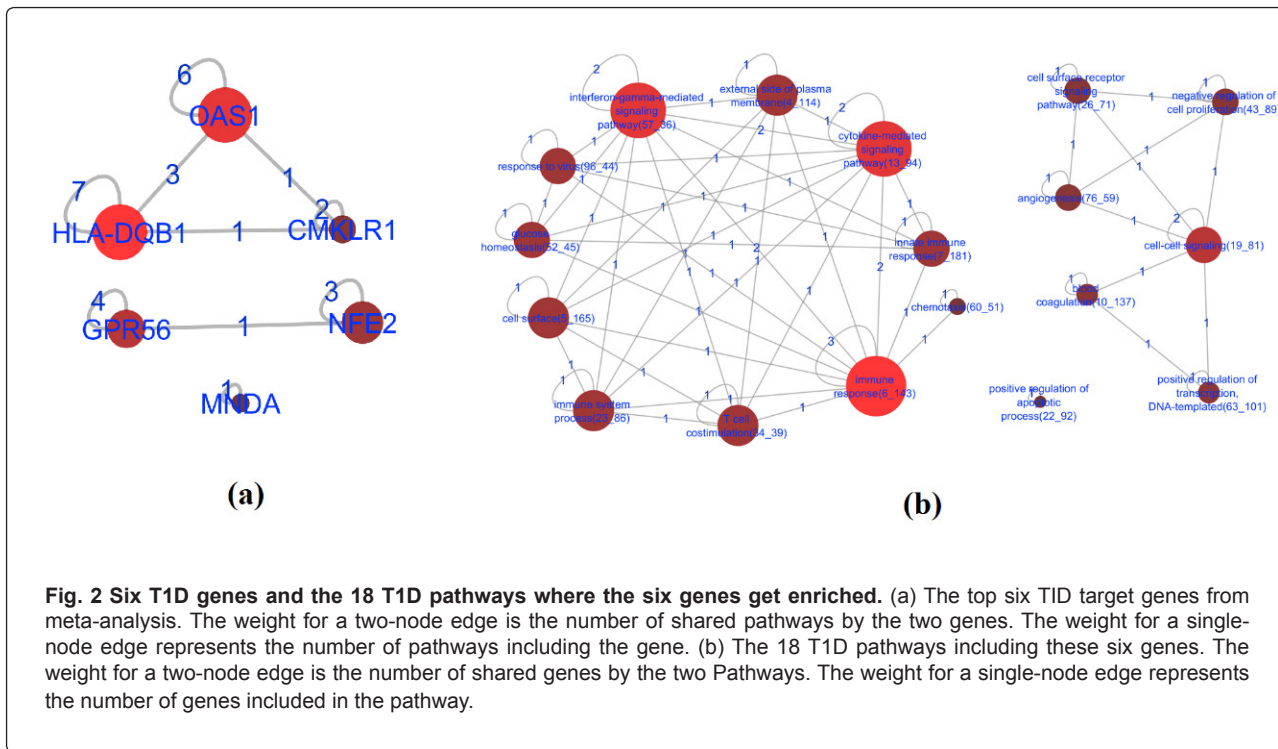
Table 2. The top 6 genes of the included studies excluded from T1D-gene relation data

Gene Correlation	EntrezGene ID	Score	Specificity	Associated Pathways	Gene Connectivity
HLA-DQB1	3119	94	5/7	external side of plasma membrane (0009897); cell surface (0009929); immune response (0006955); cytokine-mediated signaling pathway (0019221); immune system process (0002376); T cell costimulation (0031295); interferon-gamma-mediated signaling pathway (0060333)	391
CMKLR1	1240	90	5/7	immune response (0006955); chemotaxis (0006935)	165
NFE2	4778	88	6/7	blood coagulation (0007596); cell-cell signaling (0007267); positive regulation of transcription, DNA-templated (0045941)	283
OAS1	4938	88	5/7	immune response(0006955); innate immune response (0002226); cytokine-mediated signaling pathway (0019221); glucose homeostasis (0042593); interferon-gamma-mediated signaling pathway (0060333); response to virus (0009615)	386
GPR56	9289	86	5/7	cell-cell signaling (0007267); cell surface receptor signaling pathway (0007166); negative regulation of cell proliferation (0008285); angiogenesis (0001525)	251
MNDA	4332	85	6/7	positive regulation of apoptotic process (0043065)	91

Score: A gene's score is based on the statistical significance and consistency of the gene across the queried bio-sets. Specificity: A gene's specificity is the number of bio-sets in which the direction of a gene's regulation matches the selected filter. Associated Pathway: The known T1D related Pathways (T1D_GD \rightarrow Related Pathways) that contain the gene. GO ID is provided if any. Gene Connectivity: The number of known T1D related genes (T1D_GD \rightarrow Related Genes) that connect with the target gene.

Out of the six genes listed in Table 2, two were included in T1D_GD (HLA-DQB1 and OAS1), while other four were not (CMKLR1, NFE2, GPR56 and MNDA). Further study using the T1D_GD showed that, these 6 genes were enriched within multiple T1D target pathways and were connected to many other genes that were linked to T1D (see Table 2). Figure 2 (a) presents the functionally connections

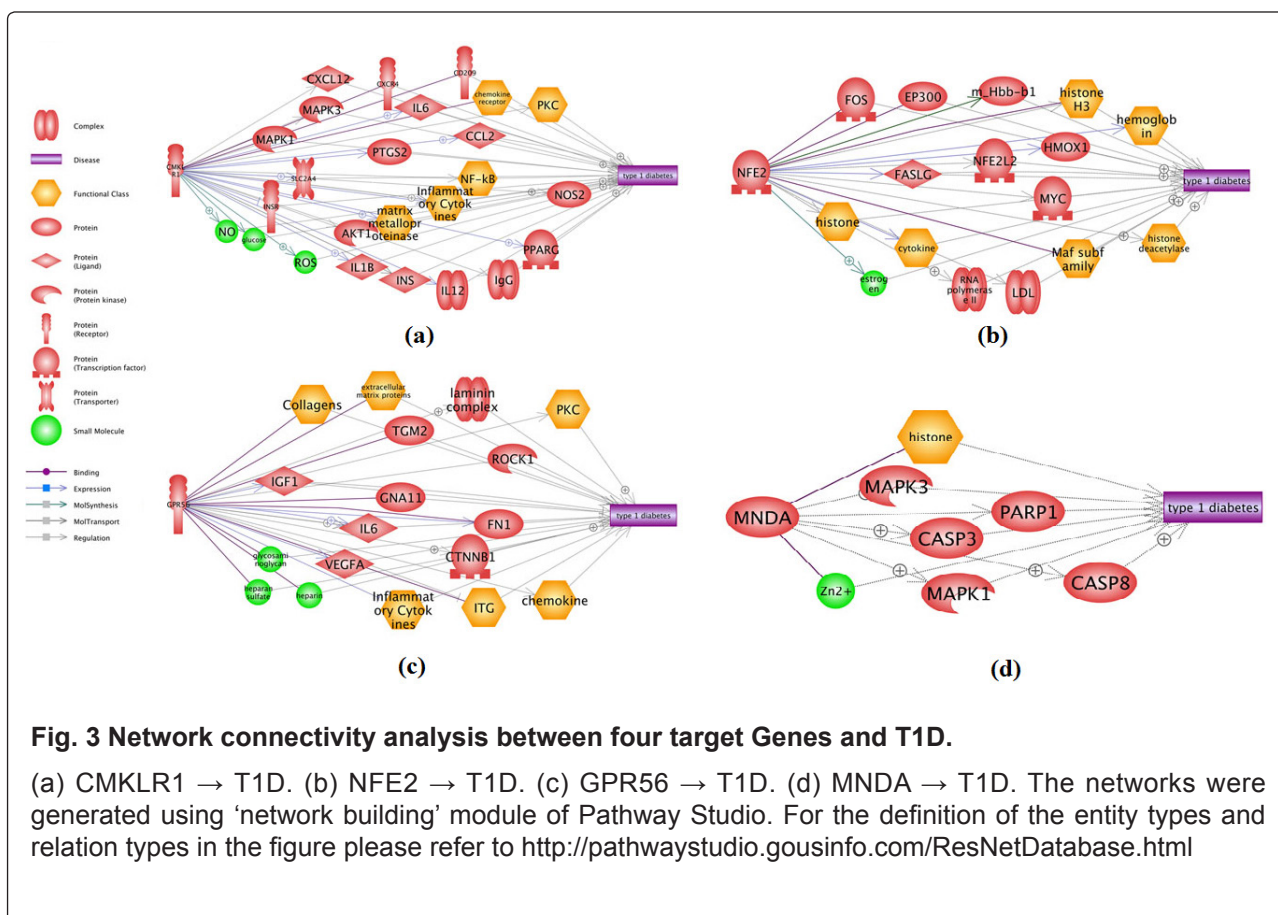
among these six genes. Fig. 2 (b) shows that 18 T1D pathways including these six genes. To note, five of these 18 pathways were among the top 10 T1D pathways (T1D_GD → Related Pathways), including external side of plasma membrane, cell surface, immune response, innate immune response and blood coagulation, as shown in Fig. 2.



3.3 Network analysis

Additional functional network connectivity analysis (NCA) using PS showed that, the four novel genes from this meta-analysis (CMKLR1, NFE2, GPR56 and MNDA) present strong functional association with T1D. These genes influence the pathogenic development of T1D through multiple pathways, as

shown in Fig. 3. Under each relation (arrow) in Fig. 3, there are support from one or more references that provide detailed description of each relation (see T1D_Meta → CMKLR1, T1D_Meta → NFE2, T1D_Meta → GPR56 and T1D_Meta → MNDA),.



4 DISCUSSION

Meta-analysis is a statistical approach that combines the results of multiple scientific studies. Although many genetic studies have been conducted to discovery genetic risk factors for T1D, combining the results from these separated studies by using meta-analysis could lead to a higher statistical power and more robust point estimate. In this study, meta-analysis was performed on seven T1D case/control bio-sets extracted from four recent studies. T1D target genes from meta-analysis were sorted by gene score, which is based on the statistical significance and consistency of the gene across the queried bio-sets. Meta-analysis results suggested six top risk genes for T1D (Score ≥ 85), and four of them are novel according to a recently updated database T1D_GD. Further analyses were conducted to study the possible correlation between T1D and these six genes, with more efforts focused on the four novel ones.

Analysis using T1D_GD showed that the two

known T1D target genes, HLA-DQB1 and OAS1, are among the top T1D_GD genes with supports from multiple independent studies (see T1D_GD → Related Genes). PEA results showed that, both these two genes and other four genes (CMKLR1, NFE2, GPR56 and MNDA) are enriched within multiple T1D pathways (T1D_GD → Related Pathways) and linked to dozens or even hundreds of other T1D genes (see Table 2 and Fig. 2). These results support the possible linkage between these genes and T1D.

Additional network connectivity analysis (NCA) revealed multiple possible functional associations between T1D and the four novel genes (see Fig. 3). It has been shown that activation of CMKLR1 could depress expression of INSR^[13], while INSR a target gene for the treatment of T1D^[14]. This suggests that CMKLR1 may play a role in the development of T1D through a CMKLR1 → INSR → T1D pathway. Previous studies have also shown that MNDA is an

inhibitor of PARP^[15]. PARP has been suggested to play a role in the development of T1D^[16]. Therefore, MNDA may serve as a protector for T1D patients. NFE2 regulates the expression of HMOX1^[17], while overexpression of HMOX1 has been shown to delay T1D^[18, 19]. Thus, NFE2 may influence the T1D pathogenesis by regulating the expression of HMOX1. Furthermore, several studies show that GPR56 can induce the secretion of IL6^[20], which has been shown to play a significant role in T1D etiopathogenesis^[21]. These observations suggest a possible GPR56 → IL6 → T1D regulatory pathway. More potential regulation pathways could be identified from the T1D_Meta database (see T1D_Meta → CMKLR1, T1D_Meta → NFE2, T1D_Meta → GPR56 and T1D_Meta → MNDA), which has been deposited into the open source 'Bioinformatics Database' (<http://database.gousinfo.com>).

In summary, this meta-analysis supported the correlation between two genes (HLA-DQB1 and OAS1) and T1D, and revealed four novel potential risk genes for the disease (CMKLR1, NFE2, GPR56 and MNDA). Further network analysis supported the meta-analysis results and identified possible functional pathways and mechanisms, thorough which these genes exert influence on T1D. Findings in this study may provide new insights into the current field of T1D genetic study.

CONFLICT OF INTERESTS

Authors claim no conflict of interests.

REFERENCES

- Centers for Disease Control and Prevention. National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2011. Available from: http://www.cdc.gov/diabetes/pubs/pdf/ndfs_2011.pdf.
- Patterson CC, Dahlquist GG, Gyürüs E, Green A, Soltész G; EURODIAB Study Group. Incidence trends for childhood type 1 diabetes in Europe during 1989–2003 and predicted new cases 2005–20: a multicentre prospective registration study. *Lancet*. 2009; 373: 2027-2033.
- Åkerblom HK, Vaarala O, Hyöty H, Ilonen J, Knip M. Environmental factors in the etiology of type 1 diabetes. *Am J Genet*. 2002; 115: 18-29.
- Knip M, Veijola R, Virtanen SM, Hyöty H, Vaarala O, Åkerblom HK. Environmental triggers and determinants of β -cell autoimmunity and type 1 diabetes. *Diabetes* 2005; 54: S125-S136.
- Karvonen M, Toumilehto J, Libman I, LaPorte R. A review of the recent epidemiological data on the worldwide incidence of type 1 (insulin-dependent) diabetes mellitus. *Diabetologia*. 1993; 36: 883-892.
- Risch N. Assessing the role of HLA-linked and unlinked determinants of disease. *Am J Hum Genet*. 1987; 40: 1-14.
- Glisic S, Klinker M, Waukau J, Jailwala P, Jana S, Basken J, Wang T, Alemzadeh R, Hagopian W, Ghosh S. Genetic association of HLA DQB1 with CD4+CD25+(high) T-cell apoptosis in type 1 diabetes. *Genes Immun*. 2009; 10(4): 334-340.
- Irvine KM, Gallego P, An X, Best SE, Thomas G, Wells C, Harris M, Cotterill A, Thomas R. Peripheral blood monocyte gene expression profile clinically stratifies patients with recent-onset type 1 diabetes. *Diabetes*. 2012; 61(5): 1281-1290.
- Yang M, Ye L, Wang B, Gao J, Liu R, Hong J, Wang W, Gu W, Ning G. Decreased miR-146 expression in peripheral blood mononuclear cells is correlated with ongoing islet autoimmunity in type 1 diabetes patients. *J Diabetes*. 2015; 7(2): 158-165.
- Ferreira RC, Guo H, Coulson RM, Smyth DJ, Pekalski ML, Burren OS, Cutler AJ, Doecke JD, Flint S, McKinney EF, Lyons PA, Smith KG, Achenbach P, Beyerlein A, Dunger DB, Clayton DG, Wicker LS, Todd JA, Bonifacio E, Wallace C, Ziegler AG. A type I interferon transcriptional signature precedes autoimmunity in children genetically at risk for type 1 diabetes. *Diabetes*. 2014; 63(7): 2538-2550.
- Nikitin A, Egorov S, Daraselia N, Mazo I. Pathway studio--the analysis and navigation of molecular networks. *Bioinform*. 2003; 19(16): 2155-2157.
- Jailwala P, Waukau J, Glisic S, Jana S, Ehlenbach S, Hessner M, Alemzadeh R, Matsuyama S, Laud P, Wang X, Ghosh S. Apoptosis of CD4+ CD25(high) T cells in type 1 diabetes may be partially mediated by IL-2 deprivation. *PLoS One*. 2009; 4(8): e6527.
- Goralski KB, Sinal CJ. Elucidation of chemerin and chemokine-like receptor-1 function in adipocytes by adenoviral-mediated shRNA knockdown of gene expression. *Method*

- Enzymol. 2009; 460: 289-312.
14. Cao Y, Li Y, Kim J, Ren Y, Himmeldirk K, Liu Y, Qian Y, Liu F, Chen X. Orally efficacious novel small molecule 6-chloro-6-deoxy-1, 2, 3, 4-tetra-O-galloyl- α -D-glucopyranose selectively and potently stimulates insulin receptor and alleviates diabetes. *J Mol Endocrinol*. 2013; 51(1): 15-26.
 15. Fotouhi-Ardakani N, Kebir DE, Pierre-Charles N, Wang L, Ahern SP, Filep JG, Milot E. Role for myeloid nuclear differentiation antigen in the regulation of neutrophil apoptosis during sepsis. *Am J Respir Crit Care Med*. 2010; 182(3): 341-350.
 16. Szabó C, Biser A, Benko R, Böttinger E, Suszták K. Poly(ADP-ribose) polymerase inhibitors ameliorate nephropathy of type 2 diabetic *Leprdb/db* mice. *Diabetes*. 2006; 55(11): 3004-3012.
 17. Wierenga AT, Vellenga E, Schuringa JJ. Convergence of hypoxia and TGF β pathways on cell cycle regulation in human hematopoietic stem/progenitor cells. *PLoS One*. 2014; 9(3): e93494.
 18. Li M, Peterson S, Husney D, Inaba M, Guo K, Kappas A, Ikehara S, Abraham NG. Long-lasting expression of HO-1 delays progression of type I diabetes in NOD mice. *Cell Cycle*. 2007; 6(5): 567-571.
 19. Biburger M, Theiner G, Schädle M, Schuler G, Tiegs G. Pivotal Advance: Heme oxygenase 1 expression by human CD4+ T cells is not sufficient for their development of immunoregulatory capacity. *J Leukoc Biol*. 2010; 87(2): 193-202.
 20. Chiang NY, Peng YM, Juang HH, Chen TC, Pan HL, Chang GW, Lin HH. GPR56/ADGRG1 Activation Promotes Melanoma Cell Migration via NTF Dissociation and CTF-Mediated G α 12/13/RhoA Signaling. *J Invest Dermatol*. 2017; 137(3): 727-736.
 21. Wedrychowicz A, Dziatkowiak H, Sztefko K, Wedrychowicz A. Interleukin-6 (IL-6) and IGF-IGFBP system in children and adolescents with type 1 diabetes mellitus. *Exp Clin Endocrinol Diabetes*. 2004; 112(8): 435-439.