Biochemical and Biophysical Research Communications 452 (2014) 213–220

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Genetic architecture of type 2 diabetes

Kazuo Hara, Nobuhiro Shojima, Jun Hosoe, Takashi Kadowaki *

The Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

ARTICLE INFO

Article history:
Received 24 July 2014
Available online 8 August 2014

Keywords:
Genome-wide association studies
Single nucleotide polymorphism
Type 2 diabetes mellitus
Risk allele frequency
Linkage disequilibrium

ABSTRACT

Genome-wide association studies (GWAS) have identified over 70 loci associated with type 2 diabetes (T2D). Most genetic variants associated with T2D are common variants with modest effects on T2D and are shared with major ancestry groups. To what extent the genetic component of T2D can be explained by common variants relies upon the shape of the genetic architecture of T2D. Fine mapping utilizing populations with different patterns of linkage disequilibrium and functional annotation derived from experiments in relevant tissues are mandatory to track down causal variants responsible for the pathogenesis of T2D.

1. Introduction

T2D is a complex, multifactorial disease characterized by hyperglycemia resulting from impaired pancreatic β-cell function and a decreased action of insulin on target tissues [1]. Heritability of T2D is estimated to range between 30% and 70%, based on the results from twin studies [2]. With the advent of high-throughput genotyping techniques, GWASs have identified over 70 loci for T2D (Fig. 1). However, single locus has very modest effect on T2D and all the reported loci seem to account only part of the heritability of T2D [3]. T2D, once called a geneticist's nightmare or headache [4,5], now becomes a mystery for geneticists [6]. Intensive efforts to discover rare or low-frequency genetic variants with larger effects are now ongoing to solve this mystery. Another issue is that genetic variants identified by GWAS usually reside in non-coding regions of the genome, which makes linking disease-associated variants to disease pathogenesis very difficult. Here, we review recent progress and offer future perspectives on the genetics of T2D.

2. Methods

After a review of the literature that were reported from decades ago up until the present, we included important papers in which researchers carried out or investigated genome wide association studies, whole-exome and whole-genome sequencing studies, functional analyses, population genetics, clinical translation and so forth to show the genetic architecture of type 2 diabetes. We included the literature that studied a non-European population as well as a European population because it would be an ideal way of fine-mapping the T2D loci. For that purpose, we introduced our researches that studied mainly an Asian population.

In addition, we showed functional annotation for T2D loci using publicly available databases. Functional annotations for the TCF7L2 locus were retrieved from publicly available databases, ENCODE [7] and isletRegulome [8], and combined. ENCODE provides valuable information on regulatory elements derived from a wide variety of cell lines. Information derived from cell lines related to islets was retrieved. The locus bound by islet-specific transcription factors, and plays a role as an active enhancer site to regulate the expression levels of TCF7L2 genes are identified. In particular by trans-ethnic meta-analysis, T2D loci have been narrow down significantly to the region that has critical role in the pathogenesis of T2D [9].

3. T2D loci identified by GWAS

Analysis of insulin receptor revealed that mutations of the gene can cause various symptoms multiple distinct syndromes including type A insulin resistance, Rabson-Mendenhall syndrome and Leprechaunism [10–12]. On the other hand, analysis of mitochondrial DNA reveals the mutations could be associated with Maternally inherited diabetes and deafness (MIDD) [13]. Although linkage analysis had successfully identified causal genes for monogenic diabetes, MODY (maturity-onset diabetes of the young) [14,15] and neonatal diabetes [16], it did not provide clear evidence regarding any individual loci is linked to T2D. In 1996, researchers pointed out the limitations of linkage analysis for detecting causal loci of complex disease and proposed the use of

* Corresponding author. Fax: +81 3 5689 7209.
E-mail address: kadowaki-3im@h.u-tokyo.ac.jp (T. Kadowaki).

http://dx.doi.org/10.1016/j.bbrc.2014.08.012
0006-291X/

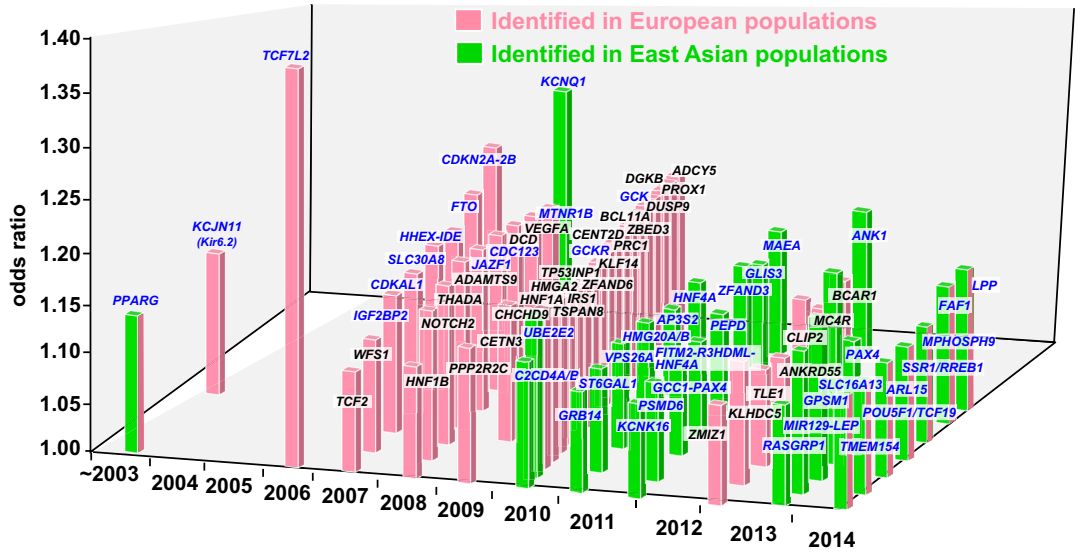

Fig. 1. The T2D loci identified mainly by GWAS. Height of the bars indicates the effect size of the index variant at the locus. Green and pink denote the population in which the association between T2D and the locus was clarified first. The loci in blue lettering were replicated in an independent study.

association tests [17]. But this proposal had to await technological developments in genotyping and information gathering on genetic variants across the genome for a decade before it was proven to work. Since the first successful GWAS on age-related macular degeneration was published in 2005 [18], the number of loci for common diseases mapped by GWAS has increased dramatically. An association study on chromosome 10q, which has been mapped as a T2D locus in the Icelandic population [19,20], identified variants in *TCF7L2* (transcription factor 7-like 2) that were consistently associated with T2D ($P = 5.4 \times 10^{-140}$) [21]. In 2007, multiple T2D loci were discovered by independent research groups. In addition to confirming the association of *TCF7L2* with T2D, they discovered *HHEX* (homeobox hematopoietically expressed) and *SLC30A8* (solute carrier family 30) in early onset-T2D [22], *CDKN2A-CDKN2A* (cyclin-dependent kinase inhibitor), *IGF2BP2* (insulin-like growth factor 2 mRNA-binding protein 2), and *CDKAL1* (CDK5 regulatory subunit associated protein 1-like 1) in late-onset T2D [23–26].

3.1. GWAS in non-European populations

We conducted a GWAS for T2D in a non-European population for the first time using a three-stage approach and reported *KCNQ1* as a T2D locus [27,28]. Based on calculations of sample size and statistical power, we genotyped at 100,000 variants in 187 Japanese T2D individuals; these genetic variations had been identified by sequencing 752 healthy Japanese individuals [29] as part of the Millennium Genome Project (Japan), which started in 2002 [30]. After conducting 2nd and 3rd follow-up tests for associations, we found strongly associated variants in intron 15 of *KCNQ1* (rs2237892 was most significantly associated with T2D: $P = 1.7 \times 10^{-42}$) with moderate effect [odds ratio (OR) = 1.40]. Another research group independently identified the same locus in addition to confirming *CDKAL1* and *IGF2BP2* as T2D loci in East Asians. *KCNQ1* encodes the pore-forming subunit of a voltage-gated potassium channel that is essential for the repolarization phase of the action potential within the cardiac muscle. Links between mutations in this gene and hereditary long QT syndrome [31] and familial atrial fibrillation are well documented [32]. The initial studies on *KCNQ1* show that individuals with risk allele have lower insulin secretion indexes and that *KCNQ1* is expressed in the human pancreas. Methodologically, the success of GWAS relies on sample size and marker density. Therefore, we used 459,359 genetic variants in 4470 T2D cases and 3071 controls to perform a larger scale Japanese GWAS and identified two previously unreported loci, *UBE2E2* and *C2CD4A-C2CD4B* [33]. *UBE2E2* encodes the ubiquitin-conjugating enzyme E2E2, which is reportedly expressed in human pancreas, liver, muscle and adipose tissue. Additionally, an ubiquitin–proteasome system reportedly plays a pivotal role in maintaining normal insulin biosynthesis, secretion and signaling, especially under conditions that increase endoplasmic reticulum stress in pancreatic β cells [34,35]. *C2CD4A* and *C2CD4B* are expressed in human pancreas, liver, muscle, and adipose tissue, as well as in a cultured insulin-secreting cell line; moreover, *C2CD4A* and *C2CD4B* expression are increased following treatments involving pro-inflammatory cytokines [36,37]. GWASs identified five T2D loci, *PTPRD*, *SRR* [38], *SPRY2*, *CDC123-CAMK1D* [39], and *PAX4* [40], in Han Chinese populations. Additionally, *GRB14*, *ST6GAL1*, *VPS26A*, *HMG20A*, *AP3S2*, *HNF4A* [41], and *SGCG* [42] were identified as loci in south Asian populations, and the intergenic region between *RND3* and *RBM43* was identified as a T2D locus in African Americans [43].

3.2. Collaborative works utilizing genotype imputation accelerate discovery of T2D loci

Each research group uses their preferred genotyping platform, which makes meta-analysis difficult. Genotype imputation can be used to combine results from studies that used different arrays or platforms, which increases the power of detecting association signals. DIAGRAM (Diabetes Genetics Replication and Meta-analysis Consortium) conducted the first meta-analysis on T2D; the sample size of this meta-analysis was 4549 cases and 5579 controls at the discovery stage [44]. Six novel loci including *JAZF1, CDC123, CAMK1D, TSPAN8, LGR5, THADA, ADAMTS9,* and *NOTCH2* were identified. The second meta-analysis by DIAGRAM had a sample size of 8130 cases and 38,987 controls; notably 12 T2D loci harboring genes that included *BCL11A, CENTD2, CHCHD9, DUSP9, HMGA2, HNF1A, KCNQ1, KLF14, PRC1, TP53INP1, ZBED3,* and *ZFAND6* were identified [45]. The third effort by DIAGRAM included 12,171 cases and 56,862 controls; eight additional loci were discovered [46]: *ZMIZ1*, *ANK1*, *KLHDC5*, *TLE1*, *ANKRD55*, *CILP2*, *MC4R,* and *BCAR1*. Among them, *ANK1* was identified as a T2D locus

independently in East Asians [47]. We participated in the Asian Genetic Epidemiology Network Type 2 Diabetes Mellitus (AGEN-T2D), which is the largest consortium on studying T2D in East Asian populations. AGEN-T2D includes approximately 55,000 individuals in the study population. The meta-analysis identified eight T2D loci that had not been in previous GWASs in European populations: *GLIS3*, *PEPD*, *FITM2-R3HDML-HNF4A*, *KCNK16*, *MAEA*, *GCC1-PAX4*, *PSMD6* and *ZFAND3* [48].

3.3. Catalogue of genetic variation: The HapMap and 1000G Project (1KG)

Genotype imputation in a study sample requires a reference panel, ideally from the same ancestry group. Such reference panels have been provided by the International HapMap project [49] and 1000 Genomes Project (1KG). The HapMap Phase II, most widely used for imputation in GWASs, provides approximately 3.8 million variants in four populations [50]. Because GWASs usually genotype much fewer variants than 3.8 million, imputation using the HapMap reference panel is expected to increase the power for detecting association signals with a higher resolution. In fact, by imputation using HapMap reference panel, a Japanese group identified *ANK1* as a T2D locus, which was not defined in the previous GWAS using genotyped variations alone [47]. 1KG was founded with the aim of collecting over 95% of all variants with minor allele frequencies $\geqslant$1% in major ethnic groups [51,52]. 1KG has collected genetic variants that are rare and not represented in the HapMap Project. Thus, it was reasoned that that genotype imputation using 1KG reference panel should increase the power for detecting rare variants associated with disease in GWAS [53]. We tested the associations between T2D and 6,209,637 genetic variants, which were directly genotyped or imputed using East Asian references from the 1KG in 5976 cases and 20,829 controls [54]. We identified three new loci with genome-wide significance, which were *MIR129-LEP* ($P = 2.55 \times 10^{-13}$, OR = 1.17), *GPSM1* ($P = 1.7 4 \times 10^{-10}$, OR = 1.15) and *SLC16A13* ($P = 7.69 \times 10^{-13}$, OR = 1.20). The 1KG will officially end this summer with completion of whole-genome sequencing of 2500 individuals from 27 populations around the world. Improvement in the quality of data, especially for rare and structural variants, and the addition of a reference panel for south Asian populations are expected to facilitate identification of many new loci for T2D.

GWASs have successfully identified over 70 loci associated with T2D, which shed light on the pathogenesis of type 2 diabetes. However, big and fundamental questions remain to be answered. Only 10% of the heritability of T2D can be explained by the loci identified thus far [55–57], which is called "missing heritability" problem.

4. Common genetic variants for T2D loci are shared among different ethnic groups

Early in the GWAS era, whether genetic variations identified in one population were relevant to other populations was not clear. We investigated the relevance of T2D loci identified in European populations to Japanese populations [58,59]. We tested a correlation between effect size and direction of genetic variants identified in European populations on T2D risk in Japanese populations and found that the correlation was generally good [54].

The PAGE (Population Architecture using Genomics and Epidemiology) consortium systematically tested transferability and generalizability of the association results from T2D GWASs of major population groups living in the United States [60]. They found marked consistency across populations with regard to the directionality of the association of most genetic variants.

With the largest sample size, the DIAGRAM confirmed the notion that common genetic variants for T2D are shared across groups [46]. They observed evidence of heterogeneity among populations at just three known loci: *TCF7L2*, *PEPD*, and *KLF14*. The similar correlation of effect size between different groups has been seen in Mexican and Hispanic Americans [61] and South Asian populations [41]. This result suggests that more common variants of T2D will be identified through combining datasets derived from different ethnic groups as much as possible.

5. Risk allele frequencies can differ substantially among populations

Although the effect size of common, T2D-associated variants is comparable across populations, risk allele frequency (RAF) for individual variants sometimes differs substantially between populations (Fig. 2). For example, RAF of the index variant in *KCNQ1* among East Asians is 0.640, while it is 0.925 among Europeans; this difference caused a weaker association in GWAS of European samples than the association estimated with East Asian samples. Recently, we identified a T2D locus at *SLC16A13* on chromosome 17 (rs312457: OR [95% CI] = 1.20 [1.14–1.26], RAF = 0.078) [54]. The top variant, rs312457, is rare in European populations (RAF = 0.004), and there was no signal in DIAGRAM. More recently, the SIGMA Consortium discovered a five-variant haplotype associated with T2D in Mexican and Mexican American populations [61].

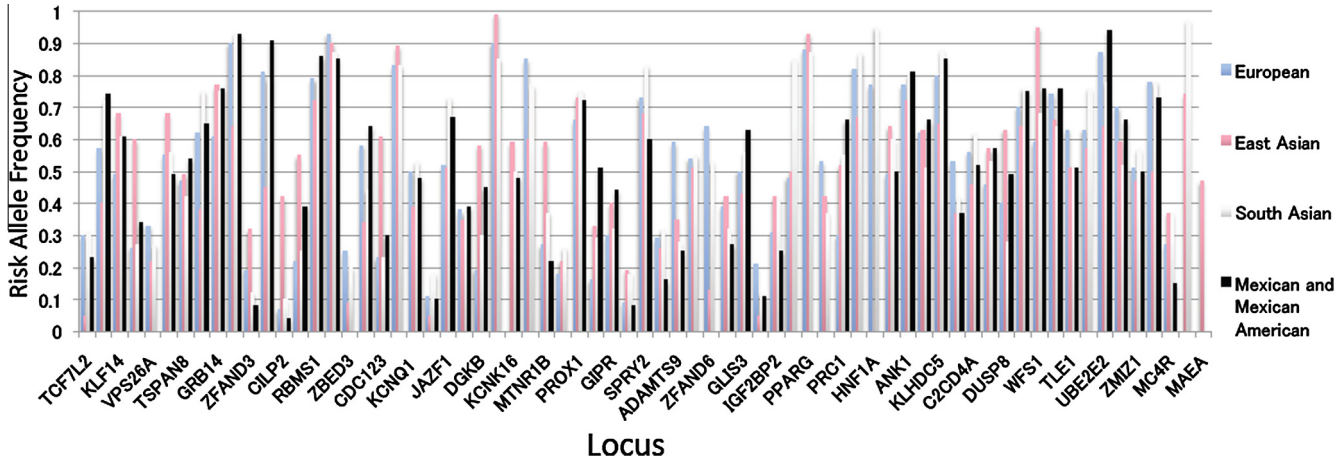

Fig. 2. Risk allele frequency (RAF) of T2-associated variants among major ancestry: European, East Asian, South Asian, and Mexican and Mexican American populations. Substantial differences in RAF were observed at some loci. The data are based on the DIAGRAMv3 [46].

The top variant (in *SLC16A13*) in the East Asian population is in perfect LD (CHB + JPN) with the variants comprising a haplotype associated with T2D found in the SIGMA study (in *SLC16A11*), suggesting that the signal at *SLC16A11* and *SLC16A13* are identical. Mexican and Hispanic Americans are both extremely heterogeneous groups admixed with European, African, and Native American ancestries. At *SLC16A11*, T2D individuals in the SIGMA study had higher proportions of Native American ancestry than healthy controls. The haplotype associated with T2D in the SIGMA is observed frequently in Native American (50%) and East Asian populations (10%), while it is rare or absent in European and African populations. These are consistent with the fact that there is no signal at this locus except with East Asian populations and Native Americans, which shares a genetic background. It would be interesting to investigate association with T2D at this locus in other Native American populations with sequencing or dense-genotype arrays. It would also be interesting to determine whether admixture mapping with samples from Hispanic Americans would detect a signal at this locus. What is surprising is that the risk haplotype might be a result of admixture between modern humans and Neanderthals, which took place after a group of modern humans migrated out of Africa. It would be interesting to study whether the substantial difference in risk allele frequency was caused by purifying selection at this locus or it was just the result of a genetic drift that followed a population bottleneck among ancestors of Native American experienced as they traveled over the Bering Strait to America. Notably, estimated mean genetic risks of T2D in populations may indicate that differences in RAFs of T2D-associated variants between populations might result from natural selection; there is significant disparity in mean genetic scores between populations for T2D variants when compared with variants in other diseases [62]. However, the conclusion that East Asians have the lowest risk of T2D is apparently not consistent with a large volume of epidemiological literature [63,64]. Disparity in T2D prevalence might be explained not by common variants alone, but by a combination of common, rare, and low-frequency variants, most of which have not yet been identified.

6. Missing heritability and genetic architecture of T2D

Although GWASs have identified over 70 loci for T2D, the collective effect of those loci can account for only a small part of the heritability of T2D. It is estimated that 63 known T2D loci taken together account for only 5.7% of variance in susceptibility to T2D among European populations [46]. Population geneticists have been seeking the source of this missing heritability, though there is an argument against the existence of missing heritability [65]. Nonetheless, there is a rationale for seeking for common variants as a source of missing heritability. Estimate based on DIAGRAM indicate that, in addition to those variants that reached genome-wide significance, another five hundred common-variants could explain approximately 50% of the genetic component of T2D [45]. Independently, Stahl et al. reported that thousands of common variants could contribute approximately half of the genetic component responsible for T2D [66]. However, it is very challenging to provide compelling evidence for associations between T2D and those variants with estimated ORs being between 1.01 and 1.11.

Most GWASs filter out low-frequency variants from an analysis to avoid spurious results derived from poor quality of genotype imputation. But, frequent updates in 1KG have been improving the quality of the reference panel; these improvements have made association studies involving low-frequency and rare variants more practicable in the last 2 years. Consequently, we have the potential to more fully examine the "goldilocks" hypothesis, which claims that there are variants with MAF of 0.5–2% (between common and rare) that have relatively large effects (OR = ~2) [67]. To test this hypothesis, we used 1KG reference panel (CHB + JPN) to perform a genome-wide association. Because concordance between directly genotyped and imputed genotypes for variants with MAF > 1% was good (94.2%), we tested associations between T2D and genetic variants with MAF > 1%. We found only one low-frequency variant (~2%) with a modest effect (OR = 2) that reached the cutoff *p*-value for follow-up analysis. Because the association was abrogated after conditioning for a genetic variation in *HHEX*, one of the established T2D genes, we did not include this variant in the follow-up analysis. We used a gene-based test to identify regions that contain low-frequency variants, each with a large effect, but we found no such region. However, as 1KG will be completed the summer of 2014 whole-genome sequence for 2500 individuals, the performance of imputation using the 1KG reference panel is expected to improve for rare and low-frequency variants. Therefore, re-imputing whole dataset and testing associations of variants with MAF > 1%, which is our ongoing effort, may yet be worthwhile. Having mentioned that, our results, along with the findings from recent whole-exome (WES) and whole-genome sequencing (WGS) studies, indicate that goldilocks variants do not have a major role in T2D. WES of 1000 T2D individuals and 1000 healthy controls from the Danish population has been done to identify rare variants with large effects on T2D. Although the study had the statistical power to detect at least one gene that harbors rare variants with large effects, they found no such gene [68]. WGS with samples from 2630 Icelanders discovered only three rare or low-frequency variants that were associated with T2D: rs76895963 in *CCND2A* (RAF = 0.0147, OR = 0.53), p.Ser539Trp of *PAM* (RAF = 0.65%, OR = 1.47), and p.Gly218Alafs*12 in *PDX1* (RAF = 0.20%, OR = 2.27) [69].

A compelling paper on the genetic architecture of T2D was published recently [70]. They created a demographic model of the European populations and defined the genetic architecture of T2D with two parameters: (1) the relationship between purifying selection and effect on T2D of causal variants and (2) the number of T2D loci. They simulated these parameters to determine which combinations of values for the two parameters would be discarded based on the results from previous GWASs and linkage studies. The simulations demonstrate that a wide range of values for the two parameters remained consistent. Based on this simulation, the two extreme scenarios were rejected: (1) oligogenic model in which a handful of loci with a very large effect can explain the heritability and (2) infinitesimal model in which huge number of loci with a very modest effect can explain the heritability. However, a continuous spectrum between the two poles remains consistent in terms of RAF and effect size: thousands of rare variants with a large effect at hundreds of T2D loci can explain about a half of genetic variance in T2D (Scenario 1) and rare variants with a relatively large effect play minor part of heritability in T2D and most of genetic variance in T2D can be attributable to common variants (Scenario 2). What is very suggestive is that exome-chip analysis including very rare variants with 55,000 samples would be able to prove which scenario is more consistent than the other (Fig. 3). Large-scale studies that involve exome-chips and focus on rare variants associated with T2D and related traits are planned, and the result will be available in the near future. If this scenario is true, disparity of T2D prevalence between populations might be explained by rare variants, which tend to be population specific. It should also be noted that T2D loci tend to harbor multiple independent variants that give rise to signal in GWAs and that identifying all such variants might substantially increase the variance explained by GWAS hits. For example, at *KCNQ1*, there are at least three independent signals: rs231362 (intron 10 of *KCNQ1* and the overlapping *KCNQ1OT1* transcript), rs2237892 (intron 15 of *KCNQ1*), and rs139647931 (intron 15 of *KCNQ1*). As for *SLC16A11*

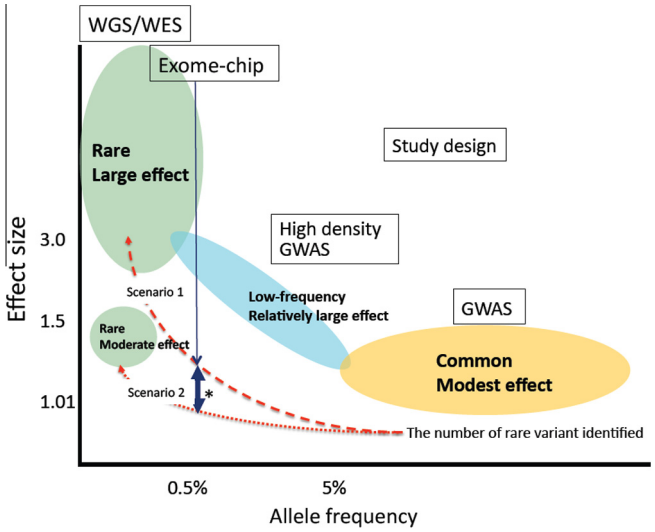

Fig. 3. Genetic architecture of T2D and study design. There have been arguments about the source of the missing heritability that cannot be explained by the common variants with modest effect identified by GWAS so far. Studies using sequencing (WGS/WES) or customized array (exome-chip) seeks for rare or low-frequency variants with larger effect compared to common variants have attempted to fill in the missing heritability. High-density GWAS using typed and imputed variants might be able to find low-frequency variants for T2D. But, success of those studies largely relies on the actual genetic architecture of T2D. *The number of rare variants that will be identified depends on the genetic architecture of T2D. It was predicted that studies using exome-chip arrays with adequate sample size might determine which scenario is more plausible: rare variants with a large effect can explain about a half of genetic variance in T2D (Scenario 1) and rare variants with a relatively large effect play minor part in heritability of T2D and most of genetic variance in T2D can be attributable to common variants with modest effect (Scenario 2) [70]. Note that the cut-off frequency for rare, low frequency, and common variants is very arbitrary and there is a continuous spectrum in frequency and effect of genetic variants. WGS: whole-genome sequencing, WES: whole-exome sequencing.

we identified recently as a novel T2D locus, a conditional analysis on the top variant in *SLC16A11* revealed second independent signal at this locus.

7. Fine mapping of reported T2D loci

For almost all T2D loci yet identified, the variant most strongly associated with T2D is a mere proxy of the causal and functional variant responsible for the association of that locus with T2D. GWASs have successfully identified a number of T2D loci mainly in European ancestry. GWAS in European populations has an advantage because of long-range linkage disequilibrium (LD); this LD is probably due to a severe population bottleneck that occurred after migration out-of Africa. However, the drawback is that a causal variant is difficult to distinguish from the many nearby variants associated with T2D due to LD with each other. This issue prompts the need to examine traits loci in different populations, which has different pattern of LD around causal variant to narrow down the region of interest. The first large-scale trans-ethnic study on T2D has been conducted by DIAGRAM [46]. Our group contributed to this effort by providing summary statistics of associations between genetic variants and T2D in East Asian populations. Using a Bayesian statistical framework that was developed by Andrew Morris, which is more powerful than the fixed effects model commonly used for meta-analysis of GWAS, the meta-analysis of three ancestor groups successfully narrowed down several known T2D loci [46]. Meta-analysis of African with European or East Asian populations would be an ideal way of fine-mapping the T2D loci. Such analysis is now underway.

8. From proxy variant to functional variant

Most T2D-associated variants fall into non-coding regions of the genome. Although it is possible that a causal variant might be a non-synonymous variant at some loci, it is currently thought that most causal variants reside within regulatory elements that influence gene expression. ENCODE (Encyclopedia Of DNA Elements) is one of the most comprehensive datasets of regulatory elements in the human genome; ENCODE includes transcription factor binding site, histone marks, and DNA methylation sites [7]. There are several lines of evidence indicating that GWAS hits are enriched with regulatory elements mapped by ENCODE. Therefore, ENCODE is a valuable resource for linking genetic variations identified in GWAS with functional and causal variants. It should be kept in mind that because limited cell types have been assayed in ENCODE, ENCODE data might lack valuable information on tissues relevant to the pathogenesis of T2D; for example, some T2D-associated genetic variants may affect gene expression in a tissue-specific manner. Recently, a paper on functional elements focusing on human islet has been published [8]. The research group chose transcription factors that are specifically expressed in human islet and used a CHIP-sequence technique to identify genomic sites bound by islet-specific transcription factors. They also used antibodies to major histone marks to conduct CHIP-sequencing, and thereby identified genomic sites that are actively transcribed. By combining the results, the research group mapped genomic sites targeted by islet specific transcription factors to drive islet-specific genes activity. They showed that transcription enhancer sites play a key role in islet-specific regulation of gene expression and hence a critical role in the pathogenesis of T2D. Therefore, identifying individual enhancer sites and the respective target genes is one of the best scenarios for fine-mapping of T2D loci. As shown in Fig. 4, by trans-ethnic meta-analysis, T2D loci have been narrow down significantly to the region that has critical role in the pathogenesis of T2D.

Genetic variant of eQTL is such a variant that is associated with gene expression level. There is a plethora of evidence that genetic variants are involved in the heritability of gene expression levels [71–74]. Thus, eQTL is a top candidate of causal variant in GWAS hit. Because different tissues and cell types have different pattern of gene expression, eQTL using relevant tissue for T2D is mandatory to understand the molecular mechanism, whereby genetic variant exerts effect on the development of T2D. In fact, it has been shown that 50–90% of eQTL are tissue dependent, and trait-associated variants tend to exert more tissue-specific effects [75,76]. Publicly available data exists, focusing on liver, adipose (subcutaneous), and skeletal muscle. It is particularly important to perform eQTL analysis in different populations and compare the results, which provide us with insight into disparity of T2D prevalence between populations as well as enable us to improve the resolution of eQTL mapping. Intensive and more comprehensive efforts to gather information from relevant tissues for the pathogenesis of T2D is urgent.

It is only recently that noncoding RNAs, such as microRNAs (miRNAs) and long noncoding RNAs (lncRNA), have been involved in the pathogenesis of common disease. miRNA that specifically expressed in human islet tends to target the genes in and near T2D loci identified by GWAS. Genetic variants that could affect expression level of miRNA could be a causal variant in T2D loci [77]. KCNQ1OT1 is a lncRNA and is involved in imprinting that regulates nearby six genes including *CDKN1C*, a regulator of beta-cell development [78]. The imprinting control region (ICR) located in intron 10 of *KCNQ1* is unmethylated on the paternal chromosome and methylated on the maternal chromosome. In accordance with this, maternally transmitted risk allele of variant in *KCNQ1* had stronger effect than paternally transmitted one [79]. Recently, it

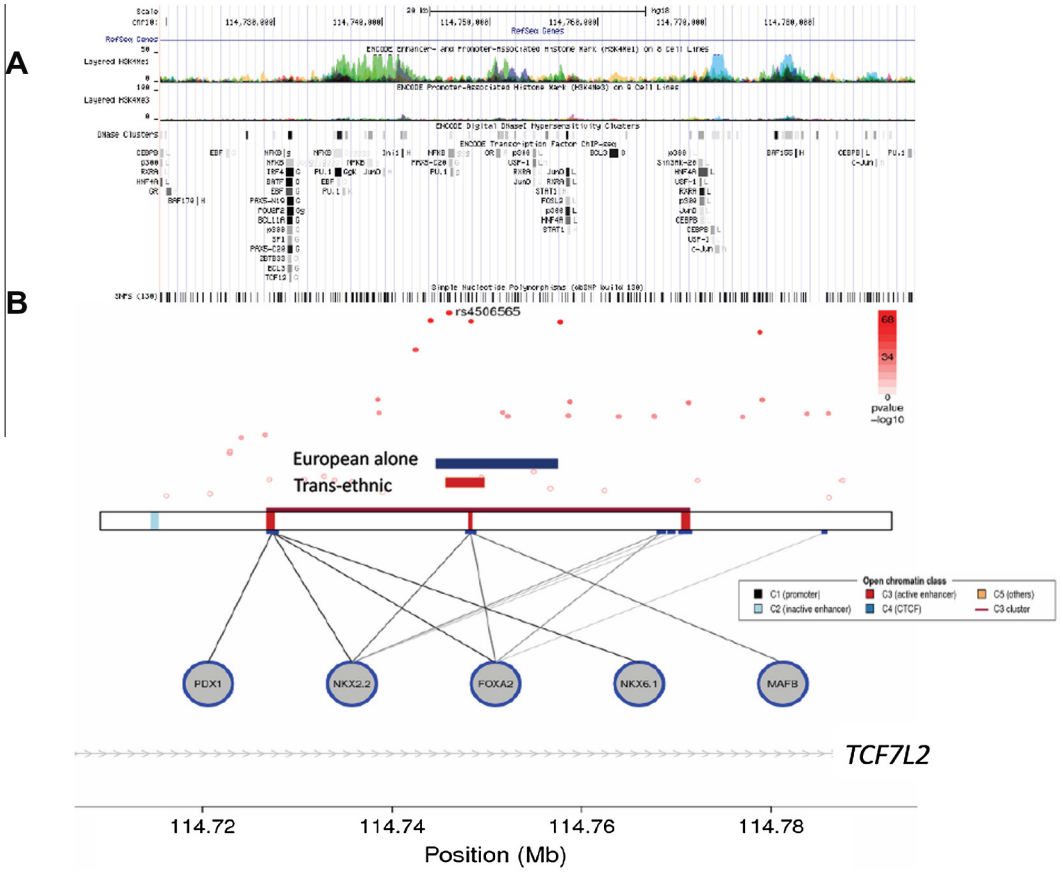

Fig. 4. Functional annotation for T2D loci using publicly available databases. Functional annotations for the TCF7L2 locus were retrieved from publicly available databases, ENCODE [7] and isletRegulome [8], and combined. (A) ENCODE provides valuable information on regulatory elements derived from a wide variety of cell lines. Information derived from cell lines related to islets was retrieved, as shown in the upper panel. (B) The lower panel demonstrated that the locus is bound by islet-specific transcription factors such as FOXA2 and NKX2.2, and plays a role as an active enhancer site (shown in vertical red bar) to regulate the expression levels of islet-specific genes. Note that the meta-analysis of major ancestry groups narrowed the region that is likely to have a causal variant in this locus (red rectangle) compared to the previous result in European populations alone (blue rectangle) [9].

was shown that individuals with risk allele of T2D-associated variant in intron 15 of *KCNQ1* had higher DNA methylation level in the ICR in fetal islet [80]. KCNQ1OT1 forms an intrachromosomal loop between differentially methylated region (DMR) and the promoter of KCNQ1. Deficiency of KCNQ1OT1 to form this loop caused loss of imprinting [81]. Taken together with the result that a genetic variant of *KCNQ1OT1* could alter the local structure of KCNQ1OT1 [82], genetic variants in *KCNQ1OT1* might affect the regulation of nearby genes such as *CDKN1C* by altering activity of imprinting at this locus. Tomizawa et al. reported a complicated mechanism whereby genetic variants in *CDKAL1* affect gene expression level of *CDKAL1* [83]. Sequence element around T2D-associated variant in *CDKAL1* affected the level of a splicing variant of *CDKAL1*, *CDKAL1-v1* which is a non-coding RNA and binds to CDKAL1-targeting miRNA competitively with mRNA of *CDKAL1*. T2D-associated variant in this locus could affect expression level of *CDKAL1* by altering the level of *CDKAL1-v1*.

Epigenetic modifications now draw a lot of attention in the field of genetics. DNA methylation is an epigenetic modification that plays an important role in regulating gene expression. Different type of cells have distinct patterns of DNA methylation and a recent study clarified that substantial parts of DNA methylation is heritable and genetic variants determine, at least in part, the pattern of DNA methylation [84]. Large-scale and systematic studies integrating genetic and epigenetic information is needed to clarify those genetic variants of T2D influence regional DNA methylation and gene expression.

9. Clinical translation and future perspective

The overarching goal of modern genetics is to translate genetic information into daily clinical practice as well as clarifying the molecular mechanism of diseases. Over the past few years, a number of prediction models of T2D using genetic variants have been proposed for better prevention of T2D. But in most of the cases, performance of the prediction model is only slightly better than that of conventional risk model incorporating sex, age, BMI, and family history of T2D [85,86]. A recent study evaluated the performance of a prediction model based on 62 known T2D loci, which is the largest number to date [87]. They found that area under the curve of the genetic risk score to predict T2D was significantly larger than the model without genetic information, but the difference was modest. Therefore, clinical usefulness of the prediction model for T2D based on genetic information is still unclear. The results of large scale exome-chip analyses on T2D and related traits are awaited to more precisely estimate T2D risk. It is also worth noting that genetic counseling using risk communication that is tailored to T2D is required in clinical settings.

References

[1] C.R. Kahn, G.C. Weir, G.L. King, et al., Diabetes: definition, genetics, and pathogenesis, Joslin's Diabetes Mellitus, 14th ed., Lippincott Williams & Wilkins, Boston, 2004, pp. 425–462.
[2] B. Newman, J.V. Selby, M.C. King, et al., Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins, Diabetologia 30 (1987) 763–768.

[3] D.A. Pyke, R. Tattersall, Clinics in endocrinology and metabolism, Vol. 6. No. 2, WB Saunders, London, 1977, pp. 285–303.
[4] H. Keen, The genetics of diabetes: from nightmare to headache, Br. Med. J. 294 (1987) 917–919.
[5] J.V. Neel, Diabetes Mellitus—a Geneticist's Nightmare: the Genetics of Diabetes Mellitus, Springer-Verlag, New York, 1976. p. 1.
[6] O. Zuk, E. Hechter, S.R. Sunyaev, et al., The mystery of missing heritability: genetic interactions create phantom heritability, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 1193–1198.
[7] ENCODE Project Consortium, The ENCODE (ENCyclopedia Of DNA Elements) Project, Science 306 (2004) 636–640.
[8] L. Pasquali, K.J. Gaulton, S.A. Rodríguez-Seguí, et al., Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants, Nat. Genet. 46 (2014) 136–143.
[9] A.P. Morris, Transethnic meta-analysis of genomewide association studies, Genet. Epidemiol. 35 (2011) 809–822.
[10] T. Kadowaki, C.L. Bevins, A. Cama, et al., Two mutant alleles of the insulin receptor gene in a patient with extreme insulin resistance, Science 240 (1988) 787–790.
[11] Y. Yoshimasa, S. Seino, J. Whittaker, et al., Insulin-resistant diabetes due to a point mutation that prevents insulin proreceptor processing, Science 240 (1988) 784–787.
[12] P. Roach, Y. Zick, D. Accili, et al., A novel human insulin receptor gene mutation uniquely inhibits insulin binding without impairing post-translational processing, Diabetes 43 (1994) 1096–1102.
[13] T. Kadowaki, H. Kadowaki, Y. Mori, et al., A subtype of diabetes mellitus associated with a mutation of mitochondrial DNA, N. Engl. J. Med. 330 (1994) 962–968.
[14] K. Yamagata, N. Oda, P.J. Kaisaki, et al., Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3), Nature 384 (1996) 455–458.
[15] K. Yamagata, H. Furuta, N. Oda, et al., Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1), Nature 384 (1996) 458–460.
[16] V. Senee, C. Chelala, S. Duchatelet, et al., Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism, Nat. Genet. 38 (2006) 682–687.
[17] K.R. Merikangas, N. Risch, Genomic priorities and public health, Science 302 (2003) 599–601.
[18] R.J. Klein, C. Zeiss, E.Y. Chew, Complement factor H polymorphism in age-related macular degeneration, Science 308 (2005) 385–389.
[19] R. Duggirala, J. Blangero, L. Almasy, et al., Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans, Am. J. Hum. Genet. 64 (1999) 1127–1140.
[20] I. Reynisdottir, G. Thorleifsson, R. Benediktsson, et al., Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34-q35.2, Am. J. Hum. Genet. 73 (2003) 323–335.
[21] S.F. Grant, G. Thorleifsson, I. Reynisdottir, et al., Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes, Nat. Genet. 38 (2006) 320–323.
[22] R. Sladek, G. Rocheleau, J. Rung, et al., A genome-wide association study identifies novel risk loci for type 2 diabetes, Nature 445 (2007) 881–885.
[23] L.J. Scott, K.L. Mohlke, L.L. Bonnycastle, A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants, Science 316 (2007) 1341–1345.
[24] R. Saxena, B.F. Voight, V. Lyssenko, et al., Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels, Science 316 (2007) 1331–1336.
[25] E. Zeggini, M.N. Weedon, C.M. Lindgren, et al., Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes, Science 316 (2007) 1336–1341.
[26] V. Steinthorsdottir, G. Thorleifsson, I. Reynisdottir, A variant in CDKAL1 influences insulin response and risk of type 2 diabetes, Nat. Genet. 39 (2007) 770–775.
[27] K. Yasuda, K. Miyake, Y. Horikawa, et al., Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus, Nat. Genet. 40 (2008) 1092–1097.
[28] H. Unoki, A. Takahashi, T. Kawaguchi, et al., SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations, Nat. Genet. 40 (2008) 1098–1102.
[29] F. Takaku, Overview on the millennium genome project, Nihon Rinsho 59 (2001) 11–18.
[30] M. Hirakawa, T. Tanaka, Y. Hashimoto, et al., JSNP: a database of common gene variations in the Japanese population, Nucleic Acids Res. 30 (2002) 158–162.
[31] Q. Wang, M.E. Curren, I. Splawski, et al., Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias, Nat. Genet. 12 (1996) 17–23.
[32] N. Neyroud, F. Tesson, I. Denjoy, et al., A novel mutation in the potassium channel gene KVLQT1 causes the Jervell and Lange-Nielsen cardioauditory syndrome, Nat. Genet. 15 (1997) 186–189.
[33] T. Yamauchi, K. Hara, S. Maeda, et al., A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B, Nat. Genet. 42 (2010) 864–868.
[34] M. Kawaguchi, K. Minami, K. Nagashima, et al., Essential role of ubiquitin–proteasome system in normal regulation of insulin secretion, J. Biol. Chem. 281 (2006) 13015–13020.
[35] M.D. López-Avalos, V.F. Duvivier-Kali, G. Xu, et al., Evidence for a role of the ubiquitin-proteasome pathway in pancreatic islets, Diabetes 55 (2006) 1223–1231.
[36] K. Warton, N.C. Foster, W.A. Gold, et al., A novel gene family induced by acute inflammation in endothelial cells, Gene 342 (2004) 85–95.
[37] N. Grarup, M. Overvad, T. Sparsø, et al., The diabetogenic VPS13C/C2CD4A/C2CD4B rs7172432 variant impairs glucose-stimulated insulin response in 5722 non-diabetic Danish individuals, Diabetologia 54 (2011) 789–794.
[38] F.J. Tsai, C.F. Yang, C.C. Chen, et al., A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese, PLoS Genet. 6 (2010) e1000847.
[39] X.O. Shu, J. Long, Q. Cai, et al., Identification of new genetic risk variants for type 2 diabetes, PLoS Genet. 6 (2010) e1001127.
[40] R.C. Ma, C. Hu, C.H. Tam, et al., Genome-wide association study in a Chinese population identifies a susceptibility locus for type 2 diabetes at 7q32 near PAX4, Diabetologia 56 (2013) 1291–1305.
[41] J.S. Kooner, D. Saleheen, X. Sim, Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci, Nat. Genet. 43 (2011) 984–989.
[42] R. Saxena, D. Saleheen, L.F. Been, et al., Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in Sikhs of Punjabi origin from India, Diabetes 62 (2013) 1746–1755.
[43] N.D. Palmer, C.W. McDonough, P.J. Hicks, et al., A genome-wide association search for type 2 diabetes genes in African Americans, PLoS One 7 (2012) e29202.
[44] E. Zeggini, L.J. Scott, R. Saxena, et al., Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes, Nat. Genet. 40 (2008) 638–645.
[45] B.F. Voight, L.J. Scott, V. Steinthorsdottir, et al., Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis, Nat. Genet. 42 (2010) 579–589.
[46] A. Mahajan, M.J. Go, W. Zhang, et al., Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility, Nat. Genet. 46 (2014) 234–244.
[47] M. Imamura, S. Maeda, T. Yamauchi, et al., A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations, Hum. Mol. Genet. 21 (2012) 3042–3049.
[48] Y.S. Cho, C.H. Chen, C. Hu, et al., Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians, Nat. Genet. 44 (2011) 67–72.
[49] The International HapMap Consortium, The International HapMap Project, Nature 426 (2003) 789–796.
[50] International HapMap Consortium, K.A. Frazer, D.G. Ballinger, et al., A second generation human haplotype map of over 3.1 million SNPs, Nature 449 (2007) 851–861.
[51] 1000 Genomes Project Consortium, G.R. Abecasis, D. Altshuler, et al., A map of human genome variation from population-scale sequencing, Nature 467 (2010) 1061–1073.
[52] 1000 Genomes Project Consortium, G.R. Abecasis, A. Auton, et al., An integrated map of genetic variation from 1092 human genomes, Nature 491 (2012) 56–65.
[53] J. Huang, D. Ellinghaus, A. Franke, et al., Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 data, Eur. J. Hum. Genet. 20 (1000) (2012) 801–805.
[54] K. Hara, H. Fujita, T.A. Johnson, et al., Genome-wide association study identifies three novel loci for type 2 diabetes, Hum. Mol. Genet. 23 (2014) 239–246.
[55] M.I. McCarthy, Genomics, type 2 diabetes, and obesity, N. Engl. J. Med. 363 (2010) 2339–2350.
[56] I. Prokopenko, M.I. McCarthy, C.M. Lindgren, Type 2 diabetes: new genes, new understanding, Trends Genet. 24 (2008) (2008) 613–621.
[57] T.A. Manolio, F.S. Collins, N.J. Cox, et al., Finding the missing heritability of complex diseases, Nature 461 (2009) 747–753.
[58] M. Horikoshi, K. Hara, C. Ito, et al., A genetic variation of the transcription factor 7-like 2 gene is associated with risk of type 2 diabetes in the Japanese population, Diabetologia 50 (2007) 747–751.
[59] H. Fujita, K. Hara, N. Shojima, et al., Variations with modest effects have an important role in the genetic background of type 2 diabetes and diabetes-related traits, J. Hum. Genet. 50 (2012) 776–779.
[60] C.A. Haiman, M.D. Fesinmeyer, K.L. Spencer, et al., Consistent directions of effect for established type 2 diabetes risk variants across populations: the population architecture using Genomics and Epidemiology (PAGE) Consortium, Diabetes 61 (2012) 1642–1647.
[61] SIGMA Type 2 Diabetes Consortium, A.L. Williams, S.B. Jacobs, et al., Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico, Nature 506 (2014) 97–101.
[62] E. Corona, R. Chen, M. Sikora, et al., Analysis of the genetic basis of disease in the context of worldwide human relationships and migration, PLoS Genet. 9 (2013) e1003447.
[63] D. Dabelea, E.J. Mayer-Davis, S. Saydah, Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009, JAMA 311 (2014) (2001) 1778–1786.
[64] L. Guariguata, D.R. Whiting, I. Hambleton, Global estimates of diabetes prevalence for 2013 and projections for 2035, Diabetes Res. Clin. Pract. 103 (2014) (2013) 137–149.

[65] O. Zuk, E. Hechter, S.R. Sunyaev, et al., The mystery of missing heritability: genetic interactions create phantom heritability, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 1193–1198.
[66] E.A. Stahl, D. Wegmann, G. Trynka, et al., Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis, Nat. Genet. 44 (2012) 483–489.
[67] J.C. Cohen, Emerging LDL therapies: using human genetics to discover new therapeutic targets for plasma lipids, J. Clin. Lipidol. 7 (2013) S1–S5.
[68] K.E. Lohmueller, T. Sparsø, Q. Li, et al., Whole-exome sequencing of 2000 Danish individuals and the role of rare coding variants in type 2 diabetes, Am. J. Hum. Genet. 93 (2013) 1072–1086.
[69] V. Steinthorsdottir, G. Thorleifsson, P. Sulem, et al., Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes, Nat. Genet. 46 (2014) 294–298.
[70] V. Agarwala, J. Flannick, S. Sunyaev, et al., Evaluating empirical bounds on complex disease genetic architecture, Nat. Genet. 45 (2013) 1418–1427.
[71] M. Morley, C.M. Molony, T.M. Weber, et al., Genetic analysis of genome-wide variation in human gene expression, Nature 430 (2004) 743–747.
[72] T. Lappalainen, M. Sammeth, M.R. Friedlander, et al., Transcriptome and genome sequencing uncovers functional variation in humans, Nature 501 (2013) 506–511.
[73] D.L. Nicolae, E. Gamazon, Zhang, et al., Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS, PLoS Genet. 6 (2010) e1000888.
[74] A.S. Dimas, S. Deutsch, B.E. Stranger, et al., Common regulatory variation impacts gene expression in a cell type-dependent manner, Science 325 (2009) 1246–1250.
[75] J. Fu, M.G. Wolfs, P. Deelen, et al., Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression, PLoS Genet. 8 (2012) e1002431.
[76] C.D. Brown, L.M. Mangravite, B.E. Engelhardt, Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs, PLoS Genet. 9 (2013) e1003649.
[77] V. Kumar, H.J. Westra, J. Karjalainen, et al., Human disease-associated genetic variation impacts large intergenic non-coding RNA expression, PLoS Genet. 9 (2013) e1003201.
[78] T. Potikha, S. Kassem, E.P. Haber, et al., P57Kip2 (cdkn1c): sequence, splice variants and unique temporal and spatial expression pattern in the rat pancreas, Lab. Invest. 85 (2005) 364–375.
[79] A. Kong, V. Steinthorsdottir, G. Masson, et al., Parental origin of sequence variants associated with complex diseases, Nature 462 (2009) 868–874.
[80] M.E. Travers, D.J. Mackay, N.M. Dekker, et al., Insights into the molecular mechanism for type 2 diabetes susceptibility at the KCNQ1 locus from temporal changes in imprinting status in human islets, Diabetes 62 (2013) 987–992.
[81] H. Zhang, M.J. Zeitz, H. Wang, et al., Long noncoding RNA-mediated intrachromosomal interactions promote imprinting at the Kcnq1 locus, J. Cell Biol. 204 (2014) 61–75.
[82] J. Wan, M. Huang, H. Zhao, et al., A novel tetranucleotide repeat polymorphism within KCNQ1OT1 confers risk for hepatocellular carcinoma, DNA Cell Biol. 32 (2013) 628–634.
[83] B. Zhou, F.Y. Wei, N. Kanai, et al., Identification of a splicing variant that regulates type 2 diabetes risk factor CDKAL1 level by a coding-independent mechanism in human, Hum. Mol. Genet. (2014). publication online on 2014 Apr 23.
[84] E. Grundberg, E. Meduri, J.K. Sandling, et al., Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements, Am. J. Hum. Genet. 93 (2013) 876–890.
[85] W. Bao, F.B. Hu, S. Rong, et al., Predicting risk of type 2 diabetes mellitus with genetic risk models on the basis of established genome-wide association markers: a systematic review, Am. J. Epidemiol. 178 (2013) 1197–1207.
[86] V. Lyssenko, M. Laakso, Genetic screening for the risk of type 2 diabetes: worthless or valuable?, Diabetes Care 36 (2013) S120–S126
[87] J.L. Vassy, M.F. Hivert, B. Porneala, et al., Polygenic type 2 diabetes prediction at the limit of common variant detection, Diabetes 63 (2014) 2172–2182.