Proceedings: Consideration of Genetics in the Design of Induced Pluripotent Stem Cell-Based Models of Complex Disease

UTA GRIESHAMMER, KELLY A. SHEPARD

California Institute for Regenerative Medicine, 210 King Street, San Francisco, CA 94107 USA

Key Words. Induced pluripotent stem cells • iPSC • Experimental models • Stem cells • Technology • Genomics

SUMMARY

The goal of exploiting induced pluripotent stem cell (iPSC) technology for the discovery of new mechanisms and treatments of disease is being pursued by many laboratories, and analyses of rare monogenic diseases have already provided ample evidence that this approach has merit. Considering the enormous medical burden imposed by common chronic diseases, successful implementation of iPSC-based models has the potential for major impact on these diseases as well. Since common diseases represent complex traits with varying genetic and environmental contributions to disease manifestation, the use of iPSC technology poses unique challenges. In this perspective, we will consider how the genetics of complex disease and mechanisms underlying phenotypic variation affect experimental design.

INTRODUCTION

Common chronic diseases, such as many cardiac, metabolic, and neurodegenerative disorders, still lack cures, are associated with extensive morbidity and mortality, and represent major economic burdens to society. The recent advent of induced pluripotent stem cell (iPSC) technology has provided new opportunities for elucidating pathological mechanisms, and in order to address these long-standing challenges, it is paramount to explore the potential of iPSCs for modeling common diseases. In the short time since their first description [1-3], iPSCs have been successfully used to model aspects of numerous human diseases [4-6]. This speedy adoption of a new technology is a testament to the community’s excitement and conviction that iPSC-based modeling will yield important new insights into disease mechanisms and in due course, reveal new molecular targets or phenotypic assays for drug development. To provide a resource for the iPSC-based study of common diseases, the California Institute for Regenerative Medicine (CIRM) is funding an iPSC Initiative with the goal of recruiting 3000 tissue donors who either suffer from a common disease or will serve as healthy controls, for deriving iPSC lines that will be banked and distributed to researchers worldwide (Table 1). This resource will include an accompanying set of de-identified demographic, medical and/or diagnostic information for each tissue donor to enable clinically meaningful assays to be designed and interpreted.

iPSCs represent a replenishable source of clinically relevant, patient-specific cell types for in vitro analyses. The rationale underlying iPSC-based modeling is that disease-contributing germline variants will be present in any type of cell, whether affected by disease or not, and will thus be carried by iPSCs derived from easily accessible cells such as blood or skin. The success of this approach depends on the ability of cells propagated in culture to mimic cellular or molecular aspects of disease, an approach that has been the basis for rational drug discovery for decades, but may greatly benefit from the use of iPSCs, differentiated into disease-relevant cell types, as they provide patient cells with the underlying disease genetics.

Common diseases represent complex traits where both environmental and genetic factors determine the course of disease, such as the timing of onset, progression and severity of symptoms. Heritability, i.e. the proportion of a disease’s manifestations that is attributable to genetic variation within a population, varies from disease to disease. Furthermore, the overall genetic contribution to complex diseases is often polygenic, consisting of few, if any, genetic variants with large effects, a handful of variants with modest effects, and/or a substantial num-
Age range
0-18
60-85

Tissue collection and iPSC generation for the CIRM iPSC bank is currently under way. Diseases were selected through a competitive process in response to a CIRM request for applications. Tissue donor numbers were proposed in the applications and some were modified during application review. Listed are the included diseases, the number of cases to be collected for each disease, and the number and age range of unrelated controls. Highlighted in grey are blue are diseases with shared control cohorts. Idiopathic pulmonary fibrosis shares controls with both the middle age and the elderly age range.

| Diseases                                           | # Cases | # Unrelated controls | Age range |
|----------------------------------------------------|---------|----------------------|-----------|
| Neurodevelopmental disabilities                    | 450a    | 200                  | 4-18      |
| Autism spectrum disorders                          | 200     |                      |           |
| Idiopathic familial dilated cardiomyopathy         | 650a    | 30                   | 0         |
| Hepatitis C                                        | 132     |                      |           |
| Nonalcoholic steatohepatitis                       | 34      | 20                   | 18-55     |
| Idiopathic pulmonary fibrosis                      | 250     |                      |           |
| Age-related macular degeneration                   | 300     | 50                   |           |
| Primary open angle glaucoma                        | 50      |                      |           |
| Proliferative diabetic retinopathy                 | 100b    | 300                  | 60-85     |
| Diabetes (no retinopathy)                          | 50b     |                      |           |
| Late onset Alzheimer’s disease                     | 235     |                      |           |

Tissue donor numbers were proposed in the applications and some were modified during application review. Listed are the included diseases, the number of cases to be collected for each disease, and the number and age range of unrelated controls. Highlighted in grey are blue are diseases with shared control cohorts. Idiopathic pulmonary fibrosis shares controls with both the middle age and the elderly age range.

| Diseases                                           | # Cases | # Unrelated controls | Age range |
|----------------------------------------------------|---------|----------------------|-----------|
| Neurodevelopmental disabilities                    | 450a    | 200                  | 4-18      |
| Autism spectrum disorders                          | 200     |                      |           |
| Idiopathic familial dilated cardiomyopathy         | 650a    | 30                   | 0         |
| Hepatitis C                                        | 132     |                      |           |
| Nonalcoholic steatohepatitis                       | 34      | 20                   | 18-55     |
| Idiopathic pulmonary fibrosis                      | 250     |                      |           |
| Age-related macular degeneration                   | 300     | 50                   |           |
| Primary open angle glaucoma                        | 50      |                      |           |
| Proliferative diabetic retinopathy                 | 100b    | 300                  | 60-85     |
| Diabetes (no retinopathy)                          | 50b     |                      |           |
| Late onset Alzheimer’s disease                     | 235     |                      |           |

A similar caveat is that disease-relevant genetic variants may be limited to affected tissues in patients with somatic mosaicism, which may, for instance, arise through clonal expansion of a cell in which a genetic alteration occurred during development, or in rare instances of chimerism, in which, for example, an organism acquires genetically distinct tissues during early development through fusion of fraternal twins. Notably, somatic mosaicism has been implicated in individuals with diseases that are of particular interest in terms of iPSC modeling, namely schizophrenia, autism, Alzheimer’s, and cardiac disease [10]. Similarly, retrotransposition, specifically in the human brain, has been hypothesized to cause genetic variation with possible implications for neurological disease [11]. In patients with somatic mosaicism, disease-relevant mutations may not be present in the source cells for iPSC derivation. Conversely, mutations that influence phenotypes in vitro may have been introduced during the reprogramming process, either due to clonal expansion of pre-existing somatic mutations in iPSC source cells or due to replication errors in culture [12], and may thus confound analysis.

| Considerations for Experimental Design |
|---------------------------------------|
| The overall genetic risk in individuals chosen for study has likely consequences for expected effect sizes and thus implications for the design of iPSC-based disease studies. Not surprisingly, most diseases modeled with iPSCs to date are highly penetrant monogenic disorders [4–6], since the genetic contribution to cellular phenotype is expected to be strong, and only few patient samples are needed to ascertain differences between cells from patients (“cases”) and those from unaffected controls. |
| iPSC models of some severe monogenic syndromes that, among other conditions, include phenotypes typical of common neurological disorders, have already suggested that relevant insights can be obtained from such studies; for instance, defects observed in iPSC-derived neurons from patients with Timothy syndrome have possible implications for autism [13]. A similar approach focuses on rare forms of complex diseases that are driven by single highly penetrant loci (genetic variants with very large effects on disease susceptibility) such as LRRK2, PARK2 or SNCA variants for Parkinson’s disease [14–16], the APP and PS1 and PS2 mutations in Alzheimer’s disease [17–19] and SCN1A mutations in epilepsy [20]. The findings in these studies support the notion that comparing individual patient samples to control samples can be sufficient to reveal statistically significant disease-specific phenotypes in a dish (Table 2) in a complex disease with strong genetic underpinnings. |
| When designing iPSC-based studies to model complex diseases with polygenic contributions and overall less pronounced heritability, including the common, sporadic forms of the neurological diseases described above, averaging across different patient samples may be necessary, although the sample sizes needed to provide statistically significant results are currently not understood. In genome wide association (GWA) studies, which are designed to detect associations between individual genetic markers and phenotypes, very large sample sizes (1000s) are often needed to reach statistical power capable of resolving small effect sizes, see [21–23]. In iPSC-based models, on the other hand, the combined effect of all disease-contribut-
ut ing alleles present in a patient, even if each variant contributes little individually, will be assayed at once, so the overall contribution of genomic variation to disease will be a driver of needed sample size.

Because heritability of a complex disease may vary tremendously amongst different patient subpopulations, it may be judicious to focus on e.g. ethnic subpopulations in which affected individuals are likely to carry strong genetic risk, thereby reducing the required sample size. Alternatively, strong genetic risk in individual patients may often be inferred from a family history of disease, or from severe disease manifestations such as early onset or extreme disease outcome. Ideally, patients with extreme genetic risk can be identified by determining the presence of single nucleotide polymorphisms (SNPs) or other genomic markers known to be associated with a disease based on GWA studies or, with recent technological advances, through whole genome sequencing.

Examples that support the validity of these approaches for reducing needed sample size already exist in the literature. For instance, in a study of schizophrenia, in which a discernable phenotype was observed in iPSC-derived neurons based on the average of four samples (Table 2), the cases presented with early onset or were from families with many affected members [24]. Since they were not compared to other cases, though, it is not known if these choices improved the ability to measure phenotypes in a dish.

In another example, iPSC-derived neurons obtained from two Alzheimer’s disease patients heterozygous for the ApoE4 risk allele exhibited abnormal phenotypes in a dish, while those from another ApoE4 carrier with the disease did not (Table 2). Interestingly, only the first two patients had manifested early onset Alzheimer’s disease [25], a difference in disease severity possibly due to differences in additional genetic risk factors, and thus expected effect sizes, that may explain the presence or absence of the observed in vitro phenotypes. Other explanations are possible, though, and much larger studies will be needed to draw firm conclusions about the cause of such differences.

Simulating environmental or other non-genetic factors that contribute to disease may enhance in vitro disease phenotypes. In a study of age-related macular degeneration (AMD), the comparison of iPSC-derived retinal pigment epithelium, the cell type affected in AMD, from unaffected and affected individuals with a protective versus a risk haplotype, respectively, did not show phenotypic differences until aging of the cells was accelerated by increasing oxidative stress [26]. Similarly, progerin-induced aging enhanced disease-related phenotypes in iPSC models of Parkinson’s disease with disease-causing mutations [27].

Table 2: iPSC-based results in familial and sporadic forms of complex diseases

| Disease                          | Affected locus | # Affected / # with in vitro phenotype | Controls | Reference |
|----------------------------------|----------------|----------------------------------------|----------|-----------|
| Parkinson’s disease, familial    | LRRK2          | 1/1                                    | 1, unrelated | [15]      |
|                                  | autosomal dominant point mutation patient homozygous |                                  |          |           |
|                                  | PARK2          | 2/2                                    | 2, unrelated | [16]      |
|                                  | autosomal recessive exon deletions |                                  |          |           |
|                                  | SNCA           | 1/1                                    | 1, unaffected sibling | [14]      |
|                                  | autosomal dominant triplication |                                  |          |           |
| Alzheimer’s disease, familial    | PS1            | 1/1                                    | 2, unrelated | [19]      |
|                                  | autosomal dominant point mutation |                                  |          |           |
|                                  | PS2            | 1/1                                    | 2, unrelated | [19]      |
|                                  | autosomal dominant point mutation |                                  |          |           |
|                                  | APP            | 2/2                                    | 2, unrelated | [17]      |
|                                  | autosomal dominant duplication |                                  |          |           |
|                                  | APP            | 2/2                                    | 3, unrelated | [18]      |
|                                  | amino acid deletion (recessive) / point mutation (dominant) |                                  |          |           |
| Alzheimer’s disease, sporadic    | unknown        | 2/1                                    | 2, unrelated | [17]      |
|                                  | unknown        | 2/1                                    | 3, unrelated | [18]      |
|                                  | ApoE3/E4       | 3/2                                    | 4, unrelated | [25]      |
|                                  | heterozygous for high risk allele E4 |                                  |          |           |
| Schizophrenia                    | unknown        | 4 / average                           | 6 / average, unrelated | [24]      |

The loci affected in patients included in iPSC-based disease modeling studies of familial forms of complex diseases are listed. This list is by no means comprehensive. The number of affected individuals included in each study ranged from 1 to 4. The phenotypes observed in vitro were reported either individually for each affected individual or as an average of all included cases. The number of controls used, and whether the control individuals were related to the cases or not, is also listed.

Abbreviations: hESC, human embryonic stem cells.
Phenotypic Variation in iPSC Models of Complex Disease—Discrete Versus Quantitative Traits

Biological variation may be observed as discrete traits that can be measured as statistically significant “on / off” or “high / low” signals, or as quantitative traits that vary continuously within a range. In complex diseases, examples of discrete traits include development of type 1 diabetes or death from myocardial infarction, while blood pressure or cholesterol levels represent quantitative traits [28]. Similarly, molecular and cellular phenotypes measured in vitro may present as discrete or as quantitative traits, and this distinction may influence experimental design.

In the case of discrete traits, an important consideration impacting design and interpretation of iPSC-based modeling studies is whether the same or different molecular pathways are malfunctioning in different individuals with a given disease. In two published iPSC-based studies of Alzheimer’s disease that were focused on familial disease due to genetic alterations in the amyloid precursor protein (APP) gene, the investigators included patients with sporadic disease. Interestingly, some of the sporadic cases displayed the same disease phenotype in vitro as those carrying an APP mutant allele, and some did not [17, 18], Table 2. One possible explanation for this difference is that the sporadic cases whose cells recapitulated the phenotype caused by an APP mutation had genetic defects affecting the same molecular pathway(s) as did the APP mutation, while the other sporadic cases may also have had a strong genetic basis, but were driven by genetic variants affecting different pathways that were not assessed in the assays chosen for study. In fact, the two familial cases included in one of the studies carried different APP mutations, and while both showed an in vitro phenotype, they differed in type of phenotype observed [18]. Such genetic variability amongst patients with the “same” disease will require smart assay design to model different modes of disease causation. Once an appropriate assay has been developed, statistically significant discrete signals may be obtained in an in vitro assay, even in individual patient samples, as long as they represent the appropriate subtype of the disease. As a matter of fact, iPSC-based assays may one day be used to stratify patient populations into subcategories of complex disease that may require different approaches to drug development and treatment based on the molecular or cellular pathways that are affected.

If heritability is relatively low, differences in discrete signals between affected versus control samples may not be readily distinguishable. Furthermore, genetically driven molecular phenotypes of a complex disease may vary along a continuum (quantitative trait), such as the activity level of an enzyme. To discern small differences between normal and affected patient iPSC-derived cell populations, large patient numbers may be required to achieve statistical significance, and multiple variables may need to be assessed simultaneously to capture a phenotype. This could be achieved by employing automated large-scale approaches involving robotic cell handling systems and automated data acquisition and analysis, using for instance omics approaches to detect global gene expression or epigenetic patterns, acquired during differentiation in vitro, that distinguish disease samples from healthy controls [29–31]. Another high throughput approach involves high content analysis, whereby large numbers of images detecting cell features, such as proliferation, survival, migration, membrane integrity, size and distribution of organelles and more, can be acquired through automated imaging and analyzed through machine learning algorithms [5, 32, 33]. Such large-scale automated systems analyses could point to common cellular or molecular disease mechanisms across the “population” of cell lines representing different individuals with a complex disease.

It is critical to realize that in cases in which the genetic contribution to disease is small, it may not be feasible to detect phenotypic differences when comparing disease iPSCs to controls, even if dozens or hundreds of different iPSC lines could be analyzed. Before embarking on iPSC-based studies of sporadic diseases with relatively low heritability, pilot studies with more highly penetrant familial forms of the disease, if possible, could be used to establish whether assays with sufficient sensitivity can be developed. Importantly, experimental variation inherent in iPSC-based models will confound analyses and should be minimized through the development of robust, reproducible assays. Pilot studies for assessing the effect of experimental variability on observed in vitro phenotypes will enable judicious selection of an appropriate number of replicate lines for analysis and will help inform determination of sample size for large scale studies.

Choice of Controls for Complex Diseases in a Dish

An important consideration for the design of iPSC models of complex diseases is the choice of controls. What constitutes an appropriate control depends on experimental design, but is not always entirely clear, and what may be considered optimal may be impractical. Thus, choosing controls for complex disease studies is usually a compromise. To increase the likelihood of identifying a disease phenotype, controls should ideally contain few, if any, of the disease risk alleles. This may be more effectively achieved by using unrelated individuals as controls, rather than unaffected blood relatives, especially if disease-associated alleles remain unknown and can, therefore, not be tracked. For instance, two siblings, one affected, one not, may have identical genetic disease risk but only the affected individual was exposed to an unknown environmental trigger. If the iPSC-based phenotype is indeed driven by the genetics of the disease, it would be the same for the affected and the unaffected sibling in this case. For this reason, GWA studies often employ unrelated individuals as controls, since unaffected blood relatives are considered ‘over-matched’ for genotypes, although the case for the use of related individuals has also been made [34–36].

When employing unrelated controls in iPSC-based studies, it is important to consider, as in GWA studies, potential effects of population structure, such as ethnicity, geography, age, and other factors that may confound inferences in case-control studies, and attempts should be made to match cases and controls accordingly. For instance, ethnically related individuals share genetic similarities beyond those associated with the disease under investigation, and genetic variants not associated with disease would contribute false positive signals in GWA studies if the cases and controls were of different ethnicities. Similarly, in iPSC-based studies, genetic variation not associated with disease may inappropriately affect in vitro phenotypes or contribute to confounding culture artifacts. It is, however, conceivable that such variation may have little effect on certain disease phenotypes in a dish, and there may be circumstances in which the inclusion of some directed cross-ethnical samples in an iPSC-based study is justified to take advantage of enhanced disease-specific phenotypic differences, especially if the disease is known to be highly heritable in one ethnic group and not
critical disease manifestation depends on recapitulating specific cell type(s) are primarily affected and if so, whether appropriate protocols to generate those cell types from iPSCs are available. As discussed, the magnitude of the genetic contribution to a complex disease involves that population structures can influence human population genetic studies, and therefore may also confound iPSC-based studies. Since some of the diseases included in the CIRM iPSC bank afflict the young, namely neurodevelopmental disorders and dilated cardiomyopathy, the use of elderly individuals as controls might introduce confounding variation that could be avoided by age-matching controls for the young case cohorts. Second, if iPSCs from the elderly retain molecular changes that accumulate with age, they may be poor controls for iPSCs from younger individuals. Interestingly, reprogramming to pluripotency has been shown to reverse age-associated markers, such as increases in formation of DNA double strand breaks and levels of mitochondrial reactive oxygen species, in cells from older individuals [27, 37, 38], although genetic changes that may have accumulated would not be reversed. The third limitation is practical in nature: the more diseases that the control cohort is intended to serve, the more diseases that would need to be excluded in each control individual. Since control individuals who are recruited into the CIRM iPSC Initiative will have to undergo testing to exclude disease, this may become a barrier to tissue donor recruitment. Instead of a single control cohort, the CIRM iPSC Initiative is implementing three shared control cohorts, age-matched to young, middle-aged, and elderly cases (see Table 1).

**Shared Controls**

A major advantage of using unrelated controls is the potential for sharing them between different disease studies [22, 23], thereby reducing the number of iPSC lines that need to be derived and banked. This approach is being implemented by CIRM for its iPSC bank of complex diseases. For those included diseases that are late onset, namely Alzheimer’s disease, idiopathic pulmonary fibrosis, and blinding eye diseases, control individuals need to be elderly to best ensure that they do not, and likely will not in the future, suffer from those diseases. With this limitation in mind, CIRM considered establishing a single control cohort, covering all included diseases, by recruiting healthy elderly individuals, with no history of the targeted diseases. However, there are three limitations to this approach that led to the rejection of this concept. The first is based on observations that population structures can influence human population genetic studies, and therefore may also confound iPSC-based studies. Since some of the diseases included in the CIRM iPSC bank afflict the young, namely neurodevelopmental disorders and dilated cardiomyopathy, the use of elderly individuals as controls might introduce confounding variation that could be avoided by age-matching controls for the young case cohorts. Second, if iPSCs from the elderly retain molecular changes that accumulate with age, they may be poor controls for iPSCs from younger individuals. Interestingly, reprogramming to pluripotency has been shown to reverse age-associated markers, such as increases in formation of DNA double strand breaks and levels of mitochondrial reactive oxygen species, in cells from older individuals [27, 37, 38], although genetic changes that may have accumulated would not be reversed. The third limitation is practical in nature: the more diseases that the control cohort is intended to serve, the more diseases that would need to be excluded in each control individual. Since control individuals who are recruited into the CIRM iPSC Initiative will have to undergo testing to exclude disease, this may become a barrier to tissue donor recruitment. Instead of a single control cohort, the CIRM iPSC Initiative is implementing three shared control cohorts, age-matched to young, middle-aged, and elderly cases (see Table 1).

**Additional Considerations**

As discussed, the magnitude of the genetic contribution to a complex disease is an important factor to consider when pursuing an iPSC-based modeling strategy, but not all complex diseases with strong genetic underpinnings will be amenable to iPSC-based modeling. Other factors need to be considered when deciding if such an approach is feasible, including whether it is known which cell type(s) are primarily affected and if so, whether appropriate protocols to generate those cell types from iPSCs are available. It is also possible that the complexities of metabolic and cellular interactions involved in disease etiology may pose a hurdle for creating a meaningful disease model in a dish, for example if a critical disease manifestation depends on recapitulating specific temporal or spatial events between cell types. As human pluripotent stem cell technology represents a relatively recent development, there are many additional scientific and technical hurdles to overcome (reviewed in [4, 39]), before efficient modeling of disease is truly attainable. These include achieving appropriate maturity of cell types, while understanding and controlling heterogeneity in iPSC-derived cell populations, and the need to develop co-cultures of different cell types if cell non-autonomous mechanisms are at play.

**Conclusion**

The genetic underpinnings of disease provide the basis for iPSC-based modeling of disease, and, by the same token, iPSCs may serve as tools to investigate the complex genetics of human disease [31, 40]. Although genetically engineered animal models can provide valuable insights into the effects of disease-causing genetic variants identified in humans, much remains to be elucidated about the causal mechanisms underlying findings from GWA studies, and iPSC-based models may be used to interrogate the functional significance of disease-associated SNPs in the context of the genome from affected individuals. Recent advances in genome wide mapping of gene regulatory regions, and the finding that many disease-associated SNPs lie within them [41, 42], provide experimental paradigms that can be tested in iPSC-based models. And recent breakthroughs in genome editing technology [43] will greatly facilitate efforts to precisely alter candidate loci in human iPSCs, even at multiple loci simultaneously, to probe their combined effect on complex disease phenotypes observed in a dish. iPSC technology may also facilitate the analysis of expression quantitative trait loci (eQTL) in complex diseases by providing access, through in vitro differentiation, to more disease-relevant cell types than previously possible [44].

In light of the fact that the mechanisms that cause common diseases are complex, and often still poorly understood, the success of iPSC-based disease in a dish models will not only depend on the availability of detailed clinical information from the participating patients, but also on the collaboration of iPSC assay developers with experts in human genetics and with clinicians who have a deep understanding of disease manifestations [45]. Designing experiments that are clinically meaningful and that go beyond verifying existing knowledge about complex disease will lead to new discoveries that may ultimately impact our understanding and treatment of complex diseases.

**Acknowledgments**

We thank Michael Yaffe (CIRM), Aarno Palotie (The Broad Institute of MIT and Harvard, Massachusetts General Hospital, and University of Helsinki), Ulrich Broeckel (Medical College of Wisconsin), and Natalie DeWitt (Baxter Laboratories, Department of Microbiology and Immunology, Stanford University) for valuable input.

**Author Contributions**

U.G.: Conception and design, manuscript writing, final approval of manuscript; K.A.S.: manuscript writing.

**Disclosure of Potential Conflicts of Interest**

The authors indicate no potential conflicts of interest.
References

1. Takahashi K, Tanabe K, Ohnuki M et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 2007;131:861–872.

2. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 2006;126:663–676.

3. Yu J, Vodyanik MA, Smuga-Otto K et al. Induced pluripotent stem cell lines derived from human somatic cells. Science 2007;318:1917–1920.

4. Onder TT, Daley GQ. New lessons learned from disease modeling with induced pluripotent stem cells. Curr Opin Genet Dev 2012;22:500–508.

5. Rajamohan D, Matsa E, Kalra S et al. Current status of drug screening and disease modelling in human pluripotent stem cells. Bioessays 2013;35:281–298.

6. Trounson A, Shepard KA, DeWitt ND. Human disease modeling with induced pluripotent stem cells. Curr Opin Genet Dev 2012;22:505–516.

7. Gratten J, Wray NR, Keller MC et al. Large-scale genomics unveils the genetic architecture of psychiatric disorders. Nat Neurosci 2014;17:782–790.

8. Lappalainen T, Montgomery SB, Nica AC et al. Epistatic selection between coding and regulatory variation in human evolution and disease. Am J Hum Genet 2011;89:459–463.

9. Marian AJ. Elements of ‘missing heritability’. Curr Opin Cardiol 2012;27:197–201.

10. Biesecker LG, Spinner NB. A genomic view of mosaicism and human disease. Nat Rev Genet 2010;11:518–530.

11. Lappalainen T, Montgomery SB, Nica AC et al. Epistatic selection between coding and regulatory variation in human evolution and disease. Am J Hum Genet 2011;89:459–463.

12. Marian AJ. Elements of ‘missing heritability’. Curr Opin Cardiol 2012;27:197–201.

13. Biesecker LG, Spinner NB. A genomic view of mosaicism and human disease. Nat Rev Genet 2010;11:518–530.

14. Lappalainen T, Montgomery SB, Nica AC et al. Epistatic selection between coding and regulatory variation in human evolution and disease. Am J Hum Genet 2011;89:459–463.

15. Marian AJ. Elements of ‘missing heritability’. Curr Opin Cardiol 2012;27:197–201.

16. Imaiizumi Y, Okada Y, Akamatsu W et al. Mitochondrial dysfunction associated with increased oxidative stress and alpha-synuclein accumulation in PARK2 iPSC-derived neurons and postmortem brain tissue. Mol Brain 2012;5:35.

17. Israel MA, Yuan SH, Bardy C et al. Probing sporadic and familial Alzheimer’s disease using induced pluripotent stem cells. Nature 2012;482:216–220.

18. Kondo T, Asal M, Tsukita K et al. Modeling Alzheimer’s disease with iPSCs reveals stress phenotypes associated with intracellular Aβ and differential drug responsiveness. Cell Stem Cell 2013;12:487–496.

19. Yagi T, Ito D, Okada Y et al. Modeling familial Alzheimer’s disease with induced pluripotent stem cells. Hum Mol Genet 2011;20:4530–4539.

20. Liu Y, Lopez-Santiago LF, Yuan Y et al. Dravet syndrome patient-derived neurons suggest a novel epilepsy mechanism. Ann Neurol 2013;74:128–139.

21. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014;511:421–427.

22. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–678.

23. Wellcome Trust Case Control Consortium, Craddock N, Hurles ME et al. Genome-wide association study in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 2010;464:713–720.

24. Brennand KJ, Simone A, Jou J et al. Modeling schizophrenia using human induced pluripotent stem cells. Nature 2011;473:221–225.

25. Duan L, Bhattacharyya BJ, Belmadani A et al. Stem cell derived basal forebrain cholinergic neurons from Alzheimer’s disease patients are more susceptible to cell death. Mol Neurodegener 2014;9:3.

26. Yang J, Li Y, Chan L et al. Validation of genome-wide association study (GWAS)-identified disease risk alleles with patient-specific stem cell lines. Hum Mol Genet 2014;23:3445–3455.

27. Miller JD, Ganat YM, Kishinevsky S et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 2011;473:43–49.

28. Lander ES, Schork NJ. Genetic dissection of complex traits. Science 1994;265:2037–2048.

29. Gibson G, Weir B. The quantitative genetics of transcription. Trends Genet 2005;21:616–623.

30. Cookson W, Liang L, Abecasis G et al. Mapping complex disease traits with global gene expression. Nat Rev Genet 2009;10:184–194.

31. Merkle FT, Eggan K. Modeling human disease with pluripotent stem cells: From genome association to function. Cell Stem Cell 2013;12:656–668.

32. Arrasate M, Finkbeiner S. Automated microscope system for determining factors that predict neuronal fate. Proc Natl Acad Sci USA 2005;102:3840–3845.

33. Daub A, Sharma P, Finkbeiner S. High-content screening of primary neurons: ready for prime time. Curr Opin Neurobiol 2009;19:537–543.

34. Li M, Boehnke M, Abecasis GR. Efficient study designs for test of genetic association using sibship data and unrelated cases and controls. Am J Hum Genet 2007;80:774–784.

35. Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000;405:847–856.

36. Visscher PM, Andrew T, Nyholt DR. Genome-wide association studies of quantitative traits with related individuals: little (power) lost but much to be gained. Eur J Hum Genet 2008;16:387–390.

37. Mahmoudi S, Brunet A. Aging and reprogramming: A two-way street. Curr Opin Cell Biol 2012;24:744–756.

38. Freije JM, Lopez-Otin C. Reprogramming aging and progeria. Curr Opin Cell Biol 2012;24:757–764.

39. Saha K, Jaenisch R. Technical challenges in using human induced pluripotent stem cells to model disease. Cell Stem Cell 2009;5:584–595.

40. Teo AK, Wagers AJ, Kulkarni RN. New opportunities: Harnessing induced pluripotency for discovery in diabetes and metabolism. Cell Metab 2013;18:775–791.

41. Ernst J, Kheradpour P, Mikkelsen TS et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 2011;473:43–49.

42. Kells M, Wold B, Snyder MP et al. Defining functional DNA elements in the human genome. Proc Natl Acad Sci USA 2014;111:6131–6138.

43. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell 2014;157:1262–1278.

44. Gupta RM, Musunuru K. Mapping novel pathways in cardiovascular disease using eQTL data: The past, present, and future of gene expression analysis. Front Genet 2012;3:232.

45. Saha K, Hurlbut JB. Disease modeling using pluripotent stem cells: Making sense of disease from bench to bedside. Swiss Med Wkly 2011;141:w13144.