



Bioinformatic and rare-variant collapsing analyses for type 1 and type 2 diabetes in the UK Biobank reveal novel pleiotropic susceptibility loci

Type 1 diabetes (T1D) is a chronic condition caused by the autoimmune destruction of pancreatic β -cells.¹ In contrast, type 2 diabetes (T2D) is characterized by impaired glucose metabolism arising from defects in insulin resistance and secretion.² More than 75 genetic loci influencing T1D risk have been identified.¹ Genome-wide association studies (GWAS) of T2D have identified over 700 risk loci.² Whole exome sequencing (WES) studies may reveal the association of rare variants to common diseases such as T1D and T2D. However, only a few large-scale WES studies have been published until Wang et al reported the relationships between rare protein-coding variants and 17 361 binary phenotypes using WES data from 269 171 UK Biobank participants (<https://azpnewas.com/>).³ Recently, Karczewski et al determined gene-based association investigating 4529 phenotypes in 394 841 UK Biobank exomes (<https://app.genebass.org/>).⁴

We used the two published UK Biobank portals (<https://azpnewas.com/> and <https://app.genebass.org/>)^{3,4} to access gene collapsing analyses of rare variation for T1D and T2D (Table 1). Ethical statements are not required for the study as no human or animal studies are involved. In order not to discard potential candidate genes we present genes with p values $<.05/20000$ genes = 2.5×10^{-6} commonly used for WES studies. Identified T1D and T2D genes were bioinformatically analyzed using the GWAS catalog (<https://www.ebi.ac.uk/gwas/>), OMIM (<https://www.omim.org/>), and GeneCards (<https://www.genecards.org/>).⁵⁻⁸ The literature was searched for identified genes using <https://pubmed.ncbi.nlm.nih.gov/>. We compared the union of the same three-digit ICD-10 codes (*International Classification of Diseases, Tenth Revision*).^{3,4} In Table 1 only the genes with genome wide significant results are shown with p values for the most significant model. One previously T1D linked gene (*HLA-DRB5*) and four novel T1D genes (*PSMB9*, *NELFE*, *SLC44A4*, and *VWA7*) were identified. For T2D four previously linked genes (*GCK*, *HNF1A*,

HNF4A, and *ANKH*) were confirmed. In addition, *GIGYF1* has recently already been linked to T2D in UK Biobank.⁹ Two novel associations were identified, the *DENND6A* and *RPS3A* genes. The identified genes were specific for each of T1D and T2D (Table 1).

Phenome-wide association studies (PheWAS) data (Table 1) could link all five identified T1D genes to other immune-mediated diseases: ankylosing spondylitis, iridocyclitis, hypothyroidism, asthma, celiac disease, sarcoidosis, psoriasis, and rheumatoid arthritis (Table 1). Thus, the five identified pleiotropic T1D genes may all contribute to the previously observed epidemiological associations between T1D and other immune-mediated diseases.¹⁰ Only the *GIGYF1* gene among the T2D linked genes was associated with a potential immune-mediated disorder (hypothyroidism) (Table 1). However, even more interesting is the association between the T2D linked *GIGYF1* gene and chronic obstructive pulmonary disease (COPD) in the PheWAS analysis (Table 1). COPD and T2D are recognized to be associated conditions with shared environmental exposures.¹¹ Treatment with the novel antihyperglycemic drugs glucagon-like peptide 1 (receptor agonists and sodium glucose transporter 2 inhibitors) have recently been associated with a reduced risk of severe exacerbations in COPD among patients with T2D.¹² Thus, the *GIGYF1* gene might contribute to the observed epidemiological association between COPD and T2D and may open novel treatments for T2D and COPD.

A limitation is that the validity of T1D is not perfect in UK Biobank. However, a diagnosis of T1D in UK Biobank may still be useful for research in large studies. Two papers about T1D in UK Biobank have been published suggesting that the UK Biobank might be useful for T1D research: one in *Lancet Diabetes & Endocrinology* and one in *Diabetes Medicine*.^{13,14} Moreover, an article by Thomas et al has shown that the accuracy of T1D and T2D tested with two different methods range from 71% to 88%.¹⁵ These articles are in line with the findings in the present

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Journal of Diabetes* published by Ruijin Hospital, Shanghai Jiaotong University School of Medicine and John Wiley & Sons Australia, Ltd.

TABLE 1 Results of gene collapsing analysis of rare variants for type 1 diabetes (T1D) and type 2 diabetes (T2D) according to ICD-10 codes (<https://azphewas.com/> and <https://app.genebase.org/>).^{3,4}

ICD-10 codes	Genebase (LOF)	Genebase (missense/LC)	Astra Zeneca portal
E10 T1D	NS	<i>PSMB9</i> 1.59×10^{-9} <i>NELFE</i> 1.74×10^{-9} <i>SLC44A4</i> 1.85×10^{-9}	* <i>HLA-DRB5</i> 5.70×10^{-7} <i>VWA7</i> 1.99×10^{-6}
E11 T2D	** <i>GCK</i> 5.86×10^{-20} <i>GIGYF1</i> 1.55×10^{-11} ** <i>HNF1A</i> 4.59×10^{-7}	** <i>ANKH</i> 1.32×10^{-12} ** <i>GCK</i> 4.43×10^{-8} <i>DENND6A</i> 5.12×10^{-7} ** <i>HNF4A</i> 1.74×10^{-6}	** <i>GCK</i> 8.18×10^{-21} <i>GIGYF1</i> 9.98×10^{-11} ** <i>HNF1A</i> 7.13×10^{-8} <i>RPS3A</i> 2.05×10^{-6}
PheWAS for identified T1D genes			
<i>PSMB9</i>	NS	Ankylosing spondylitis 3.77×10^{-8} H20 Iridocyclitis 5×10^{-8}	NS
<i>NELFE</i>	NS	M45 Ankylosing spondylitis 3.44×10^{-13} H20 Iridocyclitis 9.805×10^{-14} Iritis 2.66×10^{-7}	NS
<i>SLC44A4</i>	NS	M45 Ankylosing spondylitis 1.72×10^{-20} H20 Iridocyclitis 1.48×10^{-13}	NS
* <i>HLA-DRB5</i>	NS	E03 Other hypothyroidism 1.15×10^{-6}	J45 Asthma 7.69×10^{-10} K90.0 Coeliac disease 1.03×10^{-9} Sarcoidosis 1.32×10^{-7}
<i>VWA7</i>	NS	NS	K90.0 Coeliac disease 5.34×10^{-14} L40 Psoriasis 1.58×10^{-10} Rheumatoid arthritis 5.96×10^{-7} D41 Neoplasm of uncertain or unknown behavior of urinary organs 1.84×10^{-6} I10-I15 Hypertensive diseases 2.38×10^{-6}
PheWAS for identified T2D genes			
** <i>GCK</i>	NS	NS	H36 Retinal disorders in diseases classified elsewhere 2.86×10^{-8} Chapter IV Endocrine nutritional and metabolic diseases 9.41×10^{-7}
<i>GIGYF1</i>	E03 Other hypothyroidism 1.2×10^{-7} J44 Other chronic obstructive pulmonary disease 1.57×10^{-6}	NS	E03 Other hypothyroidism 7.12×10^{-9} E00-E07 Disorders of thyroid gland 3.47×10^{-8} E03.9 Hypothyroidism unspecified 3.84×10^{-8} Hypothyroidism myxoedema 1.45×10^{-6} R55 Syncope and collapse 2.26×10^{-6}



TABLE 1 (Continued)

ICD-10 codes	Genebase (LOF)	Genebase (missense/LC)	Astra Zeneca portal
** <i>HNF1A</i>	NS	NS	NS
** <i>ANKH</i>	NS	NS	NS
<i>DENND6A</i>	NS	NS	NS
** <i>HNF4A</i>	NS	NS	NS
<i>RPS3A</i>	ND	NS	NS

Note: Union was used to define phenotypes for <https://azpnewas.com>. Only genome-wide significant associated genes are shown (for the most significant model) that is, $p < 2.5 \times 10^{-6}$. Other disease phenotypes than diabetes were searched in the databases for the identified T1D and T2D genes, that is phenome-wide association study (PheWAS). Significance threshold was p value $< .05/20000$ genes = 2.5×10^{-6} commonly used for WES studies. For Genebase portal (<https://app.genebase.org>) the most significant p value for SKAT, SKAT-O, or burden test is shown.⁵ For Astra Zeneca portal (<https://azpnewas.com>) the most significant p value of the 12 tested models is shown.³ ICD-10, *International Classification of Diseases, Tenth Revision*; LoF, high-confidence loss of function variants indicated by LOFTEE⁴; Missense/LC, Missense variants are grouped with in-frame insertions and deletions, as well as low-confidence LoF variants filtered out by LOFTEE⁴ (The latter have a frequency spectrum consistent with missense variation and affect a set of amino acids in a similar way⁴); NS, no genome wide significant gene; ND, not determined; SKAT, sequence kernel association test; SKAT-O, optimized sequence kernel association test; WES, whole exome sequencing.

*The *HLA-DRB5* locus has been linked to T1D in genome-wide association study (GWAS) (<https://www.ebi.ac.uk/gwas/>).

**The *ANKH*, *GCK*, *HNF1A*, and *HNF4A* loci have been linked to T2D in GWAS (<https://www.ebi.ac.uk/gwas/>).

TABLE 2 Bioinformatic analysis of associated with type 1 diabetes (T1D) and type 2 diabetes (T2D) according to ICD-10 codes in the genes collapsing analysis of rare variants in the two published UK Biobank portals (<https://azpnewas.com/> and <https://app.genebase.org/>).^{3,4}

Genes	GWAS catalog*	OMIM/genecards*	Pathway analysis*
Type 1 diabetes			
<i>PSMB9c</i>	Rheumatoid arthritis (RA), autoimmune disease	Proteasome-associated auto inflammatory syndrome	Regulation of activated PAK-2p34
<i>NELFE</i>	Age-related macular degeneration	Immunodeficiency	Formation of HIV elongation complex
<i>SLC44A4</i>	Systemic lupus erythematosus (SLE), vitiligo	Deafness	Transport of cations/anions/amino acids/peptides
<i>HLA-DRB5</i>	T1D, asthma, nephrotic syndrome, RA, SLE	Pityriasis rosea, Parkinson disease	TCR signaling (Qiagen)
<i>VWA7</i>	Takayasu arteritis, asthma, Crohn prognosis	Armfield syndrome	-
Type 2 diabetes			
<i>GCK</i>	T2D, metabolic syndrome	Maturity-onset diabetes of the young	Glycolysis and trehalose degradation
<i>GIGYF1</i>	Autoimmune thyroid disease, hypothyroidism	Parkinson disease	-
<i>HNF1Ac</i>	T2D	Maturity-onset diabetes of the young	Regulation of beta-cell development
<i>ANKH</i>	T2D	Craniometaphyseal dysplasia, chondrocalcinosis	Transport of cations/anions/amino acids/peptides
<i>DENND6A</i>	Alzheimer disease	-	Vesicle-mediated transport, Rab trafficking
<i>HNF4A</i>	T2D	Maturity-onset diabetes of the young	Gene transcription, nuclear receptors
<i>RPS3A</i>	-	Diamond-Blackfan anemia, encephalopathy	Peptide chain elongation

Note: The GWAS catalog (<https://www.ebi.ac.uk/gwas/>),⁵ OMIM (<https://www.omim.org/>),⁶ and Genecards (<https://www.genecards.org/>)^{7,8} were searched for the association with relevant diseases and pathways.*

Abbreviations: GWAS, genome-wide association study; ICD-10, *International Classification of Diseases, Tenth Revision*; OMIM, Online Mendelian Inheritance in Man; TCR, T cell receptor.

*The most potentially relevant diseases and pathways are shown. For a complete list the reader is referred to the original bioinformatic resources: that is, GWAS catalog (<https://www.ebi.ac.uk/gwas/>), OMIM (<https://www.omim.org/>), and Genecards (<https://www.genecards.org/>).

study. For instance, we confirmed one previously recognized T1D gene (*HLA-DRB5*) and four previously linked T2D genes (*GCK*, *HNF1A*, *HNF4A*, and *ANKK1*). Thus, the definition used in UK Biobank for T1D and T2D could differentiate between known T1D and T2D genes, which is reassuring. Moreover, all the T1D linked genes (one old and four novel T1D genes) could in the PheWAS and the bioinformatic analysis be linked to immune-mediated disorders (Tables 1 and 2). It is well known that genetic links between many different immune-mediated disorders exist.¹⁰ Moreover, no T1D gene was linked to T2D. Thus, there was no overlap between T1D and T2D genes. Moreover, only one T2D gene could be linked in the PheWAS and bioinformatic analysis to a potential immune-mediated condition (ie, hypothyroidism). Thus, we believe the accuracy is acceptable for large studies of T1D and T2D genetics in UK Biobank.

In conclusion, rare variations in 12 genes (six novel) were associated with diabetes in the UK Biobank, T1D (five genes) and T2D (seven genes). Thus, rare variation contributes to T1D and T2D in the general population. Rare variation in all five T1D linked genes is also linked to other immune-mediated diseases in UK Biobank, whereas the T2D gene *GIGYF1* is associated with COPD.

ACKNOWLEDGEMENTS

We thank the free access to the Genebase and the AstraZeneca PheWAS portals that made this work possible (<https://azphewas.com/> and <https://app.genebase.org/>).^{3,4} This work was supported by a grant awarded to Dr Bengt Zöller by ALF-funding from Region Skåne, Sparbanken Skåne, and by the Swedish Research Council. The funders had no role in the study.

FUNDING INFORMATION

ALF-funding from Region Skåne; Sparbanken Skåne; Vetenskapsrådet

Bengt Zöller¹ 
Eric Manderstedt¹
Christina Lind-Halldén²
Christer Halldén¹

¹Center for Primary Health Care Research, Department of Clinical Sciences, Lund University and Region Skåne, Malmö, Sweden

²Department of Environmental Science and Bioscience, Kristianstad University, Kristianstad, Sweden

Correspondence

Bengt Zöller, Center for Primary Health Care Research, CRC, Skåne University Hospital, Building 28, Floor 11, Jan Waldenströms gata 35, S-205 02 Malmö, Sweden.
Email: bengt.zoller@med.lu.se

ORCID

Bengt Zöller  <https://orcid.org/0000-0002-8250-5613>

REFERENCES

- Redondo MJ, Gignoux CR, Dabelea D, et al. Type 1 diabetes in diverse ancestries and the use of genetic risk scores. *Lancet Diabetes Endocrinol.* 2022;10(8):597-608.
- DeForest N, Majithia AR. Genetics of type 2 diabetes: implications from large-scale studies. *Curr Diab Rep.* 2022;22(5):227-235.
- Wang Q, Dhindsa RS, Carss K, et al. Rare variant contribution to human disease in 281,104 UK biobank exomes. *Nature.* 2021;597(7877):527-532.
- Karczewski KJ, Solomonson M, Chao KR, et al. Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK biobank exomes. *Cell Genomics.* 2022;2(9):100168.
- Sollis E, Mosaku A, Abid A, et al. The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* 2023;51:D977-D985.
- Hamosh A, Amberger JS, Bocchini C, Scott AF, Rasmussen SA. Online Mendelian inheritance in man (OMIM®): victor McKusick's magnum opus. *Am J Med Genet A.* 2021;185:3259-3265.
- Belinky F, Nativ N, Stelzer G, et al. PathCards: multi-source consolidation of human biological pathways. *Database.* 2015;2015:bav006.
- Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics.* 2016;54:1.30.1-1.30.33.
- Deaton AM, Parker MM, Ward LD, et al. Gene-level analysis of rare variants in 379,066 whole exome sequences identifies an association of *GIGYF1* loss of function with type 2 diabetes. *Sci Rep.* 2021;11(1):21565.
- Hemminki K, Li X, Sundquist J, Sundquist K. Familial association between type 1 diabetes and other autoimmune and related diseases. *Diabetologia.* 2009;52(9):1820-1828.
- Katsiki N, Steiropoulos P, Papanas N, Mikhailidis DP. Diabetes mellitus and chronic obstructive pulmonary disease: an overview. *Exp Clin Endocrinol Diabetes.* 2021;129(10):699-704.
- Pradhan R, Lu S, Yin H, et al. Novel antihyperglycaemic drugs and prevention of chronic obstructive pulmonary disease exacerbations among patients with type 2 diabetes: population based cohort study. *BMJ.* 2022;379:e071380.
- Thomas NJ, Jones SE, Weedon MN, Shields BM, Oram RA, Hattersley AT. Frequency and phenotype of type 1 diabetes in the first six decades of life: a cross-sectional, genetically stratified survival analysis from UK biobank. *Lancet Diabetes Endocrinol.* 2018;6:122-129.
- Shang X, Hill E, Liu J, et al. Association of type 1 diabetes and age at diagnosis of type 2 diabetes with brain volume and risk of dementia in the UK biobank: a prospective cohort study of community-dwelling participants. *Diabet Med.* 2023;40:e14966.
- Thomas NJ, McGovern A, Young KG, et al. Identifying type 1 and 2 diabetes in research datasets where classification biomarkers are unavailable: assessing the accuracy of published approaches. *J Clin Epidemiol.* 2023;153:34-44. doi:10.1016/j.jclinepi.2022.10.022