ANNUAL REVIEWS

Annual Review of Pathology: Mechanisms of Disease

Systems-Wide Approaches in Induced Pluripotent Stem Cell Models

Edward Lau,[1] David T. Paik,[1] and Joseph C. Wu[1,2]

[1]Stanford Cardiovascular Institute, and Department of Medicine, Division of Cardiology, Stanford University, Stanford, California 94305, USA; email: joewu@stanford.edu

[2]Department of Radiology, Stanford University, Stanford, California 94305, USA

Annu. Rev. Pathol. Mech. Dis. 2019. 14:395–419

First published as a Review in Advance on October 31, 2018

The *Annual Review of Pathology: Mechanisms of Disease* is online at pathol.annualreviews.org

https://doi.org/10.1146/annurev-pathmechdis-012418-013046

Keywords

induced pluripotent stem cells, disease modeling, transcriptomics, proteomics, systems biology

Abstract

Human induced pluripotent stem cells (iPSCs) provide a renewable supply of patient-specific and tissue-specific cells for cellular and molecular studies of disease mechanisms. Combined with advances in various omics technologies, iPSC models can be used to profile the expression of genes, transcripts, proteins, and metabolites in relevant tissues. In the past 2 years, large panels of iPSC lines have been derived from hundreds of genetically heterogeneous individuals, further enabling genome-wide mapping to identify coexpression networks and elucidate gene regulatory networks. Here, we review recent developments in omics profiling of various molecular phenotypes and the emergence of human iPSCs as a systems biology model of human diseases.

1. INTRODUCTION

Induced pluripotent stem cells (iPSCs): cells reprogrammed from adult differentiated cells into a naive state that are subsequently capable of infinite renewal in culture and differentiation into multiple cellular lineages

Differential expression analysis: comparison of gene, protein, or metabolite molecules to identify species with statistically significant differences in abundance between two conditions

The advent of induced pluripotent stem cells (iPSCs) has transformed biomedical research (1) by allowing primary cells from individual donors to be reprogrammed to a pluripotent state that is virtually identical to embryonic stem cells (2). Reliable methods exist to dedifferentiate adult donor cells into iPSCs with minimal mutations and genome instability (3) and to further direct their differentiation into human tissue types. Current protocols can create with high efficiency and purity cells in the nervous system (4), the heart (5, 6), the liver (7), the vasculature (8, 9), and other tissues. Further work has refined differentiation protocols to derive precise cellular subtypes, including motor versus sensory neurons (4, 10) and ventricular cardiac myocytes versus pacemaker cells (11–13). These breakthroughs have sparked novel preclinical applications of iPSCs. With the en masse production of human iPSC-derived cells, preclinical cell-based or cell-free therapies have been used to restore tissue function by directly administering the cells or their secreted factors into sites of injury (14, 15). In parallel, iPSC-derived cells are used as patient surrogates in molecular screening to discover new therapeutic compounds (16, 17) or to assess pharmacokinetics and safety (18, 19). But arguably, the most transformative applications of iPSCs have been in modeling the origin and development of human diseases (20, 21). With genetic diseases in particular, iPSCs carrying the exact genetic backgrounds of patients offer an attractive route to study diseased tissues directly. Patient-specific iPSC-derived cells have been studied in cardiomyopathies (21–24) and arrhythmias (25–27), pulmonary hypertension (28), and neurodegenerative disorders (29, 30), and they have led to new insights into the cellular processes disrupted in these diseases.

In parallel, during the past decade advances in next-generation sequencing and mass spectrometry technologies have fundamentally altered how biological molecules are interrogated on a large scale (31). RNA sequencing (RNA-seq) has become ubiquitous for measuring the output of gene expression in experimental models (32, 33), and with recent technologies, it has been extended to measuring full-length transcripts (34) or gene expression in single cells (35, 36). Mass spectrometry–based proteomics, in turn, allows large-scale quantification of protein expression (37, 38), protein posttranslational modifications (39, 40), and metabolite profiles (41). Transcript expression, proteins, metabolites, and other intermediary molecular phenotypes provide mechanistic information that connects genes to traits of interest (42–44) such that potential pathogenic mechanisms can be broadly discovered in an unbiased manner from differential expression analysis in disease models. The availability of large-scale omics data has spurred systems biology approaches, which aim to connect observed phenotypes to omics changes at scale such that their underlying regulatory networks can be revealed (43, 45).

Owing to their ability to reproduce individual-specific and tissue-specific molecular expression profiles, iPSC models have proven to be particularly suited for omics experiments (46). One advantage is their ability to circumvent the inaccessibility of primary tissues, especially cardiac and neuronal cells, so that representative transcripts, proteins, and metabolites can be procured for high-throughput quantitative experiments. Discoveries can further be tested in identical live cells, such as by using gene editing techniques to introduce variants (47, 48) (**Figure 1**). More recently, omics approaches in iPSCs have been extended to large donor cohorts, in which the natural variability in human populations is leveraged to uncover the genetic architecture connecting natural variants to observable traits (49, 50).

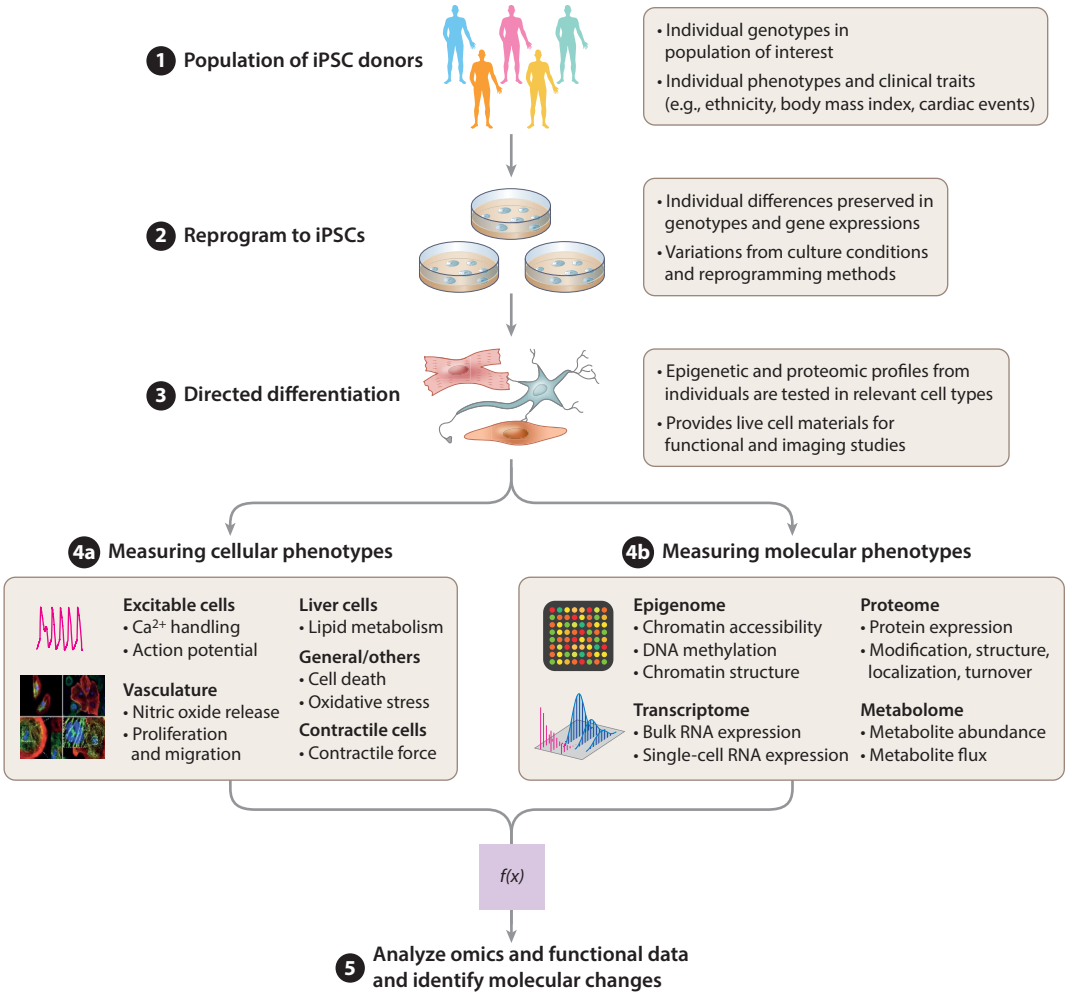

Figure 1

Omics approaches in induced pluripotent stem cell (iPSC) models of human diseases. ❶ Populations of healthy individuals and/or those with diseases donate skin or blood cells. ❷ iPSCs are derived from donors to capture their genetic backgrounds. ❸ The individual-specific iPSCs are coaxed into differentiated cells resembling primary tissues, including cardiomyocytes, neurons, and hepatocytes. ❹ The resulting iPSC-derived cells are used to profile (*a*) live cell functional phenotypes and (*b*) molecular expression. ❺ Large-scale profiling data are analyzed to discern the molecular mechanisms responsible for cellular phenotypes and disease traits.

2. USING OMICS METHODS TO QUANTIFY MOLECULAR AND CELLULAR PHENOTYPES

2.1. The Epigenome and Chromatin Structure

Unlike the genome, which is largely uniform across cells in the body, molecular phenotypes—such as chromatin landscapes, gene expression, and protein profiles—are highly dynamic and cell specific.

Epigenomic factors—including DNA methylation, transcription factor binding, enhancer–promoter contact, chromatin accessibility, and chromatin structure—have been shown to

critically regulate gene expression and may in fact be mechanistically responsible for the function of some genome-wide association study (GWAS) variants in intergenic regions. In addition, it is now known that the interphase genome folds like origami into a defined three-dimensional (3D) shape that determines cell type specification and gene expression. Two notable technologies have emerged to assay chromatin structure, including chromatin conformation capture (3C) and its derivatives, for assaying long-range interactions and chromatin topology (51), and the assay for transposase-accessible chromatin with sequencing (ATAC-seq), for identifying open and accessible chromatin regions (52). 3C and its derivatives employ formaldehyde cross-linking to physically join together DNA in proximity and then use the proximity library for high-throughput DNA sequencing to discern the 3D structure of chromatin and long-range interactions (51). These studies have generally found that human chromosomes are partitioned into Mb-sized topologically associated domains, wherein genes tend to be coregulated, and which can be classified as active or inactive. A number of extensions of these methods—including chromatin interaction analysis by paired-end tag sequencing (known as ChIA-PET) and 3C-coupled chromatin immunoprecipitation (known as Hi-ChIP)—further simplify the method and produce more targeted data by immunoprecipitation of cross-linked DNA using the particular proteins of interest to identify long-range contact associated with each protein factor (53). ATAC-seq uses a hyperactive prokaryotic Tn5 transposable element to sequence accessible chromatin regions that are open to the transposase (52). ATAC-seq has gained in popularity because it provides a sensitive method for sequencing open chromatin regions in native chromatin while requiring significantly fewer cells than previous methods (52).

In iPSC models, ATAC-seq has been applied to discover epigenomic changes occurring during differentiation processes. There appears to be a general correlation between the genome-wide transcriptome and chromatin accessibility, for instance, during directed cardiomyocyte differentiation (54), in which transcription factor binding motifs, including those for the *TBX*, *JUN*, and *STAT3* families, progressively become more open at the chromatin level to activate cardiac-specific gene regulatory networks. Moreover, the binding status of epigenetic factors, including the histone methyltransferases, has been shown to be important for regulating directed differentiation (55). In iPSC-derived neurons, ATAC-seq shows that open chromatin regions can be used to generate testable hypotheses on the function of GWAS-implicated psychiatric risk variants in noncoding regions (56).

Genome-wide association study (GWAS): large-scale association studies undertaken in a population to discover correlations between genetic variants with observable phenotypes

Chromatin conformation capture (3C): method to capture long-range interactions between two specific genomic loci by using cross-linking ligation reactions

ATAC-seq: epigenomic method that probes open chromatins to determine their accessibility for Tn5 transposase to insert adapters for sequencing

Gene regulatory network: an interconnected set of signaling genes, transcription factors, and their target genes that functions as a unit to modulate gene expression

Expression quantitative trait locus (eQTL): a quantitative trait locus that modulates the expression level of a gene as its quantitative trait

2.2. The Transcriptome and Single-Cell Gene Expression

Enabled by next-generation DNA sequencers, RNA-seq has supplanted earlier microarrays to allow for the routine quantitation of transcripts in biological samples (32, 33). The latest high-throughput sequencers from Illumina (San Diego, CA) and other manufacturers can generate billions of sequences from a single experiment thanks to the massively parallel nature of the sequencing reactions. Deducing the sequence of mRNAs with these sequencers allows for rapid quantitative assessment of the expression of tens of thousands of transcripts within a cell or tissue sample. Measurements of bulk RNA expression on a large scale are now commonly deployed to query gene expression in iPSC-derived cells. Clustering and unsupervised classification analyses are commonly used on RNA-seq data to compare diseased cells with normal cells to determine specific groups of genes or pathways that may be changed and thus implicate their potential importance in disease origin or in explaining observed cellular pathologies. RNA-seq also contextualizes the functions of gene variants in association studies and can be used for fine-mapping and identification of causal variants, as well as of the potential mechanisms by which they affect traits. Many GWAS variants function as expression quantitative trait loci (eQTLs) by affecting

transcript levels, whereas other exonic variants affect splicing ratios that can likewise be discerned using RNA-seq (57).

Single-cell RNA sequencing (scRNA-seq) is a recent development that allows transcript expression from single cells to be characterized to resolve transcriptional heterogeneity within cell populations. Its emergence was driven by technical advances made in constructing and amplifying sequencing libraries from miniscule amounts of RNA, as well as the development of microfluidic contraptions that allow for the separation of individual cells. Three major scRNA-seq approaches are in popular use. The first involves plate-based protocols that place individual cells into wells. The second involves automated microfluidic platforms that capture individual cells on microfluidic chips. The third involves a droplet-based massively parallel technique (**Table 1**) (58).

Single-cell RNA sequencing (scRNA-seq): next-generation sequencing technique that profiles the sequence and abundance of transcripts in individual single cells to measure expression heterogeneity

Plate-based techniques, such as Smart-seq, offer a fast and efficient method for analyzing 50 to 500 single cells in one experiment, with a flexible experimental setup (59). Current plate-based techniques boast increased accuracy and short processing times, and they are compatible with automation by liquid-handling robotics. They also allow cells of any morphology and size to be analyzed and can read up to 10,000 genes per single cell. C1 (Fluidigm, South San Francisco, CA), a commercially available, automated microfluidic platform, allows 96 individual cells to be captured at a time on a microfluidic chip. It offers the option to evaluate the captured cells under a microscope before reverse transcription, and it is effective for comparing homogeneous cell populations. However, the cost of reagents remains high and >10,000 cells are required as input, rendering the analysis of rare or small cell populations possible only when multiple samples are pooled. Ineffective automated sorting of cells into singlets also has been reported, during which multiplets are falsely analyzed as single cells. To overcome such limitations, cell expression by linear amplification and sequencing (CEL-seq) has been developed by combining the two technologies (60). CEL-seq applies a molecular barcode to cells at an early stage, thus lowering the cost of reagent and increasing the number of cells per sample to 500–2,000 cells. Subsequently, massively parallel scRNA-seq (known as MARS-seq) has also been developed by combining single-cell barcoding with a 384-well plate and fluorescence-activated cell sorting (known as FACS) to increase the scale and lower the costs (61). These pooled techniques allow for the isolation of various cell types and enhanced throughput. Finally, droplet-based scRNA-seq can tackle tens of thousands of single cells per sample, using barcoded complementary DNAs to label single cells encapsulated in individual droplets (62).

Insights into single-cell transcriptomes have revealed hidden heterogeneity in cell types and cell states (63, 64), decoded dynamic processes and developmental time lines (65, 66), and uncovered disease markers that are masked when cells are averaged in bulk sequencing (63, 67). In iPSC models, scRNA-seq has been used to understand the spatial and temporal heterogeneity of reprogramming (68) and differentiation (69, 70) and to identify novel surface markers for enrichment in induced cardiomyocytes (69). Machine-learning algorithms have also been used to predict the functional states of iPSC-derived neurons based on single-cell transcriptomes by using a combination of scRNA-seq and patch clamping to predict neuronal physiology and identify biomarkers of electrophysiologically active neurons (71).

2.3. The Proteome and Proteoforms

Although the central dogma of molecular biology dictates that information encoded in proteins derives from nucleic acids, the expression of proteins and, by extension, the metabolites they catalyze is poorly correlated with the expression of transcripts in many systems. In some instances, changes in the transcript explain only 10–40% of protein expression changes (72, 73). This noncorrelation is partly attributable to the fact that protein abundance is controlled by both the rates of

Table 1 Comparison of common single-cell analysis techniques

Description	Technology				
	Microfluidic (e.g., Fluidigm C1)	Droplet (e.g., Drop-seq, inDrop)[a]	Plate (e.g., Smart-seq, Smart-seq2)	Pooled approaches (e.g., CEL-seq2, MARS-seq)	Mass cytometry (e.g., cytometry by time-of-flight)
Molecule target	RNA	RNA	RNA	RNA	Protein
Single-cell separation and library construction principle	Aligns and separates cells in microfluidic channels based on size	Uses emulsion chemistry to construct separate libraries within each droplet	Deposits cells into individual wells of physical 96- or 384-well plates	Single-cell barcoding is combined with fluorescence-activated cell sorting or microfluidic-based cell sorting	Tags single cells with antibodies conjugated with heavy metal isotopes; separates by flow cytometry
Detection principle	RNA sequencing	RNA sequencing	RNA sequencing	RNA sequencing	Elemental mass spectrometry
Advantages	Works well with homogeneous populations Commercially available	Unbiased cell capture Massive parallelization of 800–10,000 cells per experiment Does not require cell sorting	Unbiased cell capture Fast and efficient way to capture 50–500 single cells Single cells can be stored in plates long term Generalizable; can analyze cells of any shape or size	Increased target cell number (500–2,000 cells) per experiment Decreased labor and reagent costs with higher sensitivity	Acquires protein-level information or cell-surface expression information
Disadvantages	High cost and high number of input cells Can analyze only up to 96–300 cells per experiment Large-scale experiments are not feasible Inadvertent multiplet capture	Reduced sensitivity and transcript recovery Restricted to analysis of cells smaller than droplet diameter	Experimental protocol can introduce technical noise Low sensitivity and high cost	Requires fluorescence-activated cell sorting prior to plating Cell number analyzable per sample is limited to 2,000 Relatively low number of genes detected	Requires compatible antibodies for targets Analyzes only 50–100 markers per cell

[a]inDrop System (1CellBio, Cambridge, MA).

production and of degradation. Whereas the rate of production scales with transcript abundance, the rate of degradation is influenced by posttranslational factors, including ubiquitin ligase activity, that cannot be easily modeled from transcripts alone (74–76). Therefore, a comprehensive model of molecular signatures requires the integration of protein- and transcript-level inquiries (76, 77). Moreover, the functional state of the proteome in a cell—such as the localization, structure, modification, and turnover dynamics of protein molecules—is fully described only in higher-dimensionality space (78); hence, there is interest in assaying these molecular parameters as well. However, proteomic analysis is a challenging endeavor. The ~20,000 genes in the human genome

can produce up to 6 million different protein species due to the addition of conformational, splicing, or chemically modified isomers, which are sometimes referred to as proteoforms (79). The dynamic range of concentration across protein species is also far greater than in transcripts, in which the most abundant protein can be 10 trillion times more abundant than the rarest species (80).

Two prevalent approaches for quantifying proteomes are affinity-based proteomics and tandem mass spectrometry (**Table 2**). For affinity-based proteomics using antibodies, mass cytometry approaches (or cytometry by time-of-flight, known as CyTOF) multiplex protein detection by attaching up to 134 metal isotopes to target-specific antibodies (81), and these are then resolved by elemental mass spectrometry (82). As with antibodies, aptamer probes operate by their specific molecular affinity to target antigens. Unlike immunoglobulins, aptamer probes are made of

Tandem mass spectrometry: analysis of proteins or metabolites that measures the accurate mass of a molecule (MS1), followed by isolation, fragmentation, and measuring the masses of its fragments

Table 2 Comparisons of proteomics technologies

Description	Technology					
	Mass spectrometry–based proteomics				Affinity-based proteomics	
Approach	Bottom-up DDA	Bottom-up DIA	Bottom-up targeted (PRM/MRM/SRM)	Top down	Antibody array	Aptamer array
Peptide or protein identification principle	Measures m/z of precursor and selects for fragmentation	Fragments m/z windows and assigns fragments to multiple peptides	Instrument programmed to monitor only specific precursor–fragment transition pairs	Measures m/z of intact proteins to discover proteoforms	Recognizes antigen epitopes using immunoglobulins	Recognizes antigen epitopes using nucleic acid
Advantages	Higher protein coverage than other mass spectrometry methods Can be adapted to protein modifications and different species or proteoforms High specificity	Higher reproducibility than DDA and higher throughput than targeted methods Can be adapted to protein modifications and different species or proteoforms High specificity	Higher sensitivity than other mass spectrometry methods Can be adapted to protein modifications and different species or proteoforms High specificity	Unique capability to identify isoforms and combinatorial modifications Can be adapted to protein modifications and different species or proteoforms High specificity	Lower cost than other affinity-based methods Ease and commercial availability High sensitivity	Higher coverage than other affinity-based methods Ease and commercial availability High sensitivity
Disadvantages	Requires hefty initial investment and access to technical expertise Sensitivity not as good as affinity-based approaches	Requires hefty initial investment and access to technical expertise Sensitivity not as good as affinity-based approaches	Requires hefty initial investment and access to technical expertise Sensitivity not as good as affinity-based approaches	Requires hefty initial investment and access to technical expertise Sensitivity not as good as affinity-based approaches	Specificity is uncertain Requires dedicated reagents for each target	Specificity is uncertain Requires dedicated reagents for each target

Abbreviations: DDA, data-dependent (shotgun) acquisition; DIA, data-independent acquisition; m/z, mass/charge; MRM, multiple reaction monitoring; PRM, parallel reaction monitoring; SRM, selected reaction monitoring.

Aptamer: oligonucleotide sequences that show binding affinity to specific shapes

specifically designed nucleic acid sequences, which fold into specific shapes to bind to targets, and they can be further modified to include hydrophobic motifs to enhance protein binding (83). An advantage of aptamers is that the probes can be easily synthesized de novo, and the selection of high-affinity probes can be automated by successive rounds of in vitro enrichment. In the past 5 years, applications have employed aptamers to analyze hundreds of samples to identify protein markers of myocardial infarcts (84) and, retrospectively, the cause of torcetrapib toxicity in a clinical trial (85). These assays are attractive for large clinical or population-based studies because they are commercially accessible and scalable, allowing more than a thousand proteins to be quantified in hundreds of samples with little prior expertise or investment needed. They have also been successfully demonstrated to yield multiprotein marker panels that can predict cardiovascular risks (86, 87).

In tandem mass spectrometry, the masses of tryptic peptides within a biological sample are measured accurately, and the peptides are fragmented in a predictable manner. The resulting tandem mass spectrum is computationally matched to a theoretical spectrum generated from genomic sequences to identify the peptide sequence. The performance of proteomics experiments has steadily increased with the availability of the advanced Orbitrap (Thermo Fisher Scientific, Waltham, MA) (88) and time-of-flight (89) mass spectrometers that feature high mass resolution and scan speed. Parallel to instrumentation advances, there has also been recent progress made in data acquisition methodologies. Conventional data-dependent (shotgun) acquisition has continued to be optimized in terms of sample processing chemistry and sampling regimens, and it can now reliably quantify virtually complete proteomes (more than 10,000 proteins), both in human cultured cells (37) and in more challenging tissues, such as from subanatomical regions of primary human heart (38). Data-independent acquisition (DIA) methods, such as sequential windowed acquisition of all theoretical fragment ion mass spectra (known as SWATH-MS), instruct the mass spectrometer to systematically queue all detectable ions in the parent mass spectrometry scan for fragmentation rather than fragmentation being triggered by the top-abundance ions. With modern fast-scanning instruments, DIA can overcome stochasticity in peptide detection and allow the same proteins to be targeted reliably in consecutive experiments (74, 90, 91). Recent DIA applications have demonstrated consistent quantification of more than 4,000 proteins across experiments in a manner that is reproducible across laboratories and operators with low coefficients of variation (92). Tandem mass spectrometry has the advantage of being a universal analyzer that can be readily applied to identify neoantigens, posttranslational modifications (79), protein interactions (93), novel isoforms (94), and proteins from any species.

In iPSC models, the power of mass cytometry to detect multiple surface protein markers in single cells has led to the discovery of markers of early stages of iPSC reprogramming that are absent both in the starting fibroblasts and reprogrammed iPSCs (95, 96). Mass spectrometry–based proteomics has also been widely utilized, including to identify cell surface glycoproteome markers for differentiating iPSC-derived hepatocytes that may be used for cell purification (97). A study in 2017 used tandem mass spectrometry to measure protein expression in 16 iPSC lines, and it showed that donor-specific differences in expression profiles remain discernible at the proteomic level (98).

2.4. The Metabolome and the Exposome

Metabolomics concerns the study of small molecules—including amino acids, α-keto acids, fatty acids, acylcarnitines, medium- and long-chain acyl-coenzyme A, and other lipids and organic acids—that are absorbed and released by different cell types in the body. The Human Metabolome Database lists more than 100,000 known metabolites, which include endogenous circulating metabolites in body fluids as well as xenobiotics, such as metabolites from foodstuffs and environmental pollutants (99). Environmental exposures, including exposures to small

molecules and other factors that can modulate disease risk and development, have been termed the exposome of an individual. The gut microbiome also interacts with the human metabolome by metabolizing small molecules in the body, some of which have been found to directly influence disease risks (100). Tandem mass spectrometry is one of the principal techniques, alongside nuclear magnetic resonance (NMR) spectroscopy, used for the large-scale discovery of metabolites and lipids. These two techniques are complementary. Although NMR can measure metabolites in a nondestructive manner in live biological specimens, mass spectrometry has higher sensitivity and throughput. When combined with stable isotope analogs of metabolites, mass spectrometry can also calculate the rate of flux of metabolites through metabolic pathways.

In iPSC models, mass spectrometry metabolomics profiling can be used to identify mechanisms in disease models, for example, to implicate oxidative stress and mitochondrial dysfunction in Pompe disease (101). For cell therapy research, metabolomic profiles of the cell culture medium of iPSC-derived cardiomyocytes under oxygen depletion revealed hypoxic responses involving ketogenesis, ketolysis, and methylglyoxal-related metabolism (102). Metabolomic profiling was also performed on a large panel of iPSC-derived hepatocytes to investigate the mechanism through which a GWAS variant operated in metabolic diseases, discovering a clear association between the risk allele at the implicated locus and variations in lipid synthesis in iPSC-derived hepatocytes. More importantly, this association was cell-type specific and was not found in iPSCs or adipocytes, corroborating the importance of measuring functional phenotypes in cells relevant to the tissue of interest (7).

2.5. The Phenome and Live Cell Functions

The complement of patient phenotypes is sometimes called the phenome to mirror the concept of the genome. The phenotypes of iPSC-derived cells can likewise be measured as the *in vitro* phenome and exposome, both of which often cannot be directly observed from patients' records or archived primary tissues. These phenotypes are useful for mapping associations with molecular profiles to identify the molecular pathways and genetic variants that control variations in functional traits.

For iPSC-derived excitable cells, such as cardiomyocytes and neurons, their action potential duration, velocity, and amplitude; depolarization curve; resting membrane potential; and individual ion currents can be measured using multielectrode arrays (103), voltage-sensitive optical reporters (104), or patch clamps (105, 106). L-type Ca^{2+} currents and intracellular Ca^{2+} stores (107) can be measured with calcium-sensitive fluorescence imaging (107, 108). These electrophysiological parameters are sensitive to individual donor differences, arrhythmias (26, 27), cytotoxicity (19), cardiomyopathies (109, 110), and neuronal pathologies (108). For contractile cells, such as iPSC-derived cardiomyocytes, the amplitude and kinetics of contraction can be measured by video imaging microscopy (111, 112) or electrical impedance (113). Their contractile force can be measured by atomic force microscopy (114), physical microposts (115), or substrate displacement (116), which decreases as expected by cardiomyopathy models (114, 117), mimicking clinical observations. Structurally, sarcomere lengths and widths can be measured by microscopy, and these have been shown to alter as expected by models of dilated cardiomyopathy (110).

Functions of the endothelium and the vasculature can be modeled by iPSC-derived endothelial cells and smooth muscle cells in tests for morphology, nitric oxide production, migration, and proliferation; these functional values can change according to endothelial dysfunction in diseases such as pulmonary arterial hypertension (28) and diabetes (118). Finally, other measurable parameters include cellular functions related to metabolism (119), energetics (22), oxidative stress (120), cell death, and proliferation.

3. USING LARGE iPSC PANELS FOR NETWORK BIOLOGY

3.1. Systems Biology Approaches Using Omics Data

Quantitative trait locus (QTL): a variant or region in the genome that is correlated with the variance of a measurable quantitative trait across individuals, such as body mass index

Coexpression networks: interconnected sets of genes that are implied to be functionally related because they covary in their expression patterns across cells, individuals, or species

Many common disorders, including coronary artery diseases and neurodegenerative diseases, arise from the interactions of complex genetic and environmental factors. Omics experiments suggest that even simple stimuli can trigger the differential expression of a plethora of genes (121), which poses questions about the overall contribution of individual targets or pathways. GWAS results further show that variants contributing to complex traits are scattered relatively evenly and ubiquitously across chromosomes in the genome, and hence, many traits may in fact be conceptualized as being omnigenic (122). This complexity has prompted interest in systems biology and network biology approaches, which take cues from the study of complex systems and attempt to model biological phenomena as the products of interacting molecular networks. Contrary to conventional single-target approaches, systems-wide approaches assume that genes function within the context of networks, in which members within a network module collectively function as a critical unit to define cell state and function (123). An example of a gene network is the cellular pluripotency network involving *NANOG*, *OCT4*, *SOX2*, and other genes. The functional state of the network cannot be reduced to any one gene (124), and the comparative analysis of a single gene is insufficient to define the network (125).

The topology of molecular networks can be inferred from large-scale surveys of their components prior to and after perturbations, and then by fitting the resulting data to suitable models (44, 121, 123). In practice, two approaches have been demonstrated in iPSC models to define relationships across genes (**Figure 2**), namely (*a*) using the association of genetic variants and molecular profiles (QTL studies) to unravel gene regulatory networks and (*b*) identifying correlated expression profiles among modules of genes (coexpression networks). Other types of biological networks, including physical protein–protein interaction networks (126, 127) and cell–cell interactions (128), also exist and await investigation in iPSC models.

3.2. Capturing Individual Variabilities in iPSC Panels

eQTL analysis finds associations between genetic variants in a population and the expression levels of transcripts (129). From these associations, eQTL studies infer the genes and variants that regulate gene expression in particular cell types. It is thought that many GWAS variants modulate the expression levels of particular genes, hence eQTLs often provide a mechanistic layer to explain how GWAS variants influence complex traits and to discern gene regulatory networks (44). Because of the accessibility of blood samples, many initial human eQTL studies were performed using transcript expression profiles from blood cells (129). Although informative, these studies did not capture all of the regulatory variants that are associated with traits of interests. Whereas many eQTLs are conserved across tissues (130), others regulate gene expression in tissue-specific manners, due in part to the unique physiological cues or chromatin landscapes in each cell type. This has motivated studies that leverage RNA-seq data sets from archived tissues. For instance, the Genotype–Tissue Expression (GTEx) Consortium data set, comprising 44 tissue types from 449 human donors (33, 131), allowed the genetic architectures of separate tissues to be examined. A drawback of archived tissues, however, is that interesting variants cannot be easily followed because live cell materials are not available from donors who have died.

Human iPSCs provide an alternative route to discern tissue-specific genetic architecture, and they also allow results to be validated in individual-specific live cells. Initial comparisons of interindividual versus intraindividual variations of RNA-seq data confirmed that iPSC lines from

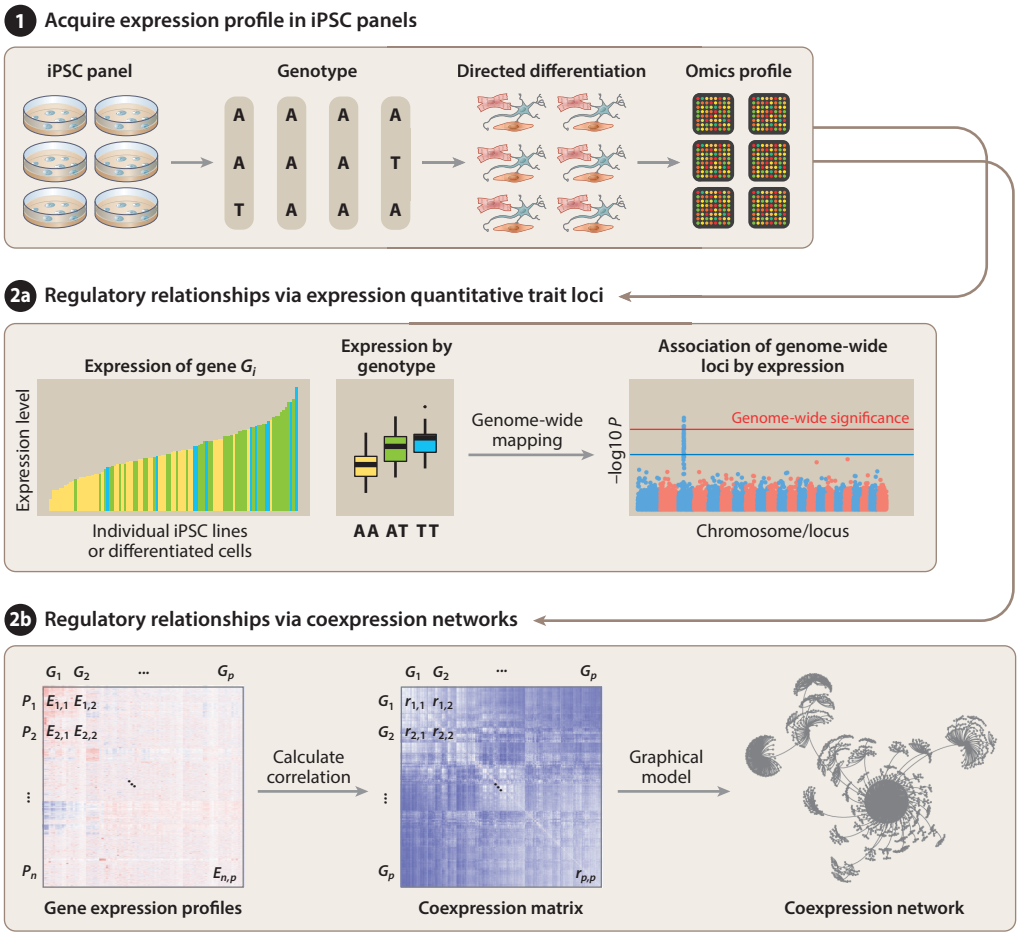

Figure 2

Identification of disease networks using quantitative trait locus (QTL) and coexpression analyses. ❶ Gene variants and expression profiles are acquired in large induced pluripotent stem cell (iPSC) panels. ❷ (*a*) QTL studies use genetic variants as causality anchors to map relationships between genes and traits (gene regulatory networks); (*b*) coexpression network modeling takes advantage of covariation in the expression profiles of functionally related genes across individuals to generate hypotheses about the underlying regulations of the genetic program(s).

individual donors can be reliably distinguished from one another using RNA-seq profiles (132–134), and individual differences in RNA-seq are sufficient to predict patient-specific phenotypes, such as responses to drug treatment (132). A study compared 12 iPSC lines from female breast cancer patients undergoing doxorubicin treatment and found that iPSC-derived cardiomyocytes recapitulated individual differences in drug cardiotoxicity that may be explained by individual RNA-seq profiles (19). It is generally observed that genetic differences between individuals are preserved in iPSCs regardless of the origin of the cells used for reprogramming (133, 134), suggesting that the architectures of gene regulatory networks may be reproduced with sufficient finesse to model individual variance in gene expression within even larger populations.

Induced pluripotent stem cell (iPSC) panels: collections of multiple iPSC lines from different donors used as a surrogate population

An early hurdle in using large iPSC panels for association mapping was the lack of consistent and scalable methods to produce a sufficient number of cell lines from large cohorts. Moreover, reprogramming and culturing conditions can create additional nongenetic sources of variance (135, 136). Although the use of iPSC models eschews some variances in primary tissue analysis, such as individual life history and environmental exposure, the overall variability of gene expression in iPSCs is thought to be greater than that in primary tissues (137). Because technical variances can mask biological signals, association studies using iPSCs require large populations to achieve sufficient power for discovery. It has been estimated that 40–80 or more lines are needed to detect regulatory variants with a large effect size in a given gene (137), and hundreds of lines are expected to be needed to discover QTLs with moderate effects. In the past 2 years, large human iPSC panels have emerged that compare iPSC lines from hundreds of donors with healthy versus diseased backgrounds. This development was enabled by advances in reprogramming and culturing protocols (138–140), quality control (141–143), and high-throughput production methods (141, 144), which increased the throughput of generating high-quality iPSCs.

Several consortia have been established to create iPSC panels from diverse donors, particularly for cardiovascular, metabolic, and neurobiology research (145). The National Heart, Lung, and Blood Institute's Next Generation Genetic Association Studies (NextGen) Consortium (146) was launched in 2011 to create iPSC libraries from thousands of multiethnic donors to study left ventricular hypertrophy, long QT syndrome, insulin resistance, and sickle cell anemia (146–148). The European Union's Innovative Medicines Initiative's StemBANCC (Stem cells for Biological Assays of Novel drugs and prediCtive toxiCology) program aims to generate 1,500 lines from 500 individuals, including healthy and diabetic donors (145). The HipSci (Human Induced Pluripotent Stem Cell Initiative) consortium, funded by the UK's Medical Research Council and the Wellcome Trust, aims to establish ~1,000 cell lines from healthy individuals and those with diseases (98) and to differentiate them into functional neurons. Complementing these efforts, more iPSC libraries have been established in biobanks at the Stanford Cardiovascular Institute, Cedars-Sinai Medical Center, the Allen Institute, and other institutes around the globe to capture cells from cohorts of patients and ensure iPSC panels are representative of diverse populations (149).

Several landmark studies have been published using iPSC panels from hundreds of individual donors. One early finding is the conclusive demonstration that iPSC lines preserve the genetic variability and architecture of individual donors (98, 137, 148, 150). Because the cohort designs of these studies involve hundreds of donors, as well as multiple iPSC clones per donor, comparisons can be made between variations in gene expression that exist across individuals (biological differences) and variations across lines from identical donors (technical variation). Comparisons of gene expression suggest that intraindividual clones cluster much more closely than interindividual lines and that interindividual variations account for 10–50% of differences in transcript expression (98, 137, 148), confirming the findings of early, smaller-scale studies (132). However, significant intraindividual heterogeneity was also evident, which explains almost 50% of gene expression variability. These residual variances are attributable, to various extents, to differences in reprogramming methods (138), culturing conditions (98, 148), batch effect (98, 148), the donor's sex (98), cell passage number (98), cell type of origin of iPSC cells (133, 151), and other unexplained sources. Curiously, the respective contributions of biological and technical variances appear to be gene specific, with donor differences being the largest contributor of variance in ~50% of genes (98). Genes with the highest intraindividual variance were strongly enriched for developmental function, whereas housekeeping genes had low variance (148), suggesting that reprogramming variability preferentially affects certain genes and that a network model may be useful for unearthing gene modules associated with reprogramming errors. A recent HipSci study further

confirmed that iPSCs preserve individual differences at the epigenomic and proteomic levels (98).

3.3. Identifying Regulatory Networks from Genetic Variants in iPSC Models

Another major finding is the demonstration that iPSC panels capture sufficient variability across individuals to power association mapping and identify eQTLs. One study of iPSC panels identified more than 4,000 eQTLs and rediscovered loci from GTEx tissue studies (148). Another study somewhat unexpectedly found that the power to discover eQTL-regulated genes in iPSCs was, in fact, comparable to that in somatic tissues of an identical sample size (98). Remarkably, up to one-third of iPSC eQTLs appear to be specific to iPSCs and are not found in other somatic tissues (98). These iPSC-specific eQTLs preferentially affect regulatory networks in stem cells and early development, including binding motifs for *NANOG* and other pluripotency factors (98, 152). They also appear to be significantly enriched in variants implicated in macroscopic GWAS traits, thus helping to connect some GWAS loci to the eQTLs of genes of interest, including telomerase reverse transcriptase in cancer (98). Moreover, up to three-quarters of these tissue-specific eQTL-regulated genes are controlled by a different set of variants in other tissues (i.e., their tissue specificity was not due to gene non-expression in other tissues) (98), strongly indicating that gene expression can be controlled by different regulatory networks in different tissues. In corroboration, many iPSC eQTLs have been found not only adjacent to the transcription start sites of the genes they regulate but also next to enhancers and promoters (148). Because enhancers and promoters are highly cell specific, this corroborates the utility of using human-specific differentiated cells to identify regulatory elements that are critical to traits of interest in relevant tissues.

Indeed, with the pluripotency of iPSCs, systems-level inquiries are readily extended to differentiated cell types to uncover genes that contribute to individual variability in tissue-specific phenotypes. The iPSCORE study provided a proof of concept that connected molecular and cellular traits in differentiated cardiovascular cells. The study used a panel of publicly available iPSCs from 222 ethnically diverse individuals, including 39 individuals with heart disease (arrhythmia or cardiomyopathies) (153). Genotyping indicated that the panel contained risk and benefit alleles for up to 95% of the single-nucleotide polymorphisms (SNPs) implicated by GWASs, including multiple SNPs associated with coronary artery disease. Following directed differentiation of seven lines into iPSC-derived cardiomyocytes, gene expression analysis showed that their profiles were clustered by genetic background, and variants were identified that associated with cellular phenotypes, including beat rates and electrophysiological measurements (153).

For metabolic traits, one NextGen study differentiated 86 iPSC lines into hepatocytes (150). RNA-seq in both cell types allowed for assessment of the tissue specificity of regulatory networks, and the study identified eQTL genes that were specific to iPSCs or iPSC-derived hepatocytes or that were common to both. Hepatocyte-specific eQTL genes are enriched for the regulation of cholesterol levels. The study further uncovered novel associations between hepatocyte-regulated genes and lipid metabolism that were not found in primary tissue analysis in the GTEx project, including associations with *CPNE1* and *ANGPTL3* (150). In parallel, one HipSci study differentiated iPSC lines from 100 donors to sensory neurons (137) and performed RNA-seq to identify eQTLs on ~3,800 genes in these neurons, approximately one-quarter of which were novel and not found in the GTEx project. In addition, RNA-seq and ATAC-seq data from the same study enabled the discovery of neuron-specific splicing QTLs and quantitative chromatin accessibility QTLs. The identified expression, splicing, and chromatin accessibility variants overlapped with GWAS hits for neuronal traits, thereby demonstrating that large-scale RNA-seq and ATAC-seq of iPSC panels can provide valuable mechanistic contexts for GWAS findings (137).

3.4. Identifying Coexpression Networks in iPSC Models

In addition to QTL mapping, coexpression analysis represents another approach for querying the structure of gene networks (43, 123). Coexpression network analysis is predicated on the assumption that, under selection, an efficient genetic program will evolve with an organization such that genes that work together physically or functionally should come to be coregulated by the cell. The analysis asks across individuals, which genes tend to have their expression levels cluster together under homeostatic versus stimulated conditions, the result of which is used to infer functionally coregulated genes (160). Further inference may be drawn by introducing constraints to the coregulated genes, either by seeding a network based on prior knowledge of the gold standard regulatory relationships or by making additional assumptions about the behavior and structure of the network (**Figure 3**) (160, 161). The constructed correlation network can be combined with probabilistic causal modeling to predict the directionality of network edges and identify driver genes. Bayesian networks are frequently employed to predict which one of a cohort

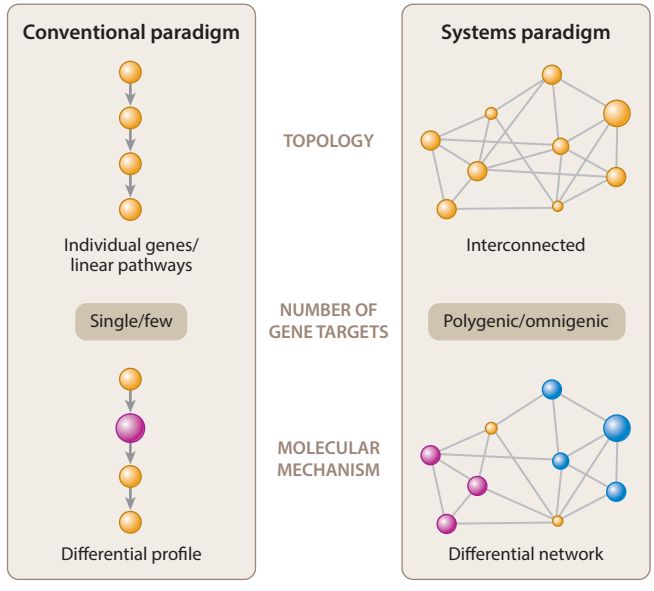

Figure 3

Properties of gene networks: genes conceptualized as linear pathways versus complex networks. In the linear pathway worldview, disease genes are identified via the conspicuous differential expression of one or a few molecular markers along known pathways (differential profile). In the complex network view, differences in subnetwork memberships can be analyzed to detect changes in networks or subnetworks between two states and to identify disease modules. Network theories have discovered a number of properties commonly found in biological networks, including gene networks (123). Gene networks are assumed to follow a scale-free topology, in which the distribution of degrees (i.e., the number of neighbors) of nodes in the network follows a power distribution. Scale-free topology has been found in various social and biological networks, is thought to emerge automatically from individual interactions (154), and underlies the resilience and evolvability of complex adaptive systems (155, 156). It also predicts the emergence of hub genes (157) and the small world phenomenon, in which most genes can be connected to one another by a small number of steps (123). One prediction of the model is that differential network topologies can also occur due to the formation or dissolution of covariation between genes, which leads to changes in node connectivity (158). In disease, changes in gene regulation can rewire networks so that genes normally associated with one subnetwork become associated with another (159).

of coexpressed genes—based on their variance across all samples—is statistically more likely to influence the expression of the other genes.

In iPSC models, consensus coexpression networks from multiple transcriptomic data sets have been leveraged to determine cell identity and guide iPSC engineering and differentiation (162). RNA-seq data from iPSC panels have also been leveraged for coexpression network analysis. In one study comparing 300 iPSC lines, the expression variance that was not explained by eQTLs was found to derive mostly from targets of the Polycomb repressive complex, suggesting that epigenetic regulation from chromatin remodeling is a major factor in reprogramming reproducibility (148). Coexpression analysis was subsequently performed to infer a probabilistic causal network, identifying a number of key drivers that causally affect the expression levels of developmental pathways in iPSCs, namely *GATA4*, *GATA6*, *EOMES*, *FOXQ1*, *CER1*, *APOA2*, and *LINC00261*, notably with the first five also being known Polycomb targets (148). In the same study, causal network inference also nominated *HOXA5* and *HOXC10* as likely key driver genes that can influence the endothelial differentiation potential of cells (148). Interestingly, the contribution of the Polycomb repressive complex to differentiation variability was corroborated in an independent comparison of gene expression profiles of iPSCs that were generated using different reprogramming methods (lentivirus, Sendai virus, episomal, mRNA, minicircles, or mRNA and microRNA) (138).

Finally, scRNA-seq is opening new avenues to coexpression network analysis. Individual cells can regulate transcription based on stochastic heterogeneity in cell state, volume, and cycle (147), and thus variations of transcripts across cells can inform coexpression even under identical genetic backgrounds. Coexpression networks can be constructed with higher throughput by considering the correlated expression of genes across many cells on the same culture dish rather than across many cell lines (163, 164). scRNA-seq can be further combined with recent advances in genome editing for massively parallel reverse genetic screens, such as in the Perturb-seq (165) or CRISPR-seq (166) approaches for modeling genetic networks. In brief, a CRISPR pooled screen is set up such that each individual cell is targeted with a random guide RNA and barcode to knock down or knock out a random gene. The resulting transcriptomic change within the cell is then measured by scRNA-seq and connected to each barcode, resulting in a massive genetic screen that simultaneously measures the consequences of the individual disruption of tens of thousands of genes. The resulting data can be modeled as a regulatory network in which the expression level of each gene is a linear combination of regulatory inputs at one or more disrupted genes.

4. CONCLUDING REMARKS

The confluence of iPSCs and omics technologies allows for the tissue-specific and patient-specific expression profiles of genes, transcripts, proteins, and metabolites to be queried on a large scale. Large-scale omics data can be further leveraged across samples to provide useful information on gene regulatory networks in relevant cell types. We anticipate that the combination of systems-wide approaches and iPSCs will benefit two primary areas of disease inquiry in the near term. First, large iPSC panels will serve as surrogates for human populations and allow systems genetics studies to identify gene–trait correlations in complex diseases (49). The construction of iPSC panels from donors of multiple ethnicities will allow association studies to more routinely cover underrepresented populations. Second, understanding the architecture of gene regulatory networks in various cell types will likely improve understanding of cell-type specification in development and disease, iPSC engineering, and deconvolution of mixture samples (162, 167). With

CURRENT LIMITATIONS AND ONGOING REFINEMENT OF iPSC MODELS

A major limitation of iPSC models is that derived cells are immature and resemble fetal cells. This is true for iPSC-derived neurons (137), hepatocytes (150), and cardiomyocytes (50). For example, iPSC-derived cardiomyocytes have underdeveloped calcium handling and contraction (116), and their gene expression profiles resemble first-trimester fetal hearts (170). This immaturity casts doubt on whether discovered gene regulatory networks are relevant to adult tissues and diseases that manifest in adulthood. Efforts to promote maturity include optimizing culturing substrates and conditions (171, 172), exposing the cells to biophysical or electrical stimulation (173), as well as supplementing the culture medium with molecular signaling cues and trophic factors (21, 174).

Another limitation is that 2D iPSC cultures do not capture tissue-level features, such as 3D geometries and cell–cell interactions (175). To address this, efforts are being made to capture essential missing interactions in tractable iPSC models using organoid and tissue-on-a-chip approaches (176). These engineering efforts attempt to create 3D cellular models that incorporate defined geometries or tissue microstructures (119, 177), and they can further incorporate multiple cell types that make up a resident tissue, including endothelial and myocyte cells (178) or astrocytes and neurons (179). These engineered systems have been demonstrated to mimic the heart, brain, liver, kidney, and other tissues (180, 181). However, a drawback is a further increase in technical variations and a decrease in throughput for large iPSC panels. Finally, the life histories and environmental exposures of individuals are not captured and need to be reintroduced in vitro. A clear consensus has yet to emerge on whether iPSCs reprogrammed from young and old individuals reset cellular age (182, 183).

continued improvements in iPSC models (see the sidebar titled Current Limitations and Ongoing Refinement of iPSC Models) as well as the falling costs of sequencing, one can envision that iPSC-based systems biology will become increasingly valuable for studies of disease mechanisms.

To make iPSC panels more accessible to biomedical researchers, we believe there is a need to improve the standardization and accessibility of cell line material and data. The influx of omics big data has prompted technical and policy advances to make biomedical data science findable, accessible, interoperable, and reusable (or FAIR). A parallel effort in iPSC democratization will include the adoption and free sharing of standardized cell lines, the adoption of standardized nomenclature, and open data from iPSC model studies. Provided that the iPSC lines used for association studies are made publicly available, iPSC panels would constitute an additive resource over time. As new sequencing techniques become more accessible in the future, new data acquired across these standard iPSC lines can be analyzed within the context of genotype and transcriptome data. This collaborative systems genetics model has yielded biological insights for preclinical studies in rodent models [such as the Hybrid Mouse Diversity Panel (168) and the Collaborative Cross (169)]. With openly accessible iPSC lines from consortiums and individual biobanks, we foresee iPSC panels will have similar impacts on human studies.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Jared Churko, Mark Chandy, June-Wha Rhee, and Blake Wu for critically reading this manuscript. This work was supported in part by the American Heart Association (grant 17MERIT336100009 to J.C.W.), a Burroughs Wellcome Fund Innovation in Regulatory Science

Award (grant 1015009 to J.C.W.), a Tobacco-Related Disease Research Program of California Award (27IR-0012 to J.C.W.), and the US National Institutes of Health (grants R01 HL128170, R01 HL130020, R01 HL141851, R01 HL141371, and R01 HL123968 to J.C.W.; F32 HL139045 to E.L.; and T32 EB009035 to D.T.P.).

LITERATURE CITED

1. Shi Y, Inoue H, Wu JC, Yamanaka S. 2017. Induced pluripotent stem cell technology: a decade of progress. *Nat. Rev. Drug Discov.* 16(2):115–30
2. Choi J, Lee S, Mallard W, Clement K, Tagliazucchi GM, et al. 2015. A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs. *Nat. Biotechnol.* 33(11):1173–81
3. Wen W, Zhang J-P, Xu J, Su RJ, Neises A, et al. 2016. Enhanced generation of integration-free iPSCs from human adult peripheral blood mononuclear cells with an optimal combination of episomal vectors. *Stem Cell Rep.* 6(6):873–84
4. Karumbayaram S, Novitch BG, Patterson M, Umbach JA, Richter L, et al. 2009. Directed differentiation of human-induced pluripotent stem cells generates active motor neurons. *Stem Cells* 27(4):806–11
5. Burridge PW, Matsa E, Shukla P, Lin ZC, Churko JM, et al. 2014. Chemically defined generation of human cardiomyocytes. *Nat. Methods* 11(8):855–60
6. Lian X, Hsiao C, Wilson G, Zhu K, Hazeltine LB, et al. 2012. Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *PNAS* 109(27):E1848–57
7. Warren CR, O'Sullivan JF, Friesen M, Becker CE, Zhang X, et al. 2017. Induced pluripotent stem cell differentiation enables functional validation of GWAS variants in metabolic disease. *Cell Stem Cell* 20(4):547–57.e7
8. Marchand M, Anderson EK, Phadnis SM, Longaker MT, Cooke JP, et al. 2014. Concurrent generation of functional smooth muscle and endothelial cells via a vascular progenitor. *Stem Cells Transl. Med.* 3(1):91–97
9. Patsch C, Challet-Meylan L, Thoma EC, Urich E, Heckel T, et al. 2015. Generation of vascular endothelial and smooth muscle cells from human pluripotent stem cells. *Nat. Cell Biol.* 17(8):994–1003
10. Chambers SM, Qi Y, Mica Y, Lee G, Zhang X-J, et al. 2012. Combined small-molecule inhibition accelerates developmental timing and converts human pluripotent stem cells into nociceptors. *Nat. Biotechnol.* 30(7):715–20
11. Protze SI, Liu J, Nussinovitch U, Ohana L, Backx PH, et al. 2017. Sinoatrial node cardiomyocytes derived from human pluripotent cells function as a biological pacemaker. *Nat. Biotechnol.* 35(1):56–68
12. Lee JH, Protze SI, Laksman Z, Backx PH, Keller GM. 2017. Human pluripotent stem cell–derived atrial and ventricular cardiomyocytes develop from distinct mesoderm populations. *Cell Stem Cell* 21(2):179–94.e4
13. Birket MJ, Ribeiro MC, Verkerk AO, Ward D, Leitoguinho AR, et al. 2015. Expansion and patterning of cardiovascular progenitors derived from human pluripotent stem cells. *Nat. Biotechnol.* 33(9):970–79
14. Neofytou E, O'Brien CG, Couture LA, Wu JC. 2015. Hurdles to clinical translation of human induced pluripotent stem cells. *J. Clin. Investig.* 125(7):2551–57
15. Mandai M, Watanabe A, Kurimoto Y, Hirami Y, Morinaga C, et al. 2017. Autologous induced stem-cell-derived retinal cells for macular degeneration. *N. Engl. J. Med.* 376(11):1038–46
16. Vazão H, Rosa S, Barata T, Costa R, Pitrez PR, et al. 2017. High-throughput identification of small molecules that affect human embryonic vascular development. *PNAS* 114(15):E3022–31
17. Liang P, Lan F, Lee AS, Gong T, Sanchez-Freire V, et al. 2013. Drug screening using a library of human induced pluripotent stem cell–derived cardiomyocytes reveals disease-specific patterns of cardiotoxicity. *Circulation* 127(16):1677–91
18. Sharma A, Burridge PW, McKeithan WL, Serrano R, Shukla P, et al. 2017. High-throughput screening of tyrosine kinase inhibitor cardiotoxicity with human induced pluripotent stem cells. *Sci. Transl. Med.* 9(377):aaf2584

19. Burridge PW, Li YF, Matsa E, Wu H, Ong S-G, et al. 2016. Human induced pluripotent stem cell–derived cardiomyocytes recapitulate the predilection of breast cancer patients to doxorubicin-induced cardiotoxicity. *Nat. Med.* 22(5):547–56
20. Chal J, Oginuma M, Al Tanoury Z, Gobert B, Sumara O, et al. 2015. Differentiation of pluripotent stem cells to muscle fiber to model Duchenne muscular dystrophy. *Nat. Biotechnol.* 33(9):962–69
21. Birket MJ, Ribeiro MC, Kosmidis G, Ward D, Leitoguinho AR, et al. 2015. Contractile defect caused by mutation in *MYBPC3* revealed under conditions optimized for human PSC-cardiomyocyte function. *Cell Rep.* 13(4):733–45
22. Li S, Pan H, Tan C, Sun Y, Song Y, et al. 2018. Mitochondrial dysfunctions contribute to hypertrophic cardiomyopathy in patient iPSC-derived cardiomyocytes with *MT-RNR2* mutation. *Stem Cell Rep.* 10(3):P808–21
23. Kodo K, Ong S-G, Jahanbani F, Termglinchan V, Hirono K, et al. 2016. iPSC-derived cardiomyocytes reveal abnormal TGF-β signalling in left ventricular non-compaction cardiomyopathy. *Nat. Cell Biol.* 18(10):1031–42
24. Wu H, Lee J, Vincent LG, Wang Q, Gu M, et al. 2015. Epigenetic regulation of phosphodiesterases 2A and 3A underlies compromised β-adrenergic signaling in an iPSC model of dilated cardiomyopathy. *Cell Stem Cell* 17(1):89–100
25. Wang Y, Liang P, Lan F, Wu H, Lisowski L, et al. 2014. Genome editing of isogenic human induced pluripotent stem cells recapitulates long QT phenotype for drug testing. *J. Am. Coll. Cardiol.* 64(5):451–59
26. Bellin M, Casini S, Davis RP, D'Aniello C, Haas J, et al. 2013. Isogenic human pluripotent stem cell pairs reveal the role of a *KCNH2* mutation in long-QT syndrome. *EMBO J.* 32(24):3161–75
27. Liang P, Sallam K, Wu H, Li Y, Itzhaki I, et al. 2016. Patient-specific and genome-edited induced pluripotent stem cell–derived cardiomyocytes elucidate single-cell phenotype of Brugada syndrome. *J. Am. Coll. Cardiol.* 68(19):2086–96
28. Gu M, Shao N-Y, Sa S, Li D, Termglinchan V, et al. 2017. Patient-specific iPSC-derived endothelial cells uncover pathways that protect against pulmonary hypertension in *BMPR2* mutation carriers. *Cell Stem Cell* 20(4):490–504.e5
29. Haidet-Phillips AM, Hester ME, Miranda CJ, Meyer K, Braun L, et al. 2011. Astrocytes from familial and sporadic ALS patients are toxic to motor neurons. *Nat. Biotechnol.* 29(9):824–28
30. Torrent R, De Angelis Rigotti F, Dell'Era P, Memo M, Raya A, Consiglio A. 2015. Using iPS cells toward the understanding of Parkinson's disease. *J. Clin. Med.* 4(4):548–66
31. Lau E, Wu JC. 2018. Omics, big data, and precision medicine in cardiovascular sciences. *Circ. Res.* 122(9):1165–68
32. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, et al. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17(1):13
33. GTEx Consort. 2017. Genetic effects on gene expression across human tissues. *Nature* 550(7675):204–13
34. Tilgner H, Jahanbani F, Blauwkamp T, Moshrefi A, Jaeger E, et al. 2015. Comprehensive transcriptome analysis using synthetic long-read sequencing reveals molecular co-association of distant splicing events. *Nat. Biotechnol.* 33(7):736–42
35. Gawad C, Koh W, Quake SR. 2016. Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* 17(3):175–88
36. DeLaughter DM, Bick AG, Wakimoto H, McKean D, Gorham JM, et al. 2016. Single-cell resolution of temporal gene expression during heart development. *Dev. Cell* 39(4):480–90
37. Bekker-Jensen DB, Kelstrup CD, Batth TS, Larsen SC, Haldrup C, et al. 2017. An optimized shotgun strategy for the rapid generation of comprehensive human proteomes. *Cell Syst.* 4(6):587–99.e4
38. Doll S, Dreßen M, Geyer PE, Itzhak DN, Braun C, et al. 2017. Region and cell-type resolved quantitative proteomic map of the human heart. *Nat. Commun.* 8(1):1469
39. Riley NM, Coon JJ. 2016. Phosphoproteomics in the age of rapid and deep proteome profiling. *Anal. Chem.* 88(1):74–94
40. Olsen JV, Mann M. 2013. Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol. Cell. Proteom.* 12(12):3444–52

41. Patti GJ, Yanes O, Siuzdak G. 2012. Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* 13(4):263–69
42. eGTEx Proj. 2017. Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. *Nat. Genet.* 49(12):1664–70
43. Parikshak NN, Gandal MJ, Geschwind DH. 2015. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* 16(8):441–58
44. Civelek M, Lusis AJ. 2014. Systems genetics approaches to understand complex traits. *Nat. Rev. Genet.* 15(1):34–48
45. Price ND, Magis AT, Earls JC, Glusman G, Levy R, et al. 2017. A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nat. Biotechnol.* 35(8):747–56
46. Matsa E, Ahrens JH, Wu JC. 2016. Human induced pluripotent stem cells as a platform for personalized and precision cardiovascular medicine. *Physiol. Rev.* 96(3):1093–126
47. Hockemeyer D, Jaenisch R. 2016. Induced pluripotent stem cells meet genome editing. *Cell Stem Cell* 18(5):573–86
48. Hotta A, Yamanaka S. 2015. From genomics to gene therapy: induced pluripotent stem cells meet genome editing. *Annu. Rev. Genet.* 49:47–70
49. Warren CR, Cowan CA. 2018. Humanity in a dish: population genetics with iPSCs. *Trends Cell Biol.* 28(1):46–57
50. Sayed N, Liu C, Wu JC. 2016. Translation of human-induced pluripotent stem cells: from clinical trial in a dish to precision medicine. *J. Am. Coll. Cardiol.* 67(18):2161–76
51. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159(7):1665–80
52. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, et al. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* 14(10):959–62
53. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, et al. 2016. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* 13(11):919–22
54. Liu Q, Jiang C, Xu J, Zhao M-T, Van Bortle K, et al. 2017. Genome-wide temporal profiling of transcriptome and open chromatin of early cardiomyocyte differentiation derived from hiPSCs and hESCs. *Circ. Res.* 121(4):376–91
55. Lee J, Shao N-Y, Paik DT, Wu H, Guo H, et al. 2018. SETD7 drives cardiac lineage commitment through stage-specific transcriptional activation. *Cell Stem Cell* 22(3):428–44.e5
56. Forrest MP, Zhang H, Moy W, McGowan H, Leites C, et al. 2017. Open chromatin profiling in hiPSC-derived neurons prioritizes functional noncoding psychiatric risk variants and highlights neurodevelopmental loci. *Cell Stem Cell* 21(3):305–18.e8
57. Li YI, van de Geijn B, Raj A, Knowles DA, Petti AA, et al. 2016. RNA splicing is a primary link between genetic variation and disease. *Science* 352(6285):600–4
58. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, et al. 2017. Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* 65(4):631–43
59. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10(11):1096–98
60. Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, et al. 2016. CEL-seq2: sensitive highly-multiplexed single-cell RNA-seq. *Genome Biol.* 17:77
61. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, et al. 2014. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343(6172):776–79
62. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8:14049
63. Papalexi E, Satija R. 2018. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* 18(1):35–45
64. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, et al. 2015. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* 33(2):155–60

65. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, et al. 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 357(6352):661–67
66. Karaiskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, et al. 2017. The *Drosophila* embryo at single-cell transcriptome resolution. *Science* 358(6360):194–99
67. Moignard V, Woodhouse S, Haghverdi L, Lilly AJ, Tanaka Y, et al. 2015. Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* 33(3):269–76
68. Hough SR, Thornton M, Mason E, Mar JC, Wells CA, Pera MF. 2014. Single-cell gene expression profiles define self-renewing, pluripotent, and lineage primed states of human pluripotent stem cells. *Stem Cell Rep.* 2(6):881–95
69. Liu Z, Wang L, Welch JD, Ma H, Zhou Y, et al. 2017. Single-cell transcriptomics reconstructs fate conversion from fibroblast to cardiomyocyte. *Nature* 551(7678):100–4
70. Paik DT, Tian L, Lee J, Sayed N, Chen IY, et al. 2018. Large-scale single-cell RNA-seq reveals molecular signatures of heterogeneous populations of human induced pluripotent stem cell–derived endothelial cells. *Circ. Res.* 123:443–50
71. Bardy C, van den Hurk M, Kakaradov B, Erwin JA, Jaeger BN, et al. 2016. Predicting the functional states of human iPSC-derived neurons with single-cell RNA-seq and electrophysiology. *Mol. Psychiatry* 21(11):1573–88
72. Liu Y, Beyer A, Aebersold R. 2016. On the dependency of cellular protein levels on mRNA abundance. *Cell* 165(3):535–50
73. Cheng Z, Teo G, Krueger S, Rock TM, Koh HWL, et al. 2016. Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress. *Mol. Syst. Biol.* 12(1):855
74. Liu Y, Borel C, Li L, Müller T, Williams EG, et al. 2017. Systematic proteome and proteostasis profiling in human trisomy 21 fibroblast cells. *Nat. Commun.* 8(1):1212
75. Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, et al. 2015. Impact of regulatory variation from RNA to protein. *Science* 347(6222):664–67
76. Chick JM, Munger SC, Simecek P, Huttlin EL, Choi K, et al. 2016. Defining the consequences of genetic variation on a proteome-wide scale. *Nature* 534(7608):500–5
77. Lau E, Cao Q, Lam MPY, Wang J, Ng DCM, et al. 2018. Integrated omics dissection of proteome dynamics during cardiac remodeling. *Nat. Commun.* 9(1):120
78. Aebersold R, Mann M. 2016. Mass-spectrometric exploration of proteome structure and function. *Nature* 537(7620):347–55
79. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, et al. 2018. How many human proteoforms are there? *Nat. Chem. Biol.* 14(3):206–14
80. Schwenk JM, Omenn GS, Sun Z, Campbell DS, Baker MS, et al. 2017. The human plasma proteome draft of 2017: building on the Human Plasma PeptideAtlas from mass spectrometry and complementary assays. *J. Proteome Res.* 16(12):4299–310
81. Han G, Chen S-Y, Gonzalez VD, Zunder ER, Fantl WJ, Nolan GP. 2017. Atomic mass tag of bismuth-209 for increasing the immunoassay multiplexing capacity of mass cytometry. *Cytometry A* 91(12):1150–63
82. Leipold MD, Obermoser G, Fenwick C, Kleinstuber K, Rashidi N, et al. 2018. Comparison of CyTOF assays across sites: results of a six-center pilot study. *J. Immunol. Methods* 453:37–43
83. Gawande BN, Rohloff JC, Carter JD, von Carlowitz I, Zhang C, et al. 2017. Selection of DNA aptamers with two modified bases. *PNAS* 114(11):2898–903
84. Jacob J, Ngo D, Finkel N, Pitts R, Gleim S, et al. 2017. Application of large scale aptamer-based proteomic profiling to "planned" myocardial infarctions. *Circulation* 137(12):1270–77
85. Williams SA, Murthy AC, DeLisle RK, Hyde C, Malarstig A, et al. 2017. Improving assessment of drug safety through proteomics: early detection and mechanistic characterization of the unforeseen harmful effects of torcetrapib. *Circulation* 137(10):999–1010
86. Ganz P, Heidecker B, Hveem K, Jonasson C, Kato S, et al. 2016. Development and validation of a protein-based risk score for cardiovascular outcomes among patients with stable coronary heart disease. *JAMA* 315(23):2532–41
87. Ngo D, Sinha S, Shen D, Kuhn EW, Keyes MJ, et al. 2016. Aptamer-based proteomic profiling reveals novel candidate biomarkers and pathways in cardiovascular disease. *Circulation* 134(4):270–85

88. Kelstrup CD, Bekker-Jensen DB, Arrey TN, Hogrebe A, Harder A, Olsen JV. 2018. Performance evaluation of the Q Exactive HF-X for shotgun proteomics. *J. Proteome Res.* 17(1):727–38
89. Garabedian A, Benigni P, Ramirez CE, Baker ES, Liu T, et al. 2017. Towards discovery and targeted peptide biomarker detection using nanoESI-TIMS-TOF MS. *J. Am. Soc. Mass Spectrom.* 29(5):817–26
90. Rosenberger G, Liu Y, Röst HL, Ludwig C, Buil A, et al. 2017. Inference and quantification of peptidoforms in large sample cohorts by SWATH-MS. *Nat. Biotechnol.* 35(8):781–88
91. Fu Q, Kowalski MP, Mastali M, Parker SJ, Sobhani K, et al. 2018. Highly reproducible automated proteomics sample preparation workflow for quantitative mass spectrometry. *J. Proteome Res.* 17(1):420–28
92. Collins BC, Hunter CL, Liu Y, Schilling B, Rosenberger G, et al. 2017. Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH–mass spectrometry. *Nat. Commun.* 8(1):291
93. Chavez JD, Lee CF, Caudal A, Keller A, Tian R, Bruce JE. 2018. Chemical crosslinking mass spectrometry analysis of protein conformations and supercomplexes in heart tissue. *Cell Syst.* 6(1):136–41.e5
94. Liu Y, Gonzàlez-Porta M, Santos S, Brazma A, Marioni JC, et al. 2017. Impact of alternative splicing on the human proteome. *Cell Rep.* 20(5):1229–41
95. Lujan E, Zunder ER, Ng YH, Goronzy IN, Nolan GP, Wernig M. 2015. Early reprogramming regulators identified by prospective isolation and mass cytometry. *Nature* 521(7552):352–56
96. Zunder ER, Lujan E, Goltsev Y, Wernig M, Nolan GP. 2015. A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. *Cell Stem Cell* 16(3):323–37
97. Mallanna SK, Cayo MA, Twaroski K, Gundry RL, Duncan SA. 2016. Mapping the cell-surface N-glycoproteome of human hepatocytes reveals markers for selecting a homogeneous population of iPSC-derived hepatocytes. *Stem Cell Rep.* 7(3):543–56
98. Kilpinen H, Goncalves A, Leha A, Afzal V, Alasoo K, et al. 2017. Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* 546(7658):370–75
99. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, et al. 2018. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* 46(D1):D608–17
100. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, et al. 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472(7341):57–63
101. Sato Y, Kobayashi H, Higuchi T, Shimada Y, Ida H, Ohashi T. 2017. Metabolomic profiling of Pompe disease–induced pluripotent stem cell–derived cardiomyocytes reveals that oxidative stress is associated with cardiac and skeletal muscle pathology. *Stem Cells Transl. Med.* 6(1):31–39
102. Zhao X, Chen H, Xiao D, Yang H, Itzhaki I, et al. 2018. Comparison of non-human primate versus human induced pluripotent stem cell–derived cardiomyocytes for treatment of myocardial infarction. *Stem Cell Rep.* 10(2):422–35
103. Gilchrist KH, Lewis GF, Gay EA, Sellgren KL, Grego S. 2015. High-throughput cardiac safety evaluation and multi-parameter arrhythmia profiling of cardiomyocytes using microelectrode arrays. *Toxicol. Appl. Pharmacol.* 288(2):249–57
104. Hochbaum DR, Zhao Y, Farhi SL, Klapoetke N, Werley CA, et al. 2014. All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nat. Methods* 11(8):825–33
105. Rajamohan D, Kalra S, Duc Hoang M, George V, Staniforth A, et al. 2016. Automated electrophysiological and pharmacological evaluation of human pluripotent stem cell–derived cardiomyocytes. *Stem Cells Dev.* 25(6):439–52
106. Deshpande A, Yadav S, Dao DQ, Wu Z-Y, Hokanson KC, et al. 2017. Cellular phenotypes in human iPSC-derived neurons from a genetic model of autism spectrum disorder. *Cell Rep.* 21(10):2678–87
107. Hwang HS, Kryshtal DO, Feaster TK, Sánchez-Freire V, Zhang J, et al. 2015. Comparable calcium handling of human iPSC-derived cardiomyocytes generated by multiple laboratories. *J. Mol. Cell. Cardiol.* 85:79–88
108. Paşca SP, Portmann T, Voineagu I, Yazawa M, Shcheglovitov A, et al. 2011. Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome. *Nat. Med.* 17(12):1657–62
109. Lan F, Lee AS, Liang P, Sanchez-Freire V, Nguyen PK, et al. 2013. Abnormal calcium handling properties underlie familial hypertrophic cardiomyopathy pathology in patient-specific induced pluripotent stem cells. *Cell Stem Cell* 12(1):101–13

110. Wyles SP, Li X, Hrstka SC, Reyes S, Oommen S, et al. 2016. Modeling structural and functional deficiencies of *RBM20* familial dilated cardiomyopathy using human induced pluripotent stem cells. *Hum. Mol. Genet.* 25(2):254–65
111. Ribeiro AJS, Schwab O, Mandegar MA, Ang Y-S, Conklin BR, et al. 2017. Multi-imaging method to assay the contractile mechanical output of micropatterned human iPSC-derived cardiac myocytes. *Circ. Res.* 120(10):1572–83
112. Hayakawa T, Kunihiro T, Ando T, Kobayashi S, Matsui E, et al. 2014. Image-based evaluation of contraction–relaxation kinetics of human-induced pluripotent stem cell–derived cardiomyocytes: correlation and complementarity with extracellular electrophysiology. *J. Mol. Cell. Cardiol.* 77:178–91
113. Scott CW, Zhang X, Abi-Gerges N, Lamore SD, Abassi YA, Peters MF. 2014. An impedance-based cellular assay using human iPSC-derived cardiomyocytes to quantify modulators of cardiac contractility. *Toxicol. Sci.* 142(2):331–38
114. Sun N, Yazawa M, Liu J, Han L, Sanchez-Freire V, et al. 2012. Patient-specific induced pluripotent stem cells as a model for familial dilated cardiomyopathy. *Sci. Transl. Med.* 4(130):130ra47
115. Rodriguez ML, Graham BT, Pabon LM, Han SJ, Murry CE, Sniadecki NJ. 2014. Measuring the contractile forces of human induced pluripotent stem cell–derived cardiomyocytes with arrays of microposts. *J. Biomech. Eng.* 136(5):051005
116. Ribeiro MC, Tertoolen LG, Guadix JA, Bellin M, Kosmidis G, et al. 2015. Functional maturation of human pluripotent stem cell derived cardiomyocytes in vitro—correlation between contraction force and electrophysiology. *Biomaterials* 51:138–50
117. Streckfuss-Bömeke K, Tiburcy M, Fomin A, Luo X, Li W, et al. 2017. Severe DCM phenotype of patient harboring *RBM20* mutation s635A can be modeled by patient-specific induced pluripotent stem cell–derived cardiomyocytes. *J. Mol. Cell. Cardiol.* 113:9–21
118. Carcamo-Orive I, Huang NF, Quertermous T, Knowles JW. 2017. Induced pluripotent stem cell–derived endothelial cells in insulin resistance and metabolic syndrome. *Arterioscler. Thromb. Vasc. Biol.* 37(11):2038–42
119. Wang G, McCain ML, Yang L, He A, Pasqualini FS, et al. 2014. Modeling the mitochondrial cardiomyopathy of Barth syndrome with induced pluripotent stem cell and heart-on-chip technologies. *Nat. Med.* 20(6):616–23
120. Ebert AD, Kodo K, Liang P, Wu H, Huber BC, et al. 2014. Characterization of the molecular mechanisms underlying increased ischemic damage in the aldehyde dehydrogenase 2 genetic polymorphism using a human induced pluripotent stem cell model system. *Sci. Transl. Med.* 6(255):255ra130
121. Keenan AB, Jenkins SL, Jagodnik KM, Koplev S, He E, et al. 2018. The Library of Integrated Network-Based Cellular Signatures NIH Program: system-level cataloging of human cells response to perturbations. *Cell Syst.* 6(1):13–24
122. Boyle EA, Li YI, Pritchard JK. 2017. An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169(7):1177–86
123. Barabási A-L, Gulbahce N, Loscalzo J. 2011. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12(1):56–68
124. Li M, Belmonte JCI. 2017. Ground rules of the pluripotency gene regulatory network. *Nat. Rev. Genet.* 18(3):180–91
125. Kolodziejczyk AA, Kim JK, Tsang JCH, Ilicic T, Henriksson J, et al. 2015. Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell* 17(4):471–85
126. Rolland T, Taşan M, Charloteaux B, Pevzner SJ, Zhong Q, et al. 2014. A proteome-scale map of the human interactome network. *Cell* 159(5):1212–26
127. Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, et al. 2017. Architecture of the human interactome defines protein communities and disease networks. *Nature* 545(7655):505–9
128. Yan G, Vértes PE, Towlson EK, Chew YL, Walker DS, et al. 2017. Network control principles predict neuron function in the *Caenorhabditis elegans* connectome. *Nature* 550(7677):519–23
129. Albert FW, Kruglyak L. 2015. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* 16(4):197–212
130. GTEx Consort. 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348(6235):648–60

131. Li X, Kim Y, Tsang EK, Davis JR, Damani FN, et al. 2017. The impact of rare variation on gene expression across tissues. *Nature* 550(7675):239–43
132. Matsa E, Burridge PW, Yu K-H, Ahrens JH, Termglinchan V, et al. 2016. Transcriptome profiling of patient-specific human iPSC-cardiomyocytes predicts individual drug safety and efficacy responses in vitro. *Cell Stem Cell* 19(3):311–25
133. Burrows CK, Banovich NE, Pavlovic BJ, Patterson K, Gallego Romero I, et al. 2016. Genetic variation, not cell type of origin, underlies the majority of identifiable regulatory differences in iPSCs. *PLOS Genet*. 12(1):e1005793
134. Rouhani F, Kumasaka N, de Brito MC, Bradley A, Vallier L, Gaffney D. 2014. Genetic background drives transcriptional variation in human induced pluripotent stem cells. *PLOS Genet*. 10(6):e1004432
135. Cahan P, Daley GQ. 2013. Origins and implications of pluripotent stem cell variability and heterogeneity. *Nat. Rev. Mol. Cell Biol.* 14(6):357–68
136. Panopoulos AD, Smith EN, Arias AD, Shepard PJ, Hishida Y, et al. 2017. Aberrant DNA methylation in human iPSCs associates with MYC-binding motifs in a clone-specific manner independent of genetics. *Cell Stem Cell* 20(4):505–17.e6
137. Schwartzentruber J, Foskolou S, Kilpinen H, Rodrigues J, Alasoo K, et al. 2018. Molecular and functional variation in iPSC-derived sensory neurons. *Nat. Genet.* 50(1):54–61
138. Churko JM, Lee J, Ameen M, Gu M, Venkatasubramanian M, et al. 2017. Transcriptomic and epigenomic differences in human induced pluripotent stem cells generated from six reprogramming methods. *Nat. Biomed. Eng.* 1:826–37
139. Chen G, Gulbranson DR, Hou Z, Bolin JM, Ruotti V, et al. 2011. Chemically defined conditions for human iPSC derivation and culture. *Nat. Methods* 8(5):424–29
140. Churko JM, Burridge PW, Wu JC. 2013. Generation of human iPSCs from human peripheral blood mononuclear cells using non-integrative Sendai virus in chemically defined conditions. *Methods Mol. Biol.* 1036:81–88
141. D'Antonio M, Woodruff G, Nathanson JL, D'Antonio-Chronowska A, Arias A, et al. 2017. High-throughput and cost-effective characterization of induced pluripotent stem cells. *Stem Cell Rep*. 8(4):1101–11
142. Tsankov AM, Akopian V, Pop R, Chetty S, Gifford CA, et al. 2015. A qPCR scorecard quantifies the differentiation potential of human pluripotent stem cells. *Nat. Biotechnol.* 33(11):1182–92
143. Danecek P, McCarthy SA, HipSci Consort., Durbin R. 2016. A method for checking genomic integrity in cultured cell lines from SNP genotyping data. *PLOS ONE* 11(5):e0155014
144. Paull D, Sevilla A, Zhou H, Hahn AK, Kim H, et al. 2015. Automated, high-throughput derivation, characterization and differentiation of induced pluripotent stem cells. *Nat. Methods* 12(9):885–92
145. Soares FAC, Sheldon M, Rao M, Mummery C, Vallier L. 2014. International coordination of large-scale human induced pluripotent stem cell initiatives: Wellcome Trust and ISSCR workshops white paper. *Stem Cell Rep*. 3(6):931–39
146. Warren CR, Jaquish CE, Cowan CA. 2017. The NextGen Genetic Association Studies Consortium: a foray into in vitro population genetics. *Cell Stem Cell* 20(4):431–33
147. Padovan-Merhar O, Nair GP, Biaesch AG, Mayer A, Scarfone S, et al. 2015. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol. Cell* 58(2):339–52
148. Carcamo-Orive I, Hoffman GE, Cundiff P, Beckmann ND, D'Souza SL, et al. 2017. Analysis of transcriptional variability in a large human iPSC library reveals genetic and non-genetic determinants of heterogeneity. *Cell Stem Cell* 20(4):518–32.e9
149. Marx V. 2015. Stem cells: disease models that show and tell. *Nat. Methods* 12(2):111–14
150. Pashos EE, Park Y, Wang X, Raghavan A, Yang W, et al. 2017. Large, diverse population cohorts of hiPSCS and derived hepatocyte-like cells reveal functional genetic variation at blood lipid–associated loci. *Cell Stem Cell* 20(4):558–70.e10
151. Sanchez-Freire V, Lee AS, Hu S, Abilez OJ, Liang P, et al. 2014. Effect of human donor cell source on differentiation and function of cardiac induced pluripotent stem cells. *J. Am. Coll. Cardiol.* 64(5):436–48

152. DeBoever C, Li H, Jakubosky D, Benaglio P, Reyna J, et al. 2017. Large-scale profiling reveals the influence of genetic variation on gene expression in human induced pluripotent stem cells. *Cell Stem Cell* 20(4):533–46.e7
153. Panopoulos AD, D'Antonio M, Benaglio P, Williams R, Hashem SI, et al. 2017. iPSCORE: a resource of 222 iPSC lines enabling functional characterization of genetic variation across a variety of cell types. *Stem Cell Rep*. 8(4):1086–100
154. Barabasi AL, Albert R. 1999. Emergence of scaling in random networks. *Science* 286(5439):509–12
155. Clune J, Mouret J-B, Lipson H. 2013. The evolutionary origins of modularity. *Proc. R. Soc. B* 280(1755):20122863
156. Albert R, Jeong H, Barabasi AL. 2000. Error and attack tolerance of complex networks. *Nature* 406(6794):378–82
157. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL. 2000. The large-scale organization of metabolic networks. *Nature* 407(6804):651–54
158. Ideker T, Krogan NJ. 2012. Differential network biology. *Mol. Syst. Biol.* 8:565
159. Hu JX, Thomas CE, Brunak S. 2016. Network biology concepts in complex disease comorbidities. *Nat. Rev. Genet.* 17(10):615–29
160. Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 9:559
161. Rau CD, Wisniewski N, Orozco LD, Bennett B, Weiss J, Lusis AJ. 2013. Maximal information component analysis: a novel non-linear network analysis method. *Front. Genet.* 4:28
162. Cahan P, Li H, Morris SA, Lummertz da Rocha E, Daley GQ, Collins JJ. 2014. CellNet: network biology applied to stem cell engineering. *Cell* 158(4):903–15
163. Fiers MWEJ, Minnoye L, Aibar S, Bravo González-Blas C, Kalender Atak Z, Aerts S. 2018. Mapping gene regulatory networks from single-cell omics data. *Brief. Funct. Genom.* 17(4):246–54
164. Chan TE, Stumpf MPH, Babtie AC. 2017. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst.* 5(3):251–67.e3
165. Dixit A, Parnas O, Li B, Chen J, Fulco CP, et al. 2016. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167(7):1853–66.e17
166. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, et al. 2016. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* 167(7):1883–96.e15
167. Uosaki H, Cahan P, Lee DI, Wang S, Miyamoto M, et al. 2015. Transcriptional landscape of cardiomyocyte maturation. *Cell Rep*. 13(8):1705–16
168. Lusis AJ, Seldin MM, Allayee H, Bennett BJ, Civelek M, et al. 2016. The hybrid mouse diversity panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. *J. Lipid Res.* 57(6):925–42
169. Flint J, Eskin E. 2012. Genome-wide association studies in mice. *Nat. Rev. Genet.* 13(11):807–17
170. van den Berg CW, Okawa S, Chuva de Sousa Lopes SM, van Iperen L, Passier R, et al. 2015. Transcriptome of human foetal heart compared with cardiomyocytes from pluripotent stem cells. *Development* 142(18):3231–38
171. Zhang ZN, Freitas BC, Qian H, Lux J, Acab A, et al. 2016. Layered hydrogels accelerate iPSC-derived neuronal maturation and reveal migration defects caused by MeCP2 dysfunction. *PNAS* 113(12):3185–90
172. Parikh SS, Blackwell DJ, Gomez-Hurtado N, Frisk M, Wang L, et al. 2017. Thyroid and glucocorticoid hormones promote functional T-tubule development in human-induced pluripotent stem cell–derived cardiomyocytes. *Circ. Res.* 121(12):1323–30
173. Ribeiro AJS, Ang YS, Fu JD, Rivas RN, Mohamed TMA, et al. 2015. Contractility of single cardiomyocytes differentiated from pluripotent stem cells depends on physiological shape and substrate stiffness. *PNAS* 112(41):12705–10
174. Tu C, Chao BS, Wu JC. 2018. Strategies for improving the maturity of human induced pluripotent stem cell–derived cardiomyocytes. *Circ. Res.* 123(5):512–14
175. Horvath P, Aulner N, Bickle M, Davies AM, Nery ED, et al. 2016. Screening out irrelevant cell-based models of disease. *Nat. Rev. Drug Discov.* 15(11):751–69
176. Tzatzalos E, Abilez OJ, Shukla P, Wu JC. 2016. Engineered heart tissues and induced pluripotent stem cells: macro- and microstructures for disease modeling, drug screening, and translational studies. *Adv. Drug Deliv. Rev.* 96:234–44

177. Abilez OJ, Tzatzalos E, Yang H, Zhao M-T, Jung G, et al. 2018. Passive stretch induces structural and functional maturation of engineered heart muscle as predicted by computational modeling. *Stem Cells* 36(2):265–77
178. Giacomelli E, Bellin M, Sala L, van Meer BJ, Tertoolen LGJ, et al. 2017. Three-dimensional cardiac microtissues composed of cardiomyocytes and endothelial cells co-differentiated from human pluripotent stem cells. *Development* 144(6):1008–17
179. Kuijlaars J, Oyelami T, Diels A, Rohrbacher J, Versweyveld S, et al. 2016. Sustained synchronized neuronal network activity in a human astrocyte co-culture system. *Sci. Rep.* 6:36529
180. Dutta D, Heo I, Clevers H. 2017. Disease modeling in stem cell–derived 3D organoid systems. *Trends Mol. Med.* 23(5):393–410
181. Liu C, Oikonomopoulos A, Sayed N, Wu JC. 2018. Modeling human diseases with induced pluripotent stem cells: from 2D to 3D and beyond. *Development* 145(5):dev156166
182. Lo Sardo V, Ferguson W, Erikson GA, Topol EJ, Baldwin KK, Torkamani A. 2017. Influence of donor age on induced pluripotent stem cells. *Nat. Biotechnol.* 35(1):69–74
183. Sharma A, Diecke S, Zhang WY, Lan F, He C, et al. 2013. The role of SIRT6 protein in aging and reprogramming of human induced pluripotent stem cells. *J. Biol. Chem.* 288(25):18439–47