nature genetics

Perspective

https://doi.org/10.1038/s41588-024-01731-9

Integrating population genetics, stem cell biology and cellular genomics to study complex human diseases

Received: 24 January 2023

Accepted: 20 March 2024

Published online: 13 May 2024

Check for updates

Nona Farbehi[1,2,3,8], Drew R. Neavin[1,8], Anna S. E. Cuomo[1,4], Lorenz Studer[3,5], Daniel G. MacArthur[4,6] & Joseph E. Powell[1,3,7]

Human pluripotent stem (hPS) cells can, in theory, be differentiated into any cell type, making them a powerful in vitro model for human biology. Recent technological advances have facilitated large-scale hPS cell studies that allow investigation of the genetic regulation of molecular phenotypes and their contribution to high-order phenotypes such as human disease. Integrating hPS cells with single-cell sequencing makes identifying context-dependent genetic effects during cell development or upon experimental manipulation possible. Here we discuss how the intersection of stem cell biology, population genetics and cellular genomics can help resolve the functional consequences of human genetic variation. We examine the critical challenges of integrating these fields and approaches to scaling them cost-effectively and practically. We highlight two areas of human biology that can particularly benefit from population-scale hPS cell studies, elucidating mechanisms underlying complex disease risk loci and evaluating relationships between common genetic variation and pharmacotherapeutic phenotypes.

Human pluripotent stem (hPS) cells have substantially impacted biological research, from basic cell biology to translational biomedicine. The inherent properties of hPS cells—self-renewability and the ability to theoretically differentiate into any cell type in the human body—make hPS cells an excellent in vitro model for studying human phenotypes and diseases. This article uses the term hPS cells to refer to both embryonic stem cells[1] and induced pluripotent stem (iPS) cells[2]. iPS cells are generated by reprogramming primary cells through forced expression of the pluripotency transcription factors (OCT3, OCT4, MYC, KLF4 and SOX2)[2,3], resulting in near-identical signatures of gene expression and epigenetic states to embryonic stem cells[2,4,5]. Thus, each hPS cell contains the genetic profile of the donor, allowing investigation of genetic drivers of human phenotypes (for example, disease susceptibility or human development).

hPS cells have primarily been used to study rare diseases[6,7], most of which have been associated with a few genetic loci, typically located in protein-coding genome regions. hPS cells have been used less frequently to study complex diseases and conditions whose risk is determined by many genetic variants and environmental factors (for example, Parkinson's disease, glaucoma and cardiovascular disease, among others)[8,9]. Although genome-wide association studies have identified many loci associated with complex diseases[10], >90% of these

[1]Garvan Weizmann Center for Cellular Genomics, Garvan Institute of Medical Research, Sydney, New South Wales, Australia. [2]Graduate School of Biomedical Engineering, University of New South Wales, Sydney, New South Wales, Australia. [3]Aligning Science Across Parkinson's Collaborative Research Network, Chevy Chase, MD, USA. [4]Centre for Population Genomics, Garvan Institute of Medical Research, University of New South Wales, Sydney, New South Wales, Australia. [5]The Center for Stem Cell Biology and Developmental Biology Program, Sloan-Kettering Institute for Cancer Research, New York, NY, USA. [6]Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, Victoria, Australia. [7]UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney, New South Wales, Australia. [8]These authors contributed equally: Nona Farbehi, Drew R. Neavin. e-mail: j.powell@garvan.org.au

https://doi.org/10.1038/s41588-024-01731-9

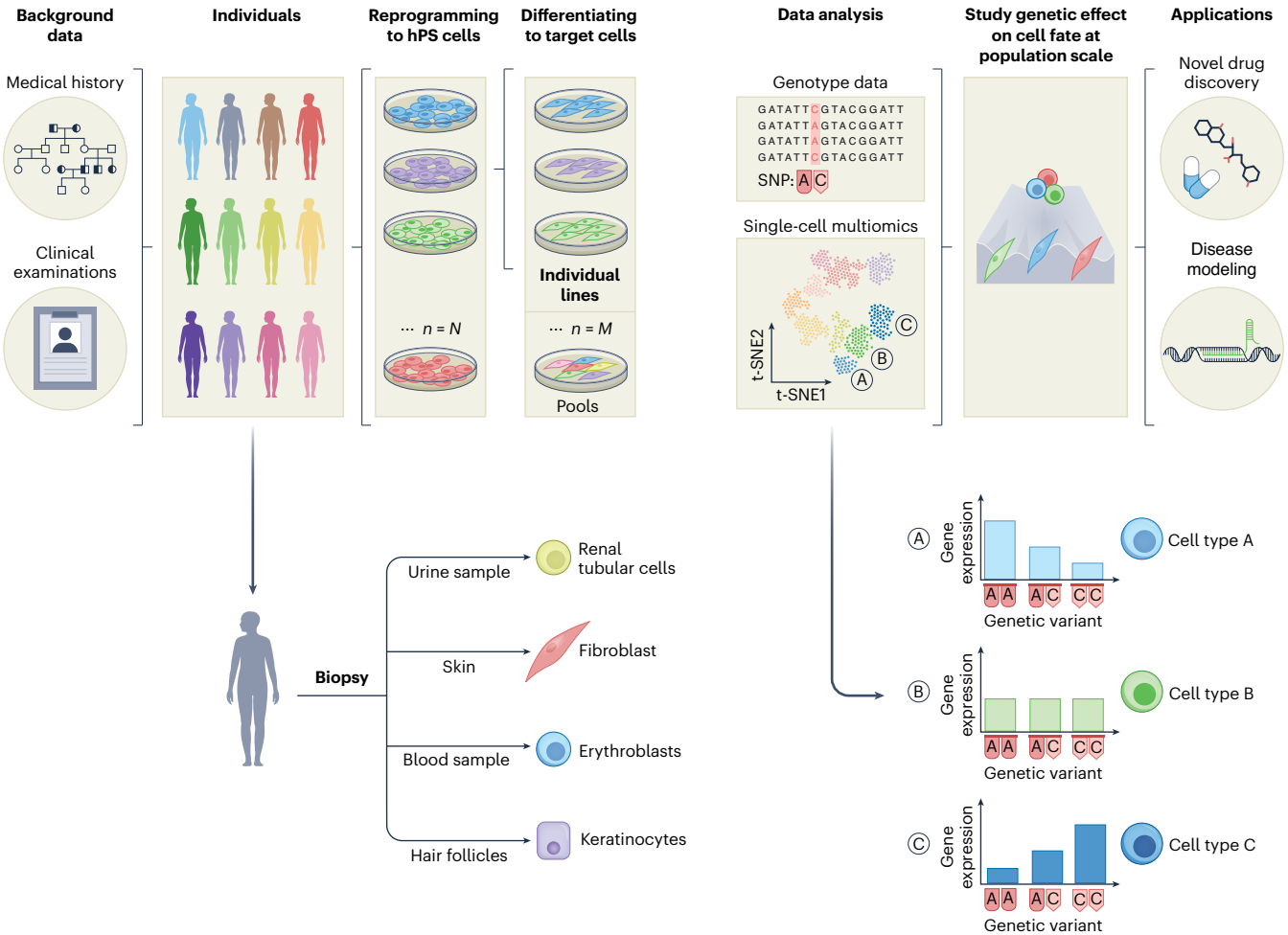


Fig. 1 | Schematic overview of population genetics studies of hPS cells to study complex human diseases. hPS cells can be generated from patient-derived somatic cells such as peripheral blood mononuclear cells and fibroblasts. hPS cell lines (1–*n*, or *N* total) can be cultured individually or pooled in a single dish (villages). An experiment can consist of *M* villages, each comprising on *N* lines. hPS cell pools combined with demultiplexing single-cell computational methods allow for cost-effective scaling of stem cell studies. By coupling single-cell genomics and patient genotype data with demultiplexing single-cell computational methods, the genetic effects of therapies on cell phenotypes can be studied, resulting in more accurate predictions of individual patient responses to different drugs. *t*-SNE, *t*-distributed stochastic neighbor embedding. Figure created with BioRender.com.

are located in intergenic regions and probably contribute to disease via changes in genome regulation[11,12]. To complicate matters further, there is growing evidence that many of these loci exert their effects in a cell-type- or context-dependent manner (for example, varying physiological or environmental stimuli)[13–16].

Although still in its early stages, the application of hPS cells to study gene regulation in complex diseases has already proven to be a powerful approach[17]. For example, hPS cells have already been used to discover the effect of noncoding regulatory mutations[18], predict drug toxicity[19] and develop cell replacement therapies[20]. However, most of these studies have been conducted on lines obtained from only a few donors (Supplementary Table 1).

Several challenges, such as experimental scalability, achieving cell maturity, accurate modeling of cell-type-specific contexts and in vivo cell physiology, have prevented the widespread use of hPS cells for population genomics studies[21]. Population genomics studies require large sample sizes to accurately reflect genetic variation and provide the statistical power to link alleles to molecular and higher-order phenotypes such as disease. Culturing and differentiating hundreds of hPS cell lines independently is time-consuming, labor-intensive and expensive. Another challenge is the variance in cell composition between cell lines (inter-line variability) and between differentiations for the same line (intra-line variability), which might impact the genomic profiles, thereby adding 'noise' to cellular phenotypes and reducing the ability to identify genetic associations[22,23]. However, the availability of large stem cell banks (for example, HipSci[24] and iPSCORE[25]), robotic handling of cell cultures and single-cell technologies[26] has helped curtail some of these limitations and provide a platform for innovative population genomics studies using hPS cells[14–16].

In this Perspective, we discuss how stem cell biology, population genetics and cellular genomics can be integrated to help resolve the functional consequences of human genetic variation. We examine some of the critical challenges that arise from the intersection of these fields and how these may be overcome to scale experiments efficiently. We highlight how pooling multiple hPS cell lines in a single dish, coupled with single-cell genomics and computational methods, can dramatically scale stem cell experiments in a cost-efficient manner, making studies of thousands of independent lines achievable (Fig. 1). We explore technological advances and tools that can facilitate drug development and pharmacogenomics using population-scale hPS cell research. Finally, we address challenges and opportunities in creating the required cell systems, data generation and analysis workflows. As this research advances, we anticipate substantial progress in resolving how genetic variants contribute to phenotypes such as disease and drug response, leading to improved medical outcomes.

Human disease modeling with stem cells, population genetics and cellular genomics

While limitations remain in using hPS cells to model in vivo cell systems, these cells provide attractive advantages. They can be used to model dynamic processes that are challenging to study at scale with in vivo systems, such as cell development, experimental manipulation to mimic cellular microenvironmental stressors and exposure to therapeutic agents[27]. They also provide cellular material for tissue systems that are challenging to access, such as neural cells[28], cardiomyocytes[29,30] or cell developmental stages[27].

As hPS cells contain a copy of the donor's DNA, experimental methods that generate genomic data across large numbers of lines provide an approach to test the effects of genetic loci under different cellular contexts. One approach has been to map molecular quantitative trait loci (QTLs) in hPS cells by intersecting standard techniques from population genetics and stem cell biology[24,31–33]. For example, Kilpinen et al. used genome-wide profiling in hPS cells from 301 donors to show that significant variation in different cell phenotypes, including differentiation capacity and cellular morphology, arises due to genetic variation[24].

Dynamic QTLs can identify context-dependent genetic effects

Statistical models of dynamic expression QTLs (eQTLs) are emerging as methods to investigate how genetic effects on cellular phenotypes vary across cell states or in different cellular contexts (such as drug exposure)[34]. They are designed to identify the interaction between genotypes (G) and cell contexts (C), which we call G × C, where the context could be a cell state or exposure to a stimulant or drug (Box 1). Compared with traditional QTL methods that test for 'static' genetic effects within a particular context, dynamic models can increase resolution and identify hidden effects, for example, a genetic effect present only when cells are exposed to a drug. They can be used to determine how a genetic effect on a molecular phenotype can change along a continuous cell state spectrum, for example, during differentiation[30]. Dynamic QTLs, particularly those with nonlinear effects, may be transient (that is, seen in cells only at specific cell states) and, therefore, will probably be missed if a study only obtains genomic data from 'mature' cell types. Sampling cells at multiple points during hPS cell differentiation and using dynamic QTL models is a valuable strategy to identify how genetic effects on genomic phenotypes vary across cell developmental lineages[27,30]. When combined with single-cell readouts, dynamic QTLs can be tested across a continuous trajectory, taking advantage of the fact that each cell will be in a slightly different state (for example, cell maturity) even when cells are collected simultaneously. To achieve this, methods are needed to accurately place a cell within a cell state landscape (or manifold space) on the basis of a defined biological property (see below). Put another way, we need to phenotype a cell's state accurately to improve the detection of dynamic genetic effects[16,35]. While it has only recently been possible to identify dynamic QTLs, results suggest that they are often located in regions without known regulatory annotations. Therefore, these loci may exert previously unappreciated regulatory effects, which could be characterized by further functional validation at relevant intermediate time points during a differentiation trajectory[12,16,27].

This example illustrates how combining hPS cell models and population genetic methods can reveal biological effects that traditional approaches would miss. Moreover, experiments can be designed to test for dynamic effects when cells are exposed to environmental challenges or drugs[28], making it possible to identify specific cell states/types whose genomic profiles vary in response to exposure and, importantly, how this varies between different individuals.

Advantages of single-cell over bulk genomic assays

Applying single-cell assays over 'bulk' approaches can increase the resolution with which the impact of human genetic variation on phenotypes can be detected. Early QTL studies in hPS cells used bulk methods, which measure molecular phenotypes averaged across many cells, to identify genetic variants associated with gene expression, methylation and chromatin accessibility[36,37]. However, advances in single-cell sequencing now allow the study and characterization of individual cell readouts across heterogeneous cells, such as those prevalent during lineage differentiation and in three-dimensional organoid models[38].

While hPS cell QTL studies using bulk sequencing methods can still have some value, they are limited because they cannot directly resolve how genomic effects vary in different cells. Single-cell methods, by definition, assay each cell separately and thus provide advantages over bulk methods for heterogeneous cultures, such as organoids, co-cultures or even targeted differentiations that contain different cell types or states[16]. Single-cell sequencing also prevents or minimizes the necessity of preselecting desired cells, for instance, by fluorescence-activated cell sorting or magnetic activated cell sorting. By avoiding sorting, cellular stresses are minimized[39], resulting in more accurate quantification of the biological state of the cells. Additionally, single-cell sequencing, predominantly single-cell RNA sequencing (scRNA-seq), allows the identification and characterization of rare or previously unidentified cell subtypes. Many single-cell computational methods exist for classifying cell types, such as placing cells in a defined manifold (distribution and organization of cells based on their gene expression profiles or other molecular features) and inferring cell states[40]. Therefore, single-cell methods are uniquely suited to identify context-dependent effects, where the relationship between genotype and environment can be investigated for individual cells[14–16].

Recently, the single-cell eQTLGen consortium[41] was founded to help identify the cellular contexts in which disease-causing genetic variants affect gene expression in individual immune cell types[42]; it has now been extended to cover many more traits and diseases. Applying single-cell methods to large-scale hPS cell studies would provide access to similar information across various cell states and types. Currently, only high-throughput scRNA-seq, single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq) and simultaneous scATAC-seq and scRNA-seq are feasible for single-cell QTL studies, including in hPS cells[13,15,41,43,44]. Both scATAC-seq and scRNA-seq data can be used in QTL models to identify how genetic variation influences these molecular phenotypes in a cell-type-dependent manner and subsequently is linked to disease risk loci to prioritize variants with regulatory effects for experimental validation[45]. As new single-cell omics methods become available[46], expanding these tests to other types of regulatory and genomic information will be possible.

Integrating genetics and stem cells for pharmacogenomics and drug development

Genetic variation significantly contributes to patient response to therapeutic agents[47,48], and therapies with genetic evidence are twice as likely to advance to market successfully[49]. However, many compounds fail clinical trials[50], partly due to differences between human and animal model physiology and the use of limited primary or immortalized in vitro cell models.

hPS cell-derived cells, including organoid systems, can be used during preclinical drug development to distinguish promising compounds from those that would fail during clinical trials[51,52] and to detect cytotoxic effects in appropriate cell types, consistent with clinical observations[53,54]. An advantage of using hPS cell lines for preclinical testing is conducting a 'clinical trial in a dish' and determining how patient cells respond to a drug on the basis of their genetic profiles[55]. Genetic variants can impact pharmacokinetics (the body's effect on the drug) and pharmacodynamics (the drug's effect on the body), often collectively referred to as pharmacogenomics, resulting in variations in therapeutic toxicity and response[47,48]. Regrettably, pharmacogenomic effects are usually not considered during clinical trials, resulting in patients who experience preventable toxicity or poor response.

https://doi.org/10.1038/s41588-024-01731-9

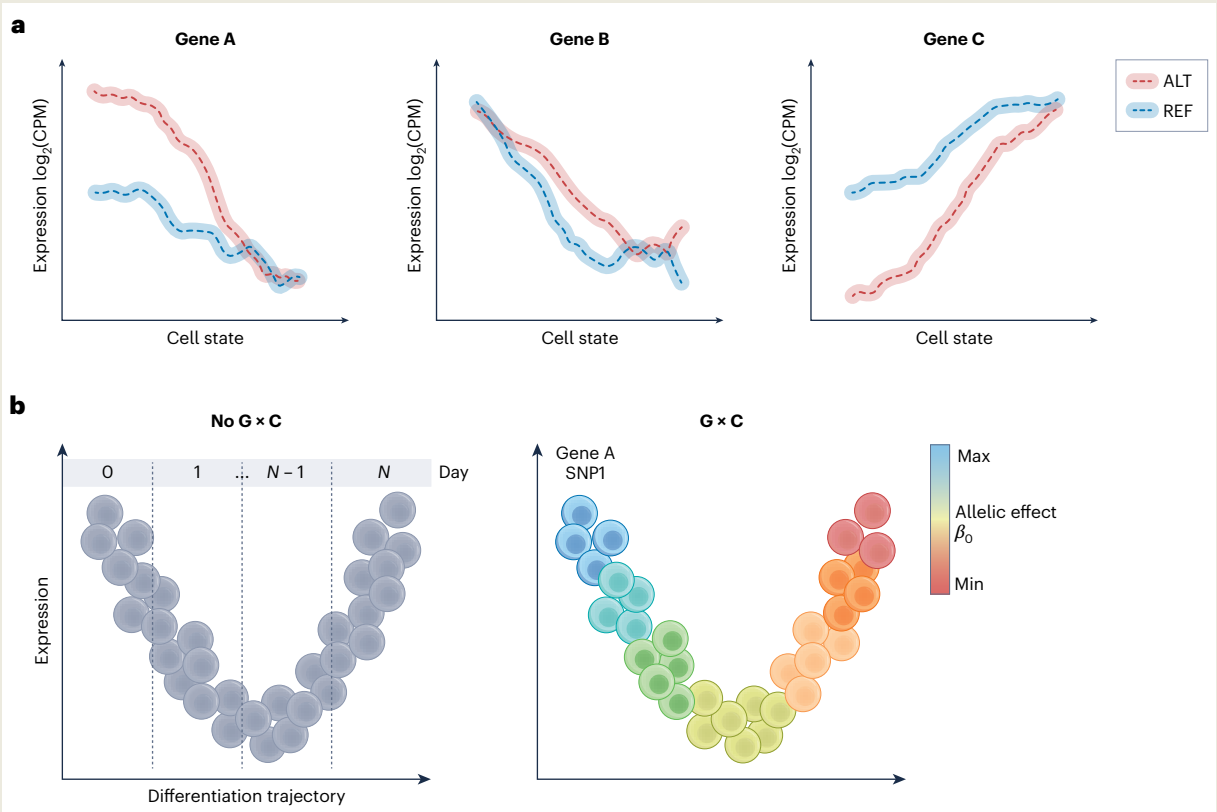
BOX 1

Representative examples of eQTLs with G × C interactions

a, Dynamic eQTL describes variants whose allelic effect on gene expression varies as cellular states change (adapted with permission from ref. 16, Springer Nature Limited). **b**, Tools for dynamic eQTL mapping allow estimation of cell-specific effect sizes, thus identifying populations of cells where genetic regulation is active, sometimes revealing opposing effects across cell states (for example, during differentiation). Figure created with BioRender.com.

Panel **a** illustrates possible relationships between SNP effects and gene expression levels across cell states. For example, the overall expression levels of gene A decrease as the cell state transitions (left to right), yet the eQTL effect, identified by the difference in expression levels of alleles, decreases. The allelic effect would be detected only at early cell states in this scenario. This is the case for the expression levels of *THUMPD1* as stem cells transition from a pluripotent to an endoderm state[16]. At the pluripotent state, *THUMPD1* expression levels are, on average, high, with a significant allelic effect observed, marked by the top expression SNP (eSNP) rs76148084[G/A]. As cells transition toward the endoderm committed state, the overall expression levels decrease, as does the allelic difference[16]. Gene C shows an opposite behavior, as allelic effects decrease across the cell state transition, while its overall expression levels increase. An alternative dynamic pattern is shown by gene B (panel **a**). Here the allelic effect on gene expression is observed only at an intermediate point in the cell state landscape. A real-world example of such a behavior is the expression of *VAT1L* during endoderm differentiation, with differences in allelic effects (marked by eSNP rs7191422[T/A]) on *VAT1L* expression observed during the intermediate mesendoderm developmental stage. This suggests that carrying the alternative T allele accelerates the decrease in expression level over developmental time[16]. Other examples of dynamic effects have been reported in the differentiation of iPS cells to cardiomyocytes, with eQTLs that vary both linearly (for example, the effect of rs11124033 on *FHL2*) and nonlinearly (for example, the effect of rs28818910 on *C15orf39*) over the cell state transition[35].

Tools for dynamic QTL mapping allow estimation of the allelic effects across cells, which can help identify the specific cell populations in which such genetic regulation is active and reveal biologically important cell states (panel **b**). An example is the dynamic allelic effects of rs113520162[A/G] on *IER3* expression as cells differentiate from iPS cells toward definitive endoderm[34]. When cells are pluripotent, each copy of the G allele is associated with increased *IER3* expression. These allelic effects disappear as cells exit pluripotency, but, as cells differentiate into the endoderm lineage, the allelic effects are reversed and the G allele is associated with a decrease in *IER3* expression. CPM, counts per million; ALT, alternative allele; REF, reference allele.



Assessing drug responses in hPS cell models could identify individual variants that can be used to inform clinical trial design. Specifically, such models could aid in determining appropriate therapeutic ranges and biomarkers to stratify patients who may experience an adverse effect or identify those unlikely to respond to a given drug.

Challenges of scaling hPS cells for population genomics studies

While hPS cells are potent tools for population genomic studies, significant challenges, such as assessing cell type maturity, accurate cell annotation, diversity and scalability, need to be addressed.

Cell maturity and accurate modeling of in vivo cell types

It has been well documented that cells differentiated using standard hPS cell protocols can have fetal cell-like characteristics and do not necessarily reach the equivalent of adult maturity. Although this is beneficial in some settings (for example, as a model of early cell development), it might limit areas of research that require more mature phenotypes (for example, aging and neurodegeneration). To address this, methods are being developed to enhance the cellular maturity of hPS cell-derived cell types, including organoid systems, biophysical, metabolic or epigenetic stimulation[56], and the addition of hormones[57] and small molecules[58].

The cell type of interest is also an important consideration. Despite progress in the development of differentiation protocols, consistently and accurately achieving cell type specificity remains an area for improvement. Currently, there is a focus on improving differentiation to limit the number of off-target cells while achieving a close model of a desired in vivo cell type. Indeed, numerous studies have evaluated desired differentiation outcomes and the genetic and physiological relevance to primary cells[59,60]. For example, comparing hPS cell-derived pancreatic β cells with in vivo data identified deficiencies in maturation and off-target cell types in the stem cell model[61]. These effects remained when these cells were transplanted into mice, although maturation continued after transplantation[61]. One strategy to reduce the number of off-target cells is to compare cellular genomic data from differentiations to inform on the most appropriate combination of medium, coatings and growth factors to enrich mature cell populations[62].

A final consideration is the removal of cell microenvironments. Three-dimensional organoids and organ-on-a-chip[21] help address this by modeling cells with closer physiological relevance to in vivo micro- and macroenvironments. These in vitro cultures also enable the perturbation of genomic effects in human cells, which is not achievable in vivo.

Accurate cell type classification and state quantification

An advantage of using single-cell hPS cells in genomic research is the ability to identify the effect of genetic variation in different cell types or states within the culture. To do so, cells must first be classified (that is, into cell types) or their state must be determined (that is, position on a differentiation trajectory). Conventional methods for cell type annotation use marker genes, often originally identified from proteomic assays. However, innovative computational methods have been developed to resolve the heterogeneity below limited-plex protein markers. One solution is to use in vivo 'reference' data to train a model and predict the cell types in hPS cell data, thereby providing a probability score for each cell's classification against a reference cell type. This requires well-annotated, 'gold-standard' in vivo data, which might not be feasible for some tissues. Cell classification probability scores, which can be considered a 'similarity score' to in vivo reference cells, can be used in screening or perturbation experiments to evaluate any changes in cell identity following treatment[63].

A related challenge in cell annotation is the variability in the differentiation potential of hPS cell lines and off-target cells[64], which can be challenging to classify[16]. While there are multiple reasons for the occurrence of off-target cells[65], we anticipate that a combination of improved differentiation protocols and better cell annotation will help reduce their adverse impact. Careful annotation of single-cell data can identify off-target cell types on the basis of a lack of alignment to in vivo references. If needed, these can be removed from subsequent data analysis. In theory, analysis of off-target cells can improve differentiation protocols by identifying differences in the genomic profiles following treatment with differentiation factors.

Diversity of hPS cells

Broadening the representation of age, gender and genetic diversity in hPS cell studies will expedite biological discoveries and promote a more equitable translation of scientific benefits. Currently, most hPS cell lines available are from donors of northern European descent. Thus, there is a need and an opportunity to diversify the ancestral genetic backgrounds of hPS cell lines. This is important because genetic variants occur at different frequencies in different ancestral populations, so a variant that is an important contributor to a specific phenotype (that is, disease or drug response) for individuals of northern European ancestry may not be informative for other ancestral groups or vice versa. For example, a recent study of iPS cell-derived megakaryocytes generated from 194 diverse individuals based on age, sex and self-reported ancestry identified differentially expressed genes and proteins between different sex and genetic ancestral groups but not for different age groups[66]. Greater inclusion of diverse hPS cell lines is crucial scientifically, morally and ethically. This will require concerted efforts from the stem cell community to engage with under-represented groups and communities to ensure ethical and equitable representation in stem cell research.

Practical challenges with scaling hPS cell experiments

A final challenge is the necessity of performing genome-wide hPS cell studies with adequate sample sizes to detect genetic effects with statistical significance. Population genomic approaches require large sample sizes, and most study designs comprise data from unrelated individuals. Fortunately, several hPS cell repositories already contain hundreds of lines, making large-scale genomic studies in hPS cells a possibility[24,25,67,68]. Indeed, QTL studies with as few as 50 donors can detect some of the large allelic effects on molecular phenotypes from common genetic variation[69]. However, the power to detect smaller genetic effects scales with sample size. Scaling hPS cell studies has been challenging owing to the time and cost associated with hPS cell culture maintenance and differentiation, including labor and expensive reagents.

Currently, most hPS cell differentiation methods maintain each line in a separate culture dish, thereby exposing them to unique microenvironments; this not only introduces batch effects but also adds to the costs, time and labor when further lines are included in an experiment. One solution would be to use pooled hPS cell systems (sometimes called villages), with multiple hPS cell lines cultured and differentiated within a single dish. These approaches allow a larger number of lines to be included in an experiment without significant increases in time or cost and can reduce technical variability such as batch effects[14,16,70–72]. However, several technical challenges must be solved when implementing pooled hPS cell culture systems. For example, the maximum number of cell lines co-cultured within a single dish has yet to be fully assessed. While there is, theoretically, no limit to the number of lines within an hPS cell pool, practical limits arise from differences in growth rates, which affect the lines that can be cultured together[71,73].

Experimental design for scaling of hPS cell models

The overarching aim of pooled hPS cell culture designs is to use their genetic profiles to identify the donor cells within the co-culture system (Fig. 2)[14,16,38,70–74]. To date, two methods to accomplish this have been developed. The first (Fig. 2b) uses single-cell approaches to assay each cell within the pool separately and then matches the genetic profiles with each hPS cell line (so-called demultiplexing)[14,16,71,73–75]. This method is ideal when experimental research questions require single-cell profiles of each hPS cell line (such as in single-cell eQTL mapping). However, it is expensive if single-cell data are not required, for example, if the aim is to link genetic variation to cell physiology phenotypes. An alternative approach is the use of 'low-pass' genome sequencing to estimate the proportion of each donor line within the pooled hPS cell culture on the basis of genetic variant allele frequencies (Fig. 2c)[70,73]. This approach helps monitor pooled hPS cell cultures over time to ensure that the proportion of each hPS cell line remains constant.

hPS cells with similar growth rates can be pooled with the relative proportions of the hPS cell lines within the selected group observed over time (Fig. 2b), which is important as hPS cell lines with higher

https://doi.org/10.1038/s41588-024-01731-9

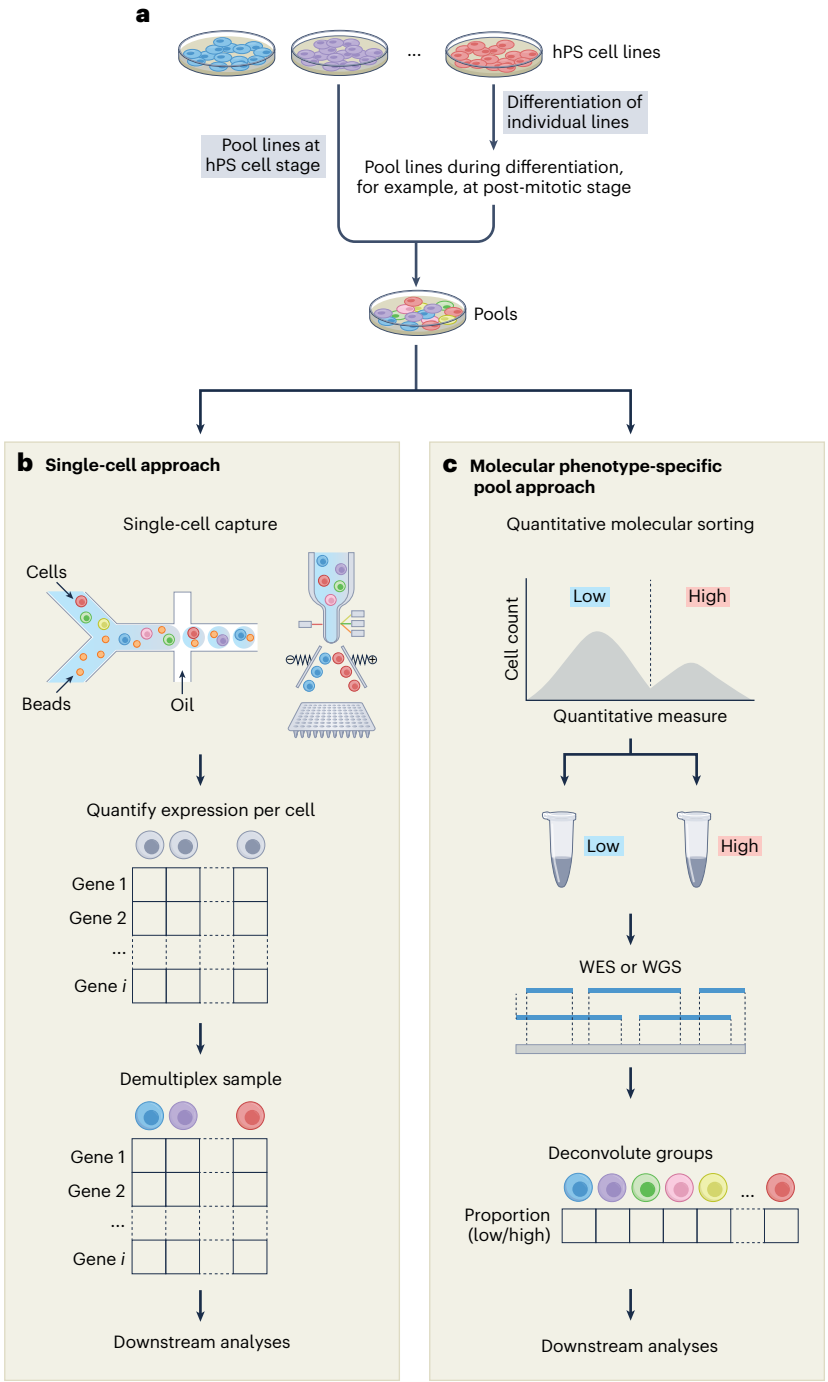


Fig. 2 | Experimental design of hPS cell line pools. a–**c**, hPS cell line pools can be generated by combining cells from multiple hPS cell lines (**a**). These pools can then be assessed with either a single-cell approach (**b**) or sorted by a molecular phenotype and combined with low-pass whole-genome sequencing (**c**). In the single-cell approach (**b**), each gene in each cell is sequenced (1) and quantified (2), followed by cell demultiplexing using genotype data (3). In the molecular sorting approach (**c**), cultured hPS cell line pools are sorted on the basis of a quantitative molecular phenotype (that is, protein level; 1). The sorted cell groups are then assessed with whole-genome or whole-exome sequencing (2) to determine the proportion of each hPS cell line in each group (3). WES, whole-exome sequencing; WGS, whole-genome sequencing. Figure created with BioRender.com.

growth rates can outcompete other lines and alter the cellular composition of lines in a pool[70,71]. Strategies to manage this involve estimating growth rates at the iPS cell stage and pooling cells with similar growth curves. However, this will not counteract lineage-dependent growth rates (Box 1), which have yet to be fully explored. However, if there are substantial differences, strategies to pool hPS cell lines on the basis of lineage-specific growth will probably be required to maximize the benefits of pooling experiments.

Another consideration is screening for known clonal mutations that increase the rate of cell proliferation[76,77]; this will be difficult to manage if these mutations arise during differentiation, but they mostly can be screened for at the iPS cell prepooling stage using sequencing techniques. Among the approaches to screen for somatic clonal mutations are digital droplet PCR (for specific mutations)[78], Sanger sequencing, single-nucleotide polymorphism (SNP) arrays and whole-genome sequencing, which have trade-offs between their sensitivity and cost.

Pooling hPS cell lines also reduces batch effects compared with the separate maintenance of hPS cell lines[16]. It is important to remember that batch effects remain between technical replicates of hPS cell pooled experiments[14,24,71]. Still, the strategies to correct batch effects differ considerably between uniculture and pooled setups. In uniculture experiments, the gold standard is to repeat an experiment multiple times and include information on technical replicates in the statistical models to directly estimate and remove batch-induced variation. In pooled experiments, batch variability can be obtained by including a few of the same lines in several pools and using these to account for culture-specific batch effects.

Moreover, there is also variation in the differentiation potential of hPS cell lines. For example, poor differentiation toward endoderm has been linked to X-chromosome reactivation[16], and the relative expression levels of some genes have been linked to poor differentiation toward neural fates[14]. In some cases, the differentiation of pooled hPS cell lines may increase the average maturity or differentiation efficiency, especially for those that involve other cells in the culture. For example, the coordinated beating of cells during differentiation into cardiomyocytes is important for the efficiency of the process. On the basis of our observations, differentiation from pooled hPS cell lines increased the differentiation efficiency of hPS cell lines that only poorly differentiated into cardiomyocytes in an isolated culture[71]. However, using pooled hPS cell cultures can only sometimes reduce such variation. For example, for differentiations that do not require coordination (for example, differentiation into dopaminergic neurons), pooling of hPS cell lines does not appear to alter differentiation efficiency compared with uniculture. Notably, co-culturing of hPS cell lines seems to have only a minimal impact on the unique transcriptomic profile of each line[15]; however, this has been directly evaluated for only a limited number of relatively short differentiations (for example, cardiac differentiation)[71]. Further evaluation across different lineages, protocols and organoid models will be vital as the field progresses.

As the pooled hPS cell culture system is relatively novel, only a few studies have used this approach[14,16,70,71]. Although nonpooled methods will probably remain of value, we anticipate that the significant experimental advantages of pooled hPS cell platforms will result in their wider adoption beyond population genomics. For example, many biological questions have been addressed using a single or a handful of hPS cell lines owing to limited access and high maintenance costs. However, using only a small number of hPS cell lines effectively ignores the influence of the interindividual diversity inherent in human populations, leading to conclusions that are not easily generalizable. Biological variability across an experimental model can be evaluated by replacing an hPS cell line with an hPS cell line pool.

Conclusions and future outlook

Genetic studies have made it possible to associate thousands of loci with phenotypes, such as disease risk, physiology and other quantitative traits, creating an increasingly comprehensive map of the relationship between genetic and phenotypic variation among humans. However, converting this catalog into actionable therapeutic strategies requires evaluation of the molecular mechanisms by which disease-associated variants impact biology and the cell types or cellular contexts in which these processes occur.

While in vivo characterization of the regulatory programs of cell types and trajectories is important for understanding cell development and disease pathogenesis, cellular models (such as hPS cells) that recapitulate human developmental processes are equally important, especially as they provide convenient, scalable platforms for experimental validation. hPS cells are a promising cellular model for assessing the consequences of human genetic variation across different lineages, developmental states and cell types that would be challenging to obtain in vivo[40]. The experimental time and cost of hPS cell maintenance and differentiation have previously limited their use in population genomics studies. However, technological advances have enabled more widespread use of hPS cell models for population genomics studies using pooling strategies and genomic readouts.

To accurately determine how genetic variation impacts cell function, statistical methods to test for genotype-by-context interactions will be necessary[79,80]. Solving this obstacle is inter-related to challenges associated with cell annotation. Application of the dynamic QTL methods presented here will be improved if cell state can be accurately inferred. Leveraging in vivo atlases comprising cell-level variation from a diverse background of genetic ancestry, disease status and demographics will be necessary[81].

We envision that the size of pooled hPS cell experiments will increase from the order of tens to potentially thousands of hPS cell lines with only relatively minimal impact on the overall costs and experimental designs. While the intersection of hPS cells, population genetics methods and cellular genomics provides a powerful platform to elucidate the effect of genetic loci on human disease and quantitative phenotypes, further progress will benefit from scalable methods for functional validation, such as CRISPR screening[67,82–85]. As some of the challenges described above are resolved and new scaling methods are developed, we expect population-scale hPS cell studies to provide an invaluable tool in advancing the discovery and translation of human genomics.

References

1. Thomson, J. A. Embryonic stem cell lines derived from human blastocysts. *Science* https://doi.org/10.1126/science.282.5391.1145 (1998).
2. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
3. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
4. Liu, G., David, B. T., Trawczynski, M. & Fessler, R. G. Advances in pluripotent stem cells: history, mechanisms, technologies, and applications. *Stem Cell Rev. Rep.* **16**, 3–32 (2020).
5. Efrat, S. Epigenetic memory: lessons from iPS cells derived from human β cells. *Front. Endocrinol.* **11**, 614234 (2020).
6. Anderson, R. H. & Francis, K. R. Modeling rare diseases with induced pluripotent stem cell technology. *Mol. Cell. Probes* **40**, 52–59 (2018).
7. Spitalieri, P., Talarico, V. R., Murdocca, M., Novelli, G. & Sangiuolo, F. Human induced pluripotent stem cells for monogenic disease modelling and therapy. *World J. Stem Cells* **8**, 118–135 (2016).
8. Passier, R., Orlova, V. & Mummery, C. Complex tissue and disease modeling using hiPSCs. *Cell Stem Cell* **18**, 309–321 (2016).
9. Warren, C. R., Jaquish, C. E. & Cowan, C. A. The NextGen genetic association studies consortium: a foray into in vitro population genetics. *Cell Stem Cell* **20**, 431–433 (2017).
10. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
11. Tak, Y. G. & Farnham, P. J. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin* **8**, 57 (2015).
12. Umans, B. D., Battle, A. & Gilad, Y. Where are the disease-associated eQTLs? *Trends Genet.* **37**, 109–124 (2021).
13. Yazar, S. et al. Single-cell eQTL mapping identifies cell type–specific genetic control of autoimmune disease. *Science* **376**, eabf3041 (2022).
14. Jerber, J. et al. Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. *Nat. Genet.* **53**, 304–312 (2021).

15. Neavin, D. et al. Single cell eQTL analysis identifies cell type-specific genetic control of gene expression in fibroblasts and reprogrammed induced pluripotent stem cells. *Genome Biol.* **22**, 76 (2021).
16. Cuomo, A. S. E. et al. Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nat. Commun.* **11**, 810 (2020).
17. Warren, C. R. et al. Induced pluripotent stem cell differentiation enables functional validation of GWAS variants in metabolic disease. *Cell Stem Cell* **20**, 547–557 (2017).
18. Kishore, S. et al. A non-coding disease modifier of pancreatic agenesis identified by genetic correction in a patient-derived iPSC line. *Cell Stem Cell* **27**, 137–146 (2020).
19. Magdy, T. et al. *RARG* variant predictive of doxorubicin-induced cardiotoxicity identifies a cardioprotective therapy. *Cell Stem Cell* **28**, 2076–2089 (2021).
20. Bourgeois, S. et al. Towards a functional cure for diabetes using stem cell-derived beta cells: are we there yet? *Cells* **10**, 191 (2021).
21. Sharma, A., Sances, S., Workman, M. J. & Svendsen, C. N. Multi-lineage human iPSC-derived platforms for disease modeling and drug discovery. *Cell Stem Cell* **26**, 309–329 (2020).
22. Volpato, V. & Webber, C. Addressing variability in iPSC-derived models of human disease: guidelines to promote reproducibility. *Dis. Model. Mech.* **13**, dmm042317 (2020).
23. Banovich, N. E. et al. Impact of regulatory variation across human iPSCs and differentiated cells. *Genome Res.* **28**, 122–131 (2018).
24. Kilpinen, H. et al. Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* **546**, 370–375 (2017).
25. Panopoulos, A. D. et al. iPSCORE: a resource of 222 iPSC lines enabling functional characterization of genetic variation across a variety of cell types. *Stem Cell Rep.* **8**, 1086–1100 (2017).
26. Chen, G., Ning, B. & Shi, T. Single-cell RNA-seq technologies and related computational data analysis. *Front. Genet.* **10**, 317 (2019).
27. Elorbany, R. et al. Single-cell sequencing reveals lineage-specific dynamic genetic regulation of gene expression during human cardiomyocyte differentiation. *PLoS Genet.* **18**, e1009666 (2022).
28. Ward, M. C., Banovich, N. E., Sarkar, A., Stephens, M. & Gilad, Y. Dynamic effects of genetic variation on gene expression revealed following hypoxic stress in cardiomyocytes. *eLife* **10**, e57345 (2021).
29. Shi, Z.-D. et al. Genome editing in hPSCs reveals *GATA6* haploinsufficiency and a genetic interaction with *GATA4* in human pancreatic development. *Cell Stem Cell* **20**, 675–688 (2017).
30. Strober, B. J. et al. Dynamic genetic regulation of gene expression during cellular differentiation. *Science* **364**, 1287–1290 (2019).
31. González, F. et al. An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* **15**, 215–226 (2014).
32. Barbeira, A. N. et al. Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol.* **22**, 49 (2021).
33. Hamazaki, T., El Rouby, N., Fredette, N. C., Santostefano, K. E. & Terada, N. Concise review: induced pluripotent stem cell research in the era of precision medicine. *Stem Cells* **35**, 545–550 (2017).
34. Cuomo, A. S. E. et al. CellRegMap: a statistical framework for mapping context-specific regulatory variants using scRNA-seq. *Mol. Syst. Biol.* **18**, e10663 (2022).
35. Cuomo, A. S. E., Nathan, A., Raychaudhuri, S., MacArthur, D. G. & Powell, J. E. Single-cell genomics meets human genetics. *Nat. Rev. Genet.* **24**, 535–549 (2023).
36. Mirauta, B. A. et al. Population-scale proteome variation in human induced pluripotent stem cells. *eLife* **9**, e57390 (2020).
37. Findley, A. S. et al. Functional dynamic genetic effects on gene regulation are specific to particular cell types and environmental conditions. *eLife* **10**, e67077 (2021).
38. Kimura, M. et al. En masse organoid phenotyping informs metabolic-associated genetic susceptibility to NASH. *Cell* https://doi.org/10.1016/j.cell.2022.09.031 (2022).
39. Llufrio, E. M., Wang, L., Naser, F. J. & Patti, G. J. Sorting cells alters their redox state and cellular metabolome. *Redox Biol.* **16**, 381–387 (2018).
40. Shen, S. et al. Integrating single-cell genomics pipelines to discover mechanisms of stem cell differentiation. *Trends Mol. Med.* https://doi.org/10.1016/j.molmed.2021.09.006 (2021).
41. van der Wijst, M. et al. The single-cell eQTLGen consortium. *eLife* **9**, e52155 (2020).
42. Soskic, B. et al. Immune disease risk variants regulate gene expression dynamics during $CD4^+$ T cell activation. *Nat. Genet.* **54**, 817–826 (2022).
43. Daniszewski, M. et al. Retinal ganglion cell-specific genetic regulation in primary open-angle glaucoma. *Cell Genomics* **2**, 100142 (2022).
44. Senabouth, A. et al. Transcriptomic and proteomic retinal pigment epithelium signatures of age-related macular degeneration. *Nat. Commun.* **13**, 4233 (2022).
45. Benaglio, P. et al. Mapping genetic effects on cell type-specific chromatin accessibility and annotating complex immune trait variants using single nucleus ATAC-seq in peripheral blood. *PLoS Genet.* **19**, e1010759 (2023).
46. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. *Nat. Rev. Mol. Cell Biol.* **24**, 695–713 (2023).
47. Weinshilboum, R. M. & Wang, L. Pharmacogenomics: precision medicine and drug response. *Mayo Clin. Proc.* **92**, 1711–1722 (2017).
48. Pirmohamed, M. Personalized pharmacogenomics: predicting efficacy and adverse drug reactions. *Annu. Rev. Genom. Hum. Genet.* **15**, 349–370 (2014).
49. Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
50. Hay, M., Thomas, D. W., Craighead, J. L., Economides, C. & Rosenthal, J. Clinical development success rates for investigational drugs. *Nat. Biotechnol.* **32**, 40–51 (2014).
51. Holmgren, G. et al. Long-term chronic toxicity testing using human pluripotent stem cell-derived hepatocytes. *Drug Metab. Dispos.* **42**, 1401–1406 (2014).
52. Kim, J.-H., Kang, M., Jung, J.-H., Lee, S.-J. & Hong, S.-H. Human pluripotent stem cell-derived alveolar epithelial cells as a tool to assess cytotoxicity of particulate matter and cigarette smoke extract. *Dev. Reprod.* **26**, 155–163 (2022).
53. Sharma, A. et al. High-throughput screening of tyrosine kinase inhibitor cardiotoxicity with human induced pluripotent stem cells. *Sci. Transl. Med.* **9**, eaaf2584 (2017).
54. Han, Y. et al. Identification of SARS-CoV-2 inhibitors using lung and colonic organoids. *Nature* **589**, 270–275 (2021).
55. Lam, C. K. & Wu, J. C. Clinical trial in a dish: using patient-derived induced pluripotent stem cells to identify risks of drug-induced cardiotoxicity. *Arterioscler. Thromb. Vasc. Biol.* **41**, 1019–1031 (2021).
56. Iwata, R. et al. Mitochondria metabolism sets the species-specific tempo of neuronal development. *Science* **379**, eabn4705 (2023).
57. Miller, J. D. et al. Human iPSC-based modeling of late-onset disease via progerin-induced aging. *Cell Stem Cell* **13**, 691–705 (2013).
58. Hergenreder, E. et al. Combined small-molecule treatment accelerates maturation of human pluripotent stem cell-derived neurons. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-023-02031-z (2024).
59. Fowler, J. L., Ang, L. T. & Loh, K. M. A critical look: challenges in differentiating human pluripotent stem cells into desired cell types and organoids. *Wiley Interdiscip. Rev. Dev. Biol.* **9**, e368 (2020).

60. Jiang, S., Feng, W., Chang, C. & Li, G. Modeling human heart development and congenital defects using organoids: how close are we? *J. Cardiovasc. Dev. Dis.* **9**, 125 (2022).
61. Tremmel, D. M. et al. Validating expression of beta cell maturation-associated genes in human pancreas development. *Front. Cell Dev. Biol.* **11**, 1103719 (2023).
62. Washer, S. J. et al. Single-cell transcriptomics defines an improved, validated monoculture protocol for differentiation of human iPSC to microglia. *Sci. Rep.* **12**, 19454 (2022).
63. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
64. Wilson, S. B. et al. DevKidCC allows for robust classification and direct comparisons of kidney organoid datasets. *Genome Med.* **14**, 19 (2022).
65. Subramanian, A. et al. Single cell census of human kidney organoids shows reproducibility and diminished off-target cells after transplantation. *Nat. Commun.* **10**, 5462 (2019).
66. Kammers, K. et al. Gene and protein expression in human megakaryocytes derived from induced pluripotent stem cells. *J. Thromb. Haemost.* **19**, 1783–1799 (2021).
67. De Sousa, P. A. et al. Rapid establishment of the European Bank for induced Pluripotent Stem Cells (EBiSC)—the Hot Start experience. *Stem Cell Res.* **20**, 105–114 (2017).
68. Morrison, M. et al. StemBANCC: governing access to material and data in a large stem cell research consortium. *Stem Cell Rev. Rep.* **11**, 681–687 (2015).
69. The GTEx Consortium The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
70. Mitchell, J. M., Nemesh, J., Ghosh, S. & Handsaker, R. E. Mapping genetic effects on cellular phenotypes with 'cell villages'. Preprint at *bioRxiv* https://doi.org/10.1101/2020.06.29.174383 (2020).
71. Neavin, D. R. et al. A village in a dish model system for population-scale hiPSC studies. *Nat. Commun.* **14**, 3240 (2023).
72. Kang, H. M. et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018).
73. Wells, M. F. et al. Natural variation in gene expression and viral susceptibility revealed by neural progenitor cell villages. *Cell Stem Cell* **30**, 312–332 (2023).
74. Neavin, D. et al. *Demuxafy*: improvement in droplet assignment by integrating multiple single-cell demultiplexing and doublet detection methods. *Genome Biol.* **25**, 94 (2024).
75. Xu, J. et al. Genotype-free demultiplexing of pooled single-cell RNA-seq. *Genome Biol.* **20**, 290 (2019).
76. Heaton, H. et al. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* **17**, 615–620 (2020).
77. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol.* **20**, 273 (2019).
78. Hindson, B. J. et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* **83**, 8604–8610 (2011).
79. Dong, X. et al. powerEQTL: an R package and shiny application for sample size and power calculation of bulk tissue and single-cell eQTL analysis. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btab385 (2021).
80. Schmid, K. T. et al. scPower accelerates and optimizes the design of multi-sample single cell transcriptomic studies. *Nat. Commun.* **12**, 6625 (2021).
81. Camp, J. G., Platt, R. & Treutlein, B. Mapping human cell phenotypes to genotypes with single-cell genomics. *Science* **365**, 1401–1405 (2019).
82. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
83. Dixit, A. et al. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866 (2016).
84. Rubin, A. J. et al. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell* **176**, 361–376 (2019).
85. Schraivogel, D. et al. Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat. Methods* **17**, 629–635 (2020).

Acknowledgements

Figures were generated with BioRender.com and further developed by A. Garcia, a scientific illustrator from Bio-Graphics. This research was supported by a National Health and Medical Research Council (NHMRC) Investigator grant (J.E.P., 1175781), research grants from the Australian Research Council (ARC) Special Research Initiative in Stem Cell Science, an ARC Discovery Project (190100825), an EMBO Postdoctoral Fellowship (A.S.E.C.) and an Aligning Science Across Parkinson's Grant (J.E.P., N.F., D.R.N. and L.S.). J.E.P. is supported by a Fok Family Fellowship.

Author contributions

All authors conceived the topic and wrote and revised the manuscript.

Competing interests

D.G.M. is a founder with equity in Goldfinch Bio, is a paid advisor to GSK, Insitro, Third Rock Ventures and Foresite Labs, and has received research support from AbbVie, Astellas, Biogen, BioMarin, Eisai, Merck, Pfizer and Sanofi-Genzyme; none of these activities is related to the work presented here. J.E.P. is a founder with equity in Celltellus Laboratory and has received research support from Illumina. The other authors declare no conflict of interest.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-024-01731-9.

Correspondence should be addressed to Joseph E. Powell.

Peer review information *Nature Genetics* thanks Kelly Frazer, Gosia Trynka and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.