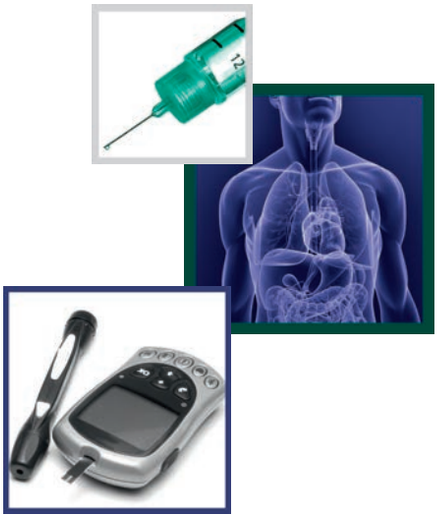
EDITORIAL

What have genome-wide association studies contributed to the understanding of the pathogenesis and future management of Type 1 diabetes?



Marina Bakay[1] Hakon Hakonarson*[1,2]

"Detailed genotype–phenotype studies are necessary before we can translate genome-wide association studies findings into meaningful benefits for patient care."

Diabetes is a major health problem and its prevalence is increasing worldwide. It is estimated that by 2030, there will be 439 million adults affected by diabetes [101]. Today diabetes impacts the lives of approximately 200 million people. Approximately 10% of all cases are Type 1 diabetes (T1D) with around a 3% increase in the incidence of T1D globally per year [1].

T1D is a complex trait that results from the interplay between environmental and genetic factors. Much evidence supports a strong genetic component associated to T1D. The epidemiological data for geographic prevalence differences is one clear indicator, with populations of European ancestry having the highest presentation rate. T1D runs strongly in families, with the sibling risk being approximately ten-times greater than in the general population [2].

It has been a challenge to isolate the susceptibility genes for diabetes. Historically, before the introduction of genome-wide association studies (GWAS), the genetic determinants of T1D consisted of five loci. The first GWAS locus was reported in 2006 [3] and the current tally of T1D GWAS loci stands at 53 [4], including rediscovery of the first five candidate gene loci.

Approximately half of the genetic risk for T1D is conferred by the genomic region harboring the HLA class II genes (primarily *HLA-DRB1*, *-DQA1* and *-DQB1* genes), which encode the highly polymorphic antigen-presenting proteins. Recent fine mapping efforts of the MHC addressed why the class II genes *HLA-DQB1* and *HLA-DRB1* cannot completely explain the association between T1D and the MHC region [5]. It turned out that most of the remaining association that could be detected was due to signals in HLA-B and HLA-A, and that the existence of other major T1D genes in the extended MHC was unlikely. Other loci established prior to GWAS are the genes encoding insulin [6], cytotoxic T-lymphocyte-associated protein 4 [7], protein tyrosine phosphatase, nonreceptor type 22 gene [8] and IL-2 receptor α [9]. However, the majority of other reported associations in the pre-GWAS era

"The advent of genome-wide association studies has changed the situation dramatically, provided pace and great benefit to the discovery of loci associated with Type 1 diabtetes, increasing the number of associated regions by a factor of ten."

[1]The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[2]The Department of Pediatrics, University of Pennsylvania School of Medicine, PA, USA
*Author for correspondence: Tel.: +1 267 426 0088; Fax: +1 267 426 0363; hakonarson@email.chop.edu

Future Medicine part of fsg

have remained debatable, where an initial report of association was not confirmed in subsequent replication attempts by other investigators, known as the 'winner's curse' [10].

The advent of GWAS has changed the situation dramatically, provided pace and great benefit to the discovery of loci associated with T1D, increasing the number of associated regions by a factor of ten. An early genome-wide single nucleotide polymorphism (SNP) genotyping approach, using only 6500 nonsynonymous SNPs [3], represented a precursor to the full GWAS approached soon after; however, it did uncover a robust association to the interferon-induced with helicase C domain 1 (*IFIH1*) gene. *IFIH1* exerts its influence through the apoptosis of virally infected cells in antiviral immune responses, which may in turn support the notion that there is a connection between viral infections and the pathogenesis of T1D [11]. Interestingly, subsequent resequencing revealed additional rarer, higher-risk-conferring variants residing within the exons of this gene [12].

The first full-scale GWAS for T1D came simultaneously from our group at The Children's Hospital of Philadelphia (PA, USA) [13] and the Wellcome Trust Case–Control Consortium (WTCCC) [14]. In our study, we examined a large pediatric cohort of European descent using the HumanHap 550 BeadChip platform (Illumina®, CA, USA). The design involved 561 cases, 1143 controls and 467 triads in the discovery stage, followed by a replication effort in 939 nuclear families. In addition to finding the usual suspects, including an impressive 392 SNPs capturing the very strong association across the MHC, we identified significant association with variation at the *KIAA0350* gene, which we replicated in an additional cohort. The WTCCC study investigated seven common complex diseases, including T1D [14], by genotyping 2000 cases and 3000 controls with approximately 500,000 SNPs using the Affymetrix® GeneChip and reported a number of novel T1D loci, including the *KIAA0350* genomic region. They confirmed these findings in a replication study in 4000 cases and 5000 controls plus nearly 3000 T1D family trios [15]. In a separate replication effort, we fast-tracked 24 SNPs at 23 distinct loci that fell just below the bar for genome-wide significance in our 2007 GWAS and established association to the 12q13 region [16]. This was the same locus as reported by the WTCCC [14] and their companion study [15]. The 12q13 region harbors several genes, including *ERBB3*, *RAB5B*, *SUOX*, *RPS26* and *CDK2*. The clarity of signals found in 2007–2009 T1D GWAS highlights the strength and consistency of the GWAS approach in contrast to traditional candidate gene and family-based studies where the consensus amongst geneticists was relatively poor [10].

Genome-wide genotyping has been relatively expensive and represents a large financial investment when leveraging large, well-powered case–control cohorts. In order to get the most from such an endeavor, GWAS investigators have chosen to combine datasets from different investigative groups in order to carry out meta-analyses. We used this data-mining approach to determine additional novel loci associated with T1D, conferring increasingly modest risks of 1.1–1.2 [17]. Through subsequent rounds of testing in an independent cohort of T1D families from Montreal (Canada), the Children's Hospital of Philadelphia (PA, USA) and the Type 1 Diabetes Genetics Consortium (USA), followed by the WTCCC dataset and the DCCT /EDIC study cohort, we observed convincing association with the genes encoding ubiquitin-associated and SH3 domain-containing protein A (UBASH3A) and broad complex-tramtrack-bric-a-brac (BTB) and cap 'n' collar (CNC) homology 2 (BACH2). In further support of our finding, the *UBASH3A* locus was subsequently implicated in T1D from a large linkage study using dense SNP genotyping data generated on affected sib pairs [18].

The meta-analysis reported by Barrett *et al.* in 2009 uncovered in excess of 40 loci, including 18 novel regions, and also confirmed a number of loci uncovered through cross-disease comparisons [19]. In addition to confirmation for already known loci, they also reported association to 1q32.1 (which harbors the interleukin genes *IL10*, *IL19* and *IL20*), Glis family zinc finger protein 3 (GLIS3; first suggested by us [17]), CD69 and IL27. These findings were further supported by our *in silico* replication efforts [20].

In our latest effort to identify additional genetic loci for T1D susceptibility, we have carried out the largest meta-analysis to date and examined associations between the disease and approximately 2.54 million SNPs in a combined cohort of 9934 cases and 16,956 controls [4]. We have uncovered three new signals associated with T1D that reached genome-wide significance. Then we performed targeted follow-up of 53 SNPs and replicated our findings in an

independent sample set of 1120 affect trios. The most significant SNP (rs539514) resided in an intronic region of the LIM domain only 7 (*LMO7*) gene on 13q22. The second most significant SNP (rs478222) resided in an intronic region of the protein EFR3 homolog B *EFR3B* gene on 2p23, and the third SNP (rs924043) lay in an intergenic region on 6q27. These latest regions add to the growing repertoire of gene networks predisposing to T1D, currently residing at 53 loci. However, additional laboratory studies are needed to identify both the causative variant and the corresponding genes.

GWAS have revolutionized the field of complex disease genetics. For the first time there is real consensus on the role of specific genetic factors underpinning common disorders. However, such genome-wide scans can lack coverage in certain regions that are difficult to genotype so it is possible that other loci with reasonable effect sizes remain to be uncovered. What is clear is that larger sample sizes and bigger meta-analyses of GWAS datasets lead to the uncovering of further loci, albeit with lower and lower effect sizes. However, it has been predicted that there are a myriad of rare variants (possibly with larger effects) contributing to disease that cannot be detected on current genotyping platforms. To uncover the remaining 'missing heritability'[21] in complex diseases such as T1D, investigators in the near future will need to work on large, high-throughput sequencing efforts involving thousands of DNA samples from affected subjects and a similar number of controls.

Novel genomic techniques, such as next-generation DNA sequencing, opened new avenues in the elucidation of genetic defects and sped up the identification of causative gene variants to systematically tackle previously intractable genetic disorders that would be missed by GWAS. Recent studies have used this approach to identify mutations for Miller syndrome [22] and Charcot–Marie–Tooth disease [23]. Although powerful, next-generation DNA sequencing is still very expensive and time-consuming. However in future, with reductions in the costs of whole genome sequencing, it will probably become the dominant method for identifying mutations.

One of the main challenges in the future will be to determine how these recently discovered SNP variants affect the expression and function of the gene products. Detailed genotype–phenotype studies are necessary before we can translate GWAS findings into meaningful benefits for patient care. However, before this occurs, we need to fully understand the various gene networks that trigger the onset of the disease and how they are impacted by risk or disease-causing mutations. Once we fully understand the relationship between disease-causing variants and the disease onset, we will be able to apply targeted genomic approaches to shut down the premature activation of the immune system and prevent the onset of T1D.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

1 EURODIAB ACE Study Group. Variation and trends in incidence of childhood diabetes in Europe. EURODIAB ACE Study Group. *Lancet* 355(9207), 873–876 (2000).

2 Clayton DG. Prediction and interaction in complex disease genetics: experience in Type 1 diabetes. *PLoS Genet.* 5(7), E1000540 (2009).

3 Smyth DJ, Cooper JD, Bailey R *et al.* A genome-wide association study of nonsynonymous SNPs identifies a Type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat. Genet.* 38(6), 617–619 (2006).

4 Bradfield JP, Qu HQ, Wang K *et al.* A genome-wide meta-analysis of six Type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet.* 7(9), E1002293 (2011).

5 Nejentsev S, Howson JM, Walker NM *et al.* Localization of Type 1 diabetes susceptibility to the MHC class I genes *HLA-B* and *HLA-A*. *Nature* 450(7171), 887–892 (2007).

6 Bell GI, Horita S, Karam JH. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* 33(2), 176–183 (1984).

7 Nistico L, Buzzetti R, Pritchard LE *et al.* The *CTLA-4* gene region of chromosome 2q33 is linked to, and associated with, Type 1 diabetes. Belgian Diabetes Registry. *Hum. Mol. Genet.* 5(7), 1075–1080 (1996).

8 Bottini N, Musumeci L, Alonso A *et al.* A functionalvariant of lymphoid tyrosine phosphatase is associated with Type I diabetes. *Nat. Genet.* 36(4), 337–338 (2004).

9 Vella A, Cooper JD, Lowe CE *et al.* Localization of a Type 1 diabetes locus in the IL2RA/CD25 region by use of tag single-nucleotide polymorphisms. *Am. J. Hum. Genet.* 76(5), 773–779 (2005).

10 Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of

genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.* 33(2), 177–182 (2003).

11 Knip M, Veijola R, Virtanen SM *et al.* Environmental triggers and determinants of Type 1 diabetes. *Diabetes* 54(Suppl. 2), S125–S136 (2005).

12 Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of *IFIH1*, a gene implicated in antiviral responses, protect against Type 1 diabetes. *Science* 324(5925), 387–389 (2009).

13 Hakonarson H, Grant SF, Bradfield JP *et al.* A genome wide association study identifies *KIAA0350* as a Type 1 diabetes gene. *Nature* 448(7153), 591–594 (2007).

14 Wellcome Trust Case Control Consortium: WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145), 661–678 (2007).

15 Todd JA, Walker NM, Cooper JD *et al.* Robust associations of four new chromosome regions from genome-wide analyses of Type 1 diabetes. *Nat. Genet.* 39(7), 857–864 (2007).

16 Hakonarson H, Qu HQ, Bradfield JP *et al.* A novel susceptibility locus for Type 1 diabetes on Chr12q13 identified by a genome-wide association study. *Diabetes* 57(4), 1143–1146 (2008).

17 Grant SF, Qu HQ, Bradfield JP *et al.* Follow up analysis of genome-wide association data identifies novel loci for Type 1 diabetes. *Diabetes* 58(1), 290–295 (2009).

18 Concannon P, Onengut-Gumuscu S, Todd JA *et al.* A human Type 1 diabetes susceptibility locus maps to chromosome 21q22.3. *Diabetes* 57(10), 2858–2861 (2008).

19 Barrett JC, Clayton DG, Concannon P *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of Type 1 diabetes. *Nat. Genet.* 41(6), 703–707 (2009).

20 Qu HQ, Bradfield JP, Li Q *et al. In silico* replication of the genome-wide association results of the Type 1 Diabetes Genetics Consortium. *Hum. Molec. Genet.* 19(12), 2534–2538 (2010).

21 Manolio TA, Rodriguez LL, Brooks L *et al.* New models of collaboration in genome-wide association studies: the Genetic Association Information Network. *Nat. Genet.* 39(9), 1045–1051 (2007).

22 Roach JC, Glusman G, Smit AF *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328(5978), 636–639 (2010).

23 Lupski JR, Reid JG, Gonzaga-Jauregui C *et al.* Whole genome sequencing in a patient with Charcot–Marie–Tooth neuropathy. *N. Engl. J. Med.* 362(13), 1181–1191 (2010).

■ Website

101 International Diabetes Federation: diabetes prevalence.
www.idf.org