CellPress
OPEN ACCESS

Trends in
Genetics

Review

Insights from multi-omics integration in complex disease primary tissues

Peter Kreitmaier,[1,2] Georgia Katsoula,[1,2] and Eleftheria Zeggini[1,3,*]

Genome-wide association studies (GWAS) have provided insights into the genetic basis of complex diseases. In the next step, integrative multi-omics approaches can characterize molecular profiles in relevant primary tissues to reveal the mechanisms that underlie disease development. Here, we highlight recent progress in four examples of complex diseases generated by integrative studies: type 2 diabetes (T2D), osteoarthritis, Alzheimer's disease (AD), and systemic lupus erythematosus (SLE). High-resolution methodologies such as single-cell and spatial omics techniques will become even more important in the future. Furthermore, we emphasize the urgent need to include as yet understudied cell types and increase the diversity of studied populations.

Highlights

Genome-wide association studies (GWAS) have revealed the genetic basis of complex diseases. Integrative studies investigating multi-omics data of disease-relevant primary tissues are needed to refine these insights.

By highlighting recent integrative multi-omics studies in relevant tissues of four distinct complex diseases (type 2 diabetes, osteoarthritis, Alzheimer's disease, and systemic lupus erythematosus), we outline the usefulness of this approach across complex disease types.

Multi-omics approaches have extended our biological understanding (e.g., functional interpretation of GWAS signals, construction of new molecular maps) and revealed potential clinically relevant insights (e.g., patient stratification, biomarker identification).

Integrating multi-omics data in complex disease primary tissues

Complex diseases are driven by a combination of multiple environmental and genetic factors. Due to their high prevalence (e.g., osteoarthritis: 40% over the age of 70 years [1]; diabetes: 6.28% of the world population [2]), complex diseases represent a substantial burden for public health systems [3]. In the context of an aging population, this burden is predicted to increase in the future, underlining the importance of developing effective and personalized treatment methods, including discovery of novel drug targets (especially for drugs that have been approved in another context, referred to as drug repurposing), the identification of biomarkers, and improved patient stratification [4].

GWAS have identified genetic risk loci implicated in complex diseases and have provided much-needed insights into their complex genetic architecture [5]. However, translating genetic findings into clinical applications remains challenging across complex diseases. Issues include the strong linkage disequilibrium between variants on risk haplotypes (the actual causal variant of a risk locus often remains elusive) or the identification of effector genes of risk variants, particularly for **variants in noncoding regions** (see Glossary).

Multi-omics data of human primary tissues provide molecular profiles of disease-relevant cell types, thus revealing insights beyond those derived from genetic studies. This molecular information will contribute to overcoming current challenges in translational efforts of complex diseases (Figure 1, Key figure). Briefly, omics data can be integrated with GWAS results to identify target genes of risk variants using causal inference (e.g., Mendelian randomization [6] or colocalization approaches [7,8]). Furthermore, omics data can improve risk variant characterization, especially for those residing in noncoding sequence. Indeed, computational intersections of GWAS with datasets generated using functional genomics techniques [e.g., **chromatin immunoprecipitation followed by sequencing (ChIP-seq)**, **assay for transposase-accessible chromatin using sequencing (ATAC-seq)**, etc.] have found that for some complex traits, risk variants tend to reside and are enriched within regulatory sequence [9–11].

[1]Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany
[2]TUM School of Medicine, Technical University of Munich, Graduate School of Experimental Medicine, Munich, Germany
[3]TUM School of Medicine, Technical University of Munich and Klinikum Rechts der Isar, Munich, Germany

*Correspondence: eleftheria.zeggini@helmholtz-muenchen.de (E. Zeggini).

CellPress
OPEN ACCESS

Key figure

Applying integrative approaches on multi-omics data of four disease-relevant primary tissues

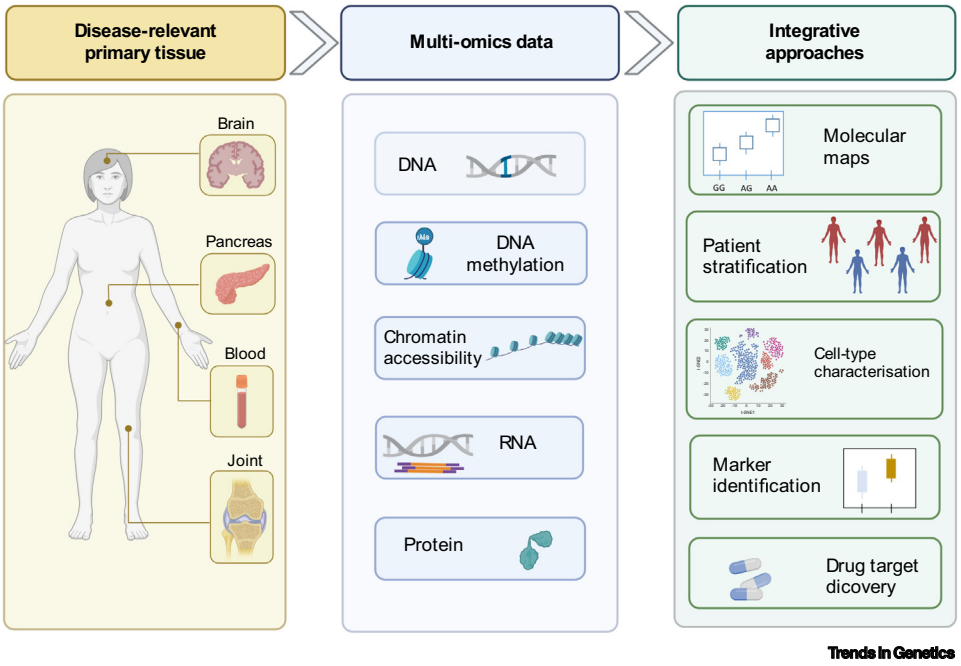

Figure 1. We describe recent findings from primary human tissues, namely joint tissues (osteoarthritis), pancreas (type 2 diabetes), brain (Alzheimer's disease), and peripheral blood (systemic lupus erythematosus). At least two omics levels were combined in integrative approaches, generating biologically relevant insights into these complex diseases. This figure was created with BioRender.com.

Functional studies have largely been enabled by recent developments in high-throughput methods that enable tissue-specific molecular profiling across several layers, such as on DNA methylation, chromatin accessibility, and transcript or protein level. Furthermore, large projects like GTEx [12], ENCODE [13], ROADMAP [14], and the Human Cell Atlas [15] have made genome-wide, tissue-specific molecular maps publicly available, thus providing well-established resources of molecular landscapes (Box 1). These large, publicly available datasets enable investigation of disease-relevant tissues across several biological, multi-omics layers [16] and provide refined insights into the link between risk factors and disease.

In addition, (multi-)omics approaches have been applied to primary samples of several disease-relevant tissue types (Box 2), that is, tissue samples collected from patients or nondisease donors. Using primary disease-relevant tissue can provide novel insights into disease mechanisms that may not have been identified when using peripheral tissues or cellular models.

This includes increased resolution into disease progression when affected primary tissues are investigated across disease stages. Here, disease-affected tissues may reflect unrelated disease processes which can be less relevant to prevention (but still relevant to treatment), whereas pre-

Glossary

450k: array technology that measures more than 450 000 methylation sites.
Assay for transposase-accessible chromatin using sequencing (ATAC-seq): a method that sequences open chromatin and used to determine genome-wide chromatin accessibility profiles and discover epigenetic regulation.
Chromatin immunoprecipitation followed by sequencing (ChIP-seq): combines ChIP with sequencing technology to identify DNA binding sites for proteins (e.g., transcription factors) genome wide. Used to explore gene regulation events.
EPIC: array technology that measures more than 850 000 methylation sites.
Genomic structural variation: includes deletions, duplications, insertions, invertions, and translocations of at least 50-base pair length.
High-throughput chromosome conformation capture (Hi-C): generates a genome-wide profile of chromatin interactions. Used to offer insights into transcriptional regulation.
HiChIP: Hi-C method that generates chromatin interaction maps, with interactions being associated with a specific (architectural) protein.
Long-read sequencing: technologies that enable live sequencing of native DNA or RNA, thus generating long reads (more than 10 kb).
Multiplexed single-cell RNA sequencing (mux-seq): droplet-based scRNA-seq approach that uses multiplexing, constituting a cost-efficient alternative to other scRNA-seq technologies.
Nuclear magnetic resonance (NMR) spectroscopy: used to measure shape and size of biological macromolecules (e.g., proteins, metabolites, nucleic acids), whereby the samples are exposed to magnetic field and a radio frequency pulse. The resonant frequencies which are specific to the molecule measured are used to detect its properties. Has several applications including molecular identification, structural and kinetics analyses.
Promoter capture Hi-C (pcHi-C): generates a genome-wide map of regions that interact with distal promoters. Used to offer insights into gene regulation and link noncoding variants to their target genes.
Quantitative trait locus (QTL): genetic variants that are significantly

Box 1. Public data resources for omics data

International collaborations have generated publicly available omics data resources which serve as reference data, for example, for functional follow-up studies of GWAS signals.

Launched in 2010, the GTEx database[i] provides a catalog of effects of genetic variants on gene expression and splicing of across 49 tissues, collected from postmortem samples of 838 individuals (version 8) [12].

ENCODE[ii] was established in 2003 with a pilot project to describe functional elements in human and mouse genomes, initially focused on 1% of the genome [72], but has expanded to the whole genome. The current version includes RNA transcription, DNA binding, chromatin modification and accessibility, DNA methylation, and replication timing data [73]. It describes 926 535 humans and 339 815 mouse candidate *cis*-regulatory elements.

Roadmap[iii] presents human epigenomic data of 111 human tissues or cell types (further provides 16 cell types from ENCODE, thus 127 in total). It comprises histone modification patterns, DNA accessibility, DNA methylation, and RNA expression [14].

The Human Cell Atlas (HCA) is an international collaboration that aims to generate reference maps of human tissues at single-cell resolution [15]. For example, one recent HCA-associated study investigated 500 000 cells and provides a single-cell reference for 400 human cell types of 24 tissues or organs [74]. The HPC data coordination platform[iv] currently provides data from more than 26 million cells of 38 000 donors (7 July 2022).

Furthermore, there are databases that provide disease-specific information. The Musculoskeletal Knowledge Portal[v] is a platform for genetic and genomic data relevant for musculoskeletal traits [75]. It currently hosts 301 datasets for 281 traits.

Similarly, there is a Type 2 Diabetes Knowledge Portal[vi] providing T2D-relevant data (349 datasets, 347 traits). Other T2D portals are Translational Human Pancreatic Islet Genotype Tissue-Expression Resource (TIGER), including omics and eQTL data from than 500 human islet samples [26] and the Diabetes Epigenome Atlas[vii] [24].

For AD, the AD Knowledge Portal is an initiative that makes AD-relevant data accessible [76].

diseased tissues may help elucidate the pathomechanism, which can be more relevant to prevention.

In this review, we cover recent insights into four complex (Box 3) diseases provided by multi-omics studies on disease-relevant primary tissue (Figure 1 and Table 1). These affect different disease-relevant tissues and pose distinct challenges: (i) T2D is a clinically heterogeneous metabolic disease for which relevant tissue samples are difficult to access. (ii) Osteoarthritis is a joint disorder for which joint tissues are challenging to access and are not included in reference databases. (iii) AD is a neurodegenerative disease that affects the brain, a complex organ that can only be studied post-mortem. (iv) SLE is an autoimmune disease with large patient heterogeneity.

Box 2. Analyses to associate multi-omics data with diseases

A standard approach to link molecular data with a disease is to conduct differential analyses, for example, between cases and controls. This is similar to the case–control approach in GWAS. In contrast to genetic studies, in which signals are estimated to play causal role in disease (and not vice versa, as genotypes are not affected by diseases because they form at conception), the changes in molecular features (RNA or protein abundances, epigenomic marks, or chromatin states) could be consequences rather than driving risk factors of the disease. Thus, differential analyses identify markers that are not necessarily causally involved in the disease of interest.

In addition, several approaches that integrate data across multiple omics layers have been developed [16]. A well-established example is the integration of genomic and gene expression data of matching samples to identify genetic variants that influence expression levels of a gene, termed expression quantitative trait loci (eQTLs), on a genome-wide scale [16]. The eQTL maps can be combined with GWAS results to identify molecular drivers (e.g., likely effector genes) through which risk variants exert their effects in disease-relevant tissue. In the context of complex diseases, these high-confidence effector genes may represent promising drug targets [77]. Other established multi-omics strategies infer information from networks [78] or estimate low-dimensional representations from multi-omics datasets, for example, to stratify samples [79].

correlated with a quantitative trait. It often refers to genetic variants associated with a molecular phenotype, such as methylation levels of a methylation site (mQTL), expression levels of a gene (eQTL) or protein (pQTL).

Sc-transposome hypersensitive sites sequencing (scTHS-seq): combines TSH-seq, a method that estimates chromatin accessibility with single-cell technology. Produces genome-wide maps of open chromatin in single-cell resolution.

Single-nucleus Droplet-based sequencing (SnDrop-seq): droplet-based method that measures transcriptomic data of single nuclei. Measuring single nuclei rather than single cells is relevant when cells cannot be separated in a single-cell resolution, such as samples that are frozen or of a specific cell type (brain, skeletal).

Variants in noncoding regions: identifying the effector genes of risk variants can be complicated since many disease-associated variants are not located in protein-, but noncoding regions of the genome. Rather than directly changing the gene product (as variants in protein-coding regions might), these variants can have a regulatory effect on the expression of a target gene, as they can reside in functional elements, such as enhancers, promoters, transcription binding sites, or noncoding RNAs.

Whole-genome bisulfite sequencing (WGBS): combines bisulfite treatment of the DNA (converting unmethylated cytosine to uracil, keeping methylated cytosine unaffected) and high-throughput DNA sequencing to measure the DNA methylation profile in a genome-wide, untargeted manner. Used to assess epigenetic regulation through methylation.

Box 3. The largest GWAS for type 2 diabetes, osteoarthritis, Alzheimer's disease, and systemic lupus erythematosus

GWAS have revealed insights into the polygenic architecture of T2D, osteoarthritis, AD, and SLE.

In T2D, the largest study to date comprised 1 339 889 individuals with 180 834 cases and 1 159 055 controls [80]. Of these, the major part was of European descent (51.1%).

For osteoarthritis, the largest GWAS investigated 826 690 individuals (177 517 cases and 649 173 controls), with more than 99.3% of European ancestry [9].

The largest AD study investigated 1 126 563 individuals (90 338 cases, 1 036 225 controls) [10]. Another recent GWAS for Alzheimer's disease included fewer individuals in total (n = 788 989), but a higher number of cases (n = 111 326) [48]. Both studies included individuals of European descent only.

The largest SLE GWAS has been performed in 208 370 individuals (13 377 SLE cases, 194 993 controls), all of which are of East Asian descent [59].

Type 2 diabetes

T2D is a complex metabolic disease affecting more than 450 million people worldwide [2] and is characterized by impairment of insulin secretion and signaling, and of carbohydrate, lipid, and protein metabolism [17,18]. The two key mechanisms in developing T2D include defective insulin secretion from the endocrine pancreatic beta cells and the lack/reduced response to insulin from insulin-sensitive tissues [19]. Over 700 genetic risk loci have been implicated in T2D to date with >90% mapped to noncoding sequences [17,18]. The majority of these variants (which explain 19% of T2D risk [20]) increase the risk of developing T2D mainly through effects on insulin secretion. To this end, studies of pancreas across multiple levels of expression are vital to gain insight into T2D molecular regulation. Despite pancreas being difficult to access, large efforts like GTEx have linked genomic variation to pancreatic gene expression. These resources constitute a valuable reference, but are not T2D specific. Therefore, multi-omics studies extending to pancreatic tissues of T2D patients have been carried out. Furthermore, islet-specific signals are not discernible when studying pancreas as a whole due to its high consistency in exocrine cells. The endocrine pancreatic islets whose dysfunction leads to T2D constitute only 1–2% of the pancreas. Therefore, multi-omics studies on pancreatic islets can shed light on the specific mechanisms of insulin secretion dysregulation in T2D.

Viñuela *et al.* studied the impact of noncoding T2D-associated variants on the expression level of proximal genes in pancreatic islet tissue from 420 nondiabetic donors [11]. This study identified 7741 cis-expression **quantitative trait loci (eQTLs**) in pancreatic islets that were replicated up to 40–73% in 44 GTEx tissues. The integration of the eQTL with epigenomic (ChIP-seq and ATAC-seq) data revealed enrichment of eQTL in active chromatin states (transcriptional start sites) and islet-specific transcription factor (TF) footprint motifs (*GLIS3*, *RFX*, and *ETS* families). Colocalization of the *eQTL* signals with variants from T2D or glycemic trait GWAS identified 47 variants with a potential causal role, highlighting *DGKB* and *TCF7L2* among the effector genes [11].

Islet-specific gene expression was further correlated with enhancer looping in pancreatic islet cells in a study by Greenwald *et al.* [21]. The authors generated a high-depth map of islet chromatin architecture of three nondiabetic donors using **high-throughput chromosome conformation capture (Hi-C)** and ATAC-seq and fine-mapped 30 known T2D signals influencing islet enhancer activity. They further identified target genes of T2D risk variants in enhancers by performing eQTL mapping, highlighting the *rs10428126* variant at the *IGF2BP2* locus as a potential causal variant for

Table 1. Overview of approach types and measured molecular omics levels in disease-relevant tissues in multi-omics studies[a]

Disease	Primary tissue	Types of approaches	Omics
Type 2 diabetes	Pancreatic islets	Bulk: **450k**, **WGBS** ATAC-seq Hi-C ChIP-seq pcHi-C RNA-seq Mass spectrometry Single-cell/nucleus: snATAC-seq scRNA-seq	 DNA methylation Chromatin accessibility Chromatin conformation Protein–DNA interactome Promoter capture chromatin conformation Transcriptomics Proteomics Single-nucleus (sn) chromatin accessibility Single-cell (sc) transcriptomics
Osteoarthritis	Cartilage	Bulk: 450k, **EPIC** ATAC-seq RNA-seq Mass spectrometry	 Methylation Chromatin accessibility Transcriptomics Proteomics
	Synovium	Bulk: RNA-seq Mass spectrometry	 Transcriptomics Proteomics
Alzheimer's disease	Brain	Bulk: ATAC-seq **HiChIP** RNA-seq ChIP-seq Mass spectrometry Single-cell/nucleus: scATAC-seq scRNA-seq **snDrop-seq** **scTHS-seq**	 Chromatin accessibility Enhancer connectome Transcriptomics Protein–DNA interactome Proteomics, phosphoproteomic, lipidomics sc chromatin accessibility sc transcriptomics sn transcriptomic sn chromatin accessibility
Systemic lupus erythematosus	Blood	Single cell **mux-seq**	 sc transcriptomics
		Bulk: RNA-seq, microarray Mass spectrometry **NMR** spectroscopy	 Transcriptomics Proteomics, metabolomics Metabolomics

[a]Abbreviations: HiChIP, Hi-C chromatin immunoprecipitation; mux-seq, multiplexed single-cell RNA sequencing; scTHS-seq, sc-transposome hypersensitive sites sequencing; snDrop-seq, single-nucleus Droplet-based sequencing; WGBS, whole-genome bisulfite sequencing.

T2D through reducing enhancer accessibility and *IGF2BP2* expression, along with compromising glucose-stimulated insulin secretion in mice [21].

Pancreatic islet enhancers have been further linked to specific gene promoters in a study from Miguel-Escalada *et al.* [22]. Using **promoter capture Hi-C (pcHi-C)** (four donors) along with ATAC-seq (13 donors), ChIP-seq (16 donors), and RNA-seq (seven donors) from nondiabetic donors, the authors identified >1300 enhancer hubs in pancreatic islets containing variants affecting insulin secretion. The authors also detected likely effector genes for 53 T2D or fasting glycemia risk loci overlapping with pancreatic islet enhancers. Among highlights were the risk

variant *rs7903146* modulating *TCF7L2* expression in beta cells and *CAMK1D* and *OPTN* regulation by an *rs11257655*-containing enhancer. Inclusion of these enhancer risk variants in polygenic risk scores (PRS) could quantify genetic risk especially for individuals with lower body mass index (BMI) (<30) mediated by islet gene regulation and insulin secretion [22]. On a related theme, Thurner *et al.* explored the pancreatic islet epigenome and described genome-wide methylation (n = 10 nondiabetic donors) and chromatin accessibility (n = 17 nondiabetic donors) alterations identifying likely causal variants for the *CDC123, ADCY5,* and *KLHDC5* loci [23].

Chiou *et al.* investigated cell-specific regulatory changes in the pancreatic islet using single-nucleus ATAC-seq (snATAC-seq) in islets of three nondiabetic donors combined with published scRNA-seq and T2D GWAS data [24]. The authors identified 12 different cell populations profiling 15 298 islet cells and proposed that T2D genetic risk is mediated through variant effects mainly on beta cells of different states (based on accessibility of the *INS* promoter) along with endocrine cell populations (mainly delta cells). The *rs231361* T2D risk variant at the *KCNQ1* locus was proposed to have state-specific effects on beta cell chromatin accessibility influencing insulin levels [24].

All the aforementioned studies have studied T2D multi-omics in pancreatic islets from postmortem donors. A first of its kind study from Wigger *et al.* profiled islet cells from living pancreatomized donors, classified along the glycemic continuum from normoglycemic to diabetic [25]. Specifically, the authors measured transcriptomics (95 donors) and proteomics (five donors) in pancreatic islets along with lipidomics (55 donors) in blood from pancreatomized donors. Data integration revealed greater heterogeneity in diabetic islet gene expression compared with nondiabetic islets, with differentially expressed genes in diabetic islets mainly involved in mitochondrial function and immune response. Furthermore, the authors identified association of expression of the glycolytic enzyme *ALDOB*, the glucose transporter *SLC2A2*, plasma ceramide levels, and ether-linked phosphatidylcholines with HbA1c levels (a marker of glycemia) proposing them as potential T2D biomarkers [25]. Lastly, this study proposed that T2D seems more likely to be the result of relaxed gene expression constraints in mature islet cells rather than a result of beta cell dedifferentiation or transdifferentiation developmental processes.

Together, these studies highlight the importance of multi-omics approaches in uncovering the regulatory mechanisms underlying genetic risk variants in T2D. Large efforts that combine publicly available different data types like the Translational Human Pancreatic Islet Genotype Tissue-Expression Resource (TIGER) [26] and Diabetes Epigenome Atlas [24] along with multiethnic association studies of T2D risk [27] pave the way for a better understanding of T2D heterogeneity.

Osteoarthritis

Osteoarthritis is a prevalent, complex musculoskeletal disorder that affects all tissues of diarthrodial joints [28]. Its most prominent feature is the degradation of cartilage. To date, GWAS have revealed approximately 150 genetic risk loci [9]. It remains unclear which genetic variants and genes drive osteoarthritis development and progression in the affected organs. As a joint disorder, relevant tissue types are difficult to access but can be collected through total joint replacement surgeries. Furthermore, osteoarthritis, initially described as disease of the joint cartilage, affects all tissues of diarthrodial joints [29]; thus, multi-omics studies can reveal tissue-specific mechanisms underlying osteoarthritis across several molecular layers. Further, public data resources have not yet included osteoarthritis-relevant joint tissue types to date. Therefore, recent multi-omics studies that focus on primary joint tissues can provide valuable resources for osteoarthritis research.

Several studies have focused on single molecular levels of osteoarthritis-affected tissues (as reviewed in [30,31]). A modest number of multi-omics studies have integrated different data types in specific genomic regions of interest [32–35] or genome wide [36–41].

Steinberg *et al.* integrated genotype with molecular data (transcriptomic and proteomics) of three osteoarthritis-relevant tissue types in matched samples, namely macroscopically intact (low-grade) and degraded (high-grade) osteoarthritis cartilage as well as from synovial tissue (a connective tissue that lines the joint capsule) in osteoarthritis-affected joints of 115 patients [37]. In each tissue, the authors characterized genetic variants that are associated with gene expression (eQTLs) or protein levels [protein quantitative trait loci (pQTLs)], providing the first genome-wide molecular QTL maps of these joint tissues. Integrating these QTL maps with results of GWAS for osteoarthritis revealed five putative effector genes in osteoarthritis primary tissues (*ALDH1A2*, *NPC1*, *SMAD3*, *FAM53A,* and *SLC44A*). GWAS signals around some of these genes colocalized with eQTLs of GTEx (which does not include cartilage data) only for small numbers of 53 tested tissues (*ALDH1A2*: ovary and tibial artery; *SMAD3*: skeletal muscle; *SLC44A2*: adrenal gland) [42], thus providing further evidence for the identified colocalization signal to be cartilage specific. In addition, the comparison of low-grade with high-grade osteoarthritis cartilage identified 409 genes linked to cartilage degeneration at both the transcriptome and proteome levels. These cross-omics signals revealed an activation of the signaling pathway 'extracellular matrix–receptor interaction' in high-grade cartilage, and an enrichment of terms related to 'extracellular space'. Genes that were differentially expressed at the transcriptome level were enriched in developmental processes (e.g., 'multicellular organism development', 'anatomical structure development'). Together with previous enrichment findings in other molecular osteoarthritis studies [36,38,39] or detected changes in chondrocyte stemness (hypertrophic differentiation) during osteoarthritis progression [43], these results provide further evidence for developmental processes being involved in osteoarthritis etiology. Furthermore, these cross-omics results were integrated in an analysis which suggested 19 compounds that could reverse disease progression in osteoarthritis cartilage at the molecular level.

Coutinho de Almeida *et al.* studied mRNA and miRNA data in low-grade and high-grade osteoarthritis cartilage from 63 patients [38]. They identified 142 miRNAs and 2387 mRNAs linked to osteoarthritis cartilage degeneration. They integrated these results using 19 samples of 15 individuals with both miRNA and mRNA data and applied a step-wise approach and selected 331 miRNA–mRNA pairs [step-wise filtering: (i) focus on negatively correlating miRNA–mRNA pairs, (ii) opposing effect direction at mRNA and miRNA level when comparing low-grade and high-grade osteoarthritis cartilage, (iii) predicted interaction, and (iv) experimental validation between miRNA–mRNA pair]. They generated a miRNA–mRNA network based on correlations and effect sizes of differential analyses. The network revealed several clusters, for example, two in which miRNA is down- or upregulated in high-grade osteoarthritis cartilage, respectively. This study provides the first miRNA–mRNA interaction map in osteoarthritis cartilage.

Together, these recent multi-omics data integration studies of primary joint tissue have revealed associations between different omics levels and have enabled a better understanding of the mechanisms underpinning biological signals across molecular levels and disease stages in osteoarthritis.

Alzheimer's disease

AD is a complex neurodegenerative disease and the most common form of dementia [44]. Pathophysiological changes of AD-affected brains include an accumulation of β-amyloid plaques and tau-containing neurofibrillary tangles in the cerebral cortex [45,46]. AD is

estimated to be driven by several causal genetic variants [47], and genetic association studies have provided insights into its complex genetic architecture [10,48]. However, some specific challenges are complicating the identification of molecular mechanisms underlying AD. Access to primary relevant tissue in the brain is challenging and limited to postmortem samples. In addition, the brain is an exceptionally heterogeneous and complex organ that consists of several regions, each including different cell types working together in an orchestrated manner. Multi-omics studies can help to better understand region- as well as cell type-specific molecular mechanisms underlying AD. Brain region-specific molecular profiles are available in public data resources (e.g., GTEx). These efforts are extended by further multi-omics studies that provide insights into AD pathology using single-cell techniques, for example, by integrating single-cell chromatin accessibility landscapes and chromatin conformation maps [49] or single nuclear transcriptomic and single-cell chromatin accessibility maps [50] in brain tissues.

Morabito *et al.* extended to AD brain samples [51]. The authors profiled matching chromatin accessibility (12 late-stage AD, eight control) and gene expression landscapes (11 late-stage AD, seven control) of 191 890 nuclei of prefrontal cortex tissue of human brains at single-cell resolution. This study identified cell type-specific, AD-relevant regulatory elements influencing genes in *cis* (e.g., the AD-relevant genes *APOE* and *CLU* in AD-linked cell type oligodendrocyte), AD-relevant TFs in glia cell populations (e.g., *SREBF1*), and a novel, integrative correlation network approach to identify clusters of coexpressed genes. The latter revealed over-representations of SREBF1 targets in oligodendrocytes, underlining the role of this TF in AD.

In addition, recent multi-omics studies in bulk data identified relevant, disease-associated molecular alterations in AD brain regions, such as gain of histone modifications H3K27ac and H3K9ac [52], *VGF* downregulation [53], and downregulation of *ATP6V1A*, which was shown to be a promising drug target [54]. Bai *et al.* provided molecular insights into AD brains (n = 90) using protein networks [55]. They integrated proteomics and phosphoproteome profiles and revealed 173 proteins linked to AD progression. Integration of further omics data prioritized AD-relevant proteins (top three: AD-linked genes *APP*, *APOE*, and *MAPT*) and pathways (e.g., amyloid and Tau pathways). A further study examined *APOE* allele-dependent differences in the molecular profile of inferior parietal lobule samples, pointing to different underlying etiological mechanisms in AD [56].

Altogether, the application of integrative multi-omics approaches has extended our insights into AD at the brain region and cell type levels.

Systemic lupus erythematosus

SLE is an autoimmune disease which affects multiple organs. It shows relapse-remitting courses and is characterized by the production of autoantibodies, leading to inflammation [57]. Autoimmune signatures can be captured in peripheral blood, where common effects of SLE include decreased white blood cell (leucopenia, lymphopenia), platelet (thrombocytopenia), or red blood cell counts [58]. To date, more than 100 SLE risk loci have been identified in GWAS [59]. SLE poses specific challenges due to its patient diversity, including heterogeneity at clinical (e.g., affected organs, disease severity, clinical manifestations) and immunological levels (heterogeneity in cytokine profiles or type I interferon responses) [60]. This patient heterogeneity complicates disease treatment as well as drug development. For example, the SLE drug belimumab showed improved treatment response in a subgroup of patients with more active SLE [61]. Multi-omics studies can help to identify patient clusters with similar molecular subtypes underlying SLE development or progression, thus contributing to a better understanding of SLE heterogeneity.

Due to the relevance for autoimmune diseases and relatively easy access, recent multi-omics SLE studies have investigated peripheral blood samples from SLE patients, focusing on mapping single-cell transcriptomes in peripheral blood mononuclear cells (PBMCs) [62], biomarker identification [63], and understanding disease-relevant molecular mechanisms [64,65].

A multi-ethnicity study measured single-cell transcriptomic data of more than 1.2 million PBMCs from 162 SLE patients and 99 healthy controls [62]. This study reports cell type-specific expression patterns, expression-based classification into SLE cases and controls, and identified shared and cell type-specific cis-eQTL. It further reports SLE patient stratification into molecular clusters, thus providing insights into SLE heterogeneity.

A recent study integrated gene expression profiles from peripheral blood (65 cases and 67 controls) as well as from purified T (32 cases and 28 controls) and B cells (38 cases and 27 controls), respectively isolated from peripheral blood [64]. Comparing the SLE patients and controls identified 750 differential expressed genes (DEGs), in total. Integrating upregulated SLE genes with TF binding data from ENCODE determined networks of coregulated genes and revealed SLE-relevant pathways (e.g., SLE interferon signature). A further integration step including (i) disease-associated genes (DAGs, identified in SLE GWAS) and (ii) publicly available protein–protein interaction networks [66] identified hierarchical regulatory processes from DAGs via TFs to differentially expressed genes in blood.

Robinson *et al.* investigated SLE heterogeneity in young patients (discovery: n = 31, replication n = 31, respective median age: 19) based on metabolomics data to investigate their cardiovascular (CV) disease risk, a major mortality cause among juvenile-onset SLE patients [63]. They determined two robust SLE patient clusters, one of which showed signs of dyslipidemia, a CV risk factor. These risk patients showed higher apolipoprotein B and A1 ratios (Apo2:ApoA1), thus suggesting it as a CV biomarker (sensitivity: 96.7%, specificity: 96.2%). Comparing high and low Apo2:ApoA1 patients identified DEGs in blood-isolated T cells (CD8+: 82 DEGs, CD4+: 417 DEGs) that were enriched in atherogenic pathways (e.g., interferon signaling). These genes also overlapped (CD8+: 23 DEGs, CD4+: 2 DEGs) with upregulated genes in T cells of atherosclerotic plaques, thus providing further insights into the link between Apo2:ApoA1 and CV risk. Higher Apo2:ApoA1 ratios were also correlated with higher SLE activities in follow-up checks (3–7 years later), highlighting its use as a clinically relevant marker.

Together, these multi-omics studies revealed SLE-relevant molecular mechanisms and helped achieve better patient stratification.

Concluding remarks

The integration of multi-omics data has refined our knowledge about molecular mechanisms that underlie disease etiology in relevant tissues. In this review, we describe recent insights into four relevant, complex disorders, obtained through integrative multi-omics approaches. These diseases affect different subsets of tissues, suggesting the importance of integrative approaches across the spectrum of complex diseases. Nevertheless, some shortcomings have to be tackled in the future (see Outstanding questions) (Figure 2).

International collaborations have created databases (Box 1) of tissue-specific information which are invaluable resources, especially in multi-omics integration studies. However, limitations in sample size, population diversity, and disease-relevant cell types remain.

Outstanding questions

Can yet understudied, disease-relevant cell types be profiled in future? Can these cell types be included in future public data sources?

How can we ensure and promote the study of more diverse population groups in genetic and multi-omics studies?

How can we apply promising, novel techniques such as single-cell and spatial omics technologies at larger scales and across multiple molecular levels to improve our understanding of dynamic and disease-relevant processes? Can we solve the technical challenges posed by these novel technologies?

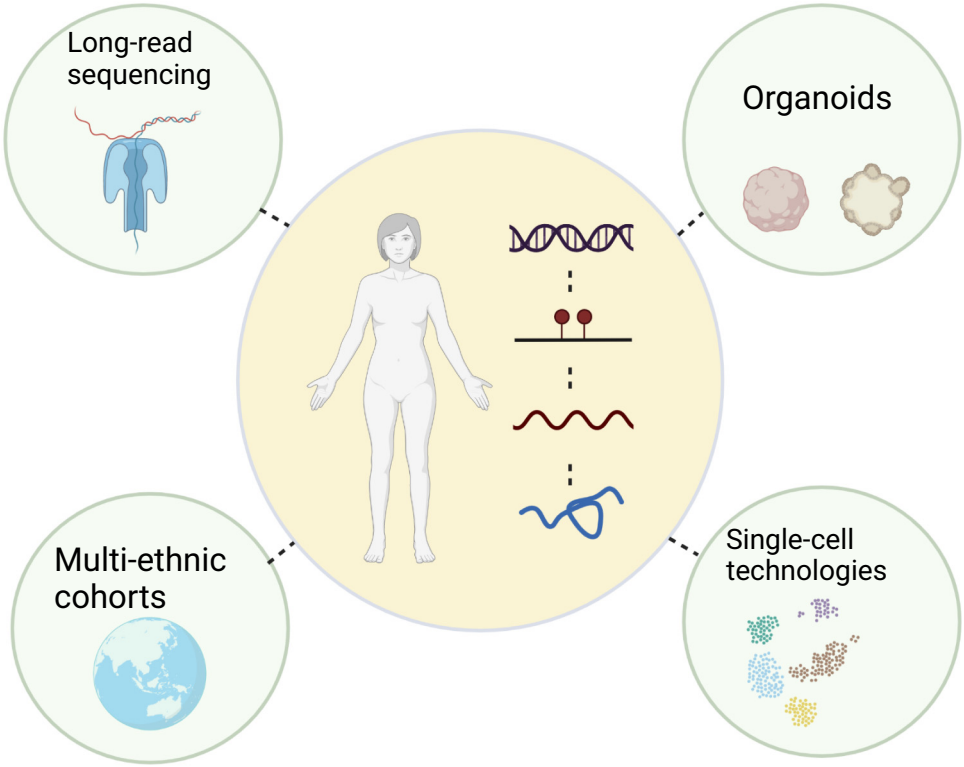

Figure 2. Overview of future directions in multi-omics research. Novel technologies (e.g., long-read sequencing, single-cell technologies, or organoid models) as well as multiethnic cohorts will help better understand the molecular profiles of human tissues. This figure was created with BioRender.com.

Most cell populations contributing to disease development remain undiscovered. Emerging single-cell and spatial multi-omics techniques that monitor more than one modality will provide higher resolutions of the characteristics of disease-relevant cell types, thus producing information beyond the inherent limitations of bulk data. To this end, emerging technologies that allow genetic screening with single-cell transcriptomics readouts, like targeted Perturb-seq, are promising to shed light on the functional genomics of complex diseases [67]. These technologies will play an important role in the future but will present new technical and analytical challenges.

Currently used sequencing technologies are primarily short read-based (e.g., up to 300 bp), limiting our ability to study disease associations on genotype (cannot measure large parts of the genome, e.g., **genomic structural variation** or highly repetitive regions like telomere, centromere) or gene expression level (limited to measuring long transcripts) [68]. Thus, **long-read sequencing** can overcome current limitations.

A number of tissue types relevant for diseases are difficult to access (e.g., joint tissues for osteoarthritis, brain tissues for AD). A promising alternative to overcome this limitation is the use of organoids. Organoids are stem cell-derived 3D *in vitro* models of human organs that share many characteristics of their corresponding human organ. [69]. They offer several advantages to other model systems such as animal models (e.g., expensive, time intensive, limited in modeling human-specific biological processes, do not reflect genetic diversity as inbred) [69,70]. Organoid models have already been generated for some complex diseases, for example, AD [71].

Further challenges going forward include the prediction of disease course and the choice on the best treatment option, particularly in early disease stages. The identification of such clinically relevant biomarkers would require well-powered studies that monitor clinical characteristics and multi-omics data across time points during disease development in appropriate sample sizes. Thus, we highlight the importance of longitudinal and time-course studies.

Genetic and molecular studies, to date, have primarily investigated cohorts that have been limited to individuals of European ancestry. This limits the general applicability of findings and produces a bias in our understanding of disease [4]. By contrast, studies across diverse and under-represented populations can generate a more complete landscape of genetic and molecular variation and help make the promise of precision medicine equally accessible to all individuals globally.

Integrative multi-omics approaches in primary disease-relevant tissues have significantly contributed to a better understanding of complex diseases. The current challenge is to further expand these previous efforts, in terms of scale, resolution, diversity, and molecular levels.

Acknowledgments

The authors are grateful to Iris Fischer and Ozvan Bocher for their assistance.

Declaration of interests

There are no conflicts of interest.

Resources

[i]https://gtexportal.org/home/

[ii]www.encodeproject.org

[iii]www.roadmapepigenomics.org/

[iv]https://data.humancellatlas.org/

[v]https://msk.hugeamp.org/

[vi]https://t2d.hugeamp.org/

[vii]https://cmdga.org/

References

1. Cui, A. *et al.* (2020) Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies. *EClinicalMedicine* 29–30, 100587
2. Khan, M.A.B. *et al.* (2020) Epidemiology of type 2 diabetes - global burden of disease and forecasted trends. *J. Epidemiol. Glob. Health* 10, 107–111
3. Vos, T. *et al.* (2020) Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 396, 1204–1222
4. Zeggini, E. *et al.* (2019) Translational genomics and precision medicine: moving from the lab to the clinic. *Science* 365, 1409–1413
5. Buniello, A. *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
6. Hemani, G. *et al.* (2018) The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 7, e34408
7. Giambartolomei, C. *et al.* (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 10, e1004383
8. Giambartolomei, C. *et al.* (2018) A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics* 34, 2538–2545
9. Boer, C.G. *et al.* (2021) Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. *Cell* 184, 4784–4818.e17
10. Wightman, D.P. *et al.* (2021) A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat. Genet.* 53, 1276–1282
11. Viñuela, A. *et al.* (2020) Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. *Nat. Commun.* 11, 4912
12. GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330
13. Dunham, I. *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74
14. Roadmap Epigenomics Consortium *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330
15. Regev, A. *et al.* (2017) The Human Cell Atlas. *eLife* 6, e27041
16. Hasin, Y. *et al.* (2017) Multi-omics approaches to disease. *Genome Biol.* 18, 83
17. Mahajan, A. *et al.* (2018) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* 50, 1505–1513
18. Scott, R.A. *et al.* (2017) An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* 66, 2888–2902
19. Galicia-Garcia, U. *et al.* (2020) Pathophysiology of type 2 diabetes mellitus. *Int. J. Mol. Sci.* 21, E6275
20. Polfus, L.M. *et al.* (2021) Genetic discovery and risk characterization in type 2 diabetes across diverse populations. *HGG Adv.* 2, 100029

21. Greenwald, W.W. *et al.* (2019) Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk. *Nat. Commun.* 10, 2078
22. Miguel-Escalada, I. *et al.* (2019) Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nat. Genet.* 51, 1137–1148
23. Thurner, M. *et al.* (2018) Integration of human pancreatic islet genomic data refines regulatory mechanisms at type 2 diabetes susceptibility loci. *eLife* 7, e31977
24. Chiou, J. *et al.* (2021) Single-cell chromatin accessibility identifies pancreatic islet cell type- and state-specific regulatory programs of diabetes risk. *Nat. Genet.* 53, 455–466
25. Wigger, L. *et al.* (2021) Multi-omics profiling of living human pancreatic islet donors reveals heterogeneous beta cell trajectories towards type 2 diabetes. *Nat. Metab.* 3, 1017–1031
26. Alonso, L. *et al.* (2021) TIGER: the gene expression regulatory variation landscape of human pancreatic islets. *Cell Rep.* 37, 109807
27. Vujkovic, M. *et al.* (2020) Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat. Genet.* 52, 680–691
28. Loeser, R.F. *et al.* (2012) Osteoarthritis: a disease of the joint as an organ. *Arthritis Rheum.* 64, 1697–1707
29. Goldring, S.R. and Goldring, M.B. (2016) Changes in the osteochondral unit during osteoarthritis: structure, function and cartilage-bone crosstalk. *Nat. Rev. Rheumatol.* 12, 632–644
30. Ratneswaran, A. *et al.* (2020) Understanding osteoarthritis pathogenesis: a multiomics system-based approach. *Curr. Opin. Rheumatol.* 32, 80–91
31. Katsoula, G. *et al.* (2022) Insights into the molecular landscape of osteoarthritis in human tissues. *Curr. Opin. Rheumatol.* 34, 79–90
32. Rice, S.J. *et al.* (2019) Prioritization of PLEC and GRINA as osteoarthritis risk genes through the identification and characterization of novel methylation quantitative trait loci. *Arthritis Rheumatol.* 71, 1285–1296
33. Rice, S.J. *et al.* (2019) Discovery and analysis of methylation quantitative trait loci (mQTLs) mapping to novel osteoarthritis genetic risk signals. *Osteoarthr. Cartil.* 27, 1545–1556
34. Parker, E. *et al.* (2021) Multi-tissue epigenetic and gene expression analysis combined with epigenome modulation identifies RWDD2B as a target of osteoarthritis susceptibility. *Arthritis Rheumatol.* 73, 100–109
35. Sorial, A.K. *et al.* (2020) Multi-tissue epigenetic analysis of the osteoarthritis susceptibility locus mapping to the plectin gene PLEC. *Osteoarthr. Cartil.* 28, 1448–1458
36. Steinberg, J. *et al.* (2017) Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis. *Sci. Rep.* 7, 8935
37. Steinberg, J. *et al.* (2021) A molecular quantitative trait locus map for osteoarthritis. *Nat. Commun.* 12, 1309
38. Coutinho de Almeida, R. *et al.* (2019) RNA sequencing data integration reveals an miRNA interactome of osteoarthritis cartilage. *Ann. Rheum. Dis.* 78, 270–277
39. Liu, Y. *et al.* (2018) Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis. *Sci. Rep.* 8, 15499
40. Kreitmaier, P. *et al.* (2022) An epigenome-wide view of osteoarthritis in primary tissues. *Am. J. Hum. Genet.* 109, 1255–1271
41. Katsoula, G. *et al.* (2022) A molecular map of long non-coding RNA expression, isoform switching and alternative splicing in osteoarthritis. *Hum. Mol. Genet.* 31, 2090–2105
42. Tachmazidou, I. *et al.* (2019) Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. *Nat. Genet.* 51, 230–236
43. Ji, Q. *et al.* (2019) Single-cell RNA-seq analysis reveals the progression of human osteoarthritis. *Ann. Rheum. Dis.* 78, 100–110
44. Schneider, J.A. *et al.* (2009) The neuropathology of probable Alzheimer disease and mild cognitive impairment. *Ann. Neurol.* 66, 200–208
45. Arnold, S.E. *et al.* (1991) The topographical and neuroanatomical distribution of neurofibrillary tangles and neuritic plaques in the cerebral cortex of patients with Alzheimer's disease. *Cereb. Cortex* 1, 103–116
46. Montine, T.J. *et al.* (2012) National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta Neuropathol.* 123, 1–11
47. Zhang, Q. *et al.* (2020) Risk prediction of late-onset Alzheimer's disease implies an oligogenic architecture. *Nat. Commun.* 11, 4799
48. Bellenguez, C. *et al.* (2022) New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* 54, 412–436
49. Corces, M.R. *et al.* (2020) Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* 52, 1158–1168
50. Lake, B.B. *et al.* (2018) Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* 36, 70–80
51. Morabito, S. *et al.* (2021) Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat. Genet.* 53, 1143–1155
52. Nativio, R. *et al.* (2020) An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. *Nat. Genet.* 52, 1024–1035
53. Beckmann, N.D. *et al.* (2020) Multiscale causal networks identify VGF as a key regulator of Alzheimer's disease. *Nat. Commun.* 11, 3942
54. Wang, M. *et al.* (2021) Transformative network modeling of multi-omics data reveals detailed circuits, key regulators, and potential therapeutics for Alzheimer's disease. *Neuron* 109, 257–272.e14
55. Bai, B. *et al.* (2020) Deep multilayer brain proteomics identifies molecular networks in Alzheimer's disease progression. *Neuron* 105, 975–991.e7
56. Lefterov, I. *et al.* (2019) APOE2 orchestrated differences in transcriptomic and lipidomic profiles of postmortem AD brain. *Alzheimers Res. Ther.* 11, 113
57. Oishi, T. *et al.* (2008) A functional SNP in the NKX2.5-binding site of ITPR3 promoter is associated with susceptibility to systemic lupus erythematosus in Japanese population. *J. Hum. Genet.* 53, 151–162
58. Fayyaz, A. *et al.* (2015) Haematological manifestations of lupus. *Lupus Sci. Med.* 2, e000078
59. Yin, X. *et al.* (2021) Meta-analysis of 208370 East Asians identifies 113 susceptibility loci for systemic lupus erythematosus. *Ann. Rheum. Dis.* 80, 632–640
60. Allen, M.E. *et al.* (2021) Leveraging heterogeneity in systemic lupus erythematosus for new therapies. *Trends Mol. Med.* 27, 152–171
61. Stohl, W. *et al.* (2017) Efficacy and safety of subcutaneous belimumab in systemic lupus erythematosus: a fifty-two-week randomized, double-blind, placebo-controlled study. *Arthritis Rheumatol.* 69, 1016–1027
62. Perez, R.K. *et al.* (2022) Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. *Science* 376, eabf1970
63. Robinson, G.A. *et al.* (2021) Increased apolipoprotein-B:A1 ratio predicts cardiometabolic risk in patients with juvenile onset SLE. *EBioMedicine* 65, 103243
64. Wang, T.-Y. *et al.* (2020) Identification of regulatory modules that stratify lupus disease mechanism through integrating multi-omics data. *Mol. Ther. Nucleic Acids* 19, 318–329
65. Liang, Y. *et al.* (2018) Coagulation cascade and complement system in systemic lupus erythematosus. *Oncotarget* 9, 14862–14881
66. Li, T. *et al.* (2017) A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat. Methods* 14, 61–64
67. Schraivogel, D. *et al.* (2020) Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat. Methods* 17, 629–635
68. Logsdon, G.A. *et al.* (2020) Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* 21, 597–614
69. Heydari, Z. *et al.* (2021) Organoids: a novel modality in disease modeling. *Biodes Manuf.* 4, 689–716
70. Kim, J. *et al.* (2020) Human organoids: model systems for human biology and medicine. *Nat. Rev. Mol. Cell Biol.* 21, 571–584
71. Cairns, D.M. *et al.* (2020) A 3D human brain-like tissue model of herpes-induced Alzheimer's disease. *Sci. Adv.* 6, eaay8828
72. ENCODE Project Consortium (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816
73. ENCODE Project Consortium (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 583, 699–710

74. Tabula Sapiens Consortium (2022) The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. *Science* 376, eabl4896
75. Westendorf, J.J. *et al.* (2022) The Musculoskeletal Knowledge Portal: improving access to multi-omics data. *Nat. Rev. Rheumatol.* 18, 1–2
76. Greenwood, A.K. *et al.* (2020) The AD Knowledge Portal: a repository for multi-omic data on Alzheimer's disease and aging. *Curr. Protoc. Hum. Genet.* 108, e105
77. King, E.A. *et al.* (2019) Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet.* 15, e1008489
78. Ogris, C. *et al.* (2021) Versatile knowledge guided network inference method for prioritizing key regulatory factors in multi-omics data. *Sci. Rep.* 11, 6806
79. Argelaguet, R. *et al.* (2020) MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* 21, 111
80. Mahajan, A. *et al.* (2022) Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat. Genet.* 54, 560–572