A large number of cardiometabolic traits with a complex but heritable component have been subjected to modern genomic techniques in an effort to understand their architecture. These analyses include traits that might be considered physiological and are our main focus here, such as heart rate, electrocardiogram (ECG) parameters, blood pressure, and those that are directly related to cardiovascular pathology such as coronary artery disease. This wealth of information has been only modestly exploited to date. In this review, we discuss the general issues that are raised in such studies and describe resources that will facilitate their investigation. Furthermore, we give exemplars illustrating specific physiological problems illuminated by the approach; however, we contend that the majority of the genetic data sets have been little utilized for physiological understanding.

Design of Genome-wide Association Studies

Before examining some specific studies, it is important to understand the general design of genetic association studies and what an association means. Throughout the human genome there is a degree of genetic variation: most commonly thought of as variations in a single base (single nucleotide polymorphisms, SNPs) but also includes insertions, deletions, and variations in copies of repeating sequence. There are currently over 78 million SNPs reported (Box 1), and these may be common or rare, present in introns or exons and in the latter case may affect the protein coding sequence or not (nonsynonymous/synonymous). Genetic variants occur on average every 1 Kb across the genome, and at some loci these alleles tend to cosegregate together in blocks of genomic DNA (haplotype blocks); this is a process known as linkage disequilibrium (LD). This is advantageous for mapping traits to a chromosomal location as a single SNP (tagSNP) can be used to impute the identity of other SNPs across a larger genomic region. TagSNPs are included on high-density SNP arrays, and these are used for genotyping of the whole genome. The genetic background of a trait can be analyzed by performing a genetic association study and these analyses can be done for either binary (case-control) or continuous traits.

Typically a genome-wide association study (GWAS) is conducted on hundreds to thousands of unrelated individuals, and these individuals are genotyped using commercial arrays with up to two million SNPs (mostly common SNPs with minor allele frequencies >5%). Then for each SNP (genotyped or imputed) the allele (a version of the variant) present is tested for association with the trait of interest. The SNP associations that are significant or suggestive after a multiple testing correction (genome-wide significance is usually defined as \( P < 5 \times 10^{-8} \)) are then usually followed up in independent samples to confirm these findings.

It is worth noting that prior to 2005 and the publication of the first GWAS, investigators generally chose a candidate gene based on known physiology for example and genotyped one or more SNPs within it and then tested for association with the trait of interest. If SNPs were in the promoter or coding region and nonsynonymous, then this was combined with an appropriate functional correlate.

The GWAS approach identifies a genomic region, and within this region there may be several genes, all being considered initially as a potential “candidate” gene. Furthermore, the most statistically significant associated SNP may be in high LD (normally defined by \( r^2 > 0.8 \)) with a large number of other SNPs, and any of these may be the actual variant that leads to...
the biological effect. The individual contribution of genetic variants found by GWAS is actually quite small, and this is a feature of many cardiometabolic traits.

**Missing Heritability**

A large number of loci have been identified for cardiometabolic traits using GWAS; however, the identified loci explain very little of the heritability of the trait. For example blood pressure has a heritability of ~30%. The loci discovered thus far collectively explain only a small proportion of the heritability (~2–3%), and the associated variants are mostly common with small effect sizes (mostly <1 mmHg systolic blood pressure and <0.5 mmHg diastolic blood pressure) (8). In contrast the heritability of Type 2 diabetes is estimated to be 64%; here the loci are estimated to explain ~10% of the heritability (47). The explanation for missing heritability for complex traits is a much-debated issue with more recent discussions suggesting that the heritability for some traits may be overestimated (43, 74). One approach to try and find missing heritability is the study of low-frequency and rare genetic variants (allele frequencies <1%) as these may have larger effect sizes. There are bespoke genotyping arrays (for example the exome chip) whose content is primarily low-frequency and rare nonsynonymous SNPs discovered by sequencing of 12,000 whole exomes or whole genomes (http://genome.sph.umich.edu/wiki/Exome_Chip_Design). These arrays have had some success in discovering new loci and have identified low-frequency variants with larger effect sizes at known GWAS loci (41). Resequencing studies of GWAS loci across cases and controls and exome sequencing of individuals with extreme phenotypes have also indicated there are some rare genetic variants with large effect sizes at GWAS loci, e.g., LDLR and APOA5 for LDL-C and ABCA1 for HDL-C (16, 61).

These studies have marginally increased the percentage of heritability explained for some traits. Missing heritability may also be explained by contributions from structural variation and epigenetic effects, and these are hypotheses that are currently under exploration.

**Annotation of GWAS Loci for Identification of Candidate Genes and Causal SNPs**

Following a GWAS, fine mapping can be deployed to refine the size of a large association interval and narrow the list of potentially causal SNPs. This is usually achieved by genotyping additional SNPs at a locus using bespoke genotyping arrays with fine mapping content such as the CardioMetabochip (71), imputing existing datasets to a 1000 Genomes reference panel (Box 1) or genotyping individuals of non-European ancestry. The LD patterns vary according to ancestry, and individuals of African ancestry have higher diversity and lower LD, permitting refinement of association signals for some loci, and genotyping of exonic variants that have functional annotation (Box 1) (7). To determine candidate genes at a GWAS locus you can generate a list of SNPs in LD (r² > 0.5) with the most associated SNP and then map these SNPs to genomic location and corresponding genes. Genes that map up to 500 kb either side of the most associated SNP should also be considered. A literature review of these genes may provide some information on the known function. To delineate the SNP(s) you would take forward into functional testing you can check if the most associated SNP is a nonsynonymous SNP or if it is in high LD with one. If this is the case, it can be postulated that the nonsynonymous SNP might affect protein function in a modest fashion, and an experimental plan can be devised to test this hypothesis. Researchers can also assess if the most associated SNP or a SNP in high LD is associated with expression levels of nearby gene(s) by performing a cis-expression quantitative trait locus (eQTL) analysis. There are online bioinformatics resources for performing these preliminary analyses in a range of different tissue and cell types (Box 1). Currently the numbers of samples from cardiovascular relevant tissues is limited. The number of GWAS SNPs that are annotated as being nonsynonymous is small, and the majority (93%) are located in introns and intergenic regions (44). It is likely that several of these variants are involved in the regulation of transcriptional activity. The ENCODE and Roadmap Epigenomics projects have revolutionized our understanding of the regulatory landscape, providing detailed information on promoter and enhancer DNA sequences across the human genome across hundreds of different cell types and tissues (Box 1). We can utilize a range of bioinformatics tools to check if GWAS SNPs or their LD proxies are located in regulatory sequences in different tissues and if the SNP(s) are predicted to affect transcription binding sites (Box 1). These online tools provide a detailed annotation of all currently known SNPs at a GWAS locus. We also need to consider if SNP(s) found in enhancer sequences are interacting with a promoter of a nearby gene or if there are long-range chromosomal interactions? The experimental methods of 4C-Seq and HiC can be utilized for this purpose (13).

Figure 1 illustrates examples of follow-up analyses for a heart rate locus on chromosome 3 (14).

**Some Case Studies in Cardiometabolic Traits Illustrating Some of the Issues in Interpretation**

As detailed above the mapping and identification of a gene from a GWAS is a challenging process and is much less certain than in hereditary monogenic disorders. The commonest justification given for the substantial investment in GWAS is identifying new genes and new biology, and processes that might lead to new targets for therapeutic innovation. How well does this stack up to closer scrutiny? We consider three potential scenarios.

**Pointing to new physiology.** Circulating blood lipid levels are important risk factors for cardiovascular disease, and the GWAS approach has yielded numerous variants associated with lipid traits and coronary artery disease (34). A recent study has reported genotyping of the exome array (Box 1) with lipid traits (24). This analysis yielded genome-wide significant associations with coding variants at 10 known lipid loci, none of which are in genes with recognized roles in lipid metabolism. Of particular interest to the wider community interested in fine mapping of GWAS loci was the observation of significant association with rs58542926, a Glu167Lys variant located in 7M6SF2. This variant is in high LD with the previously reported GWAS variant in the region. There are 21 candidate genes at this locus; further analyses of 7M6SF2 indicated high levels of expression in the liver, and “in vivo” experiments of transient overexpression of 7M6SF2 using a liver-specific adenovirus in mice led to increased total cholesterol levels and triglycerides in the mice, and transient knock-


Heart Rate GWAS locus on chromosome 3

Fig. 1. The pathway from genome-wide association study (GWAS) signal to functional analyses. A regional association plot of a GWAS locus for heart rate is shown (12). The region extending to within 500 Kb of rs9647379 [the most associated single nucleotide polymorphism (SNP)] is shown. The statistical significance of SNPs in the region are illustrated on the $-\log_{10}(P)$ scale as a function of chromosomal position (NCBI Build 37). The most associated SNP is indicated with the purple diamond, and the combined discovery and replication $P$ value is shown. The correlation of rs9647379 to other SNPs at the locus is shown on a scale from minimal (gray to blue) to maximal (red). The fine-scale recombination rate is indicated in aqua blue. The genes in the region are indicated along the x-axis. This locus has 4 candidate genes: fibronectin type III domain 3B (FNDC3B), growth hormone secretagogue receptor (GHSR), placenta specific 1-like (TMEM122), and phospholipase D1, phosphatidylinositol-specific (PLD1). Bioinformatics analysis indicates PLD1 and FNDC3B are expressed in heart tissue; there are no nonsynonymous (ns)SNPs in high linkage disequilibrium (LD) and no evidence from public databases of cis-express quantitative trait loci (eQTLs). Functional experimentation in Drosophila melanogaster and Danio rerio using RNA interference (RNAi) to knock down the expression levels of candidate genes where the orthologs were available found lower resting heart rates in D. melanogaster pupae for GHSR and PLD1 and reduced heart rates and unlooped hearts in some of the D. rerio embryos for PLD1 (12). The functional analyses indicate GHSR and PLD1 as 2 candidate genes for further studies.

down of TM6SF2 led to protein levels being significantly decreased in the liver. These data are concordant with carriers of the lysine amino acid at position 167 in TM6SF2 having decreased lipid levels. The variant was also shown to be associated with a decreased risk of myocardial infarction. The function of TM6SF2 (transmembrane 6 superfamily member 2) was previously not known; these analyses pinpoint this as the likely candidate gene at this GWAS locus, paving the way for further functional analyses.

The long QT syndrome refers to a characteristic ECG abnormality that is associated with torsade de pointes, a form of ventricular tachycardia, and sudden death. There are a number of hereditary monogenic diseases in which potassium channels ($KCNQ1$, $KCNH2$), the cardiac sodium channel ($SCN5A$), and interacting proteins are implicated as the cause (60). The QT interval is also heritable and this can be analyzed in GWAS (2, 3, 56, 74). Reassuringly, a number of genes such as $SCN5A$ and $KCNQ1$ overlapped between the monogenic diseases and the GWAS studies. However, one of the strongest associations was with a locus that included the nicotinic oxide synthase 1 adaptor protein ($NOS1AP$) gene. This prompted an investigation into $NOS1AP$ function in ventricular cardiomyocytes (9). The protein did indeed interact with NOS1 in the heart and overexpression inhibited L-type Ca$^{2+}$ channel function. In a further twist, polymorphisms in $NOS1AP$ in monogenic long QT syndromes may act as disease-modifying genes (11) and could underlie some of the predisposition to the more common drug-induced long QT syndrome (27). Finally, subsequent studies have identified the causative SNP as being within an enhancer element that affects $NOS1AP$ expression. Overexpression of $NOS1AP$ led to a shortened action potential duration and an increased conduction velocity in neonatal rat ventricular myocytes. Furthermore, $NOS1AP$ localized to the intercalated disc and a panel of genes expressed at this site were enriched and associated with the QT interval (31).

Uromodulin is a protein secreted in the urine from the kidney and was originally identified as the Tamm-Horsfall protein (54). Recently, GWAS have discovered variants in the UMOD gene to be associated with renal function, chronic kidney disease, and hypertension (36, 37, 53). There was some prior work suggesting UMOD as a candidate gene for hypertension, but there were limited genetic analyses (26, 69). Uromodulin is expressed solely in the thick ascending limb of the loop of Henle and is inserted as a glycosylphosphatidylinositol anchored protein in the apical membrane of the tubular epithelial cells. It is cleaved and released into the urine from there (52). This has prompted studies examining the interaction of uromodulin with key transport proteins in the thick ascending limb namely the Na$^+$/K$^+$/2Cl$^-$ cotransporter and the renal outer medullary potassium channel (ROM2). Experimental work revealed that uromodulin promoted the surface expression of both proteins and loss of function would potentially
promote salt loss, while increased expression might lead to salt retention and hypertension (48, 57). This hypothesis was supported by subsequent work showing common variants in the promoter of the *UMOD* gene led to increased expression of the protein and that overexpression of the protein in mice led to hypertension (70). There are no direct studies in humans indicating increased blood pressure. However, in a small observational study of “never treated” hypertensive individuals, those with the “at risk” genotype had higher blood pressures, and in a small subset of these that had treatment with a loop diuretic that blocks Na\(^+/\)K\(^+/\)2Cl\(^-\) cotransporter there was a greater decrease in blood pressure in individuals with the at-risk genotype \((n = 118)\) compared with individuals heterozygous or homozygous for the other genotypes \((n = 47)\).

**Contribution to known physiological pathways.** Blood pressure homeostasis is maintained by several systems including the central and sympathetic nervous systems, the vasculature, and the kidney, together with various hormonal regulators. Recent large meta-analyses of GWAS have led to the discovery of over 63 loci influencing blood pressure (8). A review of the SNP associations finds many that are in or near genes that are known to affect blood pressure based on prior functional and physiological experiments. GWAS have indicated several SNP associations in vasodilator and vasoconstrictor stimulators of vascular smooth muscle cells (Fig. 2). These include a SNP association upstream of endothelial nitric oxide synthase (30, 59), three SNP associations near the natriuretic peptides A and B on chromosome 1 (50, 68), a SNP near the guanylate cyclase \(\alpha\)- and \(\beta\)-subunits, and two SNPs near the C-type natriuretic peptide receptor (17, 33). In the vasoconstrictor pathways there was a SNP association in intron 1 of the angiotensinogen gene, a precursor of angiotensin II (28 –30), and a SNP association in the adrenergic beta 1 receptor, a known target for beta-blockers (33).

There has been limited functional work understanding the mechanism of GWAS SNP associations. The rs5068 (A/G) variant is associated with blood pressure and is located in the 3'-untranslated region close to the *NPPA* gene that encodes the atrial natriuretic peptide. Carriers of the G allele have increased atrial natriuretic peptide levels and a decreased risk of hypertension (51). In a recent study, the response of different genotypes to an intravenous salt challenge was compared after a week-long low- or high-salt diet (4). In both cases atrial natriuretic peptide levels were significantly higher in AG individuals before, during, and after the salt challenge. Furthermore, the authors showed that the microRNA miR-425 could influence the expression of luciferase reporter constructs containing the AA but not AG allele. Finally, the transfection of miR-425 into cardiomyocytes obtained from induced pluripotent stem cells (iPSCs) reduced *NPPA* mRNA levels and propeptide release of the precursor of atrial natriuretic peptide (4). These analyses make a compelling case for the rs5068

![Key regulators of vascular smooth muscle tone](image)

**Fig. 2.** Key mechanisms regulating vascular smooth muscle tone. Signaling events that promote vasoconstriction are indicated in red, and those that promote vasodilation are indicated in blue. The phosphophorylation (P) of myosin is a critical step in contraction of vascular smooth muscle. Vasoconstrictors such as angiotensin II and norepinephrine promote increased cystolic calcium concentration (Ca\(^2+\)); this in turn activates myosin kinase leading to myosin-actin interactions leading to vasoconstriction of the vessel. SNP associations are observed in the genes indicated in the red text (*AGT*, angiotensinogen; *ADRB1*, beta 1 adrenergic receptor). Vasodilators such as atrial natriuretic peptide (ANP) and nitric oxide via cyclic guanosine monophosphate (cGMP) activate myosin phosphatase, which leads to dephosphorylation of myosin and relaxation. Further SNP associations have been found in this pathway, also indicated in red text (endothelial nitric oxide synthase (*eNOS*); ANP, guanylate cyclase \(\alpha\)- and \(\beta\)-subunits (*GUCYA* and *GUCY B*, respectively)). Genetic associations are also observed in the natriuretic peptide receptor-C (*NPR-C*) gene that, via a separate pathway, leads to vasodilation in vascular smooth muscle. Other abbreviations: AT1-R, angiotensin 1 receptor; CNP, c type natriuretic peptide; ANP-R, atrial natriuretic peptide receptor. Figure is adapted from Ref. 39.
variant having differential effects on atrial natriuretic peptide via microRNA regulation. However, in the relatively small numbers of healthy young volunteers they studied they were not able to show a significant effect of the variant on blood pressure. It will be important to show robust changes in blood pressure according to genotype, as well as association between the genetic marker and BNP levels to move the field forward.

A more complex scenario. One of the first GWAS in 2007 reported association between SNPs at the FTO (fat mass and obesity associated) locus and obesity (21). Individuals with two copies of the risk allele had a 1.67-fold higher rate of obesity compared with individuals with no copies. Follow-up studies of FTO gene included the development of murine models to investigate the function of fto; mice overexpressing fto were found to be overweight, whereas null allele fto mice were lean (10, 19). These experiments suggested FTO as the likely candidate gene in the associated interval. The most significantly associated SNP (rs9939609) is located in intron 1 of the FTO gene. An eQTL analysis in a range of different tissues including adipose tissue, however, found no association between FTO SNPs and FTO expression levels (23, 40). This lack of association between the associated variant and gene expression has perplexed researchers searching for causal mechanism(s) based on the human GWAS results. The FTO locus is a 47 Kb region of high LD, and the obesity-associated SNPs are located in introns 1 and 2 of the FTO gene. Recently, Smemo et al. (2014) in a series of elegant experiments demonstrated that DNA sequence containing the FTO variants is in contact with the promoter sequence of IRX3, a gene located 0.5 Mb away (64). The IRX3 gene encodes a transcription factor, and currently very little is known about its function. This physical DNA interaction occurs in humans, mice, and zebrafish, and the authors showed an association between FTO genetic variants and expression levels of IRX3 in the brain. These analyses and further follow-up studies in murine models indicate that the associated variants at the FTO locus might be influencing body mass via IRX3 and not FTO.

Experimental Resources to Follow up GWAS Findings. We have proposed that the commonest scenario is for a GWAS study to point to one or more genes whose function will need to be explored, though the specifics of this might be more complex than anticipated. In this section we highlight some recent initiatives that might facilitate this endeavor.

Murine resources. One key strategy is the use of animal models, in particular the mouse, rat, and zebrafish. Importantly the genomes of each of these can be readily manipulated and large-scale international mutagenesis projects are being undertaken (Box 1). The International Mouse Phenotyping Consortium aims to generate and phenotype knockout mouse models for every gene in the genome. Each of these lines will be subjected to a broad phenotyping pipeline that includes at some centers measurement of the ECG and cardiac function by echocardiography. In an interesting early analysis of 250 lines passing through one of these pipelines, a large number of mice had some phenotype, and this was often unexpected (72). Genes without an obvious paralog (i.e., a gene that was duplicated and homologous in the same species) were more likely to be essential. Secondly, previously unstudied genes were as likely to give a phenotype as those with a prior literature. As the authors point out there was a large bias to investigation of known genes, and a majority of genes in the human genome remain completely unstudied. Furthermore, if a homozygotic knockout was lethal then heterozygotes were studied, and a potential phenotype in these was surprisingly common, suggesting haploinsufficiency is important, i.e., loss of expression of the protein from a single one of the two alleles was enough to lead to functional deficits. In many of these mice the targeting strategy enables the generation of an allele in which a critical exon is flanked by loxP sites suitable for conditional deletion in a particular tissue and/or time after interbreeding with cre recombinase expressing murine lines (63). Knockout mouse models permit the direct testing of a candidate gene for a functional role in the disease without the requirement of understanding the role of the functional polymorphisms levered from genetic studies. Mouse models overexpressing a gene of interest and models with inclusion of nonsynonymous variants that are present at some GWAS loci will also be important going forward.

In addition to the increasing availability of genetically modified mice, in vivo phenotyping capabilities have significantly advanced, and these can complement traditional ex vivo, biochemical, and single-cell studies. Many of the traits such as blood pressure represent integrative physiological processes dependent on pathways intrinsic to vascular smooth muscle and endothelium but also extrinsic events for example cardiovascular control centers in the brain (65). In particular microfabrication techniques and bespoke solutions for small mammals have allowed the real-time monitoring of ECG and blood pressure with radiotelemetry systems in conscious living animals, complex imaging using echocardiography and MRI, and measurement of in vivo electrophysiological and cardiac function (e.g., pressure-volume loops) parameters using catheters and arrays. Murine models of cardiovascular (patho)physiology have contributed greatly to the field although there are species differences. For example, the murine cardiac action potential is much shorter than that of large mammals and is determined by a different set of potassium currents (12).

Rat resources. The rat is a long-standing model system for studying cardiovascular disease (hypertension and stroke) and metabolic traits (diabetes). Monitoring the physiology of disease is thought to be easier, and for some traits (e.g., diabetes) the rat model is thought to resemble the human disease better (1). The Rat Genome Database (RGD: Box 1) provides a central resource for information on the different rat strains and includes genetic, genomic, and phenotypic data on the different models and bioinformatics tools (62). Until recently, the production of rat models with specific mutations or the generation of knockout or conditional strains was technically challenging because of issues with embryonic stem cell manipulation. These issues are now resolved, and as of Sept 2014 there were 2,998 rat strains in the RGD, and 1,977 of these were created because of issues with embryonic stem cell manipulation. These issues are now resolved, and as of Sept 2014 there were 2,998 rat strains in the RGD, and 1,977 of these were created by genome-modifying techniques. Rat models have been used to ascertain likely candidate genes at GWAS loci. Using zinc finger nuclease mutagenesis-damaging alleles were introduced into six hypertension candidate genes (agtr1, mthfr, clcn6, anp, bnp, and plod1) in the SS rat model (20). The MTHFR-NPPB locus has multiple SNPs associated with blood pressure traits (17, 50, 51). Analysis of blood pressure and renal phenotypes in each of these rat models implicated five genes to be associated with hypertension, indicating complex interactions of alleles and genes at this locus. The gene-editing approach has been also been used to assess other candidate

Physiol Genomics • doi:10.1152/physiolgenomics.00004.2015 • www.physiolgenomics.org
genes at established GWAS loci (e.g., SH2B3, PLEKHA7) (18, 58) and NR2F2, a candidate from GWAS and linkage analyses; these studies provided supportive data for each of the genes tested (38).

**Zebrafish (Danio rerio).** The argument has been made that zebrafish cardiovascular physiology is surprisingly close to that of mammals and man (15). The small size, ease of breeding, and conservation of functional domains and proteins lend themselves to genomic and chemical screens. In relation to GWAS, the zebrafish may be particularly useful when a large number of candidate genes need to be screened at a particular locus as was recently demonstrated in the study of genes implicated in a GWAS in heart rate (14). Knockdown of genes in the zebrafish is readily achievable using morpholino knockdown, but the investigation of cardiovascular function is quite specialized. However, there are some differences: the zebrafish does not have a secondary heart field and the heart has a high regenerative capacity. Furthermore, there are efforts using chemical mutagenesis to systematically disrupt every zebrafish gene as part of the Zebrafish Mutation Project (Box 1) (35).

**Human iPSCs and disease in a dish.** Animal models are much used, but direct approaches in humans and associated tissues and cells would circumvent species differences. Another technological advance is the use of human iPSCs. Adult cells, usually fibroblasts, obtained from a skin biopsy from a patient are reprogrammed to iPSCs by expressing a number of transcription factors (66). These can then be reprogrammed to a variety of tissues including cardiac myocytes (iPSC-CMs) and smooth muscle and then used as a cellular model for the relevant pathology (25). This is intellectually appealing and potentially overcomes some of the limitations of animal models. For example, it has been used by a number of groups to phenocopy the monogenic long QT syndromes. Mutations in KCNQ1 in LQT1 lead to action potential prolongation in the iPSC-CMs, and the channel complex fails to traffic to the cell membrane as we found in heterologous expression systems (46, 73). In principle such approaches could be used to investigate cardiovascular complex traits. However, the effect sizes of SNPs are much smaller than in monogenic diseases, and the effects are likely to be quite subtle. Secondly during the generation of iPSCs the cells accumulate genetic changes, particularly copy number variations (55). Finally, the iPSC-CMs have a mixed electrophysiological signature with pacemaker, atrial, and ventricular-like action potential phenotypes with a relatively depolarized resting membrane potential (6, 46, 67). The iPSC-CMs do not have a typical adult ventricular morphology and lack well developed t-tubular networks.

**Genome editing.** Recent advances in genome editing may have wide-ranging ramifications. The initial tools, zinc finger nucleases and transcription activator-like effector nucleases, were labor intensive to construct and required considerable optimization. However, the advent of CRISPR/Cas9 technology has meant the approach is technically simpler. The comparisons with other approaches and details are given in several recent articles (22, 42), and there are tools for designing appropriate strategies (see Box 1 and Ref. 45). Bespoke mouse models that mimic human mutations, for example, nonsynonymous, splice site and at transcription factor binding sites are all now technically possible and at a reduced cost compared with manipulating embryonic stem cells. There is also progress in developing the technology for editing of promoter and enhancers and introducing risk haplotypes; this will permit testing of SNPs from GWAS studies and the creation of potentially more relevant animal models. Another consequence is that it is possible to engineer genomes other than those of the traditional model organisms. This may open the door to large mammal models of cardiac disease overcoming the limitations of rodent models. Secondly, in the initial disease modeling using iPSC approaches the controls were family members without the disease. The problem is not only does the reprogramming itself introduce genetic changes but the individuals, even if siblings, are likely to have a number of genetic differences. Genome editing allows true isogenic controls by correcting for the causative mutation and is also a powerful proof of causation. This has been demonstrated in LQT2 (5). Furthermore, you needn’t necessarily use iPSCs but could genome-edit human embryonic stem cells instead to generate the relevant models.

**Studies in humans.** The above approaches are potentially highly revealing, but they do not actually directly address the link in humans between the genomic variation and the physiological trait. We have already mentioned that one of the major mechanisms is likely to be association of the causative SNP with small variations in expression through changes in regulatory networks that are significant at a population level. This can be approached with bioinformatics databases and tools (Box 1), novel tissue collections, and subsequent eQTL analyses (see above) (4, 32). An especially useful resource in the future is likely to be the UK Biobank. This is an open resource in which 500,000 individuals aged 40–60 yr of age have a core assessment measuring a number of physiological and clinical parameters. Additionally, the individuals have samples collected from them for archiving including genotyping, and their health will be followed prospectively. Furthermore, various substudies have been completed or are underway in groups of 100,000 with more specialized investigations such as exercise testing, MRI, etc. A second initiative is the electronic mining of medical records, and this can generate very large epidemiological cohorts for understanding the progression of disease. The Farr Institute interrogating medical records within the British National Health Service is an example of such an enabling resource (Box 1).

**Conclusion**

Physiology has underpinned the rational development of pharmacological agents. The pharmaceutical industry is experiencing difficulties in drug development not only in finding new targets but also in the late failure of agents in phase III clinical trials. Despite the complexity in teasing out the causative gene and mechanism from GWAS we feel that modern genomics may offer another route to target identification and validation. It is telling that a large number of proteins remain completely unstudied in animal models and human disease. Furthermore, the GWAS approach might be tailored to measure the response of populations of individuals to specific physiological challenges such as is exemplified in many of the studies underway in UK Biobank.

**APPENDIX: LIST OF USEFUL URLs**

1000 Genomes project, [http://www.1000genomes.org](http://www.1000genomes.org)

*Physiol Genomics* • doi:10.1152/physiolgenomics.00004.2015 • www.physiolgenomics.org
GenVar at the Sanger Institute, http://www.sanger.ac.uk/resources/software/genevar
Scan DB, http://www.scanb.org
Broad Institute GTEx Browser, http://www.gtexportal.org
eQTL resources at the Prichard Lab, http://eqtl.uchicago.edu
HAPLOREG, http://www.broadinstitute.org/mammals/haploreg
Exome chip, http://genome.sph.umich.edu/wiki/Exome_Chip

International Mouse Phenotyping Consortium http://www.mousephenotype.org

Rat Genome Database, http://rgd.mcw.edu
Zebrafish Mutation Project, http://www.sanger.ac.uk/Projects/D_
Haploreg, http://haploreg.project.mirbase.org

Exome chip, http://genome.sph.umich.edu/wiki/Exome_Chip

Design

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: P.B.M. and A.T. prepared figures; P.B.M. and A.T. drafted manuscript; P.B.M. and A.T. edited and revised manuscript; P.B.M.

References


