Published online 28 May 2021 *Nucleic Acids Research, 2021, Vol. 49, Web Server issue* ***W375–W387***
https://doi.org/10.1093/nar/gkab405

Mergeomics 2.0: a web server for multi-omics data integration to elucidate disease networks and predict therapeutics

Jessica Ding [1,2,†], Montgomery Blencowe [1,2,†], Thien Nghiem[1,†], Sung-min Ha[1], Yen-Wei Chen[1,3], Gaoyan Li[1,2] and Xia Yang [1,2,3,4,5,*]

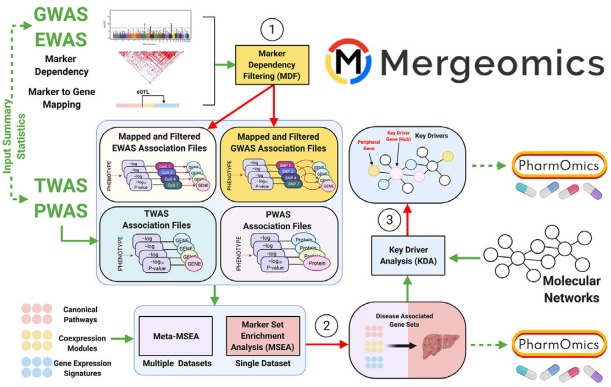
[1]Department of Integrative Biology and Physiology, University of California, Los Angeles, 610 Charles E. Young Drive East, Los Angeles, CA 90095, USA, [2]Interdepartmental Program of Molecular, Cellular and Integrative Physiology, University of California, Los Angeles, 610 Charles E. Young Drive East, Los Angeles, CA 90095, USA, [3]Interdepartmental Program of Molecular Toxicology, University of California, Los Angeles, 610 Charles E. Young Drive East, Los Angeles, CA 90095, USA, [4]Interdepartmental Program of Bioinformatics, University of California, Los Angeles, 610 Charles E. Young Drive East, Los Angeles, CA 90095, USA and [5]Institute for Quantitative and Computational Biosciences, University of California, Los Angeles, 610 Charles E. Young Drive East, Los Angeles, CA 90095, USA

Received February 28, 2021; Revised April 28, 2021; Editorial Decision April 30, 2021; Accepted May 02, 2021

ABSTRACT

The Mergeomics web server is a flexible online tool for multi-omics data integration to derive biological pathways, networks, and key drivers important to disease pathogenesis and is based on the open source Mergeomics R package. The web server takes summary statistics of multi-omics disease association studies (GWAS, EWAS, TWAS, PWAS, etc.) as input and features four functions: Marker Dependency Filtering (MDF) to correct for known dependency between omics markers, Marker Set Enrichment Analysis (MSEA) to detect disease relevant biological processes, Meta-MSEA to examine the consistency of biological processes informed by various omics datasets, and Key Driver Analysis (KDA) to identify essential regulators of disease-associated pathways and networks. The web server has been extensively updated and streamlined in version 2.0 including an overhauled user interface, improved tutorials and results interpretation for each analytical step, inclusion of numerous disease GWAS, functional genomics datasets, and molecular networks to allow for comprehensive omics integrations, increased functionality to decrease user workload, and increased flexibility to cater to user-specific needs. Finally, we have incorporated our newly developed drug repositioning pipeline PharmOmics for prediction of potential drugs targeting disease processes that were identified by Mergeomics. Mergeomics is freely accessible at http://mergeomics.research.idre.ucla.edu and does not require login.

GRAPHICAL ABSTRACT



INTRODUCTION

The advent of omics technologies has made significant strides in unveiling various disease-associated genetic and epigenetic variants, genes, proteins and metabolites. The ever-growing source of multi-omics datasets available including genomics, epigenomics, transcriptomics, proteomics and metabolomics now presents a new challenge of integrating these different data types for more meaningful and holistic interpretation of complex diseases. To conduct a comprehensive investigation of disease pathogenesis, we

*To whom correspondence should be addressed. Tel: +1 310 206 1812; Fax: +1 310 206 9184; Email: xyang123@ucla.edu
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

must consider multiple omics layers that contribute to biological complexity (1). The computational pipeline Mergeomics was developed to meet the need for multi-omics integration and functional interpretation to obtain mechanistic understanding. Mergeomics provides flexibility to incorporate the full spectrum of summary statistics (not just top hits) of individual layers of omics or multi-omics data simultaneously along with diverse functional genomics data across data types, studies and species. As such, genome-wide association studies (GWAS) as well as epigenome- (EWAS), transcriptome- (TWAS), proteome- (PWAS) and metabolome-wide association studies (MWAS) can all be accommodated.

The development of our Mergeomics tool follows the philosophy of utilizing a systems biology approach to unravel the complex interactions across molecular domains as well as cell types, tissues and organ systems that occur in disease. In particular, we are guided by the omnigenic disease model (2), which states that a large proportion of the genome likely contributes to disease pathogenesis through molecular interactions both within and between tissues. Utilizing this data-driven analysis considering the interactions among different omics layers and tissue contexts will uncover global maps to identify critical targets in disease pathogenesis, which can be followed by experimental approaches to investigate the detailed events that occur through the predicted molecules or pathways.

With the abundance of omics data available, it is unsurprising that various tools or methods have been developed to better integrate and interpret these datasets (3–5). These tools can be broadly categorized into two application categories: multi-omics biomarker predictions of diseases or subtypes (i.e. uncovering correlative or predictive but not necessarily disease-causing features) or mechanistic understanding of disease pathogenesis (i.e. regulators, molecular interactions and processes involved in disease development). Mergeomics focuses on mechanistic modeling but not predictive modeling. In terms of approaches, fusion (such as PFA (6), SNF (7), PSDF (8)), Bayesian (e.g. iCluster (9), PSDF (8), BCC (10)), correlation, multivariate (e.g. MFA (11), IntegrOmics (12), MixOmics (13)), pathway and network methods (PARADIGM (14), SNF (7), iOmicsPASS (15), MiBiOmics (16), Lemon-Tree (17), PaintOmics (18), NetICS (19), Metascape (20)) have been implemented (3–5). Mergeomics falls within the network method category that mainly focuses on understanding disease pathogenesis through uncovering multiple molecular targets within biological processes important to disease. The benefit of a network approach over other integrative options is its ability to provide biological interpretability, which is reliant not on the identification of latent structures through mathematical deconvolution but on the utilization of prior information based on molecular interactions, which can help provide clear targetable options (e.g. genes) in disease. Compared to other tools, Mergeomics not only accommodates diverse data types (GWAS, EWAS, TWAS, PWAS, MWAS) from different sources, studies, or species for a given disease, but also considers relationships between omics layers through functional genomics such as expression quantitative trait loci (eQTLs), molecular pathways, and tissue-specific gene regulatory networks to derive disease networks and predict therapeutics. Mergeomics also uses full summary statistics, not raw data or lists of top associations, as input, thereby reducing the need for raw data processing and harmonization and for pre-determining a specific cutoff to call for significant markers. Mergeomics has the ability to conduct pathway analysis and model gene regulatory networks, protein-protein interaction networks, and transcription factor networks in order to predict and visualize network regulators of disease. These unique features help maximize the utility of existing datasets and overcome limitations of other tools which utilize a narrower range of multi-omics data sources, do not provide mechanistic interpretations, or require programming skills with no intuitive web server for ease of use.

Since the release of the open source Mergeomics R package (https://bioconductor.org/packages/release/bioc/html/Mergeomics.html) (21) and web server in 2016 (22), this tool has been used to model a diverse set of diseases including cardiometabolic disorders such as non-alcoholic fatty liver disease (23), cardiovascular disease (24–26) and type 2 diabetes (27), autoimmunity including psoriasis (28) and rheumatoid arthritis (29), alcohol dependence (30), brain injury (31), Sjogren's syndrome (32) and environmental contributions to disease (33–35). Importantly, multiple validations of molecular predictions from Mergeomics with *in silico*, *in vitro* and *in vivo* studies highlight the validity and causal nature of the disease network predictions (23,27–28,31,35–40). Due to increasing demand for multi-omics integration and interpretation from scientists with different areas of expertise, we have implemented major revisions and improvements on the Mergeomics web server. Specifically, we have redesigned the user interface, simplified workflows, offered detailed tutorials and case studies, and provided more datasets and network models for utilization. The Mergeomics 2.0 web server offers the scientific community much-improved accessibility to our pipeline, caters to each user's specific goals in multi-omics studies, and addresses a broad range of biological questions, particularly emphasizing a mechanistic understanding of disease pathogenesis and prediction of potential therapeutics based on mechanistic understanding.

OVERVIEW AND UPDATES ON THE CORE FUNCTIONS OF MERGEOMICS

Overview of core functions

Mergeomics 2.0 features four core functions as previously implemented in version 1.0 with an addition of a new function. First, we provide a preprocessing tool, Marker Dependency Filtering (MDF) to remove omics marker redundancies such as linkage disequilibrium (LD) between single nucleotide polymorphisms (SNPs). Second, Marker Set Enrichment Analysis (MSEA) is used to identify omics-informed disease processes through the integrations of omics markers such as SNPs with functional genomics, canonical pathways, or co-expression networks. Third, Meta-MSEA runs MSEA on multiple datasets and conducts pathway/network level meta-analysis to retrieve consistent disease processes informed across datasets. Fourth, Key Driver Analysis (KDA) pinpoints network regulators

of disease processes based on the topology of biological networks. In Mergeomics 2.0, we added a new functional module called PharmOmics, which takes as input multi-omics-informed disease pathways or networks from Mergeomics to match with drug signatures to predict potential therapeutic drugs.

Introduction of PharmOmics into Mergeomics 2.0

We have recently developed a novel species- and tissue-specific network-based drug repositioning tool, PharmOmics, which is based on *in vivo* molecular studies of drugs (41). PharmOmics is a complementary drug repositioning tool to other existing tools, such as CMap (42) and LINCS L1000 (43), which are mostly based on *in vitro* cell line data. We provide two drug repositioning methods: network-based drug repositioning and gene overlap-based drug repositioning. Network-based drug repositioning ranks drugs based on the degree of connectivity of genes influenced by drug treatments to disease gene signatures in a given gene network model (44). Gene overlap-based drug repositioning is based on the degree of direct overlap between drug genes and disease genes. Users can directly input their disease pathway results from MSEA (genes from disease pathways are used as input) or KDA (genes from the disease network or significant key drivers (KDs) are used as input). For both MSEA and KDA, specific gene sets can be input into drug repositioning for a more refined analysis. As PharmOmics is based on gene expression studies, inputs are limited to genes or proteins. Users can also input their genes of interest into PharmOmics for drug repositioning analysis without running any other functions in Mergeomics.

Flexible workflows using the core functions

Each of the main functions of Mergeomics described above can be utilized as a standalone analysis tool or can be combined into a multi-step workflow with several different cases as portrayed in Figure 1. There are four cases or starting points that a user has the option to select. In case one, the user has one GWAS dataset and is prompted first to run MDF where they provide their association dataset, mapping data (e.g. SNP to gene), and marker dependency data (LD in the case of GWAS) to retrieve corrected SNP associations and mapping files. The MDF step is optional if the user does not wish to correct for LD, although we highly recommend this correction to avoid statistical artefacts due to LD. These results along with a gene set are fed into MSEA to uncover disease-associated pathways, which can be further analyzed in KDA to identify key regulators or PharmOmics for drug repositioning. In case two, the user has EWAS, TWAS, PWAS or MWAS data, and they are led to MSEA, where MDF and marker mapping are optional. As in the GWAS path, results from MSEA can be carried to KDA or PharmOmics. In case three, the user has multiple omics datasets and utilizes Meta-MSEA, which will run MSEA on each dataset and then conduct a meta-analysis across datasets to retrieve consistent biological processes, which can be input into PharmOmics or KDA. Finally, in case four, the user has a gene set and network of interest and can directly run KDA, which will provide KD genes and a subnetwork visualization of the top KDs, and the KDs or subnetwork can be input into PharmOmics to predict drugs.

Update on Marker Dependency Filtering (MDF)

MDF prepares input files for MSEA by correcting for dependency between omics markers and is an optional function. This preprocessing step is most commonly used for GWAS data to correct for LD between SNPs and filter out redundant SNPs, which is critical for removing redundant association signals that can result in statistical and biological artefacts in downstream analysis. Another purpose of MDF is to link the SNPs to potential downstream genes based on functional evidence, such as tissue-specific eQTLs. Correcting for dependency between other omics markers is currently seldom used. However, this feature can be utilized to correct for dependency between other types of markers (methylation sites, transcripts, etc.), if desired. MDF uses as input an association file which details markers (e.g. SNPs) and their disease association strengths (e.g. $-\log_{10}$ *P*-values or effect size, note that *P*-values are prohibited as MDF ranks larger values as stronger association strength, which is opposite of *P*-values), a mapping file used for marker to gene mapping (e.g. SNPs are mapped to genes to be enriched for gene sets), and a marker dependency file indicating the dependency between markers (e.g. LD between SNPs, to remove redundant markers) (Figure 2). The resulting corrected association and mapping files are then used as input to MSEA. MDF also allows for the selection of a top percentage of markers (50% or 25% recommended) to be considered in the analysis which reduces noise from low signal markers.

Updates to MDF include an increased number of marker to gene mapping options such as the addition of all available tissue-specific Genotype-Tissue Expression project (GTEx) (45) cis-eQTLs and splicing QTLs (cis-sQTLs) (Table 1), the ability to combine up to five mapping options, and the inclusion of LD files for all 26 populations from 1000 Genomes (1000G) (46) and methylation disequilibrium from EWAS software 2.0 (47). For analysis starting from GWAS data, MDF is a default preprocessing step, but we have included the option to skip MDF. For analytical paths starting from other omics data, users have the option to add MDF if needed.

Update on Marker Set Enrichment Analysis (MSEA)

In MSEA, full summary statistics of omics markers such as SNPs from GWAS, epigenetic sites from EWAS, genes from TWAS, proteins from PWAS, or metabolites from MWAS and their disease association values are taken as input and are integrated with functional genomics, canonical pathways, or co-expression networks to retrieve disease-associated pathways and networks. MSEA calculates and summarizes enrichment of disease/trait omics markers in sets of functionally related genes, such as canonical pathways and co-expression networks, across a range of statistical cutoffs in the full summary statistics using a chi-square-like statistic and then uses permutation to determine statistical *P*-values for the enrichment. We emphasize the importance to provide the association strength of the given

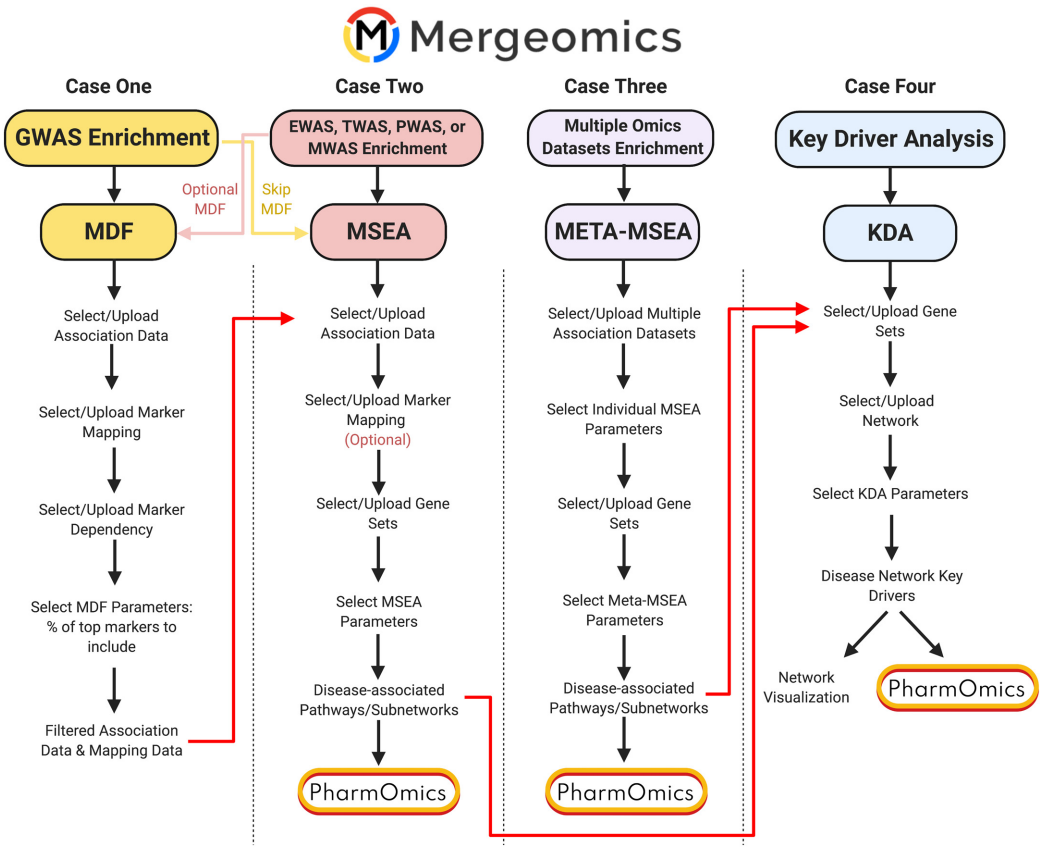


Figure 1. Workflow of Mergeomics. We provide four options on the web server to tailor to the user's data type. Case One: Individual GWAS analysis. For GWAS datasets we advise utilizing the MDF function; however, we also provide the ability to skip MDF and directly run MSEA and follow the workflow to PharmOmics or KDA. Case Two: Individual EWAS, TWAS, PWAS or MWAS analysis. In this case, we directly start at MSEA without MDF; however, we also provide the ability to utilize the MDF function if needed. From here the user can feed the MSEA results into PharmOmics or KDA. Case Three: Multi-omics analysis. If the user has multiple omics of the same type (e.g. two GWAS) or different types (e.g. TWAS and EWAS), they can utilize the Meta-MSEA function to derive disease-associated pathways and can input their results into PharmOmics or KDA. Case Four: A gene list(s) to run KDA. The user in this case can upload their gene sets of interest and upload or select a network to derive KD genes and visualize top KD subnetworks. The disease subnetwork or significant KDs can be fed into PharmOmics for drug repositioning.

marker wherein a larger number reflects greater association such as $-\log_{10}$ *P*-values or effect size to avoid incorrect downstream analysis and interpretability.

MSEA is able to analyze diverse data types, and each has different considerations of inputs which was partly described in the above MDF section (Figure 2). The output from MSEA can be interpreted as omics-informed disease pathways or networks. If GWAS is used, MSEA results can imply causal disease processes since GWAS carries causal inference. For other omics data, the MSEA results can only be interpreted as disease-associated processes but may or may not be causal. Considering GWAS along with other omics data, in our opinion, is a useful way to identify causal genes and processes. We also advise the user to take care in their interpretation of the names or annotations of pathways deemed to be significant (FDR < 0.05) as some can be misleading. Attention to the genes enriched in a given pathway derived from the input dataset should be checked in the gene details output file to confirm whether the pathway name is indeed appropriate as the genes may be more suitable or representative of another biological process. A user can conclude the analysis with results from MSEA or use the MSEA results as input to KDA with a user-defined statistical cutoff to identify network KDs of the disease processes based on molecular network topology.

In Mergeomics 2.0, we added the ability to use disease-associated gene sets derived from MSEA as input to PharmOmics for drug repositioning analysis, selecting either specific gene sets or by false discovery rate (FDR) or *P*-value threshold, to pinpoint drugs whose gene signatures align with those of the disease-associated gene sets identified by MSEA.

Update on Meta-MSEA

Meta-MSEA allows for integration of multiple datasets of the same omics type (e.g. two or more GWAS datasets) or multiple omics types (e.g. GWAS, EWAS, TWAS) and runs MSEA for each omics dataset followed by a meta-analysis. This integration reveals consistencies and differences in biological perturbation across different omics types or different studies of the same omics type.

In Mergeomics 2.0, we improved the guidance of running Meta-MSEA in regard to the differences in preprocessing of the different types of omics data. In addition, we have increased the flexibility of this analysis to allow for spe-

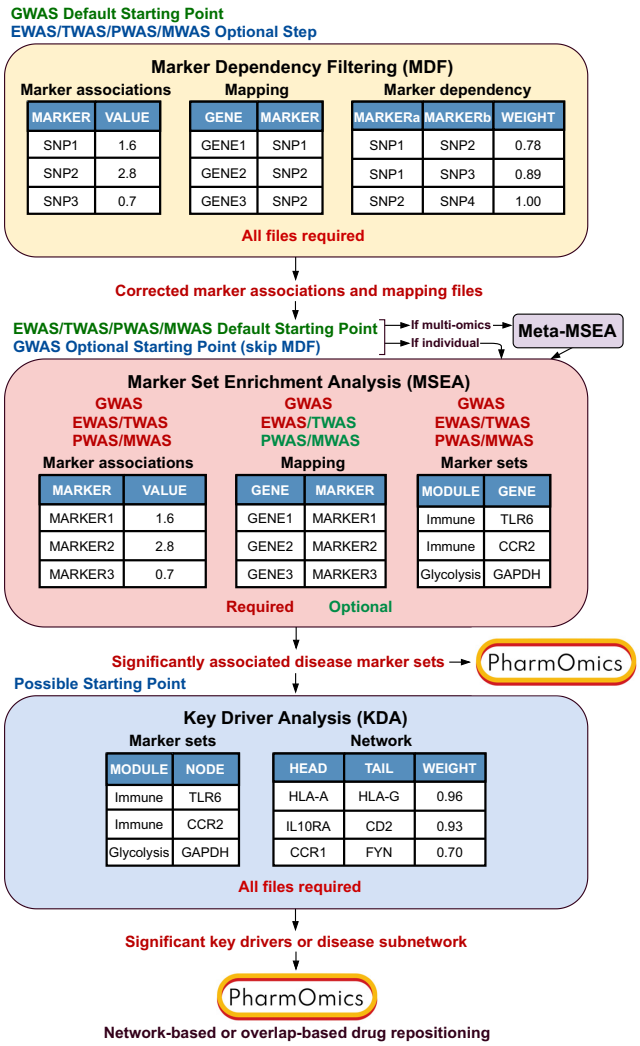


Figure 2. Mergeomics pipeline inputs. MDF is the default starting point for GWAS analysis and is an optional step for EWAS/TWAS/PWAS/MWAS. MDF requires marker-disease associations, a marker-gene mapping file, and a marker dependency file. Users with GWAS data can also skip MDF and run MSEA directly. MDF produces corrected marker-disease associations and marker-gene mapping files containing independent markers that are used for MSEA. For MSEA, required files for all datasets are the marker-disease associations and marker sets (pathway/modules). The marker to gene mapping file is required for GWAS and EWAS and optional for MWAS, TWAS and PWAS. Disease-associated marker sets from MSEA can be fed into KDA, which requires gene sets and a network. KDA can also be a starting point of analysis. Disease-associated gene sets from MSEA or KDs and disease subnetwork from KDA can be fed into PharmOmics for drug repositioning.

cific inputs and parameters for each association data. After each individual omics dataset is added, the user will be able to review which datasets have been successfully uploaded and their individual MSEA parameters with the option to add additional datasets or delete certain datasets, providing an easy way to track all the different inputs. As in the results generated from the individual MSEA, significantly associated gene sets from Meta-MSEA can be used as input to KDA or PharmOmics drug repositioning. We have also implemented user-defined individual MSEA FDR cutoffs to KDA in that the disease-associated pathways must pass all individual MSEA FDR cutoffs as well as the meta-FDR to be used in KDA, allowing the user to focus on the most consistent and robust disease processes across different datasets. In addition, we now provide heterogeneity statistics from Cochran's Q test to indicate the variability between datasets.

Update on Key Driver Analysis (KDA)

KDA identifies essential regulators of disease-associated pathways and networks, which are then visualized in the web browser using Cytoscape.js (Figure 3). KDA results can also be downloaded as network files ready to be used on Cytoscape Desktop for further customization of the network visualization. A Chi-square-like statistic, $\chi = \frac{O-E}{\sqrt{E}-\kappa}$, is used to identify genes (KDs) that are connected to a significantly larger number of disease-associated genes than what is expected by random chance. O and E represent the observed and expected numbers of disease-associated genes in a hub subnetwork, and E is estimated by $\frac{N_k N_p}{N}$ where N_p is the disease gene set size, N_k is the hub degree, and N is the full network order. KDs represent prioritized disease regulatory genes based on network topology. In numerous recent applications of Mergeomics, top KDs have been shown to be causal for diseases based on experimental evidence (23,27,36), thereby supporting their importance. KDA can be utilized as a follow up analysis to MSEA or Meta-MSEA, and it can also be used as an independent analysis using a gene list of interest and a given network as inputs. For instance, the user can upload a list of curated disease genes and choose or upload a relevant network to run KDA to identify how the disease genes interact in the network and whether there are key hub nodes in the network that regulate the disease genes.

In Mergeomics 2.0, we added the ability to visualize input gene overlap with a given network, if any, in the case that no KDs were found. The user can therefore be better informed on the reason for the lack of KD hits based on the distribution and connectivity of the input genes in the network. If few input genes are in the network or the input genes are widely dispersed in the network, KDs may not be identified. We have additionally increased the number of sample tissue-specific networks (Table 1). As we have done similarly with MSEA and Meta-MSEA, disease subnetworks or significant KDs from KDA can be used directly for PharmOmics drug repositioning, and users can further customize which processes in the subnetwork are used in drug repositioning for a more focused analysis.

DATA AND SAMPLE INPUT UPDATES

We have significantly augmented the amount of Mergeomics-ready sample files with commonly used datasets and will continue to actively update sample files to enrich data resources on a monthly basis.

In Mergeomics 2.0, we include over 20 GWAS datasets from a broader range of diseases from metabolic syndrome

Table 1. Sample resources on Mergeomics web server. Complete list in Supplementary Tables S1–S4

General data category	Data type	Specifics	Citation
Association data	GWAS	Alzheimer's disease	(71)
		Attention deficit hyperactivity disorder	(72)
		Alcohol dependence	(73)
		Body mass index	(74)
		Breast cancer	(75)
		Coronary artery disease	(76)
		Fasting glucose	(77)
		Heart failure	(78)
		High density lipoproteins (HDL)	(79)
		Low density lipoproteins (LDL)	(79)
		Major depressive disorder	(80)
		Parental lifespan	(81)
		Parkinson's disease	(82)
		Psoriasis	(83)
		Severe illness in COVID-19	(84)
		Schizophrenia	(85)
		Stroke	(86)
		Systemic lupus erythematosus	(87)
		Type 2 diabetes	(88)
		Total cholesterol	(79)
		Triglycerides	(79)
	EWAS	Birth weight	(89)
		Maternal anxiety	(90)
		Social communication	(91)
		Psoriasis	(62,63)
Marker mapping	Chromosomal distance	10kb, 20kb, 50kb	(46)
	Regulome	RegulomeDB (ENCODE)	(92)
	eQTL	49 tissue types	(45)
	sQTL	49 tissue types	(45)
Marker dependency	Linkage disequilibrium	26 populations at $r^2 > 0.5$ and >0.7	(46)
	Methylation disequilibrium	$r^2 > 0.5$	(47)
Marker sets	Canonical (knowledge based)	KEGG	(50)
		Reactome	(51)
		BioCarta	(52)
		MSigDB	(49)
		GO	(53)
		BioPlanet	(55)
		WikiPathways	(54)
	Data-driven (co-expression)	24 tissue specific modules (WGCNA/MEGENA)	(45,56–57)
Networks	Gene regulatory human and mouse composite (Bayesian)	Adipose, blood, brain, kidney, liver, muscle	(58,93–98)
	Gene regulatory (GIANT)	Adipose, blood, brain, kidney, liver, muscle	(61)
	Protein-protein interaction	STRING	(59)
	Transcription factor-target (FANTOM5)	Adipose, blood, brain, kidney, liver, muscle	(60)

to psychiatric disorders (Table 1; detailed data sources and links in Supplementary Table S1). For omics dependency filtering options, we have added the full array of LD data from 26 human populations studied in 1000G (46) with LD above 0.5 and 0.7 for SNP filtering to remove redundant SNPs in high LD and have also provided an example methylation disequilibrium data file for correction of EWAS data. For SNP to gene mapping options, we have added all tissue-specific cis-eQTL and cis-sQTL mapping files from the GTEx version 8 (q-value < 0.05) (45), which inform on the SNPs associated with gene expression level changes (eQTL) or differential splicing (sQTL). In addition, we offer ENCODE regulatory gene mapping (48) and various chromosomal location-based mapping options (Table 1; Supplementary Table S2). Moreover, we have increased the number of curated pathways from version 1 to include all gene sets from Molecular Signatures Database (MSigDB) (49) such as KEGG (50), Reactome (51), Biocarta (52) canonical pathways, chemical and genetic perturbation, microRNA and transcription factor targets, and cell type marker signatures, Gene Ontology (53), Wikipathways (54) and Bioplanet (55), among others (Table 1; Supplementary Table S3). To complement knowledge-based pathways, we include our data-driven tissue-specific co-expression network modules utilizing GTEx transcriptome datasets and co-expression network construction tools MEGENA (56) and WGCNA (57) (Table 1; details of data sources, methods, and parameters used to construct networks in Supplementary Table S3). Finally, we have constructed tissue-specific Bayesian gene regulatory networks (58) and include them as sample networks on the web server. We also provide human protein-protein interaction networks (59), transcription factor networks (60) and GIANT networks (61) (Table 1; Supplementary Table S4). Sample files are available to download from our sample resources page (http://mergeomics.research.idre.ucla.edu/samplefiles.php), and further clarification on correct formatting of input data is detailed on the web server and in Figure 2.

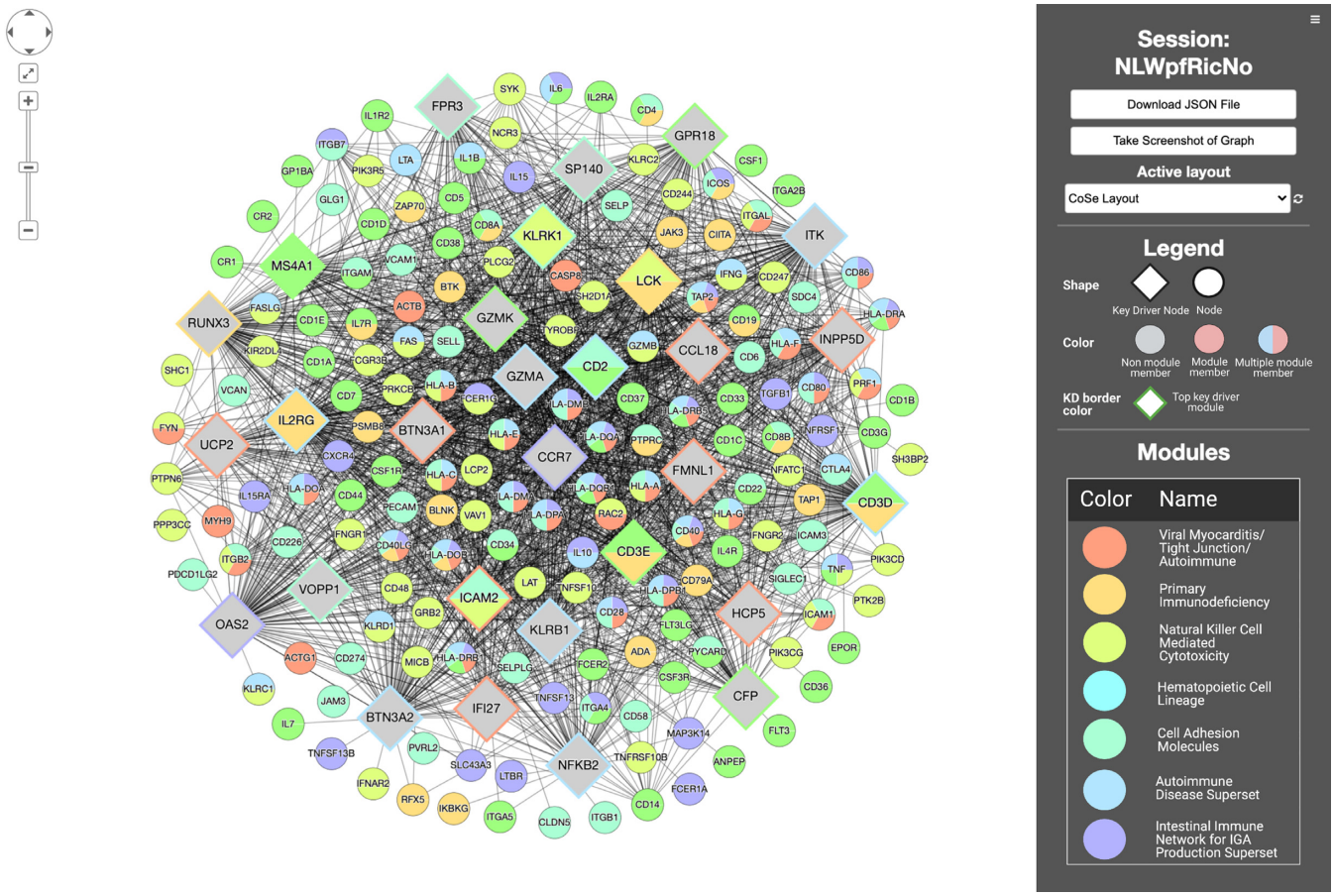


Figure 3. Top KDs network visualization. Screenshot of the in-browser interactive network visualization (using Cytoscape.js) directed from the KDA results page. The colors of the nodes represent member genes of a disease-associated pathway. The diamond shaped nodes represent KD genes, where the border color represents the top pathway that is regulated by the KD. If a node has multiple colors, it is part of two or more disease-associated pathways, and if a node is grey, it does not belong to the disease pathways (non-member genes) but is present in the input network.

GENERAL UPDATES

We have completely redesigned the user interface for a much more intuitive guidance of the use of the pipeline for different omics data types. To start the pipeline, users are presented with four workflow options in regard to their data: (i) GWAS, (ii) EWAS, TWAS, PWAS or MWAS, (iii) multiple of the same or different types of omics data and (iv) a gene set list (user can run KDA or PharmOmics). The separation of GWAS from other omics datasets is for the additional need to correct for LD and link SNPs to candidate genes through MDF, which is not required or is optional for other omics datasets. For EWAS, a marker to gene mapping file is required if the user uploads epigenetic markers such as CpG probes. For MWAS, a metabolite to gene mapping file is optional but not required if the user uses metabolite sets as the marker sets to be tested. Marker mapping is not needed for TWAS and PWAS as the markers (genes and proteins) match the gene sets. This workflow design clearly delineates what is needed for each specific data type, which is more intuitive for the user. We have also improved the fluidity and presentation of the pipeline workflow as each collapsible step appears below the previous in a vertical format so that the user can revisit input files, parameters, and results of previous steps in the pipeline and choose to rerun a step at any point in the pipeline. A workflow map with navigation links is also generated on the left sidebar to help visualize the steps taken and downstream paths.

We have improved the system that allows users to return to their session where results of analyses can be revisited or continued onto the next step using a unique tracking ID number that is valid for up to 48 hours after the start of their session. The user can also choose to have their results emailed upon completion of the analysis, which is not mandatory but is recommended because the tracking ID allows the user to reload their session and retrieve completed jobs in case a crash occurs. Because later steps of the pipeline, KDA and PharmOmics, can be run independently, downloadable result files from MSEA and KDA can be uploaded directly to the desired next step in the analysis (e.g. MSEA to KDA/PharmOmics or KDA to PharmOmics).

In addition, we have improved case-specific responsiveness of the web server to better inform the user such as error-checking of user uploaded files to ensure the file is formatted correctly and providing feedback on user results such as whether the results are substantial enough to be used in the next step of the analysis. Across all applications of Merge-

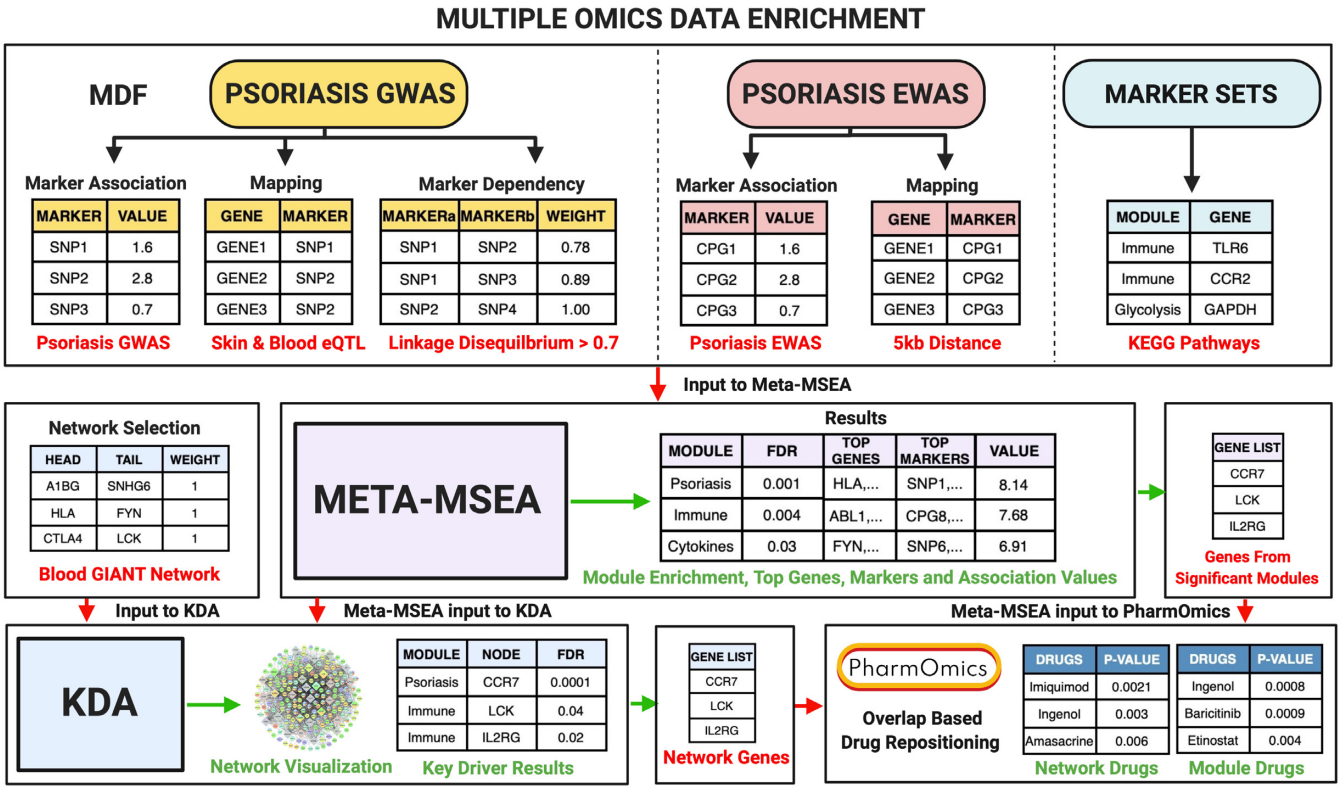


Figure 4. Meta-MSEA use case study overview. To showcase the function and output of the web server, we utilized multiple human psoriasis GWAS and EWAS data and ran the multiple omics data workflow (Case 3 in Figure 1, Meta-MSEA). Firstly, we uploaded the psoriasis GWAS data, mapped the SNPs to genes using a combined skin and blood eQTL file, and filtered for LD > 0.7 to remove redundant SNPs in LD. Next, we uploaded our psoriasis EWAS association datasets and mapped the CpG sites to genes based on a 5 kb distance. Finally, we uploaded KEGG pathways with a psoriasis control set. Pathway enrichment results are produced, and each pathway's top genes, markers, and corresponding association values are displayed. Psoriasis-associated pathways are used as input into KDA as well as PharmOmics drug repositioning (using genes from significant pathways/modules). In the KDA, along with the Meta-MSEA input, we chose the blood GIANT network option and ran the KDA providing KD results and visualization (Figure 3) and additionally utilized the network genes as an input into PharmOmics. Finally, two sets of drug repositioning results were produced using gene overlap-based drug repositioning in PharmOmics: one based on the genes of significant pathways from the Meta-MSEA results and the other based on the KDA subnetwork genes.

omics 2.0 we have provided an improved review of analysis inputs and parameters and new interactive tables with pagination, sorting, and search features (Figure 5). We also implemented real-time runtime analysis output and progress updates, and this job log including any errors that occurred is available for download at the conclusion of the analysis. Finally, we have improved multi-device usage including on tablets and phones such that it can be appropriately viewed on different screen sizes. We further improved the tutorial to explain input file preparation, parameter setting, and the underlying methods of each computational function and provide video tutorials to demonstrate the different pipeline options.

USE CASE: IDENTIFYING PATHOGENIC PATHWAYS AND NETWORKS FOR PSORIASIS BASED ON MULTI-OMICS DATA

The use case described here utilizes publicly available GWAS and EWAS data to perform Meta-MSEA and subsequently KDA to find pathogenic pathways and regulators of psoriasis (Figure 4). All data used in this example are provided as sample data on the web server which can be downloaded (http://mergeomics.research.idre.ucla.edu/samplefiles.php). GWAS of psoriasis was obtained from dbGAP database (www.ncbi.nlm.nih.gov/gap) with accession phs000019.v1.p1, and two EWAS of psoriasis were obtained from GEO (GSE31835 and GSE63315) (62,63). For preprocessing of the GWAS data, we use the top 50% of SNPs ranked by $-\log_{10}$ *P*-value and correct for LD between SNPs using MDF with the psoriasis GWAS summary statistics as the marker associations, combined skin and blood eQTLs as the SNP to gene mapping, and the 1000G CEU LD structure containing SNPs with $r^2 > 0.7$ as the marker dependency file. For the EWAS data, CpG sites are mapped to adjacent genes within 5 kb. Next, we chose canonical pathways from the KEGG database and a positive control gene set from the NHGRI-EBI GWAS catalog (64) for psoriasis as the pathways or marker sets to be examined. We ran Meta-MSEA across the GWAS and two EWAS datasets. At the conclusion of Meta-MSEA, a set of results files and a summary table display are generated on the webpage detailing the pathways ranked by meta *P*-value and their top markers and corresponding mapped genes (Figure 5A; Supplementary Table S5).

Figure 5. Output files from Meta-MSEA, KDA, and PharmOmics based on the case study of psoriasis outlined in Figure 4. Tables are interactive with pagination, search, and sort functions. Result files are downloadable from links on the webpage above the output tables (not shown). (**A**) Example Meta-MSEA output from the psoriasis use case. The table shown details the significance of association of each pathway/module and the top markers and corresponding association strengths that contributed to the module association. There are two additional tables which can be displayed by clicking on the tabs to the right of 'Module Results' at the top. The second table shows the significance and details of merged modules after merging redundant pathways (termed 'Supersets'), and these non-overlapping gene sets are used as input to KDA. The third table shows the individual significance values for each omics dataset included in this Meta-MSEA of one GWAS and two EWAS of psoriasis. (**B**) Example KDA output from the psoriasis use case. The table shown records the significance of KDs, the pathways/modules that they regulate based on network topology, and details of the local subnetwork such as the number of KD subnetwork genes and number of pathway/module gene overlap with the KD subnetwork. Merged pathways/modules are represented by the term 'Superset', which means they are comprised of multiple redundant (significant gene overlap) pathways. (**C**) Example PharmOmics drug repositioning output using a gene overlap-based analysis between disease pathways and drug signatures. Gene overlap-based drug repositioning queries all tissue- and species-specific meta-analyzed and dose/time segregated gene signatures of drugs in our PharmOmics database as well as all L1000 drug signatures. The table shown gives the dataset source of the drug signature, the method of differential gene expression analysis, details of the drug study including species, tissue or cell line, whether the study was done *in vitro* or *in vivo*, the dose and time regimen, the Jaccard score, and statistical significance of the gene overlap between the input psoriasis related genes from Meta-MSEA and the drug signatures.

As shown in Figure 5A, 'Cytokine cytokine receptor interaction', 'Graft versus host disease' and 'Natural killer cell mediated cytotoxicity' were three of the top pathways identified among others. Following Meta-MSEA, KDA was run with default parameters using non-redundant supersets (pathways that were merged due to significant overlap in gene members) significantly associated with psoriasis from Meta-MSEA and a blood GIANT Bayesian gene regulatory network (61) (chosen due to the relevance of the immune system to psoriasis) to identify KDs of the disease related gene sets. At the conclusion of the KDA, a table is produced on the webpage listing the KDs and significance of enrichment of psoriasis-associated gene sets in their network neighborhood (Figure 5B; Supplementary Table S6). For example, *ICAM2* is identified as the KD for the viral myocarditis/tight junction/autoimmune pathway, and *CD2* is identified as a KD for the Autoimmune Disease Superset. By default, the top five KDs and their local subnetworks from each gene set is included in the interactive subnetwork visualization in the browser (Figure 3).

With addition of the PharmOmics pipeline to the Mergeomics web server, we ran two drug repositioning analyses: one directly from the MSEA results and the other considering the whole subnetwork derived from the KDA (Figure 5C; Supplementary Table S7). In this case study, we do not consider gene expression direction changes (upregulation or downregulation) in psoriasis and therefore will simply be utilizing genes involved in disease without considering if they are protective or pathogenic; thus, our predicted drug list will contain drugs that can induce as well as drugs that can potentially treat psoriasis. In addition, PharmOmics interrogates all drug signatures regardless of the tissue or species, and the user can choose to focus on the relevant drug studies for their given dataset. For example, we mainly focused on drugs that were studied in integument tissue, due to its relevance to psoriasis. In the top 10 repositioned drugs derived from psoriasis associated gene sets from Meta-MSEA, we find 8/10 to have prior association with a role in psoriasis pathogenesis (Imiquimod (65)) or treatment including broad options suggesting classes of drugs such as anti-inflammatory, immunosuppressant, JAK inhibitors, and anti-rheumatic drugs and more specific options such as Baricitinib (66), Ingenol (67), and Etinostat (68) (Figure 5C; Supplementary Table S7). Similarly, using the psoriasis subnetwork from the KDA highlights Imiquimod and Ingenol within the top 10 drugs, and the remainder of the results are broad categories such as JAK inhibitors, anti-inflammatory drugs, and anti-rheumatic drugs (Supplementary Table S8), each of which are actively being investigated in the treatment of psoriasis (69,70). The predicted drugs can form new hypotheses for experimental testing.

FUTURE DIRECTIONS

The web server will continue to actively incorporate the most up-to-date public resources including multi-omics association data, functional genomics data such as eQTLs or protein QTLs (pQTLs), knowledge-based pathways, gene co-expression networks, and gene regulatory networks on a monthly basis. We will also include single cell networks when available to understand the gene regulatory connections within a given cell type or between cell types rather than across a whole tissue, which will offer higher resolution molecular mechanisms of disease pathogenesis. Cell type level association data derived from single cell omics studies can be used in the current platform. We will also continue incorporating additional analytical functions into the web server such as different forms of meta-analysis that can be conducted within the Meta-MSEA tool as well as adding new features to better accommodate analysis of data types that are currently not considered or well tested, such as gut microbiome and spatial transcriptomics data.

CONCLUSION

Thanks to advancements in technologies, the number of multi-omics data (GWAS, EWAS, TWAS, PWAS and MWAS) increases exponentially. The systems biology approach to interrogate multi-tissue multi-omics data has become a promising method to understand biology in a data-driven way and sheds light on the hidden mechanisms. However, the computational knowledge and skills required to perform such integrative analysis are often considered as a hurdle to many biologists. Therefore, the Mergeomics web server was developed to lower this barrier to enable fellow researchers to dive into multi-omics systems biology. The current update, Mergeomics 2.0, is a versatile web-based tool that provides multi-omics data integration using a pathway- and network-based approach. The improvements we made support a wide range of pre-calculated networks and data for all steps of the pipeline to fulfill a variety of needs and research purposes. In addition, the new user interface presents a more intuitive and flexible environment that greatly improves its ease of use. In addition to a detailed tutorial, each step of the pipeline contains embedded guidance to facilitate the user experience. We believe that the Mergeomics 2.0 and systematics approach applied here will accelerate our understanding of complex diseases and guide therapeutics.

DATA AVAILABILITY

Sample resources are available on our sample resources page on the Mergeomics web server (http://mergeomics.research.idre.ucla.edu/samplefiles.php), and the R package for Mergeomics can be found on (https://bioconductor.org/packages/release/bioc/html/Mergeomics.html).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Dr Yuqi Zhao for guiding the use case example. We thank Yanning Zuo, Russell Littman, Neil Hsu, Jenny Cheng, I-Hsin Tseng, Caden McQuillen, Yutian Zhao, Zara Saleem, Hengjian Li and Carissa Zhu for their suggestions and testing of the web server.

FUNDING

X.Y. is supported by NIH R01 [NS117148, NS111378, DK117850, HL145708, HL147883, HD100298]; Montgomery Blencowe is supported by the American Heart Association Predoctoral Fellowship; Sung-min Ha is supported by the UCLA QCBio Collaboratory Postdoc Fellowship. Funding for open access charge: National Institutes of Health [DK117850, HD100298, HL145708, HL147883, NS111378, NS117148].
Conflict of interest statement. None declared.

REFERENCES

1. Yang,X. (2020) Multitissue multiomics systems biology to dissect complex diseases. *Trends Mol. Med.*, **26**, 718–728.
2. Boyle,E.A., Li,Y.I. and Pritchard,J.K. (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell*, **169**, 1177–1186.
3. Subramanian,I., Verma,S., Kumar,S., Jere,A. and Anamika,K. (2020) Multi-omics data integration, interpretation, and its application. *Bioinform Biol. Insights*, **14**, 1177932219899051.
4. Graw,S., Chappell,K., Washam,C.L., Gies,A., Bird,J., Robeson,M.S. 2nd and Byrum,S.D. (2020) Multi-omics data integration considerations and study design for biological systems and disease. *Mol. Omics*, **17**, 170–185.
5. Huang,S., Chaudhary,K. and Garmire,L.X. (2017) More is better: recent progress in multi-omics data integration methods. Front Genet., **8**, 84.
6. Shi,Q., Zhang,C., Peng,M., Yu,X., Zeng,T., Liu,J. and Chen,L. (2017) Pattern fusion analysis by adaptive alignment of multiple heterogeneous omics data. *Bioinformatics*, **33**, 2706–2714.
7. Wang,B., Mezlini,A.M., Demir,F., Fiume,M., Tu,Z., Brudno,M., Haibe-Kains,B. and Goldenberg,A. (2014) Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods*, **11**, 333–337.
8. Yuan,Y., Savage,R.S. and Markowetz,F. (2011) Patient-specific data fusion defines prognostic cancer subtypes. *PLoS Comput. Biol.*, **7**, e1002227.
9. Shen,R., Mo,Q., Schultz,N., Seshan,V.E., Olshen,A.B., Huse,J., Ladanyi,M. and Sander,C. (2012) Integrative subtype discovery in glioblastoma using iCluster. *PLoS One*, **7**, e35236.
10. Lock,E.F. and Dunson,D.B. (2013) Bayesian consensus clustering. *Bioinformatics*, **29**, 2610–2616.
11. de Tayrac,M., Le,S., Aubry,M., Mosser,J. and Husson,F. (2009) Simultaneous analysis of distinct Omics data sets with integration of biological knowledge: multiple factor analysis approach. *BMC Genomics*, **10**, 32.
12. Le Cao,K.A., Gonzalez,I. and Dejean,S. (2009) integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics*, **25**, 2855–2856.
13. Rohart,F., Gautier,B., Singh,A. and Le Cao,K.A. (2017) mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.*, **13**, e1005752.
14. Vaske,C.J., Benz,S.C., Sanborn,J.Z., Earl,D., Szeto,C., Zhu,J., Haussler,D. and Stuart,J.M. (2010) Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics*, **26**, i237–i245.
15. Koh,H.W.L., Fermin,D., Vogel,C., Choi,K.P., Ewing,R.M. and Choi,H. (2019) iOmicsPASS: network-based integration of multiomics data for predictive subnetwork discovery. *NPJ Syst Biol Appl*, **5**, 22.
16. Zoppi,J., Guillaume,J.F., Neunlist,M. and Chaffron,S. (2021) MiBiOmics: an interactive web application for multi-omics data exploration and integration. *BMC Bioinformatics*, **22**, 6.
17. Bonnet,E., Calzone,L. and Michoel,T. (2015) Integrative multi-omics module network inference with Lemon-Tree. *PLoS Comput. Biol.*, **11**, e1003983.
18. Hernandez-de-Diego,R., Tarazona,S., Martinez-Mira,C., Balzano-Nogueira,L., Furio-Tari,P., Pappas,G.J. Jr and Conesa,A. (2018) PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data. *Nucleic Acids Res.*, **46**, W503–W509.
19. Dimitrakopoulos,C., Hindupur,S.K., Hafliger,L., Behr,J., Montazeri,H., Hall,M.N. and Beerenwinkel,N. (2018) Network-based integration of multi-omics data for prioritizing cancer genes. *Bioinformatics*, **34**, 2441–2448.
20. Zhou,Y., Zhou,B., Pache,L., Chang,M., Khodabakhshi,A.H., Tanaseichuk,O., Benner,C. and Chanda,S.K. (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.*, **10**, 1523.
21. Shu,L., Zhao,Y., Kurt,Z., Byars,S.G., Tukiainen,T., Kettunen,J., Orozco,L.D., Pellegrini,M., Lusis,A.J., Ripatti,S. *et al.* (2016) Mergeomics: multidimensional data integration to identify pathogenic perturbations to biological systems. *BMC Genomics*, **17**, 874.
22. Arneson,D., Bhattacharya,A., Shu,L., Makinen,V.P. and Yang,X. (2016) Mergeomics: a web server for identifying pathological pathways, networks, and key regulators via multidimensional data integration. *BMC Genomics*, **17**, 722.
23. Chella Krishnan,K., Kurt,Z., Barrere-Cain,R., Sabir,S., Das,A., Floyd,R., Vergnes,L., Zhao,Y., Che,N., Charugundla,S. *et al.* (2018) Integration of multi-omics data from mouse diversity panel highlights mitochondrial dysfunction in non-alcoholic fatty liver disease. *Cell Syst.*, **6**, 103–115.
24. Chen,L., Yao,Y., Jin,C., Wu,S., Liu,Q., Li,J., Ma,Y., Xu,Y. and Zhong,Y. (2019) Integrative genomic analysis identified common regulatory networks underlying the correlation between coronary artery disease and plasma lipid levels. *BMC Cardiovasc. Disord.*, **19**, 310.
25. Hartman,R.J.G., Owsiany,K., Ma,L., Koplev,S., Hao,K., Slenders,L., Civelek,M., Mokry,M., Kovacic,J.C., Pasterkamp,G. *et al.* (2021) Sex-stratified gene regulatory networks reveal female key driver genes of atherosclerosis involved in smooth muscle cell phenotype switching. *Circulation*, **143**, 713–726.
26. Liu,Y., Lu,P., Wang,Y., Morrow,B.E., Zhou,B. and Zheng,D. (2019) Spatiotemporal gene coexpression and regulation in mouse cardiomyocytes of early cardiac morphogenesis. *J. Am. Heart Assoc.*, **8**, e012941.
27. Shu,L., Chan,K.H.K., Zhang,G., Huan,T., Kurt,Z., Zhao,Y., Codoni,V., Tregouet,D.A., Cardiogenics,C., Yang,J. *et al.* (2017) Shared genetic regulatory networks for cardiovascular disease and type 2 diabetes in multiple populations of diverse ethnicities in the United States. *PLoS Genet.*, **13**, e1007040.
28. Zhao,Y., Jhamb,D., Shu,L., Arneson,D., Rajpal,D.K. and Yang,X. (2019) Multi-omics integration reveals molecular networks and regulators of psoriasis. *BMC Syst. Biol.*, **13**, 8.
29. Jung,S.M., Park,K.S. and Kim,K.J. (2020) Deep phenotyping of synovial molecular signatures by integrative systems analysis in rheumatoid arthritis. *Rheumatology (Oxford)*, doi:10.1093/rheumatology/keaa751.
30. Drake,J., McMichael,G.O., Vornholt,E.S., Cresswell,K., Williamson,V., Chatzinakos,C., Mamdani,M., Hariharan,S., Kendler,K.S., Kalsi,G. *et al.* (2020) Assessing the role of long noncoding rna in nucleus accumbens in subjects with alcohol dependence. *Alcohol. Clin. Exp. Res.*, **44**, 2468–2480.
31. Meng,Q., Zhuang,Y., Ying,Z., Agrawal,R., Yang,X. and Gomez-Pinilla,F. (2017) Traumatic brain injury induces genome-wide transcriptomic, methylomic, and network perturbations in brain and blood predicting neurological disorders. *EBioMedicine*, **16**, 184–194.
32. Min,H.K., Moon,S.J., Park,K.S. and Kim,K.J. (2019) Integrated systems analysis of salivary gland transcriptomics reveals key molecular networks in Sjogren's syndrome. *Arthritis Res. Ther.*, **21**, 294.
33. Diamante,G., Cely,I., Zamora,Z., Ding,J., Blencowe,M., Lang,J., Bline,A., Singh,M., Lusis,A.J. and Yang,X. (2021) Systems toxicogenomics of prenatal low-dose BPA exposure on liver metabolic pathways, gut microbiota, and metabolic health in mice. *Environ. Int.*, **146**, 106260.
34. Zhang,G., Byun,H.R., Ying,Z., Blencowe,M., Zhao,Y., Hong,J., Shu,L., Chella Krishnan,K., Gomez-Pinilla,F. and Yang,X. (2020) Differential metabolic and multi-tissue transcriptomic responses to fructose consumption among genetically diverse mice. *Biochim. Biophys. Acta Mol. Basis Dis.*, **1866**, 165569.
35. Shu,L., Meng,Q., Diamante,G., Tsai,B., Chen,Y.W., Mikhail,A., Luk,H., Ritz,B., Allard,P. and Yang,X. (2019) Prenatal bisphenol a

exposure in mice induces multitissue multiomics disruptions linking to cardiometabolic disorders. *Endocrinology*, **160**, 409–429.
36. Blencowe,M., Ahn,I.S., Saleem,Z., Luk,H., Cely,I., Makinen,V.P., Zhao,Y. and Yang,X. (2021) Gene networks and pathways for plasma lipid traits via multitissue multiomics systems analysis. *J. Lipid Res.*, **62**, 100019.
37. Zhao,Y., Blencowe,M., Shi,X., Shu,L., Levian,C., Ahn,I.S., Kim,S.K., Huan,T., Levy,D. and Yang,X. (2019) Integrative genomics analysis unravels tissue-specific pathways, networks, and key regulators of blood pressure regulation. *Front Cardiovasc Med*, **6**, 21.
38. Hui,S.T., Kurt,Z., Tuominen,I., Norheim,F., Richard,C.D., Pan,C., Dirks,D.L., Magyar,C.E., French,S.W., Chella Krishnan,K. *et al.* (2018) The genetic architecture of diet-induced hepatic fibrosis in mice. *Hepatology*, **68**, 2182–2196.
39. Meng,Q., Ying,Z., Noble,E., Zhao,Y., Agrawal,R., Mikhail,A., Zhuang,Y., Tyagi,E., Zhang,Q., Lee,J.H. *et al.* (2016) Systems nutrigenomics reveals brain gene networks linking metabolic and brain disorders. *EBioMedicine*, **7**, 157–166.
40. Makinen,V.P., Civelek,M., Meng,Q., Zhang,B., Zhu,J., Levian,C., Huan,T., Segre,A.V., Ghosh,S., Vivar,J. *et al.* (2014) Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease. *PLos Genet.*, **10**, e1004502.
41. Chen,Y.-W., Diamante,G., Ding,J., Nghiem,T.X., Yang,J., Ha,S.-m., Cohn,P., Arneson,D., Blencowe,M., Garcia,J. *et al.* (2021) PharmOmics: a species- and tissue-specific drug signature database and online tool for drug repurposing. bioRxiv doi: https://doi.org/10.1101/837773, 30 March 2021, preprint: not peer reviewed.
42. Lamb,J. (2007) The Connectivity Map: a new tool for biomedical research. *Nat. Rev. Cancer*, **7**, 54–60.
43. Subramanian,A., Narayan,R., Corsello,S.M., Peck,D.D., Natoli,T.E., Lu,X., Gould,J., Davis,J.F., Tubelli,A.A., Asiedu,J.K. *et al.* (2017) A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*, **171**, 1437–1452.
44. Cheng,F., Desai,R.J., Handy,D.E., Wang,R., Schneeweiss,S., Barabasi,A.L. and Loscalzo,J. (2018) Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat. Commun.*, **9**, 2691.
45. Consortium,G.T. (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.
46. Auton,A., Brooks,L.D., Durbin,R.M., Garrison,E.P., Kang,H.M., Korbel,J.O., Marchini,J.L., McCarthy,S., McVean,G.A. and Abecasis,G.R. (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
47. Xu,J., Zhao,L., Liu,D., Hu,S., Song,X., Li,J., Lv,H., Duan,L., Zhang,M., Jiang,Q. *et al.* (2018) EWAS: epigenome-wide association study software 2.0. *Bioinformatics*, **34**, 2657–2658.
48. Consortium,E.P. (2004) The ENCODE (ENCyclopedia Of DNA Elements) project. *Science*, **306**, 636–640.
49. Liberzon,A., Birger,C., Thorvaldsdottir,H., Ghandi,M., Mesirov,J.P. and Tamayo,P. (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.*, **1**, 417–425.
50. Kanehisa,M., Araki,M., Goto,S., Hattori,M., Hirakawa,M., Itoh,M., Katayama,T., Kawashima,S., Okuda,S., Tokimatsu,T. *et al.* (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, **36**, D480–D484.
51. Fabregat,A., Jupe,S., Matthews,L., Sidiropoulos,K., Gillespie,M., Garapati,P., Haw,R., Jassal,B., Korninger,F., May,B. *et al.* (2018) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **46**, D649–D655.
52. Nishimura,D. (2001) BioCarta. *Biotech. Softw. Internet Rep.*, **2**, 117–120.
53. The Gene Ontology,C. (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.*, **47**, D330–D338.
54. Slenter,D.N., Kutmon,M., Hanspers,K., Riutta,A., Windsor,J., Nunes,N., Melius,J., Cirillo,E., Coort,S.L., Digles,D. *et al.* (2018) WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.*, **46**, D661–D667.
55. Huang,R., Grishagin,I., Wang,Y., Zhao,T., Greene,J., Obenauer,J.C., Ngan,D., Nguyen,D.T., Guha,R., Jadhav,A. *et al.* (2019) The NCATS BioPlanet - an integrated platform for exploring the universe of cellular signaling pathways for toxicology, systems biology, and chemical genomics. Front. Pharmacol., **10**, 445.
56. Song,W.M. and Zhang,B. (2015) Multiscale embedded gene co-expression network analysis. *PLoS Comput. Biol.*, **11**, e1004574.
57. Langfelder,P. and Horvath,S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.
58. Zhu,J., Lum,P.Y., Lamb,J., GuhaThakurta,D., Edwards,S.W., Thieringer,R., Berger,J.P., Wu,M.S., Thompson,J., Sachs,A.B. *et al.* (2004) An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet. Genome Res.*, **105**, 363–374.
59. Szklarczyk,D., Gable,A.L., Nastou,K.C., Lyon,D., Kirsch,R., Pyysalo,S., Doncheva,N.T., Legeay,M., Fang,T., Bork,P. *et al.* (2021) The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.*, **49**, D605–D612.
60. Marbach,D., Lamparter,D., Quon,G., Kellis,M., Kutalik,Z. and Bergmann,S. (2016) Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat. Methods*, **13**, 366–370.
61. Greene,C.S., Krishnan,A., Wong,A.K., Ricciotti,E., Zelaya,R.A., Himmelstein,D.S., Zhang,R., Hartmann,B.M., Zaslavsky,E., Sealfon,S.C. *et al.* (2015) Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.*, **47**, 569–576.
62. Roberson,E.D., Liu,Y., Ryan,C., Joyce,C.E., Duan,S., Cao,L., Martin,A., Liao,W., Menter,A. and Bowcock,A.M. (2012) A subset of methylated CpG sites differentiate psoriatic from normal skin. *J. Invest. Dermatol.*, **132**, 583–592.
63. Gu,X., Nylander,E., Coates,P.J., Fahraeus,R. and Nylander,K. (2015) Correlation between reversal of DNA methylation and clinical symptoms in psoriatic epidermis following narrow-band UVB phototherapy. *J. Invest. Dermatol.*, **135**, 2077–2083.
64. Buniello,A., MacArthur,J.A.L., Cerezo,M., Harris,L.W., Hayhurst,J., Malangone,C., McMahon,A., Morales,J., Mountjoy,E., Sollis,E. *et al.* (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, **47**, D1005–D1012.
65. van der Fits,L., Mourits,S., Voerman,J.S., Kant,M., Boon,L., Laman,J.D., Cornelissen,F., Mus,A.M., Florencia,E., Prens,E.P. *et al.* (2009) Imiquimod-induced psoriasis-like skin inflammation in mice is mediated via the IL-23/IL-17 axis. *J. Immunol.*, **182**, 5836–5845.
66. Menter,A., Disch,D., Clemens,J., Janes,J., Papp,K. and Macias,W. (2014) *J. Am. Acad. Dermatol.*, **70**, AB162.
67. Martin,G., Strober,B.E., Leonardi,C.L., Gelfand,J.M., Blauvelt,A., Kavanaugh,A., Stein Gold,L., Berman,B., Rosen,T. and Stockfleth,E. (2016) Updates on psoriasis and cutaneous oncology: Proceedings from the 2016 MauiDerm meeting based on presentations by. *J. Clin. Aesthet. Dermatol.*, **9**, S5–S29.
68. McLaughlin,F. and La Thangue,N.B. (2004) Histone deacetylase inhibitors in psoriasis therapy. *Curr. Drug Targets Inflamm. Allergy*, **3**, 213–219.
69. Kwatra,S.G., Dabade,T.S., Gustafson,C.J. and Feldman,S.R. (2012) JAK inhibitors in psoriasis: a promising new treatment modality. *J. Drugs Dermatol.*, **11**, 913–918.
70. Rendon,A. and Schäkel,K. (2019) Psoriasis pathogenesis and treatment. *Int. J. Mol. Sci.*, **20**, 1475.
71. Marioni,R.E., Harris,S.E., Zhang,Q., McRae,A.F., Hagenaars,S.P., Hill,W.D., Davies,G., Ritchie,C.W., Gale,C.R., Starr,J.M. *et al.* (2018) GWAS on family history of Alzheimer's disease. *Translational Psychiatry*, **8**, 99.
72. Middeldorp,C.M., Hammerschlag,A.R., Ouwens,K.G., Groen-Blokhuis,M.M., Pourcain,B.S., Greven,C.U., Pappa,I., Tiesler,C.M.T., Ang,W., Nolte,I.M. *et al.* (2016) A genome-wide association meta-analysis of attention-deficit/hyperactivity disorder symptoms in population-based pediatric cohorts. *J. Am. Acad. Child Adolesc. Psychiatry*, **55**, 896–905.
73. Olfson,E. and Bierut,L.J. (2012) Convergence of genome-wide association and candidate gene studies for alcoholism. *Alcohol Clin. Exp. Res.*, **36**, 2086–2094.
74. Locke,A.E., Kahali,B., Berndt,S.I., Justice,A.E., Pers,T.H., Day,F.R., Powell,C., Vedantam,S., Buchkovich,M.L., Yang,J. *et al.* (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature*, **518**, 197–206.
75. Rashkin,S.R., Graff,R.E., Kachuri,L., Thai,K.K., Alexeeff,S.E., Blatchins,M.A., Cavazos,T.B., Corley,D.A., Emami,N.C., Hoffman,J.D. *et al.* (2020) Pan-cancer study detects genetic risk

variants and shared genetic basis in two large cohorts. *Nat. Commun.*, **11**, 4423.
76. Nikpay,M., Goel,A., Won,H.-H., Hall,L.M., Willenborg,C., Kanoni,S., Saleheen,D., Kyriakou,T., Nelson,C.P., Hopewell,J.C. *et al.* (2015) A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics*, **47**, 1121–1130.
77. Manning,A.K., Hivert,M.F., Scott,R.A., Grimsby,J.L., Bouatia-Naji,N., Chen,H., Rybin,D., Liu,C.T., Bielak,L.F., Prokopenko,I. *et al.* (2012) A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.*, **44**, 659–669.
78. Shah,S., Henry,A., Roselli,C., Lin,H., Sveinbjörnsson,G., Fatemifar,G., Hedman,Å.K., Wilk,J.B., Morley,M.P., Chaffin,M.D. *et al.* (2020) Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.*, **11**, 163.
79. Willer,C.J., Schmidt,E.M., Sengupta,S., Peloso,G.M., Gustafsson,S., Kanoni,S., Ganna,A., Chen,J., Buchkovich,M.L., Mora,S. *et al.* (2013) Discovery and refinement of loci associated with lipid levels. *Nat. Genet.*, **45**, 1274–1283.
80. Coleman,J.R.I., Peyrot,W.J., Purves,K.L., Davis,K.A.S., Rayner,C., Choi,S.W., Hübel,C., Gaspar,H.A., Kan,C., Van der Auwera,S. *et al.* (2020) Genome-wide gene-environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK Biobank. *Mol. Psychiatry*, **25**, 1430–1446.
81. Timmers,P.R., Mounier,N., Lall,K., Fischer,K., Ning,Z., Feng,X., Bretherick,A.D., Clark,D.W. and eQTLGen ConsortiumeQTLGen Consortium and Shen,X. *et al.* (2019) Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *Elife*, **8**, e39856.
82. Blauwendraat,C., Heilbron,K., Vallerga,C.L., Bandres-Ciga,S., von Coelln,R., Pihlstrom,L., Simon-Sanchez,J., Schulte,C., Sharma,M., Krohn,L. *et al.* (2019) Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and alpha-synuclein mechanisms. *Mov. Disord.*, **34**, 866–875.
83. Nair,R.P., Duffin,K.C., Helms,C., Ding,J., Stuart,P.E., Goldgar,D., Gudjonsson,J.E., Li,Y., Tejasvi,T., Feng,B.J. *et al.* (2009) Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. *Nat. Genet.*, **41**, 199–204.
84. Pairo-Castineira,E., Clohisey,S., Klaric,L., Bretherick,A.D., Rawlik,K., Pasko,D., Walker,S., Parkinson,N., Fourman,M.H., Russell,C.D. *et al.* (2020) Genetic mechanisms of critical illness in COVID-19. *Nature*, **591**, 92–98.
85. Schizophrenia Working Group of the Psychiatric Genomics, C. (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**, 421–427.
86. Hahn,J., Fu,Y.P., Brown,M.R., Bis,J.C., de Vries,P.S., Feitosa,M.F., Yanek,L.R., Weiss,S., Giulianini,F., Smith,A.V. *et al.* (2020) Genetic loci associated with prevalent and incident myocardial infarction and coronary heart disease in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. *PLoS One*, **15**, e0230035.
87. Wang,Y.F., Zhang,Y., Lin,Z., Zhang,H., Wang,T.Y., Cao,Y., Morris,D.L., Sheng,Y., Yin,X., Zhong,S.L. *et al.* (2021) Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. *Nat. Commun.*, **12**, 772.
88. Fuchsberger,C., Flannick,J., Teslovich,T.M., Mahajan,A., Agarwala,V., Gaulton,K.J., Ma,C., Fontanillas,P., Moutsianas,L., McCarthy,D.J. *et al.* (2016) The genetic architecture of type 2 diabetes. *Nature*, **536**, 41–47.
89. Kupers,L.K., Monnereau,C., Sharp,G.C., Yousefi,P., Salas,L.A., Ghantous,A., Page,C.M., Reese,S.E., Wilcox,A.J., Czamara,D. *et al.* (2019) Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. *Nat. Commun.*, **10**, 1893.
90. Sammallahti,S., Cortes Hidalgo,A.P., Tuominen,S., Malmberg,A., Mulder,R.H., Brunst,K.J., Alemany,S., McBride,N.S., Yousefi,P., Heiss,J.A. *et al.* (2021) Maternal anxiety during pregnancy and newborn epigenome-wide DNA methylation. *Mol. Psychiatry*, doi:10.1038/s41380-020-00976-0.
91. Rijlaarsdam,J., Cecil,C.A.M., Relton,C.L. and Barker,E.D. (2021) Epigenetic profiling of social communication trajectories and co-occurring mental health problems: a prospective, methylome-wide association study. *Dev. Psychopathol.*, doi:10.1017/S0954579420001662.
92. Boyle,A.P., Hong,E.L., Hariharan,M., Cheng,Y., Schaub,M.A., Kasowski,M., Karczewski,K.J., Park,J., Hitz,B.C., Weng,S. *et al.* (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.*, **22**, 1790–1797.
93. Emilsson,V., Thorleifsson,G., Zhang,B., Leonardson,A.S., Zink,F., Zhu,J., Carlson,S., Helgason,A., Walters,G.B., Gunnarsdottir,S. *et al.* (2008) Genetics of gene expression and its effect on disease. *Nature*, **452**, 423–428.
94. Derry,J.M., Zhong,H., Molony,C., MacNeil,D., Guhathakurta,D., Zhang,B., Mudgett,J., Small,K., El Fertak,L., Guimond,A. *et al.* (2010) Identification of genes and networks driving cardiovascular and metabolic phenotypes in a mouse F2 intercross. *PLoS One*, **5**, e14319.
95. Wang,S.S., Schadt,E.E., Wang,H., Wang,X., Ingram-Drake,L., Shi,W., Drake,T.A. and Lusis,A.J. (2007) Identification of pathways for atherosclerosis in mice: integration of quantitative trait locus analysis and global gene expression data. *Circulation Res.*, **101**, e11–e30.
96. Yang,X., Schadt,E.E., Wang,S., Wang,H., Arnold,A.P., Ingram-Drake,L., Drake,T.A. and Lusis,A.J. (2006) Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.*, **16**, 995–1004.
97. Schadt,E.E., Molony,C., Chudin,E., Hao,K., Yang,X., Lum,P.Y., Kasarskis,A., Zhang,B., Wang,S., Suver,C. *et al.* (2008) Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.*, **6**, e107.
98. Tu,Z., Keller,M.P., Zhang,C., Rabaglia,M.E., Greenawalt,D.M., Yang,X., Wang,I.M., Dai,H., Bruss,M.D., Lum,P.Y. *et al.* (2012) Integrative analysis of a cross-loci regulation network identifies App as a gene regulating insulin secretion from pancreatic islets. *PLoS Genet.*, **8**, e1003107.