Cardiometabolic diseases represent a common complex disorder with a strong genetic component. Currently, genome-wide association studies (GWAS) have yielded some 755 single-nucleotide polymorphisms (SNPs) encompassing 366 independent loci that may help to decipher the molecular basis of cardiometabolic diseases. Going from a disease SNP to the underlying disease mechanisms is a huge challenge because the associated SNPs rarely disrupt protein function. Many disease SNPs are located in noncoding regions, and therefore attention is now focused on linking genetic SNP variation to effects on gene expression levels. By integrating genetic information with large-scale gene expression data, and with data from epigenetic roadmaps revealing gene regulatory regions, we expect to be able to identify candidate disease genes and the regulatory potential of disease SNPs.

**Genetics of Complex Disease**
Cardiometabolic diseases have become one of the most common conditions of this century, affecting more than one billion people worldwide. In particular, a Western lifestyle, characterized by obesity, leads to an increased susceptibility to diabetes and cardiovascular diseases (CVD), and a better insight into the molecular and genetic etiology of these diseases is urgently needed. Until recently, gene discovery mainly relied on the identification of nonsynonymous mutations showing Mendelian segregation patterns [1–3]. Well-known examples include the low-density lipoprotein (LDL) receptor (LDLR), ABCA1 (ATP binding cassette A1), and PCSK9 (proprotein convertase subtilisin/kexin type 9) gene loci, but many other genes have been found [4,5]. These discoveries depend on mutations in candidate genes showing severe phenotypic effects segregating in families. However, such mutations remain relatively rare, and cannot explain the differences in the susceptibility to cardiometabolic diseases seen in the general population. **Genome-wide association studies** (GWAS, see Glossary) are yielding more comprehensive knowledge of the mechanisms underlying cardiometabolic risk in the general population. GWAS are unbiased and do not make use of a priori knowledge of established pathways and mechanisms. Although it is likely that GWAS will identify the established CVD genes (thereby validating this approach), such studies are equally capable of finding many loci that have not yet been linked to CVD. This review will show how far the genetics of cardiometabolic disease has come, and how we can move forward using genomic methods to prioritize candidate genes and functional variants.

**Breakthroughs That Led to the Golden Age of GWAS**
Before the development of GWAS in 2007, associations between single candidate genes and diseases were difficult to identify and were often plagued with the winner’s curse. Many initial
Box 1. The Genetic Basis of the GWAS Approach

GWAS analysis is based on the concept that genetic variation shows considerable linkage disequilibrium (LD) (Figure I). This implies that a given SNP is tightly correlated to a large number of other SNPs. Such an LD region usually encompasses a small genomic region harboring anything from 0 to 10 more genes, as well as many functional and regulatory units such as enhancers [84]. GWAS is based on testing a single SNP from regions of LD (so-called tag SNPs) to mark the regions in the genome showing disease association. However, such associations cannot distinguish ‘causal’ SNPs from the ‘bystander’ SNPs to which they are closely correlated. The HapMap consortium [7] paved the way for large-scale GWAS by mapping the LD landscape and developing a genome-wide set of tag SNPs. This has greatly simplified the detection of associations with common diseases because only a subset of the millions of SNPs in the human genome need to be tested. Typically, a GWAS analysis involves some 500K–1000K SNPs, thereby interrogating more than 80% of all the common variants known in the human genome [84]. Both common and low-frequency SNPs that are not genotyped directly can be inferred by the process of imputation, which requires adequate reference genomes such as those obtained through the 1000 Genomes Project and the Genome of The Netherlands [65–67]. It is, however, still difficult to identify the disease-predisposing genes and the underlying molecular mechanisms because GWAS identify a series of SNPs associated with the disease, while LD makes it difficult to pinpoint the functional candidate genetic variant.

Figure I. Concept of Linkage Disequilibrium and Tag SNPs. Combinations of SNPs within linkage disequilibrium (LD) blocks that are found in a chromosome and transmitted together define haplotypes. In this figure, a LD block consisting of nine SNPs is part of five frequent haplotypes. Genotyping of three tag SNPs reveals all the haplotypes. Thus, the use of tag SNPs enables efficient GWAS analysis.

associations could not be repeated in later studies and showed gradually evaporating effect sizes during replication studies [6]. With the completion of the human genome sequence, two further breakthroughs laid the basis for the success of the GWAS approach (Box 1). One was the completion of the HapMap project [7], which yielded the information to define the tag SNPs (Box 1) that formed the basis of the technological breakthrough leading to the development of single-nucleotide polymorphism (SNP) array platforms (e.g., Affymetrix, Illumina). With this technology in place, a large-scale application became available to test the hypothesis that common variants explain the phenotypic variation seen in complex traits such as CVD [8]. The Wellcome Trust Case Control Consortium (WTCCC) was established to launch such an effort, and analyzed seven common traits using information from 17 000 subjects [9]. Their results represented the second major breakthrough. Compared to today’s standards, the yield was modest because only 24 SNPs were detected across all diseases \( P < 5 \times 10^{-8} \). However, the study marked a turning point because it demonstrated that robust associations between genetic loci and traits could be found, that the ‘common variant hypothesis’ was true, and that the approach was worth scaling-up. Despite the relatively low effect sizes, most of the loci were indeed replicated in later studies, and in some cases the WTCCC study was able to replicate previous findings.

Glossary

Chromatin conformation capture (3C): a technique used to study 3d structures of chromatin that occur in living cells as a result of DNA–DNA interaction between different chromosomal regions. It involves formaldehyde crosslinking of cells followed by chromatin isolation and digestion with a restriction enzyme. The fragments are then ligated into rings and the crosslinks are reversed. The abundance of these recombinant fragments indicates the interaction frequency and specificity of the two ligated regions.

Epigenetics: the study of heritable changes in phenotype caused by factors other than DNA sequence variation. Research in the field of epigenetics is now uncovering changes in many human diseases, including metabolic diseases. DNA methylation, histone modification, and ncRNA-mediated gene regulation are currently being described as epigenetic regulators.

Expression quantitative trait loci (eQTLs): genetic variants (e.g., SNPs) associated with the expression levels of genes. eQTLs that are linked to the expression levels of nearby genes are referred to as cis-eQTLs and those that are linked to genes that lie further away are termed trans-eQTLs.

Genome-wide association study (GWAS): an assessment of thousands of common genetic variants (SNPs) in different individuals to test if any genetic variant is associated with a particular phenotype. GWAS mainly uses SNPs as genetic markers to test for associations between SNPs and phenotypes, such as seen in human complex diseases.

Mendelian randomization: a genetic study design that takes advantage of the randomization of genetic information to examine the causal relationship between a modifiable exposure and an outcome.

Metabochip: a custom-made genotyping array designed to test some 200 000 SNPs of interest to finely map metabolic traits and CVD-associated loci.

Metabolic quantitative trait loci (mQTL): a genetic variant (e.g., SNP) associated with the metabolite levels.

Noncoding RNAs (ncRNAs): RNA transcripts that do not encode a
Some scientists may be disappointed with the small effect-sizes observed: even when multiple associated SNPs were combined and more of the phenotype was explained [3,10], the variants did not provide sufficient statistical power to better predict the occurrence of disease. While the use of these loci for predictive testing is still not feasible, the genetic findings have already led to novel insights into disease etiology. For example, TCF7L2 (transcription factor 7-like 2; T cell specific, HMG box), a gene with relatively large effect-sizes associated with type 2 diabetes (T2D), was initially predicted to play a role in β cells in humans, but subsequent knockout experiments in mice showed that it is involved in controlling metabolic genes in the liver [11]. Whereas TCF7L2 was remarkable in showing a relative strong effect-size, most GWAS loci have (much) weaker effects. Therefore, a good inventory and subsequent prioritization of the loci for functional analysis are urgently needed.

Identification of Cardiometabolic Disease SNPs

We provide here an overview of the independent loci currently associated with obesity [based on body mass index (BMI), waist-to-hip ratio (WHR), obesity (case/control)], plasma lipids [LDL, high-density lipoprotein (HDL), very low density lipoprotein (VLDL)], intermediate-density lipoprotein (IDL), triglycerides (TG), total cholesterol (TC)], diabetes-related traits [T2D, glucose (GLU), insulin (INS)], homeostatic model assessment β (HOMA-β), and CVD [coronary artery disease (CAD), coronary heart disease (CHD), myocardial infarction (MI), ischemic stroke (IS), carotid intima-media thickness (CIMT), atherosclerotic plaques]. These phenotypes are interrelated, thus we should expect individual disease SNPs to be associated with more than one phenotype. SNPs are considered to be associated when \( P < 5 \times 10^{-8} \), which is the threshold for genome-wide significance. This \( P \) value is very strict, but we should keep in mind that the number of phenotypes and SNPs being investigated is large, requiring robust thresholds to avoid chance findings. Furthermore, reports of genetic associations are generally not accepted for publication in high-ranking journals until they can show appropriate validation in independent replication studies.

Recent studies for obesity [12,13], plasma lipids [14], diabetes [15], and CVD [16] typically analyzed patient cohorts of some 100 000 individuals or more. Moreover, the numbers of SNPs tested were between 300 000 and one million per individual (and >2.5 million after imputation).

Some of these studies made use of the Metabochip [17], a custom genotyping array of approximately 200 000 SNPs which allows fine-mapping and analysis of the GWAS loci associated with metabolic and CVD traits. The Metabochip harbors a set of common SNPs marking the haplotypes, as well as a very large set of rare variants. Such an array will help to identify causal variants, or narrow down the number of potential causal variants, from a pool of loci associated with the trait of interest.

The Gene Map for Cardiometabolic Disease

For this review we gathered cardiometabolic disease SNPs from the Catalog of Published Genome-Wide Association studies, and supplemented these with the results of some recent studies that were not yet included in the catalog. Our information compiled data from 125 manuscripts, and we obtained 755 unique SNPs encompassing 366 independent loci (see Figure 1, and Tables S1 and S2 in the supplementary material online). When plotting the 366 loci to the 22 autosomes and the X-chromosome (Figure 2; for an extended map for all chromosomes see Figure S1), we observed a considerable overlap between the loci associated with obesity, plasma lipids, diabetes-related traits, and CVD (Figure S1). Some 60 loci harbor SNPs that are linked to two or more of the major phenotypes (Figure 3A). Another 38 loci are associated with CVD but without showing association to these risk factors. If we can exclude that these 38 loci are associated with other established risk factors, such as inflammation or blood pressure, then the 38 loci may represent important opportunities to identify entirely novel risk factors. Conversely, many loci are associated with obesity, diabetes, and/or lipid risk factors
for CVD, but are not associated directly with CVD. In these cases, the associations may not be sufficiently strong. This phenomenon has been observed for diabetes-related trait SNPs, showing only moderate effects on CVD risk that often remain unnoticed [18]. It may also be necessary to develop other computational methods that take the genetic complexity of CVD into account.

DEPICT pathway analysis (see the section Network-Based Approaches to Prioritizing Disease Genes in GWAS Loci for details) has identified 27 genes (loci) as shared between two or more traits. Some of the well-known pleiotropic genes detected by DEPICT include both coding and noncoding genes, for example, PCSK9 and apolipoprotein B (APOB) (lipids and CHD), ANRIL (antisense noncoding RNA on the INK4 locus/CDKN2B-AS1) (T2D and CHD), and RPPGRIP1 (retinitis pigmentosa GTPase regulator interacting protein 1) (obesity, lipids, and T2D) (Figure S1). Thus, SNPs can indeed exert pleiotropic effects. Li et al. used a systematic approach to detect 15 pleiotropic associations between lipids and glucose traits [19]. Furthermore, loci containing 56 genes and associated with CAD, BMI, blood pressure, lipids, and T2D were found to be pleiotropic [20]. Loci showing pleiotropic effects are challenging to interpret, as the question arises whether it is a true pleiotropic effect, caused by the same gene through a
Figure 2. Loci on Chromosome 6 Associated with Cardiometabolic Disease. The physical locations of independent loci associated to cardiometabolic traits are represented by circles; each major phenotype is given in a different color. This figure shows a locus on the HLA (human leukocyte antigen) region to be associated with all four major phenotypes. The region encompasses many immune genes such as HLA-C, HLA-B, MICA (MHC class I polypeptide-related A), and MICB. cis-eQTL mapping identified MICB as a candidate casual gene in this region (Table S2 in the supplementary material online) and trans-eQTL mapping identified TMEM154 (transmembrane protein 154) as a trans-affected gene. An extended figure with similar annotation of loci on all chromosomes is shown in Figure S1. Only loci with P values of \(< 5.0 \times 10^{-8}\) are plotted. Genes predicted by eQTL (italic) or DEPICT (bold), or both (italic and bold), are shown as potential candidate genes in the associated loci; the color of the genes corresponds to the phenotype color. However, the names of the genes do not necessarily represent the disease-predisposing genes. CHD, coronary heart disease; T2D, type 2 diabetes.

Identification of Disease-Predisposing Genes from GWAS Loci
One strategy to identify the disease-predisposing genes at risk loci is to carry out large-scale DNA sequence analysis. The underlying idea is that loci harboring common variations associated with diseases could unmask the disease-predisposing genes because these would harbor more rare variants with relatively strong effects. A good example is the ABCA1 gene for Tangier disease, for which mutations have been found in affected families. This is a recessive trait characterized by extremely low HDL levels [23]. However, rare nonsynonymous variants in ABCA1 have also been identified for total cholesterol serum levels [24]. Similarly, other genes in the lipoprotein metabolism pathway show both severe and mild mutations [1], and the latter could only be identified by GWAS. Unfortunately, large-scale sequencing of the GWAS loci for cardiometabolic disease did not lead to the identification of large numbers of novel genes, nor to mutations with strong effect-sizes in the established disease-predisposing genes. This is also true for other diseases [25]. Strikingly, whole-genome sequencing in 9793 early-onset MI cases did identify rare mutations, but only in LDLR and apolipoprotein A-V (APOA5) [26]. Another study that embarked on whole-genome sequencing in families expected to have monogenic inheritance failed to identify mutations in 32 of 41 pedigrees because the large number of rare variants identified precluded a good interpretation of the data [27]. The main issue here was probably of a statistical nature. The detection of rare variants is often limited by the sample sizes, which often lack sufficient statistical power to implicate the variants on the basis of association evidence. DNA variants that are very rare (e.g., seen in a single person) need to be aggregated with other
Shared Loci Reveal the Central Role of Lipid Metabolism in Cardiometabolic Diseases

Figure 3. (A) Loci that overlap with at least two major phenotypes are shown as a heatmap (connected to Table S2 in the supplementary material online). The y-axis shows the chromosomal locations of the shared loci and the x-axis is labeled with the phenotypes. Colored boxes are the shared loci (red for obesity, blue for lipid traits, yellow for T2D-related traits, and green for CVD). (B) A circos plot shows the affected pathways that are common to at least three different phenotypes. The REACTOME pathways were extracted from DEPICT pathway enrichment analysis. Only pathways that were common to three or more phenotypes are shown. Lipid metabolic pathways are in blue to show that these make up the majority of the pathways commonly affected in cardiometabolic disease. (C) A model to describe the central role of lipid traits and the relationship between the different phenotypes. Bold arrows indicate strong connections; the broken arrow indicates a weaker connection. CHD, coronary heart disease; CVD, cardiovascular disease; PPAR, peroxisome proliferator activated receptor; T2D, type 2 diabetes.

rare variants from the same gene to be tested for an association with a phenotype. Furthermore, distinguishing functional from neutral missense mutations will help improve the statistical power issues.

In addition, the functional annotation of GWAS SNPs using methods to detect DNase hypersensitivity sites in the genome can provide information on the regulatory potential of variants and
has revealed that >93% of disease SNPs overlap gene regulatory regions [28]. These observations suggest that it is well worth exploring whether disease SNPs act by affecting gene expression rather than by disrupting protein function.

Expression QTL (eQTL) Mapping for Prioritizing Candidate Disease Genes

Disease SNPs are often associated with altered gene expression levels (expression quantitative trait loci; eQTLs), in other words these loci contribute to variation in the expression levels of mRNAs. Such eQTLs can be detected by testing the association between a risk allele and the mRNA level of transcripts on a population scale. eQTLs are either in cis (cis-eQTL), where the disease SNP is located near the affected gene (e.g., within 1 Mb), or in trans (trans-eQTL), where the SNP is located far away from the affected gene (e.g., more than 5 Mb, or even on a completely different chromosome from the SNP). Previous eQTL studies have shown that ~50% of the disease-associated SNPs affect levels of expression of nearby genes in blood [12,29,30]. Identification of eQTLs for CVD SNPs has already been pivotal to the discovery of sortilin1 [31], a protein involved in cholesterol homeostasis. Of the 755 cardiometabolic SNPs, 40% affect expression levels of genes located within a 250 kb region of the SNPs based on the blood eQTL browser [32], and thereby yielded a list of potential disease-predisposing genes (see Table S2). This number may even be an underestimate because eQTLs are frequently tissue-specific in their effect [33]. In addition, levels of gene expression may also depend on context and might only be detected after induction or at a specific developmental stage [34–36]. Most eQTLs have been defined in easily obtainable blood leukocytes, but there is a need for more exhaustive analyses in a wide variety of tissues. This is currently being addressed by the Genotype–Tissue Expression (GTEx) project [37].

The power of eQTL mapping is that it can also detect disease-predisposing genes located outside regions of linkage disequilibrium. For example, eQTL mapping of the FTO (fat mass and obesity associated) locus, that is associated with obesity, identified IRX3 (iroquais homeobox 3) as a disease-predisposing gene [38]. A considerable amount of data had been collected to prove the role of FTO in obesity – given that the risk SNP was mapped in the FTO gene itself – but eQTL mapping in brain tissue showed the profound effects of the risk SNP on the level of expression of IRX3, which is a gene located 1 Mb away. This suggests a long-range interaction between the risk SNP and IRX3 [38]. After deleting IRX3 in a rodent model, the body weight of Iox3-deficient mice dropped by 25–30%, providing functional evidence for IRX3 as the obesity-predisposing gene in the FTO locus [38].

In addition to mapping cis-eQTLs, identifying trans-eQTLs will help to reveal downstream pathways affected by the disease-associated SNPs. In a locus on 11q12.2, that is associated with metabolic syndrome [39], levels of human metabolites [40], T2D [41], and cardiac conduction and rhythm disorder [42], TMEM258 (transmembrane protein 258), FADS1 (fatty acid desaturase 1), and FADS2 were identified as disease-predisposing genes in cis, but were also significantly associated with the expression of LDLR in trans [32,43]. LDLR encodes the LDL receptor and contains common variants that are also associated with lipid levels [10], highlighting the well-established role of lipid metabolism pathways in cardiometabolic diseases. After systematically extracting trans-eQTLs for all 755 cardiometabolic disease-SNPs using the blood eQTL browser [32], nearly 15% of the SNPs were found to affect gene expression in trans (Table S2). Interestingly, a locus on chromosome 12q24.12, associated with lipid traits and CHD, affects the expression levels of SH2B3 (SH2B adaptor protein 3) and ALDH2 (aldehyde dehydrogenase 2 family, mitochondrial) in cis, while the expression level of STAT1 is affected in trans (Table S2). In addition, the locus on chromosome 16p11.2, associated with obesity-related traits, also affects STAT1 expression in trans (Table S2). STAT1 encodes ‘signal transducer and activator of transcription 1’, a regulator of type 1 interferon signaling. These SNPs are associated with three different phenotypes and the convergence of their regulatory
effects on the innate immune signaling strongly suggests they play a key role in the inflammatory component in cardiometabolic diseases. Therefore, identifying such trans-eQTLs can provide new biological insights into common pathways involved in the pathogenesis of cardiometabolic diseases.

One of the limitations of early eQTL studies is that they were based on microarrays, and only contained protein-coding genes, thus completely ignoring 65% of the annotated human genome transcribed into ncRNAs [44]. It has become clear that ncRNAs (both miRNA and long ncRNAs) are involved in many biological processes, mainly as regulators of gene expression. It has been shown that some of the cardiometabolic disease-associated SNPs physically overlap with ncRNAs [45,46] and that other disease-associated SNPs, including CHD and T2D SNPs, can also affect the expression of ncRNAs [47]. The large reduction in sequencing costs means that RNA sequencing (RNA-seq) is now quickly replacing microarrays as a means to assess genome-wide transcript abundance. Sequence data offer a number of advantages over microarray data: first, RNA-seq provides quantification of the global transcriptome at a high resolution. Second, it also captures efficiently all transcripts, including the less abundant ncRNAs. Third, RNA-sequencing allows allele-specific and transcript isoform-specific expression analyses, which can then be correlated to genetic variants [48] to identify disease SNPs that affect allele-specific [49] or transcript isoform-specific expression [50]. The number of publicly-available RNA-seq samples is increasing exponentially, which means that future eQTL studies should be able to employ this rich resource to study the effects of disease SNPs in the tissues relevant to a particular disease and to aid translation of disease associations to function.

Importantly, genetic data can be linked to human metabolite profiles, leading to the discovery of metabolic quantitative trait loci (mQTLs); the most comprehensive study to date reported 145 mQTLs [40]. Of these, 14 loci overlap with the genetic map we present here (ABO, ANGPTL3, APOA5, CETP, FUT2, GCKR, LACTB, LIPC, LIPG, NAT2, PDXDC1, SH2B3, SPTLC3, and FAD genes), providing new information on potential mechanisms. Integrating the eQTLs and mQTLs is expected to greatly facilitate further gene discovery.

Network-Based Approaches to Prioritizing Disease Genes in GWAS Loci

Current eQTL studies are mainly using expression data from hematological samples, whereas tissues more relevant to cardiometabolic disorders, such as arterial smooth muscle cells, vascular smooth muscle cells (SMCs), or foam cells, have not yet been queried to identify eQTLs. Indeed, studies have shown that SMCs transdifferentiate to foam cells, suggesting that they are likely to be more crucial in human atherosclerotic lesions than are macrophages [51]. Notably, key metabolic tissues, such as liver, muscle, adipose tissue, and β cells, as well as brain, also need to be included in eQTL analyses. It should be evident that current eQTL studies are still unable to prioritize disease-predisposing genes that affect only specific cell types, or in a particular tissue context. Given that no significant eQTL effect in blood [43] can be found for nearly 60% of cardiometabolic disease-associated SNPs, we need more annotation strategies. Such tools are being developed based on the potential functional relevance of the genes located in GWAS-associated loci [52]. The general approach for such gene prioritization methods has been to systematically search for commonalities in functional annotations between genes from different associated loci, derived either from text mining [53] or based on protein–protein interaction evidence [54]. These methods have helped to prioritize new disease-predisposing genes and pathways, especially for immune-mediated diseases [55,30]. However, at present these methods still suffer from the incomplete annotation of genes and pathways, and the results are skewed towards well-studied genes. The coexpression-based method “DEPICT” [56] (see ‘The Gene Map for Cardiometabolic Disease’ section), which uses the large amount of publicly-available gene expression data and predicts functional connections between genes, can help to prioritize disease-predisposing genes in an unbiased manner. Applying DEPICT to our data
helped to prioritize genes for 617 of 755 cardiometabolic SNPs (354 with a significant nominal P value; Table S2), including the 300 SNPs that failed in the eQTL analysis. Identifying the correct disease-predisposing gene in every locus is important because these genes will serve as core gene sets to reveal interconnected biological networks. For example, by performing pathway enrichment analysis using genes prioritized by DEPICT for each of the four phenotypic groups (specified in Figure 1), we saw a significant enrichment for genes belonging to lipid metabolic pathways that were common to at least three different phenotypes (Figure S3B, Key Figure). Thus, one could think of a model where there is a strong connection between lipid traits with obesity, diabetes-related traits, and CVD, although the genetic connection between diabetes-related traits and CVD is much weaker (Figure 3C). It was not surprising to find lipid metabolism as a key pathway that connects these four major phenotypes because lipid trait-associated loci are the most commonly shared regions between cardiometabolic phenotypes, whereas there is less overlap between loci associated with diabetes-related traits and CVD (Figure 3A).

Identifying such patterns helps to pinpoint the key mediators that are influenced by disease-associated variation. With the assumption that SNP effects on gene expression mediate phenotype variation, it was interesting that a recent study by the Framingham Heart Study integrated eQTLs from more than 5000 individuals with extensive phenotype data, including blood lipid levels, glucose levels, metabolites, and BMI. This enabled the investigators to identify crucial functional networks involved in CVD [43]. Their study provided important insights, but took genes prioritized by eQTLs from only 40% of the loci. It would probably be more informative if we can apply computation-based gene prioritization as a complementary method to eQTLs to pinpoint potentially disease-predisposing genes in most disease loci, and then to perform systems genetics analysis to integrate the multidimensional datasets.

Concluding Remarks and Future Perspectives
Genetic studies have gained an enormous momentum, of which we have only seen the beginning so far. Combining epidemiological findings with genetic data led to the design of Mendelian randomization studies, which were instrumental in changing the concept that raised HDL levels protect against CVD [57]. Missing heritability still remains an issue, and there are several other questions (see Outstanding Questions). However, the delineation of shared phenotypes enabled by GWAS is of great interest, such as the recent insight into the relationship between height and CAD [58]. Other remarkable findings using large-scale sequencing include the identification of inactivating mutations in the NPC1L1 (Niemann–Pick disease C1-like 1) gene that protect against CHD because they help to reduce plasma LDL levels [59]. In general, whole-exome and whole-genome sequencing are useful for identifying mutations in genes that have already been implicated in CVD. However, we need to perform gene expression analyses to understand the mechanisms behind the associations identified in GWAS studies because more than 90% of SNPs are expected to have a regulatory role. In this respect, it is important to study a wide range of cell types, as is now being facilitated by GTEx [37]. Notably, for the FTO locus, the major breakthroughs came when it was discovered that a brain-eQTL for IRX3 causes obesity [38], and with the more recent finding that both IRX3 and the nearby IRX5 regulate adipocyte thermogenesis to control obesity [60]. In addition, gene annotation tools such as DEPICT can be instrumental in further narrowing down the number of candidate genes, as shown by Ghosh et al. [61]. Lastly, once potential disease-causing candidate genes have been defined, robust evidence for their role in CVD is needed. This can be achieved using rodent models that can be efficiently manipulated with CRISPR/Cas9 technology to obtain cell type-specific knock-outs as well as specific gene mutations [62]. The CRISPR/Cas9 system also allows studies in human cells and in the extremely promising organs-on-chips approach [63]. Such studies, coupled to analysis methods aimed at understanding the regulome, will provide exciting insights into the regulatory circuits perturbed by genetic disease variation and may well lead on to new therapeutic options for CVD.
References

19. Li, N. et al. (2014) Pleiotropic effects of lipid genes on plasma glucose, HDL1c and HDMA-IR levels. Diabetes 63, 1–47
29. Fehrmann, R.S.N. et al. (2011) Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. PLoS Genet. 7, e1002197