Systems biology in cardiovascular disease: a multiomics approach

Abhishek Joshi[1,2], *Marieke Rienks*[1], *Konstantinos Theofilatos*[1] *and Manuel Mayr*[1] ✉

Abstract | Omics techniques generate large, multidimensional data that are amenable to analysis by new informatics approaches alongside conventional statistical methods. Systems theories, including network analysis and machine learning, are well placed for analysing these data but must be applied with an understanding of the relevant biological and computational theories. Through applying these techniques to omics data, systems biology addresses the problems posed by the complex organization of biological processes. In this Review, we describe the techniques and sources of omics data, outline network theory, and highlight exemplars of novel approaches that combine gene regulatory and co-expression networks, proteomics, metabolomics, lipidomics and phenomics with informatics techniques to provide new insights into cardiovascular disease. The use of systems approaches will become necessary to integrate data from more than one omic technique. Although understanding the interactions between different omics data requires increasingly complex concepts and methods, we argue that hypothesis-driven investigations and independent validation must still accompany these novel systems biology approaches to realize their full potential.

Technical advances in the past two decades have led to an increase in the generation of biological data. Gene and RNA sequencing, proteomics, metabolomics, lipidomics, the advent of microbiome studies, and computational collation of clinical data have required increasingly sophisticated analytical techniques to derive meaningful conclusions. The application of systems biology approaches allows investigators to integrate large datasets and undertake analyses across experimental and theoretical models. In this Review, we describe the growing number of omic techniques, together with their strengths and weaknesses. We examine areas in cardiovascular medicine in which these techniques have been applied, highlighting exemplary studies.

Systems biology to embrace complexity

Systems biology, as a formal discipline, is a philosophical position running counter to reductionist thinking. Put simply, systems biology attempts to explain biology in terms of interacting components. The action of an enzyme on a substrate is a system in the simplest terms, insofar as more than one component is considered and their interactions are crucial to the understanding of the model. These apparently simple systems have long been understood without the requirement for complex computation or high-throughput measurements. However, the advent of technologies that can measure many distinct components of a system at once has led to more complex analyses, often requiring computational support. Systems biology is now a field predicated on the generation of large datasets measuring many analytes, in which interpretation requires modern computational approaches. Not a single, universally accepted definition of systems biology exists. For the purposes of this Review, we consider a study to be classified as systems biology if it fulfils two criteria: that more than one component is measured and that the interactions between components are essential to the conclusions.

Although other investigators have suggested more detailed or stringent definitions of systems biology, mandating computational analyses or specific data types (and it is true that this approach has become increasingly central), our loose definition allows for an understanding of the philosophical rather than technical approach (FIG. 1).

Omics technologies

Although only 2% of the human genome codes for protein, the rest of the non-coding genome has emerged to be as functional and dynamic. Alongside an estimated 20,000 protein-coding genes, tens of thousands of human non-coding RNAs (ncRNAs) have been identified[1]. Many regions of the genome regulate other genes, either through direct contact or via ncRNAs or DNA-binding or RNA-binding proteins[2]. Long ncRNAs (lncRNAs), microRNAs (miRNAs) and other ncRNA species seem to perform regulatory functions on protein expression, either through direct interaction with

[1]*King's British Heart Foundation Centre, King's College London, London, UK.*

[2]*Bart's Heart Centre, St. Bartholomew's Hospital, London, UK.*

✉*e-mail: manuel.mayr@kcl.ac.uk*

https://doi.org/10.1038/s41569-020-00477-1

Key points

- Cardiovascular diseases are complex states, with the effect of the environment usually being far greater than that of the genetic status of an individual; therefore, the understanding of cardiovascular disease requires investigation of many biological levels.
- Omics techniques generate very large, complex and non-linear datasets, which mandate a systems biology approach, that is, the understanding of a biological process through examining the interactions between heterogeneous components.
- Current systems biology approaches applying network theories or machine learning to single-platform omics data have helped to make some progress in understanding cardiovascular disease but caveats remain.
- Integrated multiomics approaches explain the interactions between omics dimensions and are likely to require new ontological approaches to describe their findings.
- The lure of obtaining large datasets should not replace thoughtful, well-designed experiments; investigators must understand the technical and biological limitations of omics approaches alongside their strengths.

DNA or proteins (as for lncRNAs[3]) or with other mRNAs (as for miRNAs[4]). The number of human proteins is not equivalent to the number of protein-coding genes because a single gene can have several splice variants[5,6] and, when translated, proteins are subject to post-translational modifications, which can differ according to the prevailing conditions[7]. These complexities can be interrogated using omics technologies, which can generate quantitative measurements of RNAs, proteins or small molecules and can help to describe metabolic states[8], pharmacological responses[9–11] or disease phenotypes, particularly in combination with the use of electronic health records (EHRs)[12–15] (FIG. 2). The availability of multiomics data over the past decade has encouraged the use of systems approaches[16], in particular because of the capacity of these approaches to deal with multidimensional, heterogeneous data[17,18]. A discussion of the techniques and exemplary studies follows, and key studies are summarized in TABLE 1.

Genomics. Single-nucleotide polymorphisms (SNPs) are base pair differences in the genome between individuals. SNPs can be detected using sequencing approaches and can be statistically linked to phenotypes to identify genetic associations with diseases in traditional genome-wide association studies (GWAS). GWAS have been undertaken for cardiovascular diseases (CVDs), including acute coronary syndromes[19–21], atherosclerosis[22], atrial fibrillation[23], sudden cardiac death[24] and heart failure[25,26]. So far, 934 SNPs have been associated with CVD[27]. Studies have also found genetic interactions between known risk factors and subsequent coronary artery disease (CAD) and have moved beyond the predominantly white cohorts used in early studies[28–30]. Despite this success in identifying linkages, most GWAS discoveries have yet to be translated into benefits for patients. For example, perhaps the most robust finding with regards to CVD, an association between the 9p21 locus and CAD and myocardial infarction, remains unexplained in terms of its mechanism despite intense study[31–39]. As with many other risk variants, the 9p21 locus is within a non-coding region of the genome. The most likely candidate, ANRIL, is an antisense ncRNA with unknown biological function in CVD, although circular RNAs from this region have been implicated[40–42].

Standalone GWAS do not consider the complex interaction between genes. Large-scale studies provide increasing statistical predictive power and can uncover the contribution of rare variants but the trade-off is the small effect size compared with that of more common variants. However, analysing the results of several gene–phenotype linkages in parallel, for example, as part of a polygenic risk score, might begin to illustrate these interactions. In the CARDIoGRAMplusC4D consortium, 1.7 million SNPs were assessed for linkage with CAD and used to generate a genomic risk score[43]. This score provided an estimated risk of incident CAD based on an individual's genetic profile and was validated in the UK Biobank. A standard deviation increase in the genomic score conferred a hazard ratio of 1.71 and remained predictive after corrections for common environmental and clinical risk factors[43]. This large effect on the risk of CAD is similar to that of lifestyle factors such as smoking and might form the basis of future clinical decision-making if superiority to existing risk scores can be shown. Among participants with a high genetic risk of CAD, a favourable lifestyle alone could nearly halve their relative risk of CAD[44,45]. The large effect of the environment and behaviour (such as smoking, diet, exercise and socioeconomic status) on CVD are not traditionally captured in genomic analyses[46,47]. When both genetic and clinical risk scores were combined in the same study, susceptible individuals frequently had high polygenic and clinical risk scores in conjunction. Conversely, polygenic risk scores outperformed clinical risk scores in young individuals, potentially serving as early markers of clinical risk factors that might manifest in later life[48].

A major issue for future polygenic risk scores is the description of confounding conferred by population structure in the cohorts analysed owing to unaccounted-for relatedness within study populations, as demonstrated by two near-simultaneously published studies investigating genetic determinants of height[49,50]. Furthermore, whether polygenic risk score approaches can account for the gene–gene interactions, or epistasis, is unclear, which further increases the complexity of potential interactions and for which few studies in CVD are available[51]. Nevertheless, genomic analysis is the most mature field among the omics approaches and other techniques should be expected to have their own, as yet unidentified, biases and limitations.

The use of genetic variants to represent environmentally modifiable exposures as 'instrumental variables' creates the possibility of making causal inferences with observational data. The application of the instrumental variable method in GWAS has been referred to as Mendelian randomization and can be considered analogous to randomized controlled trials. Mendelian randomization studies are based on dividing participants into two groups of approximately equal size based on their genetic score, a number based on the variation in multiple genetic loci and their associated weights. Each variant included in the genetic score is inherited randomly and independently of other variants included in the score. Therefore, the number of disease-related alleles

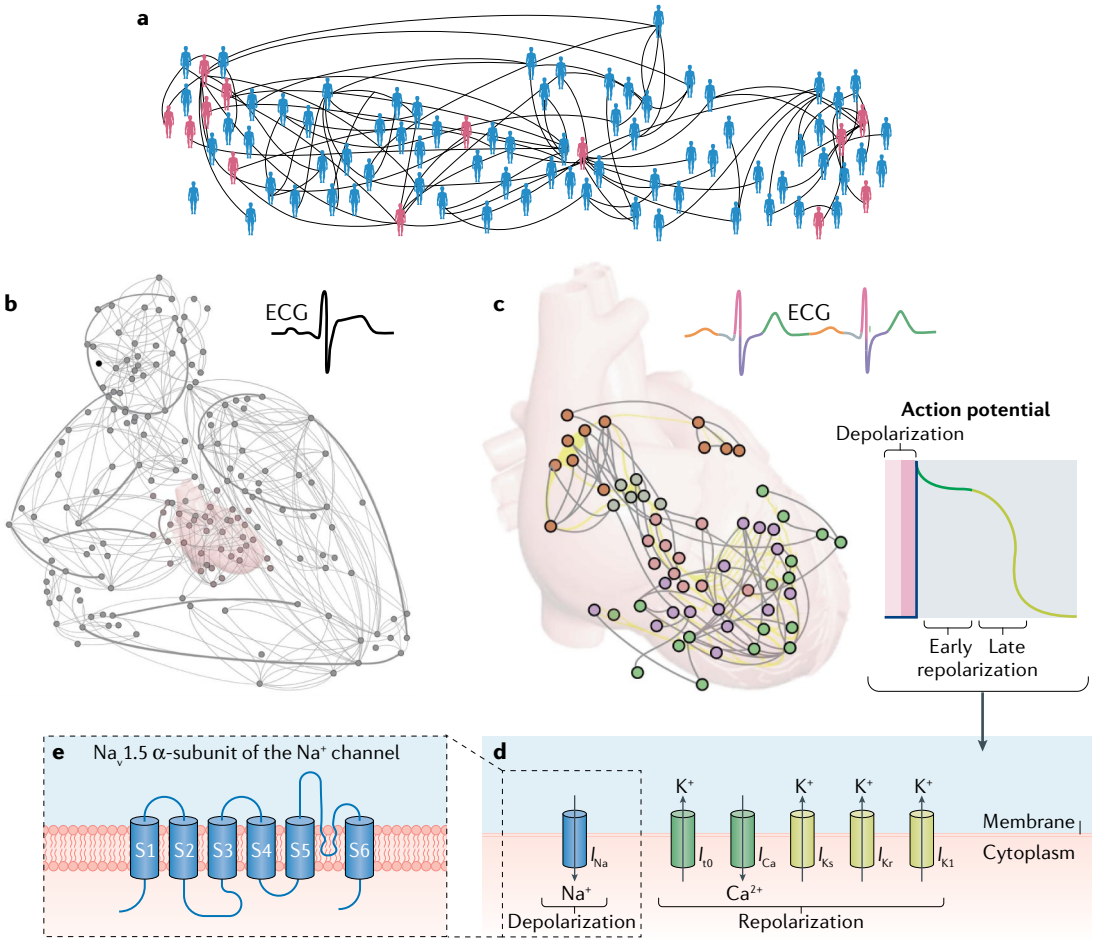


Fig. 1 | **Levels of explanation in systems biology.** Diseases and their biology can be explained by a systems approach, at different levels of explanation. **a**,**b** | At the level of clinical medicine, the systems approach is well established, integrating clinical signs, symptoms and investigations, such as electrocardiograms (ECGs), to describe disease states and syndromes. **c** | In the example in the figure, the interaction of chest pain, clinical examination and ECG changes describes the 'emergent' syndrome of acute myocardial infarction. ECG changes are the product of integration of the spatiotemporal properties of varying action potentials acting as components of the system. Electrophysiology ablation therapies can target combinations of action potentials at this level. **d** | The action potential is the result of combinations of various channel components, combining to generate its phases. Diseases affecting these channels or variations in these channels, either at the level of the proteins or the arrangement of the various channels, will affect the initiation, shape and duration of the action potential. **e** | Finally, the Na^+ channel is a system of protein subunit components that act to control Na^+ flow across the cell plasma membrane. Changes in the protein structure caused by genetic variation or disease states will alter the function of the channel. Scientific research at this level focuses on basic science discovery. A pharmacological study requires an understanding of the molecular effects of the drug (panels **d**, **e**), surrogate and safety end points (panels **b**, **c**), and individual outcomes data (panel **a**). I_{Ca}, Ca^{2+} current; I_{K1}, inward-rectifier K^+ current; I_{Kr}, rapid delayed-rectifier K^+ current; I_{Ks}, slow delayed-rectifier K^+ current; I_{Na}, Na^+ current; I_{to}, transient outward K^+ current.

that a person inherits should be random, allowing for causal inferences of randomly allocated genetic variants with disease[52]. Mendelian randomization studies have been proposed in addition to randomized controlled trials to identify promising therapeutic targets for CVD[53,54]. Multivariate Mendelian randomization can assess the contribution of more than one gene variant and is likely to be increasingly used, particularly in complex diseases such as CVD[55,56]. Many studies have investigated the genetic determinants of blood lipid levels, driven by both the clinical importance and the availability of validated, reliable measurements. A GWAS performed primarily in the Million Veteran Program cohort identified racial differences in the allele frequency of SNPs linked to blood lipid levels but strong agreement in the effect size within races[57]. Analysis of low-frequency genetic variants in non-European populations identified potentially new causal loss-of-function SNPs, including in *PCSK9*, and a Mendelian randomization approach to the EHR phenome demonstrated the presence of causal SNPs in the 'druggable' genome[57] (FIG. 3). Predicted loss-of-function genetic variants can identify human 'knockouts' to improve the search for drug targets[58]. Mendelian randomization studies have also been used to rule out causal factors for CVD, in principle allowing more rational approaches to drug discovery and testing. For example, whereas the plasma level of C-reactive protein is a well-established biomarker for incident CAD,

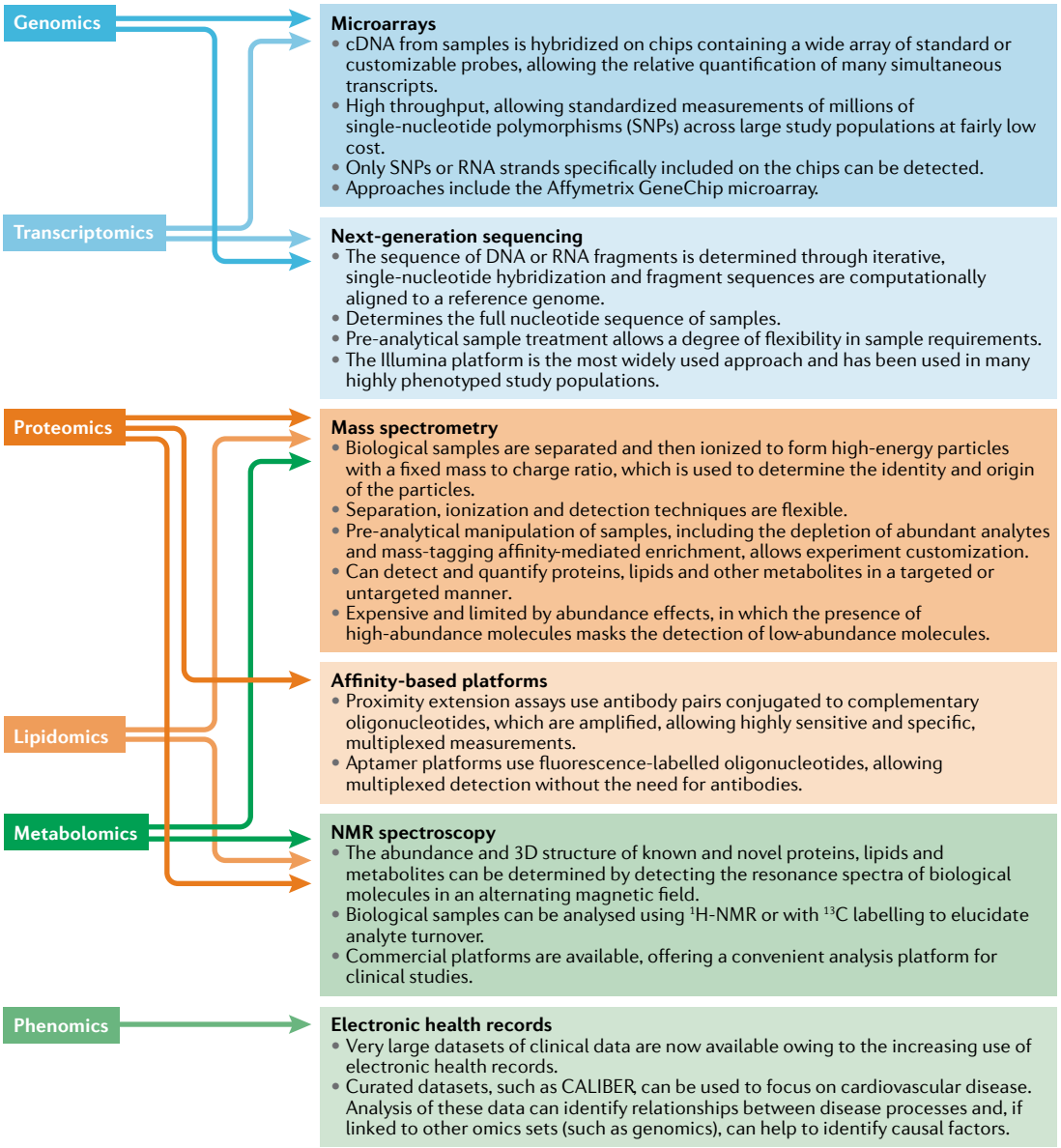


Fig. 2 | **Omics platforms used in systems biology.** Multiple analytical platforms can be used to interrogate samples and systems at a molecular level. In general, nucleotide-based experiments, such as genome-wide association studies, Mendelian randomization and transcriptomics studies, use quantitative PCR, microarrays and next-generation sequencing technologies. Next-generation sequencing and microarray approaches tend to offer discovery opportunities, whereas quantitative PCR is frequently used as a validation step. Proteomics, lipidomics and metabolomics approaches share common technologies, such as mass spectrometry and nuclear magnetic resonance (NMR) spectroscopy. Proteomics approaches also include affinity-based platforms that are based on antibody pairs (such as the Olink proximity extension assay) and aptamers (SomaScan; SomaLogic). Phenomics requires access to often-unstructured electronic health record databases.

a Mendelian randomization study determined that these levels were not causal and, therefore, C-reactive protein was not considered to be a promising target for drug therapy for CAD[59]. Platforms facilitating Mendelian randomization analyses across datasets from multiple polygenic risk scores are now publicly available[60]. This combination of genomic techniques might help to identify genetic relationships between disparate diseases. Interestingly, for CVD, genomic profiles linked with lifestyle factors, such as educational attainment, had protective effects[61].

Other models of analysing genomic data exist. The omnigenic model of complex diseases posits that disease risk might be largely driven by genes with no direct relevance to disease[62]. Instead, perturbations in peripheral genes are propagated through regulatory networks (BOX 1) to a much smaller number of core genes with direct effects. According to this model, if disease risk is defined as a trait, the genetic contribution for this complex trait can be divided into direct effects from core genes and indirect effects from peripheral genes that regulate core gene expression[63]. This model, in particular

Table 1 | **Exemplary studies of multiomics approaches to cardiovascular disease**

Study (year)	Sample or disease	Omics technique	Systems analysis	Main findings	Ref.
Inouye et al. (2010)	Blood lipids, lipoproteins and inflammation markers ($n = 518$)	Genomics, transcriptomics, NMR metabolomics and lipidomics; comparisons with affinity-based measurements of inflammatory biomarkers	WGCNA and NEO analysis of normalized gene expression and NMR data; conditional co-expression of analytes across quintiles of core module expression	A core leukocyte–lipid network was identified, the expression of which correlated with 83 metabolites and canonical markers of inflammation; the leukocyte–lipid network, containing genes related to basophils and mast cells, was implicated in close control of serum HDL levels	190
Greene et al. (2015)	987 publicly available genomic and transcriptomic datasets	Validated informatic findings in human smooth muscle cells	Regularized Bayesian integration to identify tissue-specific gene expression networks; NetWAS analysis linking SNP networks	Predicted and confirmed a regulated gene network responding to IL-1β stimulation	65
Talukdar et al. (2016)	Multi-tissue sampling in patients with CAD; samples from the STAGE cohort	Microarray SNPs and gene expression	WGCNA gene expression network analysis; association of networks with CAD; enrichment with previously identified SNPs associated with CAD; predictive Bayesian network analysis	Identified key driver genes in regulatory gene networks; confirmed the phenotypic relevance of key gene drivers in in vitro experiments	95
Miller et al. (2016)	CAD and coronary artery smooth muscle cells and coronary biopsy samples	ATAC-seq epigenomic analysis and ChIP sequencing in perturbed smooth muscle cells in culture; enrichment for CAD from existing GWAS analyses; validation in explanted coronary arteries and external datasets	Functional annotation using ENCODE; roadmap validation using eQTL cohorts	Identified causal gene loci involved in state-dependent differentiation of smooth muscle cells in atherosclerosis	93
Liao et al. (2016)	mRNA–protein complex pulldown from a mouse cardiomyocyte cell line in vitro	Mass spectrometry to identify the RNA–protein interactome	Referencing RBP with the OMIM database; combined discovery using RBDmap and spectral searches	Broader characterization of the RNA-binding proteome; discovery of overlap of the RNA–protein interactome with proteins associated with CVD; identification of the Rossmann fold as an RNA-binding motif; implication of metabolic proteins in RNA-binding and therefore in gene expression	137
Langley et al. (2017)	Carotid endarterectomy samples	Proteomic and transcriptomic analysis of atherosclerotic plaques; validation of a tissue-derived biomarker panel in plasma in two prospective, community-based studies	Differential proteomic profiles were determined using NSAF-PLGEM; Limma was used to identify differentially expressed transcripts; co-expression network analysis identified a biomarker risk cluster	A nine-protein biomarker panel that captured inflammation in atherosclerotic plaques was identified; the biomarker panel predicted progression to manifest CVD in plasma samples of two prospective community-based cohorts with a 10-year follow-up	116
Vilne et al. (2017)	Atherosclerosis in a mouse model of gene-switch controlled hypercholesterolaemia	RNA microarray data from aortas of *Ldlr*$^{-/-}$*Apob*$^{100/100}$*Mttp*$^{flox/flox}$*Mx1*–Cre (Reversa) mice over a prolonged time course, with induced hypercholesterolaemia	Gene ontology-based identification of mitochondrial genes; WGCNA co-expression analysis of gene expression; identification of transcription factor binding sites; selection of top genes using an external dataset	Hypercholesterolaemia regulates the expression of *Esrra* and *Ppargc1a*, which are central to a network of genes involved in the regulation of mitochondrial biogenesis and are differentially regulated in atherosclerotic plaques; activation of this gene network leads to rapid expansion of atherosclerosis	98

Table 1 (cont.) | **Exemplary studies of multiomics approaches to cardiovascular disease**

Study (year)	Sample or disease	Omics technique	Systems analysis	Main findings	Ref.
Klarin et al. (2018)	Blood lipid levels in a large, phenotyped patient cohort	GWAS of blood lipid levels; transcriptomic data from multiple tissues and cohorts; PheWAS using ICD-9 codes	Meta-analysis of multiple cohorts; multivariate Mendelian randomization of lipid-related SNPs associated with abdominal aortic aneurysms	Identified racial differences in the incidence of SNPs associated with lipid phenotypes; discovered a novel loss-of-function SNP correlated with adverse lipid profile and CAD; suggested a new potential therapeutic indication for PCSK9 inhibitors for abdominal aortic aneurysms	57
Lau et al. (2018)	Cardiac remodelling in wild-type mice from different genetic backgrounds in a model of hypertrophic cardiomyopathy	Deuterium-labelled mass spectrometry to analyse protein turnover; publicly available transcriptome data; interactome analysis using the STRING database	Unsupervised and interaction-enriched hierarchical and network clustering; functional enrichment using the Ensembl database	Identified contra-directional changes in protein turnover and gene expression; identified clusters of proteins with similar turnover profiles; showed that interacting proteins turn over at similar rates	188
Walter et al. (2018)	Cellular fractions from injured and healthy myocardium from *Cx3cr1*$^{GFP/+}$ mice	RNA sequencing of macrophages isolated from the myocardium over a time course after iatrogenic myocardial injury	Unsupervised k-means clustering; partial deconvolution to determine cell-type mix; previous knowledge network analysis and Boolean dynamical model to determine steady states	Identified differential cluster expression in injured and healthy hearts; macrophages after myocardial infarction do not conform to the canonical M1/M2 subtypes; network analysis determined three steady-state gene-expression profiles; microRNA–mRNA networks control gene expression in myocardial healing	99

ATAC-seq, assay for transposase-accessible chromatin using sequencing; CAD, coronary artery disease; ChIP, chromatin immunoprecipitation; CVD, cardiovascular disease; ENCODE, Encyclopedia of DNA Elements; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; ICD-9, International Classification of Diseases 9th revision; Limma, linear models for microarray data; NEO, network edge orientation; NetWAS, network-wide association study; NMR, nuclear magnetic resonance; NSAF-PLGEM, normalized spectral abundance factor–power law global error model; OMIM, Online Mendelian Inheritance in Man; PheWAS, phenome-wide association study; RBP, RNA-binding protein; SNP, single-nucleotide polymorphism; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins; WGCNA, weighted correlation network analysis.

the key concept of core genes, has been challenged and remains controversial[64]. An attempt to describe these networks used genomic data combined with information from data mining of publicly available datasets to determine the tissue profile of correlated networks of genes[65]. This study used previous knowledge of gene and protein expression in hundreds of smaller experiments and applied a GWAS-like approach, dubbed NetWAS, to link the expression of these networks with SNPs, potentially identifying causal genes that control the expression of networks of genes in specific tissues. A network controlled by IL-1β, considered to be a key cytokine in CVD, was validated in further external datasets[65].

The new frontier is to establish the biological function of the strongest loci in CVD, with the hope that the information obtained from GWAS analyses will provide new treatment targets. For example, a differentially expressed gene network was identified in the peripheral blood samples of patients with CAD[66]. Subsequent enrichment analysis of this network with the use of the CARDIoGRAM database[67] and other consortia databases implicated lipid metabolism and inflammatory processes in CAD pathogenesis. For other diseases, such as atrial fibrillation, the clinical applicability of knowledge from genomics is less obvious. GWAS have identified risk loci for QRS duration[68,69], PR interval[70] and sudden cardiac death[24,71]. However, the underlying aetiology of atrial fibrillation is complex and at least partially inherited[72]. GWAS have identified risk loci for atrial fibrillation[23], but studies in humans, especially of gene expression, often focus on atrial fibrillation after cardiac surgery because of the availability of atrial appendage tissue[73,74] and might therefore not be relevant to the highly prevalent lone atrial fibrillation.

Epigenomics and transcriptomics: gene expression, eQTLs and ncRNAs. Rather than comparing genomes with phenotypes, epigenomics considers interactions between genes. Studies of the expression of genes through quantifying relevant mRNA in a tissue, or transcriptomics, can give an indication of the differential expression of these genes and model their

Enrichment analysis
A set of bioinformatics and statistical techniques that identify classes of molecules (such as genes or proteins) that are over-represented in a large dataset and might have an association with a functional term, a biological pathway or disease phenotypes.

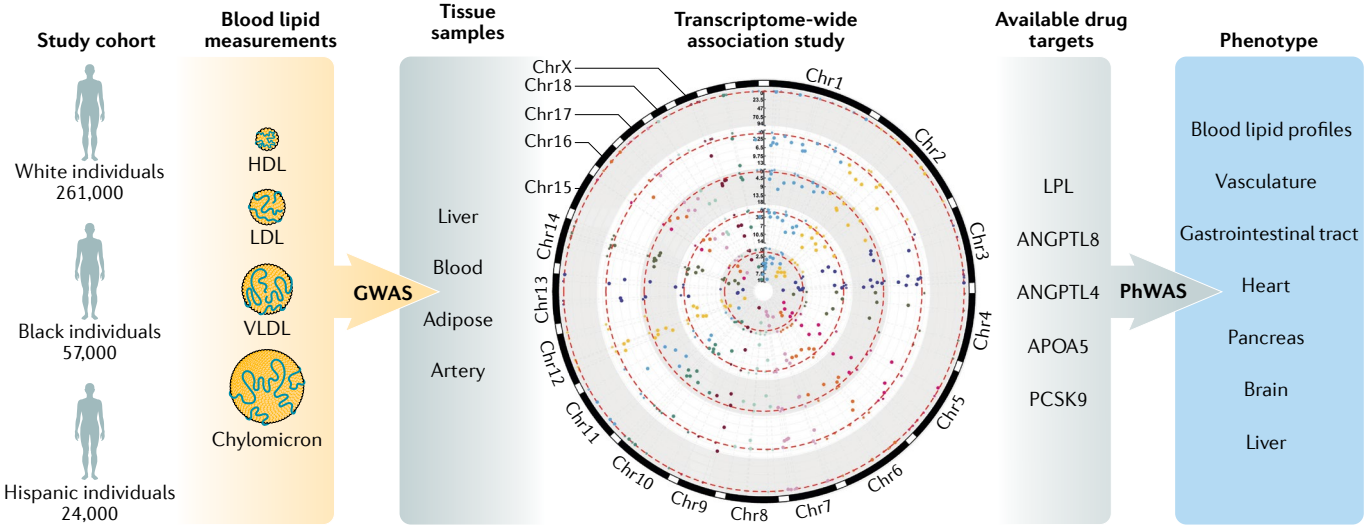


Fig. 3 | **Characterization of genetic determinants of blood lipid phenotypes.** The figure shows a schematic representation of the experimental workflow in a study that used data from ~300,000 multi-ethnic participants of the Million Veteran Program to assess the genetic determinants of the blood lipid phenotype[57]. The study investigated genomic data and blood lipid measurements in a highly phenotyped cohort, including racial background. Genome-wide association studies (GWAS) identified novel single-nucleotide polymorphisms (SNPs), some of which were further investigated in transcriptomic analyses of tissue samples from healthy and disease states. Specific SNPs in druggable regions of the genome were identified and used in a phenome-wide association study (PhWAS) of electronic health records data to identify the effect of the SNPs on disease states as identified by International Classification of Disease 9th revision codes. This analysis confirmed the importance of the identified biological axes in cardiometabolic disease and genomic analyses suggested a new potential therapeutic indication for PCSK9 inhibitors for abdominal aortic aneurysms. The circus plot shows, from the inner to the outer circle, the $\log_{10}$ *P* values of the 118 most significant SNPs for LDL, HDL, triglycerides, total cholesterol and coronary artery disease. The size of the dots is proportional to the $\log_{10}$ *P* value calculation and different colours have been assigned for each chromosome. The dashed red line in each circle denotes a *P* value threshold of 0.05. ANGPTL4, angiopoietin-like 4; ANGPTL8, lipasin; APOA5, apolipoprotein A5; LPL, lipoprotein lipase.

interactions. This method has allowed the identification of expression quantitative trait loci (eQTLs) through comparison of SNPs with mRNA expression profiles[75–80]. eQTLs are regions of the genome that seem to regulate the expression of other genes at *cis* (adjacent) or *trans* (distant) locations. Initially identified through microarray assessment of immortalized cell lines and later with the use of massively parallel RNA-sequencing (RNA-seq) in tissues from large, phenotyped study populations, eQTLs that regulate a broad range of coding and non-coding regions of the genome have been identified[81–84]. For example, a GWAS found that a SNP on chromosome 1p13 was consistently associated with raised LDL levels and myocardial infarction[36]. A study using eQTL data confirmed this finding and showed that the SNP at the 1p13 locus altered the expression of *SORT1* in the liver[85]. Studies in animal models of RNA interference and viral overexpression confirmed that *Sort1* is involved in the regulation of plasma LDL levels[85,86]. Crucially, this series of studies included biological hypothesis testing and validation. Early studies using microarrays have been supplanted by studies using RNA-seq (FIG. 2) and informatic approaches, such as TopHat, can reduce false discovery rates[87–89]. Emerging approaches, such as modelling the effects of allelic expression (the relative expression of maternal and paternal haplotypes), might increase the sensitivity of the analyses but have not yet been applied to CVD[90].

Expression quantitative trait loci
(eQTL). Places on the genome in which polymorphisms explain a significant proportion of the variation in mRNA expression levels.

Other epigenomic approaches measure the degree to which the genome is available for transcription in tissue samples. Hi-C techniques determine the spatial orientation and association of regions of the genome in the nucleus, inferring direct gene–gene interactions[91]. The assay for transposase-accessible chromatin using sequencing (ATAC-seq) is used to identify chromatin regions that are available for transcription[92]. In a study of coronary artery smooth muscle cells in vitro, ATAC-seq identified regions of the genome with altered transposase accessibility under pathological conditions[93]. Combining these findings with the publicly available CARDIoGRAMplusC4D SNP database identified the accessible genomic regions most strongly associated with CAD. The genes in these regions were validated in other available databases as being potentially causative but further interrogation of the mechanism remains to be completed[93].

As described above, GWAS have identified SNPs that seem to contribute to CAD and acute coronary syndromes[19–21,26] but the identified risk loci contribute to only 10–20% of the expected inherited risk of CAD[94]. One reason might be that genes acting in networks, rather than as individual risk loci, potentiate these risks. In a study using transcriptomic data from diseased and healthy tissue from 612 patients with CAD from the STAGE study, the investigators created undirected gene co-expression networks and were able to infer 30 causal modules that included a gene

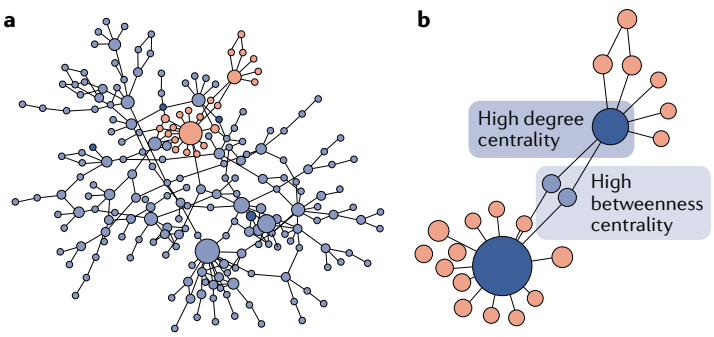
Box 1 | **Network theory and analysis**

Networks are computational models used in many systems biology investigations[220,221]. Networks allow the quantitative organization of complex biological information. In the network presented in the figure (panel **a**), circles represent individual components of the system, known as nodes[222]. These nodes can be genes, proteins or any other component. Connecting lines are called edges, which represent relationships between individual nodes. Edges can contain information about the strength, type and direction of this relationship from experimental data, adding predictive power to Bayesian computational methods or artificial neural networks[171,223–225]. Nodes have specific, calculated properties, such as centrality (describing the number of connecting nodes), betweenness (describing the degree to which a node communicates with the rest of the network) and accessibility (describing the individual node's relationship to the rest of the network.)

Biological networks are 'scale-free' networks, they are composed of a few highly connected hubs that are central to the biological process and many sparsely connected nodes that seem to have more modest effects on the system under study. Disease-causing nodes in cardiovascular genomic systems are usually peripheral nodes with low connectivity[226,227]. This observation is mirrored in the relatively weak contributions of individual genes to the overall burden of cardiovascular disease, in which entire networks seem to have a greater contribution.

Modules or clusters within a network are closely related nodes sharing topographical, functional or disease characteristics. Modules or clusters share edges and form edges with one another's neighbours[228]. Modules or clusters are assumed to work together in a subsystem (in pink in panel **a** and in isolation from the main network in panel **b**). Objects that are found to share statistically similar expression patterns (that is, are highly expressed or lowly expressed together) are assigned a network edge with a value describing this link in a co-expression network. Substrates that share differential expression values are described as being differentially co-expressed and are linked with an edge in a differential expression network[229].

The co-expression or differential expression data do not determine causality. Furthermore, the concept of attaching significance to changes in the relative concentration of molecules is potentially misleading. A gene with tightly regulated expression might be central to a system but have no differential expression in a large metabolic perturbation. This gene would be neutral in a differential expression analysis but detected in more traditional knockout or dose–response models.



or eQTL previously linked to CAD in GWAS[95]. These modules tended to have novel key drivers, suggesting that small differences in gene expression, which are detectable only when subject to network analysis, result in significant phenotypic changes. Three of these CAD-associated modules predicted atherosclerosis burden in a mouse model and individual silencing of four of the key driver genes (*AIP*, *DRAP1*, *POLR2I* and *PQBP1*) in cultured foam cells reduced cholesterol ester accumulation[95]. The identified key driver genes are involved in RNA processing, suggesting that gene–gene interactions are an important contributing factor to CAD in this analysis. Only one of the key driver genes (*PQBP1*) has been subsequently investigated, again in a systems approach studying the effects of anti-retroviral drugs on the development of foam cells[96]. The stage is set for further, hypothesis-driven investigation into these discoveries.

Similar work combining genotyping and RNA-seq was undertaken in six types of tissue from 600 patients with well-characterized CAD in the STARNET study[97]. Using data from GWAS and gene-regulatory studies, the researchers identified and enriched *cis*-eQTLs and generated overlapping causality networks for each tissue type and for a range of diseases, including CAD and Alzheimer disease, with the use of statistical correlation-based techniques to identify causal mutations. A core network of 33 genes regulated by risk SNPs seemed to have regulatory effects across all diseases[97]. For example, SNPs regulating *PCSK9* expression in abdominal fat (but not in the liver) were linked with raised LDL-cholesterol levels in plasma. This finding awaits further experimental validation, especially because *PCSK9* expression levels in adipose tissue are minimal compared with those in the liver[83]. The STAGE and STARNET studies have the benefit of including an analysis of more than one human tissue in a phenotyped cohort and the integration of existing GWAS and cross-species data. The strength of the STARNET study comes from the use of disease-relevant tissues to obtain expression data (for example, the analysis included atherosclerotic artery samples) rather than blood samples. Crucially, the novel findings still require validation in biological experimental models.

Omics applications to animal models can offer mechanistic insight. In a longitudinal study of atherosclerosis-prone mice with conditional hypercholesterolaemia, microarray transcriptomics analysis identified 12 networks of differentially expressed genes[98]. As mice aged under hypercholesterolaemic conditions, the expression of nuclear-encoded mitochondrial gene networks decreased, particularly during the stages of rapid atherosclerosis progression. This effect was rescued by genetically lowering plasma LDL-cholesterol levels. Co-expression network analysis led to the identification of the *Esrra–Ppargc1a* module as a cholesterol-sensitive 'switch' for mitochondrial gene expression[98]. Although this investigation is exemplary in identifying modules of genes and in using external data sources to enrich the conclusions, the small number of animals (a total of 23 across five experimental conditions) compared with larger studies such as STARNET reduces its power to detect anything but the largest effects. Importantly, these effects might potentially be due to the infiltration of other cell types during plaque progression, rather than due to changes in the native cells.

Indeed, successful transcriptomic studies usually require careful isolation of specific cell types. A transcriptomic study of isolated macrophages from a time course series of infarcted and healing mouse hearts detected changing proportions of gene expression identifying M1 or M2 subtypes[99]. Using gene-expression networks informed by previous knowledge of expression patterns, three steady-state combinations of gene-expression profiles were identified, corresponding to phases of infarction, healing and recovery. Further network analysis, including miRNA expression levels, suggested that these states were regulated and maintained by a dynamic miRNA–mRNA

network[99]. Pre-analytical processing of RNA-seq data can offer novel insights into ncRNAs. For example, the ribose-quencing technique, in which ribosome-bound mRNA is used to infer protein translation, has been used to describe the cardiac translatome and to identify microproteins translated from lncRNAs and circular RNAs[100], which were previously considered to be translationally silent.

Genetic variants identified in transcriptomic studies that do not show an association in GWAS should not be discounted for further study. Disease risk is often mediated by multiple cell types, in which the same components might have different, interacting functions. For example, miR-126 is a crucial miRNA for vascular function and endothelial integrity[101]. miR-126 is not only highly abundant in endothelial cells[102] but is also present in megakaryocytes and a SNP altering miR-126 levels was shown to be positively correlated with platelet reactivity[103,104]. Therefore, miR-126 remains an important regulator of at least two central biological processes related to CVD, endothelial cell function and platelet reactivity. However, a miR-126-related genotype has not been reported as conferring risk of CVD in any GWAS analysis to date. Given that cellular networks vary according to cell type, the quantitative effect of a variant would be an average of its effect size in each cell type[62].

Proteomics: mass spectrometry and affinity-based approaches. Mass spectrometry allows the multiplexed discovery and quantification of proteins in a biological sample via comparison with known protein databases, quantification of tagged proteins[105,106], detection of newly synthesized proteins[107,108], and detection of conformational changes[109] and post-translational modifications[110]. Current limitations include cost, detection bias in favour of high-abundance proteins and incomplete coverage of the mammalian proteome[111]. These challenges can be partially overcome through the depletion of high-abundance plasma proteins[112], the further pre-fractionation of samples or through targeted proteomics approaches that analyse only the proteins of interest[111]. A targeted proteomics approach overcomes stochastic sampling and contributes to data completeness albeit at the expense of proteome coverage. Data-independent acquisition is a novel proteomics method that combines the advantages of targeted proteomics with the broader coverage of discovery proteomics by generating a 'digital fingerprint' of all accessible proteins in a sample[113,114]. This advance facilitates the high-throughput analysis of hundreds to thousands of proteins in a given sample, with improved data completeness. Nonetheless, throughput and costs as well as assay drifts can make applying mass spectroscopy-based proteomics workflows to very large cohorts prohibitive.

Other technologies, such as SomaScan (SomaLogic) and proximity extension assay (Olink) technologies, benefit from their applicability at scale but were initially designed for a single sample type: human plasma. The Olink proximity extension assays combine the specificity of antibody techniques with DNA amplification steps. Proximity extension assays rely on the binding of two antibodies to a single protein, bringing their conjugated, complementary DNA strands into proximity and amplifying and detecting the resulting double-stranded DNA. Antibody–DNA pairs can be multiplexed, currently allowing for the discovery of a total of 1,536 proteins in plasma[115]. Unlike mass spectrometry, the amplification steps in the Olink proximity extension assay allow less abundant proteins, including cytokines or chemokines, to be better detected. The use of antibody–DNA pairs increases specificity. Our group has used the Olink platform techniques alongside mass spectrometry to investigate potential plasma biomarkers for incident CVD and atrial fibrillation[116,117]. Combinations of strategies, rather than reliance on a single platform, offer a deeper, more quantitative coverage (FIG. 4).

The SomaLogic technology uses unique, protein-binding oligonucleotides (slow off-rate modified aptamers or SOMAmers) to increase the depth of proteomic analysis. The use of the SomaScan assay in plasma profiling in CVD illustrates the complexity of systems approaches[118]. This technique harnesses aptamers to detect plasma proteins in multiplex[119]. Panels of 1,000–5,000 protein binders are available, offering the convenient, scalable detection of proteins, but the technique encounters challenges when attempting to discover novel biomarkers owing to the limited selection of targets. A study measuring 1,129 plasma proteins after alcohol septal ablation in patients with hypertrophic cardiomyopathy identified the established biomarkers of myocardial injury troponin I, creatine kinase muscle–brain and myoglobin[120]. A follow-up study in the same cohort using a platform with 4,783 proteins confirmed that troponin I had the greatest fold-change after planned myocardial injury[121]. By contrast, our group used mass spectrometry for the discovery of novel biomarkers of myocardial infarction, identifying cardiac myosin-binding protein C as a promising and novel biomarker of myocardial injury, which was missed by the aptamer approach[122–124].

A SomaScan investigation of a cohort with a well-phenotyped cardiovascular risk profile also identified well-known biomarkers for cardiovascular risk (such as C-reactive protein, apolipoprotein E (apoE) and matrix metalloproteinase 12 (MMP12))[125]. Similarly, Olink panels were used in combination with a supervised machine learning approach to identify the circulating protein profile in patients with or without high-risk coronary atherosclerotic plaques, confirming elevated MMP12 levels as a risk marker for CAD[115]. Surprisingly, a large study using the SomaScan platform identified causal SNPs (protein quantitative trait loci (pQTLs)) that affected the plasma protein abundance of MMP12 and inferred a protective effect of MMP12 against coronary heart disease[126]. These apparently conflicting findings have not yet been resolved but highlight the probable environmental rather than genetic determination of the levels of many plasma biomarkers and the limited variation in plasma concentrations that might be explained by pQTLs.

Future work will lead towards evidence-based CVD biomarker panels that can then be tested in larger cohorts. For example, application of the SomaScan panel to plasma from five large, independent cohorts and the use of a supervised machine learning algorithm identified a 13-protein biomarker panel to predict the

Mass spectrometry
A category of analytical tool that is used to measure the mass-to-charge ratio of one or more molecules present in a sample. Typically, mass spectrometry can be used to identify and quantify unknown compounds and to determine the structure and chemical properties of molecules.

Data-independent acquisition
Mode of data collection in mass spectrometry, in which all precursor ions within defined mass to charge windows are fragmented sequentially, rather than the selection and fragmentation of the most abundant precursor ions in data-dependent acquisition approaches.

Aptamers
Molecules, for example, oligonucleotides, that bind to a specific target molecule and are used to perform high-throughput proteomics quantification such as in the SomaScan (SomaLogic) platform.

Protein quantitative trait loci
(pQTL). Genomic loci that explain a significant proportion of the variation in the quantities of a protein in a set of biosamples.

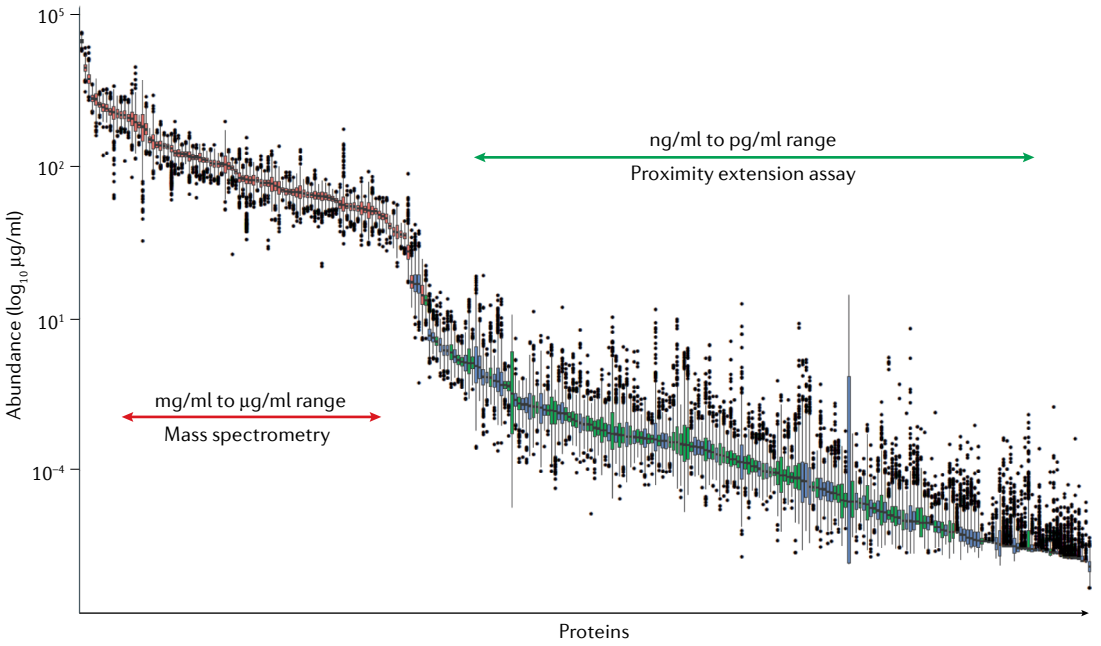


Fig. 4 | **Combined approaches to proteomics analysis.** The presence of high-abundance proteins, such as serum albumin or immunoglobins, in the highly complex plasma matrix directly affects the sensitivity of mass spectrometry to detect lower-abundance proteins such as cytokines. Combining measurement techniques by adding an antibody-based proximity extension assay (Olink) to mass spectrometry increases the depth of proteome coverage. However, although mass spectrometry trades a flexible dynamic range and broad coverage for bias in favour of abundance, affinity-based approaches are highly standardized but only for a subset of the proteome. Combinations of both analytical platforms can facilitate more comprehensive conclusions.

probability of a future cardiovascular event that achieved an area under the curve of 0.69 in the validation cohort, similar to but not significantly better than assessment of standard CVD risk factors[127]. In a more complex study design, a comparison of the measurements of the same 1,129 proteins across three genotyped study cohorts determined that the expression levels of 268 proteins could be determined by SNPs[128]. Using these data, pQTLs were generated to directly impute protein levels in large GWAS cohorts based on the genotype alone. These pQTLs were used to impute protein measurements in a study population with the use of features selected from EHR databases, leading to the identification of 55 pQTLs that predicted 89 clinical diagnoses. External validation through direct measurements of pQTL targets in patients with clinically determined indices of atherosclerosis confirmed C-type lectin domain family 1 member B and platelet-derived growth factor receptor-β as biomarkers for atherosclerosis. This study combines genomic, proteomic and EHR data to identify new potential pathways to understand CVD; however, recognizing the limitations of each step is important. Early assessment of the SomaScan platform found that 27% of the aptamers were more specific for targets in chicken plasma than for human proteins[129]. Furthermore, around 14% of the aptamers bind and detect more than one protein or isoform, and 32% have differential binding, and thereby quantification, owing to the protein sequence variation caused by SNPs. Therefore, aptamer binding does not necessarily measure the true protein concentration[126]. Finally, the reliance on EHR-derived phenomics has been shown to be subject to substantial potential bias and inaccuracy, particularly related to incentivized areas of medicine[15,130,131]. Despite these limitations, the apparent convenience of outsourced protein measurements, without the need for training and generating experimental expertise in-house, is likely to lead to many studies using the Olink and SomaLogic proteomic panels in cohorts of increasing size. For example, the application of the SomaScan panel to plasma from 4,263 individuals and using age alone as an independent variable allowed the identification of 'waves' of co-expressed proteins (a 'proteomic clock') and demonstrated the relationships between these waves and CVD[132].

The raw abundance of proteins in a given sample does not take into consideration the interactions between proteins, which are crucial to understanding the biological landscape. Protein–protein interaction experiments have densely mapped the interactome with the use of high-throughput yeast two-hybrid experiments[133,134]. Enriching GWAS studies with interactome data has helped to identify putative gene–protein networks of interacting proteins related to hypertension[135]. Protein–protein interaction experiments previously relied on the hybridization of physically interacting proteins expressed in modified yeast to generate a detectable colour change. A high number of putative interactions between pairs of proteins have been discovered and are available in large public databases. More recently, information on the protein interactome and 3D structure of proteins has become available through harnessing proximity-based crosslinking mass spectrometry[136]. For example, a study using an unlabelled mass spectrometry approach to analyse RNA–protein complexes generated through ultraviolet-induced

Interactome
The whole set of molecular interactions in a particular sample, tissue or cell in a specific organism or phenotype.

crosslinking in cultured mouse cardiomyocytes identified RNA-binding proteins that can influence gene expression, many of which were found to have roles in intermediate metabolism, highlighting previously unidentified protein-to-gene regulation[137].

Metabolomics and lipidomics. Metabolomic studies are traditionally aimed at identifying small molecules but often also include lipid and lipoprotein species[138]. Lipidomic approaches involve the assessment of molecular lipid species rather than lipid classes and these lipid profiles have been associated with CVD risk factors[139–141]. The clinical success of the PCSK9 inhibitor class of drugs[142] has further proven the LDL hypothesis, in which lowering plasma LDL levels reduces the risk of CVD. Many large epidemiological databases and studies, including the UK Biobank, take advantage of a commercial nuclear magnetic resonance (NMR) metabolomics platform. This metabolomics platform uses a proprietary technology that is based on two ^{1}H-NMR analyses per sample, each optimized to detect different classes of biomolecules at different concentration ranges[143]. One study applied this NMR platform to analyse plasma from genotyped participants from the PROSPER study of pravastatin, embedded in a wider population study of eight large genomics cohorts[144]. The metabolomic effects of a loss-of-function *PCSK9* variant, as determined by Mendelian randomization, and those of therapy with statins were strikingly similar when assessing lipid and lipoprotein species. This finding suggests that physiological responses to changes in plasma LDL-cholesterol levels are shared between statin and PCSK9-inhibitor therapies.

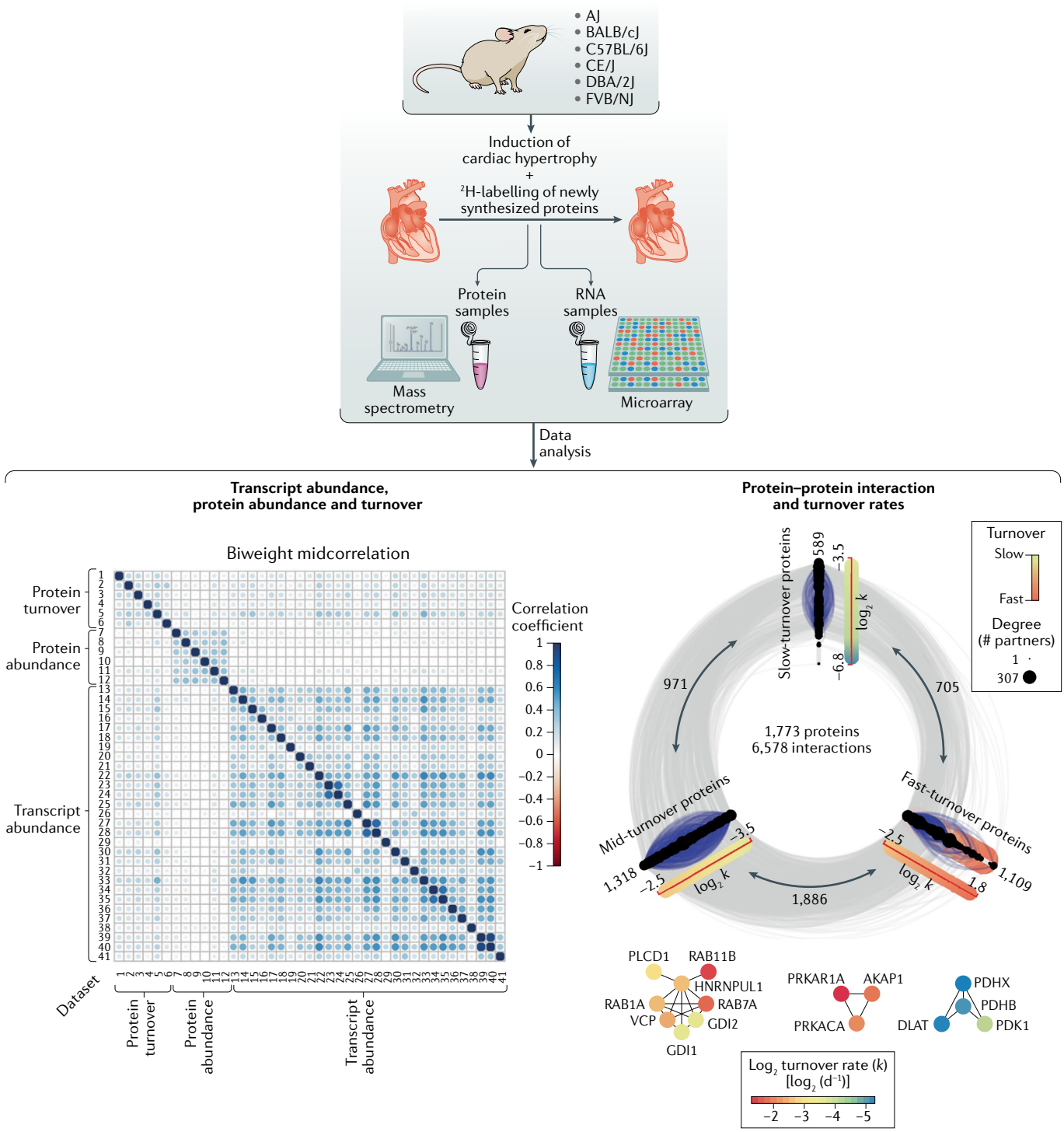
Our group undertook a mass spectrometry lipidomic analysis of plasma from a prospective study cohort and identified a cluster of lipids that crossed traditional lipid-class boundaries and conferred CVD risk[145]. We then used agglomerative hierarchical clustering to analyse a novel, targeted mass spectrometry panel of apolipoproteins in the same cohort and found that high plasma levels of apolipoprotein components of triglyceride-rich lipoproteins (apoC-II, apoC-III and apoE) were strongly linked with adverse CVD outcomes[146]. An antisense RNA therapy targeting apoC-III has shown promising results in a phase II study[147] and, in addition to reducing plasma levels of apoC-III, also lowered apoC-II and apoE concentrations[146]. The association between triglyceride-rich lipoproteins and the risk of CAD was replicated in other study cohorts with the use of NMR[148]. The emergence of triglyceride-rich lipoproteins as mediators of CVD risk has led to the consideration of all lipoproteins containing apoB as being associated with the risk of CAD[149,150]. This signal was also seen in a multivariate Mendelian randomization analysis of the UK Biobank, in which SNPs linked to increased apoB levels in plasma conferred the greatest risk of CVD and plasma triglyceride levels were also identified as being potentially causal[151]. Other studies have used a mass spectrometry-based lipidomics approach to analyse plasma, identifying ceramide species as a lipid subclass whose levels were associated with the risk of CVD and were modified by statin therapy[140,152].

Nuclear magnetic resonance (NMR). A physical phenomenon in which nuclei in a strong constant magnetic field are perturbed by a weak oscillating magnetic field and respond by producing an electromagnetic signal with a frequency characteristic of the magnetic field at the nucleus. This physical phenomenon is used in NMR spectroscopy, which is a technique for determining the structure of organic compounds with applications in lipoprotein profiling and metabolomics.

Microbiomics: the TMAO story. The microbiome comprises the total genetic content of the microbiota of all commensal and pathogenic microorganisms residing at a particular site in a host. Measurement of the microbiome in a sample depends on the accurate DNA or RNA sequencing of all prokaryotic, fungal and viral microorganisms on or within the host, which is a dynamic rather than fixed system[153,154]. Gut microbiota-derived metabolites have been implicated in the progression of atherosclerosis and the risk of thrombosis[155,156]. The apparent ubiquity of microbiota-related findings was noted and questioned in 2020 (REF.[157]). As this new field matures, a more critical approach will be applied to microbiome research.

The poster child for the potential importance of the gut microbiota in CVD is the discovery and validation of trimethylamine-*N*-oxide (TMAO). Discovery of this microbiota-derived metabolite of choline as a biomarker and potential treatment target for atherosclerosis integrates observational and interventional data across multiple studies and systems. Using an in-house proton NMR platform to analyse aortas from atheroprone apoE-deficient mice, we initially described that TMAO was present in advanced, but not early, atherosclerotic plaques[158]. Subsequent studies used targeted methods to measure TMAO in the plasma of patients with stable or unstable CVD. An unbiased mass spectrometry-based metabolomics approach in a cohort of patients with CAD identified a cluster of metabolites that predicted subsequent adverse cardiovascular outcomes[159]. One of these metabolites was confirmed to be TMAO. In a second study, the same group detected elevated TMAO levels in plasma from healthy participants who received heavy isotope-labelled phosphatidylcholine together with choline-rich foods[160]. Furthermore, TMAO was undetectable in plasma after the suppression of intestinal microbiota with antibiotic therapy, confirming that dietary choline, metabolized by commensal bacteria, was the source of TMAO. Elevated plasma TMAO levels were associated with an increased risk of adverse cardiovascular events over a 3-year follow-up period[160]. The mechanisms by which TMAO contributes to CVD risk have been investigated. Patients with recent ischaemic stroke who had raised plasma TMAO levels had more circulating pro-inflammatory intermediate monocytes, a phenotype linked to CVD[161], although this study did not demonstrate causality. Further work has identified platelets as potential mediators of the TMAO-associated risk of thrombosis[162] and a non-lethal inhibitor of bacterial enzymes reduced the diet-induced increases in TMAO levels in mice and reduced platelet hyper-reactivity[163,164].

Clinical validation of the concept of a causal role of TMAO in CVD has met challenges. In studies including further statistical analyses, plasma TMAO levels have been shown to be associated with renal disease[165]. Chronic kidney disease is a known risk factor for CVD and the direction of causation between these three entities is not yet established. Two studies published in 2020 have investigated the predictive value of plasma TMAO levels for the risk of CVD. A longitudinal study in 760 women found that increases in plasma TMAO levels over a decade were associated with incident CVD but the study relied on the self-reported rather than measured



determination of chronic kidney disease[166]. A nested case–control study in a clinical trial on the antiplatelet agent ticagrelor confirmed a relationship between TMAO levels in plasma and the risk of renal disease and recapitulated the association between elevated TMAO levels and a higher risk of adverse cardiovascular outcomes in patients with previous myocardial infarction[167]. However, the study did not find a significant difference in the effect of ticagrelor across plasma TMAO quartiles and the association between TMAO levels and CVD risk was not significant after adjustment for the traditional biomarkers of CVD, again questioning the causal role of TMAO in CVD. Other effects of the gut microbiota on CVD have been investigated. Statin therapy was shown to reduce the Bact2 enterotype associated with obesity and to be associated with an altered inflammatory phenotype[168].

Phenomics: systems at the human level. The curation and combination of large datasets from EHR allow the generation of interrogable resources[12,14,169]. When other clinical data are included, such as laboratory or radiological

◀ Fig. 5 | **Integrated omics analysis of proteome dynamics during cardiac remodelling.** The figure shows the experimental outline of a study that used an integrated omics approach to assess the proteome dynamics during the development of pathological cardiac hypertrophy in mice[188]. The investigators induced pathological cardiac hypertrophy by isoprenaline stimulation in different mouse strains, with concurrent administration of deuterium oxide in the drinking water to label cardiac proteins during the cardiac remodelling process. Protein and RNA samples were analysed by mass spectrometry and RNA microarray analysis, respectively. Analysis of the data allowed the correlation of protein and transcript pairs and the identification of the dynamics of protein turnover in cardiomyocytes during cardiac remodelling. Protein networks were identified by co-turnover rates, providing a temporal understanding of the pathological processes underlying cardiac hypertrophy. Adapted with permission from REF.[188].

results, computational techniques can be applied to discover associations or predictors of disease. A study conducted in 2017 linked four large clinical databases in the CALIBER research platform to investigate bleeding and antiplatelet strategies after myocardial infarction[170]. The study demonstrated the existence of a dynamic risk profile within the first year after myocardial infarction, recommending ongoing risk assessment for antiplatelet therapy. Importantly, non-health factors, such as changes in reimbursement criteria, can bias EHR data. An analysis of EHR-coded diagnoses leading to the implantation of coronary stents found that the reporting of unstable angina as an indication for the procedure increased significantly in regions where 'appropriate use' criteria had been applied but not in a region where these criteria were not in place[131].

Machine learning, in which mathematical algorithms provide novel insights or predictions that are based on large datasets, is predicted by some investigators to become central to the analysis of the large, highly complex data generated by omics techniques[171–175]. A combination of supervised and unsupervised machine learning methods was used to identify differences in electrocardiogram (ECGs) to predict death in an population of patients with acute CAD[176]. A study using an unsupervised approach identified three subgroups among patients with heart failure with preserved ejection fraction on the basis of clinical data[177]. A supervised learning algorithm using echocardiography data achieved a 96.2% accuracy for classifying either constrictive or restrictive physiology[178]. A supervised learning algorithm applied to coronary CT angiography and clinical data from 10,030 patients with suspected CAD outperformed traditional Framingham risk scores and other imaging-derived prognostic indicators[179]. Finally, a deep neural network approach to analyse >2 million ECGs from a single database outperformed clinical risk scores and clinicians in predicting 1-year mortality[180]. Machine learning approaches in cardiology have so far been primarily applied to the interpretation of clinical rather than biological data, where machine learning largely supports or replaces the clinicians' performance of repetitive or laborious tasks[181]. Novel biological insights that are based purely on machine learning have so far been elusive. The application of these techniques to very large cross-platform omics approaches might help to identify disease processes or entities that escape traditional models.

Deep neural network
A form of an artificial neural network with many hidden layers, used in classification, regression, clustering and other machine learning applications.

Multiomics approaches to CVD

A strength of a systems biology approach is the potential to include measurements from multiple platforms. TABLE 1 highlights exemplary studies of cross-platform omics approaches combined with computational approaches to reach robust experimental conclusions.

Inherited cardiomyopathies offer a rich substrate for investigation with a systems approach, because the aetiology of these conditions is founded on an abnormal genome and systems analysis can therefore take advantage of the full range of omic datasets. More than 30 genes have been reported to contribute to inherited forms of dilated cardiomyopathy, with a substantial overlap with genes associated with hypertrophic cardiomyopathy and with genes with a range of proposed cellular functions[26,182–185]. However, these genes were not differentially expressed in patients with dilated cardiomyopathy[186,187]. Mutations linked to cardiac disease are often found within sarcomeric proteins and might alter protein interactions rather than gene expression. Furthermore, protein turnover rates identified by deuterium-labelled peptides did not demonstrate a universally positive correlation of protein turnover and gene expression in mouse models of hypertrophic cardiomyopathy[188] (FIG. 5). By contrast, differentially expressed miRNAs and proteins were identified in early manifestations of an established mouse model of dilated cardiomyopathy with the use of microarray techniques and were validated with RT-qPCR and shotgun proteomic approaches[189]. The action of the identified miRNAs was further investigated through analysis of target protein expression and the study investigators suggest that these miRNAs regulate apoptotic pathways by altering specific protein targets. This study demonstrates how mechanistic elucidation must use data from multiple omic dimensions.

Multiomics analyses of plasma samples have offered insight into the complex relationship between apparently distinct regulatory axes. An early example of combining omics measurements, in this case whole-blood RNA-seq with targeted NMR metabolomics, identified a gene-expression cluster that incorporated genes related to both immune cells and lipid biology and linked the gene module expression with previously identified protein biomarkers for cardiovascular risk[190]. The degree of gene clustering varied as the plasma levels of the individual lipids changed, describing the dynamics of a putative network between lipid-producing cells and immune cells[190]. Whole-blood transcriptomics and metabolomics were combined in a further study, identifying cross-omic correlations between mRNA and metabolite levels[191]. A weighted bipartite network analysis allowed the identification of previously known regulatory pathways. Gene ontology enrichment of the transcriptome identified crosstalk between pathways, again including lipid–immune interactions[192]. Mass spectrometry-based lipidomics of plasma from three large study populations was combined with Mendelian randomization to determine five new causal SNPs for variations in plasma lipid levels, with polyunsaturated lipids demonstrating the strongest genetic control, whereas traditional lipid classes had a relatively weak inheritability[193]. GWAS investigating associations with algorithmically derived covariate clusters of metabolites found nine new gene loci, including *SERPINA1* and *AQP9*, to be associated with specific metabolic clusters containing LDL and

intermediate-density lipoprotein (IDL) particles in the case of *SERPINA1* and a broad range of metabolites in the case of the *AQP9* locus[194]. Fine mapping of these genes identified SNPs contributing to this effect. A higher expression of *Aqp9* in the liver was associated with a larger atherosclerotic lesion area in atherosclerotic *Apoe*$^{-/-}$ mice and the expression of both *SERPINA1* and *AQP9* was higher in human atherosclerotic plaques than in control samples. This study combined a genomic and metabolomic approach with validation of the findings in the relevant tissue. A study using the METSIM cohort applied multivariate analyses of gene–metabolite relationships and identified 13 genes (*APOA5*, *APOC1*, *CELSR2*, *CETP*, *DOCK7*, *FADS2*, *GALNT2*, *GCKR*, *LDLR*, *LIPC*, *LPL*, *PCSK9* and *TRIB1*) that were responsible for 75% of the genetic associations with metabolite variability[195]. Most of these genes code for well-known regulators of lipid metabolism but interactions with age and the use of statins were identified.

Our group undertook a multiomics analysis of human atherosclerotic plaques by combining gene-expression microarray and proteomic mass spectrometry data and applying a number of validation techniques[116]. We identified a tissue-based biomarker panel that included pro-inflammatory molecules that, when measured in plasma, outperformed traditional risk factors and plasma C-reactive protein levels as predictors of CVD in two independent cohorts. This approach using tissue material from humans with the relevant disease, in which the biomarker candidates should be most enriched, along with subsequent validation of findings in plasma could be a paradigm for multiomics research. Of note, fold-changes in mRNA levels were far smaller than those in protein levels, requiring a tenfold increase in sample size to achieve adequate statistical power. In the future, studies with large cohorts, such as the GLOBAL study ($n = 10{,}000$), will integrate multiomics data from individuals who have been closely phenotyped (in the case of the GLOBAL study with coronary CT angiography[196]).

A major challenge for multiomics research is the integration of datasets derived from different analytical techniques. Most multiomics approaches use cross-platform omics data to validate one another when investigating a specific phenotype. This approach is useful but does not exploit the full potential of the multiomics data. Despite promising computational methods that are explicitly designed for multiomics techniques[197], the lack of standardization across platforms and the limitations of the existing multiomics techniques leave room for improvement in exploring multiomics analysis. For example, more emphasis should be given to the disagreements between the different levels of molecular information (such as overexpressed transcripts corresponding to reduced protein levels) to reveal the mechanisms that explain these inconsistencies. Network analysis of complex datasets (BOX 1) can incorporate data from more than one experiment and, indeed, more than one sample type. Identifying multiomics modules or constructing networks from other networks might help to enrich analyses in the future but is currently methodologically immature[198,199].

Fine mapping
Process by which a trait-associated region from a genome-wide association study is analysed to identify the particular genetic variants that are most likely to be causal.

From tissue homogenates to single cells

Although most CVDs arise on the background of a normal genome, the system is composed from a number of different cell types that will differentially contribute to measurements of omics profiles and whose relative abundance is still undetermined[200]. Multiomics approaches must respect the biology of the sample under analysis. For example, fibroblasts and endothelial cells comprise a small proportion of the cardiac mass, but these cells are likely to be the ones with the greatest contribution to nuclear content and therefore determine much of the cardiac mRNA and ncRNA content. Conversely, cardiomyocytes would contribute the most to the cardiac protein pool. Standalone analysis of whole-heart RNA and protein abundance would draw potentially inaccurate correlations as a result of these biases.

Microfluidics techniques will enable the omic analysis of single cells selected from a tissue, in theory allowing the generation of a complete spatiotemporal model of a cell or tissue[201–203]. Careful selection of cell and tissue types as well as of time during the disease course or under experimental conditions, might allow a detailed and integrated understanding of the cardiovascular system to be obtained but necessitates a thorough understanding of the biology at play and of the measurement techniques in use. For example, droplet RNA-seq is applicable to single-cell techniques owing to the amplification-based detection of reads, whereas single-cell proteomics is currently limited to multiplexed antibody assays or mass cytometry, necessarily reducing the number and increasing the bias of simultaneous measurements[204–206]. However, single-cell selection depends on the cell in question surviving the relevant isolation technique. More stable cells, such as lymphocytes or fibroblasts, might be robust enough to be collected in fluorescence-assisted cell-sorting experiments, but more labile cells, including neutrophils, are likely to degranulate or activate. Conversely, cardiomyocytes can be too large for conventional droplet RNA-seq. Approaches including the relative location of a single cell in a tissue block and different time points along the course of differentiation have demonstrated the overlapping and fluid nature of traditional cell-type categories. For example, unsupervised clustering analysis of single-cell RNA-seq data from fibroblasts differentiating to induced cardiomyocytes demonstrated a degree of heterogeneity among starting populations of fibroblasts and the consequent suppression of non-cardiomyocyte subgroups as differentiation occurred[207]. These approaches are increasingly applicable to 3D tissue samples[208]. Assuming this heterogeneity exists in vivo, the logical conclusion of the increasing resolution of both assays and sample selection is the generation of new categories of cell types that will, in turn, have to be subjected to empirical and reductionist experimental validation. This work has begun in a study using post-mortem human samples in which single-nuclei RNA-seq identified heterogeneity in cell subtypes between cardiac chambers and between sexes and revealed 20 subclusters of cell types in the heart[209]. However, the use of post-mortem samples complicates the analysis because of variations in RNA degradation rates.

Finally, exemplars of a systems biology approach in CVD that model single cellular or organelle function are available. Genome-scale metabolic networks constructed from gene ontology annotations and manual curation of experimental data can be used for metabolic-control analysis of metabolic pathways[210,211]. This technique has been applied to study cardiomyocytes, generating a metabolic model of the heart that simulated the interactions between 368 metabolic functions and can model the effects of perturbations in important metabolites such as glucose or fatty acids[212]. The application of flux balance analysis, in which the interaction between enzymes, metabolites and cellular function is modelled, identified a group of reactions predicted to be central to mitochondrial responses to hypoxia[213]. These putative reactions were mapped to SNPs that are prevalent in high-altitude dwelling populations.

More than the sum of its parts?

> *"To whatever degree we might imagine our knowledge of the properties of the several ingredients of a living body to be extended and perfected, it is certain that no mere summing up of the separate actions of those elements will ever amount to the action of the living body itself." (J. S. Mill A System of Logic Bk III, Ch. 6, §1)*[214]

The concept of emergent properties is fundamental to systems biology. Recall the systems description of the ECG (FIG. 1). A full understanding of the surface ECG requires knowledge only of the location and nature of the action potentials, which in turn require knowledge only of the ion channels, and below this level, only knowledge of the proteins of the channel and so on. However, it could be argued that the ECG itself contains information that could not be readily extracted from mere knowledge of each individual ion channel. For example, the PR interval can be measured only at the level of the ECG. The PR interval is an example of an emergent property, in which real, measurable and reproducible properties can be detected only at high levels of complexity and are not apparent at more basic levels. Thus, the core network of genes identified across multiple diseases in the STARNET studies could be considered as a list of genes, but their interaction en bloc with the remaining biological system might qualify them to be a single unit of observation, to be measured and studied as a network rather than as a collection of genes. Such thinking would require new, agreed and biologically relevant parameters for the measurement of these objects — a non-trivial task.

Current methodology in this regard is immature; no agreed parameters for biologically relevant networks have been set, let alone the ontology required to describe variations of the networks or the interactions between them. One salient finding from omics approaches is that biological systems are frequently highly redundant, with many parallel, regulating and interacting pathways. Single-molecule interventions might not be the only silver bullet cure to complex diseases. The use of more than one antiplatelet agent in current clinical practice already outlines the beginnings of a systems approach to CVD. The technical and methodological challenges stated here are likely to be overcome with time. However, proponents of emergent properties in medicine should describe what an intervention that treats a network or module would look like. Is there a therapy that could, even in principle, treat the ECG as an ECG and not its underpinning biology? In the same way that no therapy 'treats' the ECG, because the ECG is an integrated measurement of electrical activity, any intervention that would 'treat' a network would in fact target the network's components.

A second important philosophical warning must be made regarding the increasing enthusiasm for the application of more powerful analytical and computational methods to personalize medical approaches[215]. Although these approaches will in no doubt increase the amount of data available for analysis, the preceding decades of epidemiology research have highlighted the effects of practically (rather than in principle) immeasurable events on health and disease outcomes[216]. This high degree of apparently random variation might be theoretically captured but the degree to which a single individual could benefit from population-based approaches remains undetermined.

Conclusions

The role of systems biology in cardiology is attracting increasing attention[217,218]. A huge proliferation of reports has occurred in the past two decades as publicly available datasets have been published online and high-throughput analysis of tissue has become more widespread. However, the future for cardiovascular medicine is unlikely to be in reanalysing data but in leveraging informatics and gene-editing techniques to interrogate the function of GWAS-identified SNPs or the manipulation of networks, at least in animal models[219]. Combined with subsequent omics interrogations and careful experimental design, these techniques offer the potential to advance our understanding of gene-to-disease pathways.

Systems biology is an approach that provides essential tools for the analysis of the complex, multidimensional datasets generated by the omics technologies and, more widely, the move towards big data in health care. Omics experiments are themselves subject to bias and overinterpretation and cross-validation of the different technological platforms is essential. Investigators must carefully choose which publicly available dataset, if any, to include in a systems analysis. The use of network theory and machine learning can yield impressive results, but these methods are not yet standardized. The studies reviewed here are exemplars of high quality, crucially because they validate emergent findings from systems biology with empirical models. In the few successful cases, careful experimental design, including interventional studies and clinical trials, is required alongside the insights offered by bioinformatics analysis of omics approaches. We argue that, although appealing to emergent properties is tempting in order to capture these new findings in more simple concepts, we would side with the English philosopher William of Ockham when he states: "It is futile to do with more things that which can be done with fewer".

Published online 18 December 2020

Flux balance analysis
Mathematical method for simulating metabolic processes in genome-scale reconstructions of metabolic networks.

1. Yates, A. D. et al. Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688 (2020).
2. Gagniuc, P. & Ionescu-Tirgoviste, C. Gene promoters show chromosome-specificity and reveal chromosome territories in humans. *BMC Genomics* **14**, 278 (2013).
3. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10**, 155–159 (2009).
4. Barwari, T., Joshi, A. & Mayr, M. MicroRNAs in cardiovascular disease. *J. Am. Coll. Cardiol.* **68**, 2577–2584 (2016).
5. Wilhelm, M. et al. Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).
6. Shang, L. L. et al. Human heart failure is associated with abnormal C-terminal splicing variants in the cardiac sodium channel. *Circ. Res.* **101**, 1146–1154 (2007).
7. Rosas, P. C. et al. Phosphorylation of cardiac myosin-binding protein-C Is a critical mediator of diastolic function. *Circ. Hear. Fail.* **8**, 582–594 (2015).
8. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **17**, 451–459 (2016).
9. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727–730 (2002).
10. Russ, A. P. & Lampel, S. The druggable genome: an update. *Drug Discov. Today* **10**, 1607–1610 (2005).
11. Griffith, M. et al. DGIdb: mining the druggable genome. *Nat. Methods* **10**, 1209–1210 (2013).
12. Kohane, I. S. Using electronic health records to drive discovery in disease genomics. *Nat. Rev. Genet.* **12**, 417–428 (2011).
13. Denny, J. C. Chapter 13: mining electronic health records in the genomics Era. *PLoS Comput. Biol.* **8**, e1002823 (2012).
14. Jensen, P. B., Jensen, L. J. & Brunak, S. Mining electronic health records: towards better research applications and clinical care. *Nat. Rev. Genet.* **13**, 395–405 (2012).
15. Morley, K. I. et al. Defining disease phenotypes using national linked electronic health records: a case study of atrial fibrillation. *PLoS ONE* **9**, e110900 (2014).
16. Berger, B., Peng, J. & Singh, M. Computational solutions for omics data. *Nat. Rev. Genet.* **14**, 333–346 (2013).
17. Huang, S.-S. C. & Fraenkel, E. Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. *Sci. Signal.* **2**, ra40 (2009).
18. Auffray, C. et al. From genomic medicine to precision medicine: highlights of 2015. *Genome Med.* **8**, 12 (2016).
19. Kathiresan, S. et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* **41**, 334–341 (2009).
20. Schunkert, H. et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
21. Reilly, M. P. et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet* **377**, 383–392 (2011).
22. Vargas, J. D. et al. Common genetic variants and subclinical atherosclerosis in the multi-ethnic study of atherosclerosis. *Circulation* **128**, 230–236 (2013).
23. Ellinor, P. T. et al. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat. Genet.* **44**, 670–675 (2012).
24. Kao, W. H. L. et al. Genetic variations in nitric oxide synthase 1 adaptor protein are associated with sudden cardiac death in US white community-based populations. *Circulation* **119**, 940–951 (2009).
25. Villard, E. et al. A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur. Heart J.* **32**, 1065–1076 (2011).
26. Meder, B. et al. A genome-wide association study identifies 6p21 as novel risk locus for dilated cardiomyopathy. *Eur. Heart J.* **35**, 1069–1077 (2014).
27. Piñero, J. et al. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* **2015**, bav028 (2015).
28. Liu, D. J. et al. Exome-wide association study of plasma lipids in >300,000 individuals. *Nat. Genet.* **49**, 1758–1766 (2017).
29. Lu, X. et al. Exome chip meta-analysis identifies novel loci and East Asian–specific coding variants that contribute to lipid levels and coronary artery disease. *Nat. Genet.* **49**, 1722–1730 (2017).
30. Zhao, W. et al. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nat. Genet.* **49**, 1450–1457 (2017).
31. Larson, M. G. et al. Framingham Heart Study 100K project: genome-wide associations for cardiovascular disease outcomes. *BMC Med. Genet.* **8**, S5 (2007).
32. Shen, G.-Q. et al. Four SNPs on chromosome 9p21 in a South Korean Population Implicate a Genetic Locus that confers high cross-race risk for development of coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.* **28**, 360–365 (2008).
33. Shen, G.-Q. et al. Association between four SNPs on chromosome 9p21 and myocardial infarction is replicated in an Italian population. *J. Hum. Genet.* **53**, 144–150 (2008).
34. Schunkert, H. et al. Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. *Circulation* **117**, 1675–1684 (2008).
35. Anderson, J. L. et al. Genetic variation at the 9p21 locus predicts angiographic coronary artery disease prevalence but not extent and has clinical utility. *Am. Heart J.* **156**, 1155–1162.e2 (2008).
36. Samani, N. J. et al. Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357**, 443–453 (2007).
37. Helgadottir, A. et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491–1493 (2007).
38. Patel, R. S. et al. Genetic variants at chromosome 9p21 and risk of first versus subsequent coronary heart disease events: a systematic review and meta-analysis. *J. Am. Coll. Cardiol.* **63**, 2234–2245 (2014).
39. Patel, R. S. et al. Association of chromosome 9p21 with subsequent coronary heart disease events: A GENIUS-CHD study of individual participant data. *Circ. Genomic Precis. Med.* **12**, 161–172 (2019).
40. Holdt, L. M. et al. Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nat. Commun.* **7**, 12429 (2016).
41. Holdt, L. M. et al. ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. *Arterioscler. Thromb. Vasc. Biol.* **30**, 620–627 (2010).
42. Lo Sardo, V. et al. Unveiling the role of the most impactful cardiovascular risk locus through haplotype editing. *Cell* **175**, 1796–1810.e20 (2018).
43. Inouye, M. et al. Genomic risk prediction of coronary artery disease in 480,000 adults. *J. Am. Coll. Cardiol.* **72**, 1883–1893 (2018).
44. Abraham, G. et al. Genomic prediction of coronary heart disease. *Eur. Heart J.* **37**, 3267–3278 (2016).
45. Khera, A. V. et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N. Engl. J. Med.* **375**, 2349–2358 (2016).
46. Cornelis, M. C. et al. The gene, environment association studies consortium (GENEVA): maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. *Genet. Epidemiol.* **34**, 364–372 (2010).
47. Murcray, C. E., Lewinger, J. P. & Gauderman, W. J. Gene-environment interaction in genome-wide association studies. *Am. J. Epidemiol.* **169**, 219–226 (2009).
48. Mars, N. et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat. Med.* **26**, 549–557 (2020).
49. Sohail, M. et al. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife* **8**, e39702 (2019).
50. Berg, J. J. et al. Reduced signal for polygenic adaptation of height in UK biobank. *eLife* **8**, e39725 (2019).
51. Li, Y. et al. Statistical and functional studies identify epistasis of cardiovascular risk genomic variants from genome-wide association studies. *J. Am. Heart Assoc.* **9**, e014146 (2020).
52. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).
53. Ference, B. A. et al. Mendelian randomization study of ACLY and cardiovascular disease. *N. Engl. J. Med.* **380**, 1033–1042 (2019).
54. Ray, K. K. et al. Safety and efficacy of bempedoic acid to reduce LDL cholesterol. *N. Engl. J. Med.* **380**, 1022–1032 (2019).
55. Sanderson, E., Davey Smith, G., Windmeijer, F. & Bowden, J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int. J. Epidemiol.* **48**, 713–727 (2019).
56. Zuber, V., Colijn, J. M., Klaver, C. & Burgess, S. Selecting likely causal risk factors from high-throughput experiments using multivariable Mendelian randomization. *Nat. Commun.* **11**, 29 (2020).
57. Klarin, D. et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat. Genet.* **50**, 1514–1523 (2018).
58. Minikel, E. V. et al. Evaluating drug targets through human loss-of-function genetic variation. *Nature* **581**, 459–464 (2020).
59. Timpson, N. J. et al. C-reactive protein and its role in metabolic syndrome: Mendelian randomisation study. *Lancet* **366**, 1954–1959 (2005).
60. Richardson, T. G., Harrison, S., Hemani, G. & Smith, G. D. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *eLife* **8**, e43657 (2019).
61. Tillmann, T. et al. Education and coronary heart disease: Mendelian randomisation study. *BMJ* **358**, j3542 (2017).
62. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
63. Liu, X., Li, Y. I. & Pritchard, J. K. Trans effects on gene expression can drive omnigenic inheritance. *Cell* **177**, 1022–1034.e6 (2019).
64. Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J. & Visscher, P. M. Common disease is more complex than implied by the core gene omnigenic model. *Cell* **173**, 1573–1580 (2018).
65. Greene, C. S. et al. Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.* **47**, 569–576 (2015).
66. Huan, T. et al. A systems biology framework identifies molecular underpinnings of coronary heart disease. *Arterioscler. Thromb. Vasc. Biol.* **33**, 1427–1434 (2013).
67. Preuss, M. et al. Design of the Coronary Artery Disease Genome-wide Replication and Meta-analysis (CARDIoGRAM) study. *Circ. Cardiovasc. Genet.* **3**, 475–483 (2010).
68. Holm, H. et al. Several common variants modulate heart rate, PR interval and QRS duration. *Nat. Genet.* **42**, 117 (2010).
69. Chambers, J. C. et al. Genetic variation in SCN10A influences cardiac conduction. *Nat. Genet.* **42**, 149–152 (2010).
70. Pfeufer, A. et al. Genome-wide association study of PR interval. *Nat. Genet.* **42**, 153 (2010).
71. Noseworthy, P. A. & Newton-Cheh, C. Genetic determinants of sudden cardiac death. *Circulation* **118**, 1854–1863 (2008).
72. Lubitz, S. A. et al. Association between familial atrial fibrillation and risk of new-onset atrial fibrillation. *JAMA* **304**, 2263 (2010).
73. Lin, H. et al. Gene expression and genetic variation in human atria. *Heart Rhythm.* **11**, 266–271 (2014).
74. Kertai, M. D. et al. Genome-wide association study of new-onset atrial fibrillation after coronary artery bypass grafting surgery. *Am. Heart J.* **170**, 580 (2015).
75. Gupta, R. M. & Musunuru, K. Mapping novel pathways in cardiovascular disease using eQTL data: the past, present, and future of gene expression analysis. *Front. Genet.* **3**, 232 (2013).
76. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* **10**, 184–194 (2009).
77. Brem, R. B. et al. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
78. Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. *Nat. Rev. Genet.* **7**, 862–872 (2006).
79. Michaelson, J. J., Loguercio, S. & Beyer, A. Detection and interpretation of expression quantitative trait loci (eQTL). *Methods* **48**, 265–276 (2009).
80. Lappalainen, T., Scott, A. J., Brandt, M. & Hall, I. M. Genomic analysis in the age of human genome sequencing. *Cell* **177**, 70–84 (2019).
81. Stranger, B. E. et al. Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).

82. Lappalainen, T. et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
83. Ardlie, K. G. et al. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
84. GTEx Consortium et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
85. Musunuru, K. et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714–719 (2010).
86. Linsel-Nitschke, P. et al. Genetic variation at chromosome 1p13.3 affects sortilin mRNA expression, cellular LDL-uptake and serum LDL levels which translates to the risk of coronary artery disease. *Atherosclerosis* **208**, 183–189 (2010).
87. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
88. Wu, P. Y. et al. Cardiovascular transcriptomics and epigenomics using next-generation sequencing challenges, progress, and opportunities. *Circ. Cardiovasc. Genet.* **7**, 701–710 (2014).
89. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
90. Mohammadi, P. et al. Genetic regulatory variation in populations informs transcriptome analysis in rare disease. *Science* **366**, 351–356 (2019).
91. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
92. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
93. Miller, C. L. et al. Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat. Commun.* **7**, 12092 (2016).
94. Eichler, E. E. et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* **11**, 446–450 (2010).
95. Talukdar, H. A. et al. Cross-tissue regulatory gene networks in coronary artery disease. *Cell Syst.* **2**, 196–208 (2016).
96. Frades, I. et al. Systems pharmacology identifies an arterial wall regulatory gene network mediating coronary artery disease side effects of antiretroviral therapy. *Circ. Genomic Precis. Med.* **12**, 262–272 (2019).
97. Franzén, O. et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827–830 (2016).
98. Vilne, B. et al. Network analysis reveals a causal role of mitochondrial gene activity in atherosclerotic lesion formation. *Atherosclerosis* **267**, 39–48 (2017).
99. Walter, W. et al. Deciphering the dynamic transcriptional and post-transcriptional networks of macrophages in the healthy heart and after myocardial injury. *Cell Rep.* **23**, 622–636 (2018).
100. van Heesch, S. et al. The translational landscape of the human heart. *Cell* **178**, 242–260.e29 (2019).
101. Wang, S. et al. The endothelial-specific microRNA miR-126 governs vascular integrity and angiogenesis. *Dev. Cell* **15**, 261–271 (2008).
102. Zampetaki, A. et al. Plasma microRNA profiling reveals loss of endothelial miR-126 and other microRNAs in type 2 diabetes. *Circ. Res.* **107**, 810–817 (2010).
103. Willeit, P. et al. Circulating microRNAs as novel biomarkers for platelet activation. *Circ. Res.* **112**, 595–600 (2013).
104. Kaudewitz, D. et al. Association of microRNAs and YRNAs with platelet function. *Circ. Res.* **118**, 420–432 (2016).
105. Savitski, M. M. et al. Measuring and managing ratio compression for accurate iTRAQ/TMT quantification. *J. Proteome Res.* **12**, 3586–3598 (2013).
106. Werner, T. et al. Ion Coalescence of neutron encoded TMT 10-plex reporter ions. *Anal. Chem.* **86**, 3594–3601 (2014).
107. Mann, M. Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* **7**, 952–958 (2006).
108. Ong, S.-E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protoc.* **1**, 2650–2660 (2007).
109. Gstaiger, M. & Aebersold, R. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat. Rev. Genet.* **10**, 617–627 (2009).
110. Witze, E. S., Old, W. M., Resing, K. A. & Ahn, N. G. Mapping protein post-translational modifications with mass spectrometry. *Nat. Methods* **4**, 798–806 (2007).
111. Kusebauch, U. et al. Human SRMAtlas: a resource of targeted assays to quantify the complete human proteome. *Cell* **166**, 766–778 (2016).
112. Tu, C. et al. Depletion of abundant plasma proteins and limitations of plasma proteomics. *J. Proteome Res.* **9**, 4982–4991 (2010).
113. Schubert, O. T., Röst, H. L., Collins, B. C., Rosenberger, G. & Aebersold, R. Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat. Protoc.* **12**, 1289–1294 (2017).
114. Doerr, A. DIA mass spectrometry. *Nat. Methods* **12**, 35 (2015).
115. Bom, M. J. et al. Predictive value of targeted proteomics for coronary plaque morphology in patients with suspected coronary artery disease. *EBioMedicine* **39**, 109–117 (2019).
116. Langley, S. R. et al. Extracellular matrix proteomics identifies molecular signature of symptomatic carotid plaques. *J. Clin. Invest.* **127**, 1546–1560 (2017).
117. Willeit, K. et al. Association between vascular cell adhesion molecule 1 and atrial fibrillation. *JAMA Cardiol.* **2**, 516 (2017).
118. Smith, J. G. & Gerszten, R. E. Emerging affinity-based proteomic technologies for large-scale plasma profiling in cardiovascular disease. *Circulation* **135**, 1651–1664 (2017).
119. Rohloff, J. C. et al. Nucleic acid ligands with protein-like side chains: Modified aptamers and their use as diagnostic and therapeutic agents. *Mol. Ther. Nucleic Acids* **3**, e201 (2014).
120. Ngo, D. et al. Aptamer-based proteomic profiling reveals novel candidate biomarkers and pathways in cardiovascular disease. *Circulation* **134**, 270–285 (2016).
121. Benson, M. D. et al. Application of large-scale aptamer-based proteomic profiling to planned myocardial infarctions. *Circulation* **137**, 1270–1277 (2017).
122. Jacquet, S. et al. Identification of cardiac myosin-binding protein C as a candidate biomarker of myocardial infarction by proteomics analysis. *Mol. Cell. Proteom.* **8**, 2687–2699 (2009).
123. Marjot, J. et al. Quantifying the release of biomarkers of myocardial necrosis from cardiac myocytes and intact myocardium. *Clin. Chem.* **63**, 990–996 (2017).
124. Kaier, T. E. et al. Direct comparison of cardiac myosin-binding protein C with cardiac troponins for the early diagnosis of acute myocardial infarction. *Circulation* **136**, 1495–1508 (2017).
125. Ganz, P. et al. Development and validation of a protein-based risk score for cardiovascular outcomes among patients with stable coronary heart disease. *JAMA* **315**, 2532 (2016).
126. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
127. Williams, S. A. et al. Plasma protein patterns as comprehensive indicators of health. *Nat. Med.* **25**, 1851–1857 (2019).
128. Mosley, J. et al. Probing the virtual proteome to identify novel disease biomarkers. *Circulation.* **138**, 2469–2481 (2018).
129. Christiansson, L. et al. The use of multiplex platforms for absolute and relative protein quantification of clinical material. *EuPA Open Proteom.* **3**, 37–47 (2014).
130. Wei, W.-Q. et al. Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLoS ONE* **12**, e0175508 (2017).
131. Wadhera, R. K. et al. Temporal trends in unstable angina diagnosis codes for outpatient percutaneous coronary interventions. *JAMA Intern. Med.* **179**, 259–261 (2019).
132. Lehallier, B. et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019).
133. Rual, J.-F. et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
134. Vidal, M., Cusick, M. E. & Barabasi, A. L. Interactome networks and human disease. *Cell* **144**, 986–998 (2011).
135. Jensen, M. K. et al. Protein interaction-based genome-wide analysis of incident coronary heart disease. *Circ. Cardiovasc. Genet.* **4**, 549–556 (2011).
136. O'Reilly, F. J. & Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat. Struct. Mol. Biol.* **25**, 1000–1008 (2018).
137. Liao, Y. et al. The cardiomyocyte RNA-binding proteome: links to intermediary metabolism and heart disease. *Cell Rep.* **16**, 1456–1469 (2016).
138. Shah, S. H. & Newgard, C. B. Integrated metabolomics and genomics: systems approaches to biomarkers and mechanisms of cardiovascular disease. *Circ. Cardiovasc. Genet.* **8**, 410–419 (2015).
139. Hoefer, I. E. et al. Novel methodologies for biomarker discovery in atherosclerosis. *Eur. Heart J.* **36**, 2635–2642 (2015).
140. Mundra, P. A. et al. Large-scale plasma lipidomic profiling identifies lipids that predict cardiovascular events in secondary prevention. *JCI Insight* **3**, e121326 (2018).
141. Huynh, K. et al. High-throughput plasma lipidomics: detailed mapping of the associations with cardiometabolic risk factors. *Cell Chem. Biol.* **26**, 71–84.e4 (2019).
142. Karatasakis, A. et al. Effect of PCSK9 inhibitors on clinical outcomes in patients with hypercholesterolemia: a meta-analysis of 35 randomized controlled trials. *J. Am. Heart Assoc.* **6**, e006910 (2017).
143. Soininen, P. et al. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* **134**, 1781 (2009).
144. Sliz, E. et al. Metabolomic consequences of genetic inhibition of PCSK9 compared with statin treatment. *Circulation* **138**, 2499–2512 (2018).
145. Stegemann, C. et al. Lipidomics profiling and risk of cardiovascular disease in the prospective population-based bruneck study. *Circulation* **129**, 1821–1831 (2014).
146. Pechlaner, R. et al. Very-low-density lipoprotein–associated apolipoproteins predict cardiovascular events and are lowered by inhibition of APOC-III. *J. Am. Coll. Cardiol.* **69**, 789–800 (2017).
147. Gaudet, D. et al. Antisense inhibition of apolipoprotein C-III in patients with hypertriglyceridemia. *N. Engl. J. Med.* **373**, 438–447 (2015).
148. Würtz, P. et al. Metabolite profiling and cardiovascular event risk. *Circulation* **131**, 774–785 (2015).
149. Nordestgaard, B. G. & Varbo, A. Triglycerides and cardiovascular disease. *Lancet* **384**, 626–635 (2014).
150. Ference, B. A. et al. Association of triglyceride-lowering LPL variants and LDL-C–lowering LDLR variants with risk of coronary heart disease. *JAMA* **321**, 364 (2019).
151. Richardson, T. G. et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLoS Med.* **17**, e1003062 (2020).
152. Ng, T. W. K. et al. Association of Plasma ceramides and sphingomyelin with VLDL apoB-100 fractional catabolic rate before and after rosuvastatin treatment. *J. Clin. Endocrinol. Metab.* **100**, 2497–2501 (2015).
153. Gilbert, J. A. et al. Current understanding of the human microbiome. *Nat. Med.* **24**, 392–400 (2018).
154. Org, E. et al. Relationships between gut microbiota, plasma metabolites, and metabolic syndrome traits in the METSIM cohort. *Genome Biol.* **18**, 70 (2017).
155. Tang, W. H. W. & Hazen, S. L. The gut microbiome and its role in cardiovascular diseases. *Circulation* **135**, 1008–1010 (2017).
156. Jie, Z. et al. The gut microbiome in atherosclerotic cardiovascular disease. *Nat. Commun.* **8**, 845 (2017).
157. Walter, J., Armet, A. M., Finlay, B. B. & Shanahan, F. Establishing or exaggerating causality for the gut microbiome: lessons from human microbiota-associated rodents. *Cell* **180**, 221–232 (2020).
158. Mayr, M. et al. Proteomic and metabolomic analyses of atherosclerotic vessels from apolipoprotein E-deficient mice reveal alterations in inflammation, oxidative stress, and energy metabolism. *Arterioscler. Thromb. Vasc. Biol.* **25**, 2135–2142 (2005).
159. Klipfell, E. et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
160. Tang, W. H. W. et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* **368**, 1575–1584 (2013).
161. Haghikia, A. et al. Gut microbiota-dependent trimethylamine N-oxide predicts risk of cardiovascular events in patients with stroke and is related to proinflammatory monocytes. *Arterioscler. Thromb. Vasc. Biol.* **38**, 2225–2235 (2018).
162. Zhu, W. et al. Gut microbial metabolite TMAO enhances platelet hyperreactivity and thrombosis risk. *Cell* **165**, 111–124 (2016).

163. Roberts, A. B. et al. Development of a gut microbe–targeted nonlethal therapeutic to inhibit thrombosis potential. *Nat. Med.* **24**, 1407–1417 (2018).
164. van Mens, T. E., Büller, H. R. & Nieuwdorp, M. Targeted inhibition of gut microbiota proteins involved in TMAO production to reduce platelet aggregation and arterial thrombosis: a blueprint for drugging the microbiota in the treatment of cardiometabolic disease? *J. Thromb. Haemost.* **17**, 3–5 (2019).
165. Manor, O. et al. A multi-omic association study of trimethylamine N-oxide. *Cell Rep.* **24**, 935–946 (2018).
166. Heianza, Y. et al. Long-term changes in gut microbial metabolite trimethylamine N-oxide and coronary heart disease risk. *J. Am. Coll. Cardiol.* **75**, 763–772 (2020).
167. Gencer, B. et al. Gut microbiota-dependent trimethylamine N-oxide and cardiovascular outcomes in patients with prior myocardial infarction: a nested case control study from the PEGASUS-TIMI 54 Trial. *J. Am. Heart Assoc.* **9**, e015331 (2020).
168. Vieira-Silva, S. et al. Statin therapy is associated with lower prevalence of gut microbiota dysbiosis. *Nature* **581**, 310–315 (2020).
169. Chan, S. Y. & Loscalzo, J. The emerging paradigm of network medicine in the study of human disease. *Circ. Res.* **111**, 359–374 (2012).
170. Pasea, L. et al. Personalising the decision for prolonged dual antiplatelet therapy: development, validation and potential impact of prognostic models for cardiovascular events and bleeding in myocardial infarction survivors. *Eur. Heart J.* **38**, 1048–1055 (2017).
171. Needham, C. J., Bradford, J. R., Bulpitt, A. J. & Westhead, D. R. A primer on learning in bayesian networks for computational biology. *PLoS Comput. Biol.* **3**, e129 (2007).
172. Shilaskar, S. & Ghatol, A. Feature selection for medical diagnosis: evaluation for cardiovascular diseases. *Expert Syst. Appl.* **40**, 4146–4153 (2013).
173. Sanz, J. A. et al. Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system. *Appl. Soft Comput.* **20**, 103–111 (2014).
174. Libbrecht, M. W. & Noble, W. S. Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* **16**, 321–332 (2015).
175. Tarca, A. L. et al. Machine learning and its applications to biology. *PLoS Comput. Biol.* **3**, e116 (2007).
176. Liu, Y. et al. Beatquency domain and machine learning improve prediction of cardiovascular death after acute coronary syndrome. *Sci. Rep.* **6**, 34540 (2016).
177. Shah, S. J. et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation* **131**, 269–279 (2015).
178. Sengupta, P. P. et al. Cognitive machine-learning algorithm for cardiac imaging. *Circ. Cardiovasc. Imaging* **9**, e004330 (2016).
179. Motwani, M. et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur. Heart J.* **38**, 500–507 (2016).
180. Raghunath, S. et al. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nat. Med.* **26**, 886–891 (2020).
181. Dey, D. et al. Artificial intelligence in cardiovascular imaging: jacc state-of-the-art review. *J. Am. Coll. Cardiol.* **73**, 1317–1335 (2019).
182. McNally, E. M. et al. Genetic mutations and mechanisms in dilated cardiomyopathy. *J. Clin. Invest.* **123**, 19–26 (2013).
183. Hershberger, R. E., Hedges, D. J. & Morales, A. Dilated cardiomyopathy: the complexity of a diverse genetic architecture. *Nat. Rev. Cardiol.* **10**, 531–547 (2013).
184. Stark, K. et al. Genetic association study identifies HSPB7 as a risk gene for idiopathic dilated cardiomyopathy. *PLoS Genet.* **6**, e1001167 (2010).
185. Norton, N. et al. Exome sequencing and genome-wide linkage analysis in 17 families illustrate the complex contribution of TTN truncating variants to dilated cardiomyopathy. *Circ. Cardiovasc. Genet.* **6**, 144–153 (2013).
186. Camargo, A. & Azuaje, F. Identification of dilated cardiomyopathy signature genes through gene expression and network data integration. *Genomics* **92**, 404–413 (2008).
187. Liu, Y. et al. RNA-seq identifies novel myocardial gene expression signatures of heart failure. *Genomics* **105**, 83–89 (2015).
188. Lau, E. et al. Integrated omics dissection of proteome dynamics during cardiac remodeling. *Nat. Commun.* **9**, 120 (2018).
189. Isserlin, R. et al. Systems analysis reveals down-regulation of a network of pro-survival miRNAs drives the apoptotic response in dilated cardiomyopathy. *Mol. Biosyst.* **11**, 239–251 (2015).
190. Inouye, M. et al. An immune response network associated with blood lipid levels. *PLoS Genet.* **6**, e1001113 (2010).
191. Inouye, M. et al. Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol. Syst. Biol.* **6**, 441 (2010).
192. Bartel, J. et al. The human blood metabolome-transcriptome interface. *PLoS Genet.* **11**, e1005274 (2015).
193. Tabassum, R. et al. Genetic architecture of human plasma lipidome and its link to cardiovascular disease. *Nat. Commun.* **10**, 4329 (2019).
194. Inouye, M. et al. Novel loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. *PLoS Genet.* **8**, e1002907 (2012).
195. Gallois, A. et al. A comprehensive study of metabolite genetics reveals strong pleiotropy and heterogeneity across time and context. *Nat. Commun.* **10**, 4788 (2019).
196. Voros, S. et al. Precision phenotyping, panomics, and system-level bioinformatics to delineate complex biologies of atherosclerosis: Rationale and design of the 'Genetic Loci and the Burden of Atherosclerotic Lesions' study. *J. Cardiovasc. Comput. Tomogr.* **8**, 442–451 (2014).
197. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biol.* **18**, 83 (2017).
198. Choobdar, S. et al. Assessment of network module identification across complex diseases. *Nat. Methods* **16**, 843–852 (2019).
199. Bennett, L., Kittas, A., Muirhead, G., Papageorgiou, L. G. & Tsoka, S. Detection of composite communities in multiplex biological networks. *Sci. Rep.* **5**, 10345 (2015).
200. Pinto, A. R. et al. Revisiting cardiac cellular composition. *Circ. Res.* **118**, 400–409 (2016).
201. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* **17**, 175–188 (2016).
202. Tang, F. et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
203. Wang, D. & Bodovitz, S. Single cell analysis: the new frontier in 'omics'. *Trends Biotechnol.* **28**, 281–290 (2010).
204. Thul, P. J. et al. A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).
205. Spitzer, M. H. & Nolan, G. P. Mass cytometry: single cells, many features. *Cell* **165**, 780–791 (2016).
206. Sun, Z. et al. A Bayesian mixture model for clustering droplet-based single-cell transcriptomic data from population studies. *Nat. Commun.* **10**, 1649 (2019).
207. Liu, Z. et al. Single-cell transcriptomics reconstructs fate conversion from fibroblast to cardiomyocyte. *Nature* **551**, 100–104 (2017).
208. Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
209. Tucker, N. R. et al. Transcriptional and cellular diversity of the human heart. *Circulation* **142**, 466–482 (2020).
210. Cascante, M. et al. Metabolic control analysis in drug discovery and disease. *Nat. Biotechnol.* **20**, 243–249 (2002).
211. Duarte, N. C. et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl Acad. Sci. USA* **104**, 1777–1782 (2007).
212. Karlstädt, A. et al. CardioNet: a human metabolic network suited for the study of cardiomyocyte metabolism. *BMC Syst. Biol.* **6**, 114 (2012).
213. Edwards, L. M. et al. Genome-scale methods converge on key mitochondrial genes for the survival of human cardiomyocytes in hypoxia. *Circ. Cardiovasc. Genet.* **7**, 407–415 (2014).
214. J. S. Mill *A System of Logic* Bk III, Ch. 6, §1
215. Califf, R. M. Future of personalized cardiovascular medicine: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* **72**, 3301–3309 (2018).
216. Smith, G. D. Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice. *Int. J. Epidemiol.* **40**, 537–562 (2011).
217. Trachana, K. et al. Taking systems medicine to heart. *Circ. Res.* **122**, 1276–1289 (2018).
218. Leopold, J. A. & Loscalzo, J. Emerging role of precision medicine in cardiovascular disease. *Circ. Res.* **122**, 1302–1315 (2018).
219. Lempiäinen, H. et al. Network analysis of coronary artery disease risk genes elucidates disease mechanisms and druggable targets. *Sci. Rep.* **8**, 3434 (2018).
220. Chan, S. Y. & Loscalzo, J. The emerging paradigm of network medicine in the study of human disease. *Circ. Res.* **111**, 359–374 (2012).
221. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
222. Ma'ayan, A. Introduction to network analysis in systems biology. *Sci. Signal.* **4**, tr5 (2011).
223. Schadt, E. E. & Lum, P. Y. Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. Reverse engineering gene networks to identify key drivers of complex disease phenotypes. *J. Lipid Res.* **47**, 2601–2613 (2006).
224. Cheng, L. et al. Recurrent neural network for non-smooth convex optimization problems with application to the identification of genetic regulatory networks. *IEEE Trans. Neural Netw.* **22**, 714–726 (2011).
225. McGeachie, M. J. et al. CGBayesNets: conditional gaussian bayesian network learning and inference with mixed discrete and continuous data. *PLoS Comput. Biol.* **10**, e1003676 (2014).
226. Sarajlic´, A., Janjic´, V., Stojkovic´, N., Radak, D. J. & Pržulj, N. Network topology reveals key cardiovascular disease genes. *PLoS ONE* **8**, e71537 (2013).
227. Dewey, F. E. et al. Gene coexpression network topology of cardiac development, hypertrophy, and failure. *Circ. Cardiovasc. Genet.* **4**, 26–35 (2011).
228. Ravasz, E. et al. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
229. de la Fuente, A. From 'differential expression' to 'differential networking' - identification of dysfunctional regulatory networks in diseases. *Trends Genet.* **26**, 326–333 (2010).

Acknowledgements
A.J. is a British Heart Foundation (BHF) Clinical Research Training Fellow (FS/16/32/32184). M.M. is a BHF Chair Holder (CH/16/3/32406) and receiving BHF programme grant support (RG/16/14/32397). Research by M.M. was made possible through support from the BIRAX Regenerative Medicine Initiative and funding from the EU Horizon 2020 Research and Innovation Programme under Marie Skłodowska-Curie grant agreement No 813716 (TRAIN-HEART), the Leducq Foundation (13CVD02 and 18CVD02), the Excellence Initiative VASCage (Centre for Promoting Vascular Health in the Ageing Community, project number 868624) of the Austrian Research Promotion Agency FFG (COMET program – Competence Centers for Excellent Technologies) funded by the Austrian Ministry for Transport, Innovation and Technology, the Austrian Ministry for Digital and Economic Affairs and the federal states Tyrol (via Standortagentur), Salzburg, and Vienna (via Vienna Business Agency), and the National Institute of Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS (National Health Service) Foundation Trust and King's College London in partnership with King's College Hospital. We thank Raimund Pechlaner (Medical University of Innsbruck, Austria) for providing Fig. 4.

Author contributions
All authors researched data for the article, discussed its content, wrote the manuscript, and reviewed and edited it before submission.

Competing interests
M.M. is a named inventor on patents for cardiovascular biomarkers. The other authors declare no competing interests.

Peer review information
Nature Reviews Cardiology thanks M. Ala-Korpela and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

RELATED LINKS
CALIBER research: https://caliberresearch.org/portal
TopHat: https://ccb.jhu.edu/software/tophat/index.shtml