Stem cells, genome editing and the path to translational medicine

Frank Soldner¹ and Rudolf Jaenisch¹,²,³
¹The Whitehead Institute, 455 Main Street, Cambridge, MA 02142, USA
²Department of Biology, Massachusetts Institute of Technology, 31 Ames Street, Cambridge, MA 02139, USA

Summary

The derivation of human embryonic stem cells (hESCs) and the stunning discovery that somatic cells can be reprogrammed into human induced pluripotent stem cells (hiPSCs) holds the promise to revolutionize biomedical research and regenerative medicine. In this review, we focus on disorders of the central nervous system and explore how advances in human pluripotent stem cells (hPSCs) coincide with evolutions in genome engineering and genomic technologies to provide realistic opportunities to tackle some of the most devastating complex disorders.

Introduction

The first report of human pluripotent stem cells (hPSCs) remains a key discovery with the promise to fundamentally reshape regenerative medicine and the study of complex human diseases (Thomson, 1998). These cells termed human embryonic stem cells (hESCs) were derived directly from the human pre-implantation embryo and are generally considered the functional in vitro equivalent to the pluripotent cells of the blastocyst, thought to be the founder cells of the embryo proper (Reubinoff et al., 2000; Thomson, 1998). Once removed from the blastocyst, hESCs can be maintained in culture for an extended period and, comparable to their in vivo counterpart, have the developmental potential to differentiate into any somatic cell type. The derivation of hESCs has captured the imagination of biomedical researchers and the public likewise based on the promise to provide an essentially unlimited supply of human somatic cells for basic research and regenerative medicine.

In vitro differentiation of hESCs has already revolutionized our ability to study early human development and cell-type specification in a cell culture dish and provides suitable cells for cell replacement therapies. However, broad application of hESCs remains challenging due to technical difficulties like immune rejection after transplantation of non-autologous cells and...
ethical concerns associated with the use of human embryos for research. The discovery that transient expression of a few transcription factors (Oct4, Sox2, KLF4 and c-Myc) is sufficient to reprogram somatic cells to a pluripotent state (Maherali et al., 2007; Takahashi and Yamanaka, 2006; Wernig et al., 2007) allows the derivation of human induced pluripotent stem cells (hiPSCs) (Takahashi et al., 2007; Yu et al., 2007) and resolves many limitations associated with hESCs. The basic biological properties of hiPSCs are highly similar to hESCs, opening up the exciting opportunity for cell transplantation therapies using patient-derived autologous cells that should evade immune rejections. Clinical trials are already underway for a number of diseases, including macular degeneration (Mandai et al., 2017; Mehat et al., 2018; Schwartz et al., 2015; Song et al., 2015), ischemic heart disease (Menasché et al., 2018), diabetes (Viacyte; https://clinicaltrials.gov: NCT03163511) and spinal cord injury (Asterias Biotherapeutics; https://clinicaltrials.gov: NCT02302157). Despite these promising initial trials, the routine and safe application of cell replacement therapies for a broader spectrum of degenerative diseases is likely years away (Trounson and DeWitt, 2016).

Beside the enthusiasm for cell replacement therapies, hiPSC technology also provides the unique opportunity to establish cellular models of human diseases. Studying human diseases has been limited by the lack of relevant model systems that combine known genetic elements with disease-associated phenotypic readouts, especially for common medical conditions with no well-defined genetic etiology or Mendelian inheritance patterns, such as obesity, heart disease, diabetes and sporadic neurodegenerative disorders. The complex etiology of sporadic diseases, thought to result from the interaction between genetic and non-genetic (lifestyle and environmental) risk factors, has impeded understanding the molecular mechanisms of complex diseases (Lander, 2011; McClellan and King, 2010). Following in vitro differentiation, patient-derived hiPSCs provide somatic cells, which carry all the genetic elements implicated in disease development, allows the study of underlying genetic aberrations in human diseases-relevant cell types (Shi et al., 2016; Soldner and Jaenisch, 2012).

The development of hPSCs technology coincides with recent advances in genomic technologies and genome engineering. Genome-scale genetic and epigenetic information have provided valuable insights into the genetic basis of complex diseases and identified genetic and epigenetic variations associated with many human disorders. Unfortunately, most risk-associated variants have no established biological function or clinical relevance (McClellan and King, 2010). Therefore, translating these associations to mechanistic insights into disease development and progression remains a fundamental challenge. The remarkable progress of genome editing technologies, in particular the simplicity and versatility of the CRISPR/Cas system (Komor et al., 2017), has vastly increased our ability to modify the genome in human cells and enables us to begin to systematically dissect the functional effect of genetic variants.

In this review, we will summarize recent progress, limitations and potential solutions in using hPSCs technology to model complex human diseases focusing on neurological diseases, which pose unique challenges. We will highlight how integrating hPSC technology with CRISPR/Cas genome editing and genome-scale genetic and epigenetic information can
systematically dissect the function of disease-associated risk alleles to provide a genetically controlled system to study sporadic diseases in culture. We will discuss the unique challenges to capture the full complexity of brain disorders including differentiation, maturation and aging of neuronal cell types and the analysis of non-cell autonomous interactions in complex organoids in vitro and interspecies chimeric in vivo systems. We will end by reviewing strategies to implement such systems for phenotypic genetic and compound screens for drug development.

**HPSC-based models for monogenetic diseases**

The identification of genes linked to familial forms of diseases such as cystic fibrosis, sickle cell anemia and diabetes fundamentally changed our understanding of many diseases by revealing insights into the molecular and cellular pathogenesis (Lander, 2011; McClellan and King, 2010). Detailed knowledge of disease-causing mutations allows us to develop effective model systems that combine known genetic elements with disease-associated phenotypic readouts. In the past, typical experimental systems were either based in vitro on primary human cells derived from patients with known mutations or in vivo using genetically engineered animal model systems. Unfortunately, studying diseases of the nervous system using primary cells has been challenging due to limited access to live brain tissue and the inability to reliably maintain and expand human neuronal cells such as post-mitotic neurons in culture. In contrast, genetically engineered animal models have provided remarkable insight into the development and function of the brain and the pathophysiology of neurological diseases; however, currently available rodent models for neurodevelopmental or neurodegenerative diseases recapitulate some but not all of the relevant features of the human disease (Bolker, 2017; Jucker, 2010), one of the major reasons why many therapeutic approaches show promise in preclinical disease models but fail in subsequent clinical trials.

hPSC technology now empowers the study of human brain physiology and neurological disease development in culture by providing a nearly unlimited access to all somatic cell types of the brain (Shi et al., 2016; Soldner and Jaenisch, 2012). The approach entails the derivation of hPSCs with known disease-associated mutations followed by in vitro differentiation into disease-relevant cell types. Differentiation into all major cell types of the CNS including a wide variety of specific neuronal subtypes, such as astrocytes, oligodendrocytes and more recently microglial cells (Muffat et al., 2016), can be induced with great precision (reviewed in (Tao and Zhang, 2016)). Using such in vitro generated cells in disease modeling may uncover alterations in molecular and cellular phenotypes associated with a specific disease, which can subsequently reveal detailed mechanistic insights and guide the development of novel therapeutics.

Initial hPSC models used hESC lines derived from blastocysts that carried mutations associated with monogenetic diseases or chromosomal abnormalities (Ilic and Ogilvie, 2016). Routine use of preimplantation genetic diagnosis (PGD), a procedure to detect genetic defects in embryos prior to transfer in utero, provides access to donated embryos with congenital defects, which would be otherwise discarded. To date, more than 100 hESC lines have been derived from genotyped embryos (Ilic and Ogilvie, 2016), including cell lines carrying mutations implicated in neurological disorders like spinal muscular atrophy,
spinocerebellar ataxia, fragile X syndrome and Huntington disease. However, hESC based models remain challenging given the ethical and regulatory issues.

HiPSCs technology has resolved many of the ethical and technical issues in disease modeling and has rapidly evolved over the last decade. Following in vitro differentiation, patient-derived hiPSCs provide disease-relevant somatic cells, which carry all the genetic elements implicated in the disease reflecting the genetic spectrum of the patient population. While discussing all of these studies is beyond the scope of this work, excellent recent reviews summarize hiPSC diseases models for neurodevelopmental (Ardhanareeswaran et al., 2017), neuropsychiatric (Hoffman et al., 2018) and neurodegenerative diseases including Alzheimer’s disease (Sullivan and Young-Pearse, 2017) and Parkinson’s disease (Shi et al., 2016; Zhang et al., 2017), motor neuron and Huntington’s disease. The spectrum of phenotypes, which can be assessed in hiPSC-based in vitro models include a wide range of molecular, metabolic, cellular and electrophysiological analysis. Such in vitro models may allow studying even complex disease-relevant phenotypes, such as pathologies associated processes in embryonic development, cellular maturation, non-cell autonomous effects between diverse neuronal cell populations, synaptic transmission and neuronal network function. HiPSC disease modeling has confirmed disease pathologies gained in other model systems and provided remarkable, novel molecular and cellular insights into many disorders.

Importantly, hiPSC-based drug discovery has identified novel compounds that can reverse disease-associated phenotypes in vitro and are currently under clinical investigation to treat diseases such as amyotrophic lateral sclerosis (McNeish et al., 2015), tauopathies in Alzheimer’s disease, progressive supranuclear palsy (Bright et al., 2015) and spinal muscular atrophy (Naryshkin et al., 2014). Two important observations emerge from this work. Robust and disease-relevant phenotypes in cell culture often occur (i) in neurodevelopmental or early age of onset disorders and (ii) in monogenetic disorders with highly penetrant mutations and severe, rapidly progressive well-defined disease-associated pathology and clinical manifestation. However, many of the most prevalent and devastating neurological disorders such as Alzheimer’s and Parkinson’s disease have no well-defined genetic etiology and manifest only late in life with slow progression over many years, a potentially serious complication for studying the complexity of late age of onset neurodegenerative disorders.

It is generally recognized that individual hiPSCs display a high degree of variability in their biological properties such as the propensity to differentiate into functional cell types, which is independent of genotype or disease status (Liang and Zhang, 2013; Soldner and Jaenisch, 2012). This critically limits the ability to identify robust disease-associated phenotypes by comparing patient-derived cells with unrelated controls (Figure 1). Such system-innate variability proves challenging for age-related disorders, as the disease-associated phenotypes typically progress slowly over many years. Due to the rather limited time in culture, in vitro phenotypes are expected to be rather mild and subtle. Reprogramming-associated genetic and epigenetic alterations, the lack of robust in vitro differentiation protocols and genetic background variations constitute the main sources of phenotypic cell to cell variation (Liang and Zhang, 2013; Soldner and Jaenisch, 2012) and remain a major limitation of the current hiPSC approach.
Genetic engineering of disease models

Even mutations that cause highly penetrant human diseases are influenced by genetic background and show significant phenotypic and clinical variation like disease penetrance or age of onset. Using concepts from genetically engineered mouse models, genetic engineering of hPSCs should provide an elegant solution to study the effect of disease-associated mutations in a genetically controlled context. However, classical gene targeting approaches using homologous recombination were surprisingly inefficient in human cells. Advances in genome editing technologies using engineered site-specific nucleases, such as mecanucleases, zinc finger nucleases (ZFNs) (Hockemeyer et al., 2009; Urnov et al., 2005), transcription activator-like effector-based nucleases (TALEN) (Boch et al., 2009; Hockemeyer et al., 2011) and the CRISPR/Cas9 system (Cong et al., 2013; Jinek et al., 2012; 2013; Mali et al., 2013), have vastly increased our ability to modify the genome with unprecedented flexibility (Komor et al., 2017). In particular, the simplicity and versatility of the CRISPR/Cas9 system to efficiently modify the genome in human cells, even at multiple loci simultaneously, allows the engineering of genetically controlled hPSC based models, which differ only at known genetic disease-causing variants.

Testing specific disease-associated variants in an otherwise identical isogenic context should detect even subtle molecular or cellular phenotypes masked by variability in genetic background (Figure 1A). As a proof of principle, we used ZFNs to either seamlessly correct Parkinson’s disease-associated mutations in the SNCA gene in patient-derived hiPSCs or to insert similar variants into wild-type hESCs (Soldner et al., 2011). These isogenic hPSC lines provided an experimental system with a controlled genetic background (Figure 1A), which allowed us to identify mitochondrial dysfunction and nitrosative and oxidative stress as molecular events in the pathogenesis of Parkinson’s disease (Chung et al., 2013; Ryan et al., 2013). Engineering of isogenic hPSCs by either correcting the disease-associated mutation in patient-derived hiPSCs or insertion into wild-type hPSCs has become standard to distinguish disease-associated effects from background variation. Examples include Parkinson’s and Alzheimer’s disease (Paquet et al., 2016; Reinhardt et al., 2013; Schöndorf et al., 2014), hereditary motor and sensory neuropathy (Murakami et al., 2017), frontotemporal lobar degeneration (Imamura et al., 2016), Huntington disease (Xu et al., 2017), tetrahydrobiopterin metabolism disorder (Ishikawa et al., 2016), fragile X syndrome (Xie et al., 2016) and the functional analysis of neurorexin-1 in neuropsychiatric diseases (Pak et al., 2015). Limitations to this approach are off-target cleavage of site-specific nucleases and undesired genetic alterations in cell survival pathways such as P53 mutations as a consequence of genome engineering associated genotoxic stress and isolating single cell clones (Haapaniemi et al., 2018; Ihry et al., 2018; Merkle et al., 2017). Also, effects of genetic modifiers in a given genetic background may mask or alter relevant disease-associate phenotypes.

HPSC models of complex diseases – genetic considerations

Monogenetic familial diseases are rare in neurological disorders. Most cases are sporadic with no well-defined genetic etiology and inheritance patterns. Epidemiology and population genetics suggest that sporadic diseases result from complex interactions between genetic and non-genetic factors. Even mutations that cause highly penetrant human diseases are influenced by genetic background and show significant phenotypic and clinical variation like disease penetrance or age of onset. Using concepts from genetically engineered mouse models, genetic engineering of hPSCs should provide an elegant solution to study the effect of disease-associated mutations in a genetically controlled context. However, classical gene targeting approaches using homologous recombination were surprisingly inefficient in human cells. Advances in genome editing technologies using engineered site-specific nucleases, such as mecanucleases, zinc finger nucleases (ZFNs) (Hockemeyer et al., 2009; Urnov et al., 2005), transcription activator-like effector-based nucleases (TALEN) (Boch et al., 2009; Hockemeyer et al., 2011) and the CRISPR/Cas9 system (Cong et al., 2013; Jinek et al., 2012; 2013; Mali et al., 2013), have vastly increased our ability to modify the genome with unprecedented flexibility (Komor et al., 2017). In particular, the simplicity and versatility of the CRISPR/Cas9 system to efficiently modify the genome in human cells, even at multiple loci simultaneously, allows the engineering of genetically controlled hPSC based models, which differ only at known genetic disease-causing variants.

Testing specific disease-associated variants in an otherwise identical isogenic context should detect even subtle molecular or cellular phenotypes masked by variability in genetic background (Figure 1A). As a proof of principle, we used ZFNs to either seamlessly correct Parkinson’s disease-associated mutations in the SNCA gene in patient-derived hiPSCs or to insert similar variants into wild-type hESCs (Soldner et al., 2011). These isogenic hPSC lines provided an experimental system with a controlled genetic background (Figure 1A), which allowed us to identify mitochondrial dysfunction and nitrosative and oxidative stress as molecular events in the pathogenesis of Parkinson’s disease (Chung et al., 2013; Ryan et al., 2013). Engineering of isogenic hPSCs by either correcting the disease-associated mutation in patient-derived hiPSCs or insertion into wild-type hPSCs has become standard to distinguish disease-associated effects from background variation. Examples include Parkinson’s and Alzheimer’s disease (Paquet et al., 2016; Reinhardt et al., 2013; Schöndorf et al., 2014), hereditary motor and sensory neuropathy (Murakami et al., 2017), frontotemporal lobar degeneration (Imamura et al., 2016), Huntington disease (Xu et al., 2017), tetrahydrobiopterin metabolism disorder (Ishikawa et al., 2016), fragile X syndrome (Xie et al., 2016) and the functional analysis of neurorexin-1 in neuropsychiatric diseases (Pak et al., 2015). Limitations to this approach are off-target cleavage of site-specific nucleases and undesired genetic alterations in cell survival pathways such as P53 mutations as a consequence of genome engineering associated genotoxic stress and isolating single cell clones (Haapaniemi et al., 2018; Ihry et al., 2018; Merkle et al., 2017). Also, effects of genetic modifiers in a given genetic background may mask or alter relevant disease-associate phenotypes.

HPSC models of complex diseases – genetic considerations

Monogenetic familial diseases are rare in neurological disorders. Most cases are sporadic with no well-defined genetic etiology and inheritance patterns. Epidemiology and population genetics suggest that sporadic diseases result from complex interactions between genetic and non-genetic factors. Even mutations that cause highly penetrant human diseases are influenced by genetic background and show significant phenotypic and clinical variation like disease penetrance or age of onset. Using concepts from genetically engineered mouse models, genetic engineering of hPSCs should provide an elegant solution to study the effect of disease-associated mutations in a genetically controlled context. However, classical gene targeting approaches using homologous recombination were surprisingly inefficient in human cells. Advances in genome editing technologies using engineered site-specific nucleases, such as mecanucleases, zinc finger nucleases (ZFNs) (Hockemeyer et al., 2009; Urnov et al., 2005), transcription activator-like effector-based nucleases (TALEN) (Boch et al., 2009; Hockemeyer et al., 2011) and the CRISPR/Cas9 system (Cong et al., 2013; Jinek et al., 2012; 2013; Mali et al., 2013), have vastly increased our ability to modify the genome with unprecedented flexibility (Komor et al., 2017). In particular, the simplicity and versatility of the CRISPR/Cas9 system to efficiently modify the genome in human cells, even at multiple loci simultaneously, allows the engineering of genetically controlled hPSC based models, which differ only at known genetic disease-causing variants.

Testing specific disease-associated variants in an otherwise identical isogenic context should detect even subtle molecular or cellular phenotypes masked by variability in genetic background (Figure 1A). As a proof of principle, we used ZFNs to either seamlessly correct Parkinson’s disease-associated mutations in the SNCA gene in patient-derived hiPSCs or to insert similar variants into wild-type hESCs (Soldner et al., 2011). These isogenic hPSC lines provided an experimental system with a controlled genetic background (Figure 1A), which allowed us to identify mitochondrial dysfunction and nitrosative and oxidative stress as molecular events in the pathogenesis of Parkinson’s disease (Chung et al., 2013; Ryan et al., 2013). Engineering of isogenic hPSCs by either correcting the disease-associated mutation in patient-derived hiPSCs or insertion into wild-type hPSCs has become standard to distinguish disease-associated effects from background variation. Examples include Parkinson’s and Alzheimer’s disease (Paquet et al., 2016; Reinhardt et al., 2013; Schöndorf et al., 2014), hereditary motor and sensory neuropathy (Murakami et al., 2017), frontotemporal lobar degeneration (Imamura et al., 2016), Huntington disease (Xu et al., 2017), tetrahydrobiopterin metabolism disorder (Ishikawa et al., 2016), fragile X syndrome (Xie et al., 2016) and the functional analysis of neurorexin-1 in neuropsychiatric diseases (Pak et al., 2015). Limitations to this approach are off-target cleavage of site-specific nucleases and undesired genetic alterations in cell survival pathways such as P53 mutations as a consequence of genome engineering associated genotoxic stress and isolating single cell clones (Haapaniemi et al., 2018; Ihry et al., 2018; Merkle et al., 2017). Also, effects of genetic modifiers in a given genetic background may mask or alter relevant disease-associate phenotypes.
non-genetic risk factors including age, lifestyle, and environment (Lander, 2011; McClellan and King, 2010). This multi-faceted complexity of sporadic diseases complicates generating genetically defined disease-relevant cellular and animal models impeding our understanding of the underlying molecular mechanisms. While rare monogenetic forms may provide mechanistic insights into mechanisms, common pathological or clinical features between familial and sporadic forms may be a poor predictor for common underlying causes because of the highly variable genetic and phenotypic spectrum of sporadic diseases.

The evolution of sequencing and genotyping technologies allows dissecting the genetic basis of sporadic forms of common diseases in a systematic and cost effective manner (Lander, 2011). Genome-scale sequencing and large-scale genome-wide association studies (GWASs) have identified sequence variants such as single nucleotide polymorphisms (SNPs), deletions and insertions associated with a wide variety of neurological diseases including Parkinson’s disease (Chang et al., 2017), Alzheimer’s disease (Lambert et al., 2013), schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), bipolar disorders (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013) and autism spectrum disorders (Gaugler et al., 2014). These studies suggest a fluent spectrum of genetic risk with disease reflecting the individual genetic architecture (Pihlstrøm et al., 2017). Prevailing models assume that disease susceptibility arises through a combination of multiple rare or common risk variants rather than determined by strict Mendelian inheritance (Lander, 2011; McClellan and King, 2010). While rare genetic alterations and de-novo mutations with large effects usually occur predominantly in neurodevelopmental and psychiatric diseases, common alleles with low effect size may be more significant for neurodegenerative disorders.

Risk variants pinpoint genomic loci and genes that may play key roles in disease pathogenesis, but functional interpretation of GWAS datasets remains challenging. Sequence variants merely are statistical associations with disease risk and have no established biological function (Lander, 2011; McClellan and King, 2010). Most GWAS loci contain numerous disease-associated risk variants in linkage disequilibrium (LD) spanning a wide genomic area with multiple target genes complicating the identification of relevant disease genes. Further, the majority of disease-associated genetic variants reside in the non-coding part of the genome, which largely prevents any functional analysis in non-human cellular or animal model systems given the limited sequence conservation of non-coding elements. Advancing from disease association to mechanistic insights into disease development and progression, however, is central to our understanding of complex diseases (Ward and Kellis, 2012).

**HPSC models of complex diseases - conventional approach**

hiPSCs have fundamentally changed our ability to investigate the pathogenesis of complex diseases. Without detailed knowledge of the underlying genetics, patient-derived hiPSCs after in vitro differentiation provide a nearly unlimited supply of disease-relevant somatic cells for in vitro modelling that carry all the genetic elements implicated in disease development. Comparing neurons derived from patients with schizophrenia or autism spectrum disorders to age-matched unaffected individuals (Figure 1B) identified disease-
relevant phenotypic changes providing proof-of concept to modeling sporadic diseases (reviewed in Ardhanareswaran et al., 2017; Hoffman et al., 2018). While initially surprising, considering the patients’ genetic heterogeneity, such robust disease-associated phenotypes may result from diverse genetic variants that converge on common developmental pathways such as cortical development, synaptic function and epigenetic processes (Hoffman et al., 2018). To reduce variability and facilitate the identification of disease-associated phenotypes, large-scale hiPSC consortia have significantly increased the number of independent samples (HD iPSC Consortium, 2012; Soares et al., 2014).

Stringent patient selection based on genetic or clinical features may reduce heterogeneity in the patient population and significantly increase the power of hiPSC based studies (Figure 1B) (Hoekstra et al., 2017). One approach uses a polygenic risk score, that summarizes the number of GWAS identified risk alleles and the respective effect size for genetically informed patient stratification (Lewis and Vassos, 2017). Stratification of patients with schizophrenia by polygenic risk scores (Hoffman et al., 2018) or patients with autism spectrum disorders by common clinical features such as macrocephaly or microcephaly have facilitated the identification of robust diseases- associated molecular and cellular phenotypes (Marchetto et al., 2015; Mariani et al., 2017). A similar approach using genetic risk information has been successfully applied to neurodegenerative diseases to stratify patients, which carry rare, high-risk mutations in GBA (glucosylceramidase beta) associated with increased risk to develop Parkinson’s disease (Schöndorf et al., 2014) or specific APOE (Huang et al., 2017) or SORL1 (sortilin-related receptor, L(DLR class) A repeats containing) alleles (Young et al., 2015) associated with increased risk to develop Alzheimer’s disease.

**HPSC models of complex diseases - population genetics approach**

Common disease-associated risk alleles for late age of onset disorders typically have a small effect size, which often precludes observations in cellular models. One common strategy to dissect risk loci is to correlate disease-associated sequence variations with phenotypic or molecular quantitative traits. A quantitative trait locus (QTL) is a specific region in the genome where a particular sequence variant correlates with the variation in a quantitative trait. Although gene expression (eQTLs) is the most systematically analyzed trait (GTEx Consortium, 2015), a wide spectrum of molecular features can identify QTLs, such as the expression of microRNAs, noncoding RNAs, alternative splicing, protein expression, the abundance of metabolic products (mQTLs) or epigenetic modifications including chromatin accessibility, histone modifications, DNA methylation (meQTLs) and transcription factor binding (Ward and Kellis, 2012).

The promise of genome-scale eQTL mapping to gain insight into the functional effect of common risk variants initiated collaborative efforts to provide high-quality transcriptome datasets from genotyped individuals. The main databases that include brain samples are the Genotype-Tissue Expression (GTEx) project (GTEx Consortium, 2015), Braincloud (Colantuoni et al., 2011), CommonMind (Senthil et al., 2017), the PsychENCODE project (PsychENCODE Consortium, 2015) and the UK brain expression consortium (UKBEC; http://www.braineac.org). Although cis-acting effects of sequence variants on gene
expression have been identified as major factors for phenotypic variation and disease susceptibility (GTEx Consortium, 2015; Lee and Young, 2013), this approach has many limitations in the CNS, in particular access to large numbers of high quality brain samples for sufficient statistical power to reliably identify eQTLs with small effect size. Cis-regulatory effects on gene expression are extremely cell-type specific, so resolving cellular heterogeneity within brain tissue (e.g. variable contribution from neurons, astrocytes, oligodendrocytes, microglia, epithelial cells and blood cells) represents a significant challenge to eQTL mapping. It remains to be seen if emerging single cell technologies, such as single nuclei FACS sorting from post mortem tissue and single-nuclei RNA-sequencing (snRNA-seq), will resolve some limitations (Habib et al., 2017).

An alternative approach to resolve cellular heterogeneity complements brain QTL analysis with in vitro hiPSC-derived datasets. Many collaborative initiatives established large hiPSC cell repositories from healthy subjects and patients with a wide spectrum of monogenic and complex diseases (Soares et al., 2014). A consortium supported by the National Heart, Lung, and Blood Institute’s Next Generation Association Studies (NextGen) and the Human Induced Pluripotent Stem Cell Initiative (HipSci) tested an idea from hiPSC-based QTL association studies in vitro (Figure 1B) (Warren et al., 2017a). Initial “population genetics in a dish” analysis combined more than 2000 hiPSCs and investigated sources of variation of gene expression in hiPSCs. Genetic background was the main source of variability allowing identification of numerous genetic variants that contribute to variation and cell-line specific regulation of transcription (DeBoever et al., 2017; Kilpinen et al., 2017). However, comparing transcriptional variation within and across individuals indicated a nearly equal amount of non-genetic variability, partly driven by Polycomb repression complex target genes, which may reflect reprogramming-associated variability (Carcamo-Orive et al., 2017).

The subsequent analysis of eQTLs and chromatin accessibility in hiPSC-derived sensory neurons (Schwartzentruber et al., 2017), and eQTLs and metabolic alterations (mQTLs) in hiPSC-derived hepatocyte-like cells (Pashos et al., 2017; Warren et al., 2017b) showed a significant overlap between primary tissues and in vitro identified QTLs that may complement in vivo datasets with novel associations not observed in primary tissues or patients. The unique advantage of this approach is that in vitro differentiated somatic cells not only serve as discovery platform for QTLs but at the same time provide an experimental system to perform functional cellular genomic experiments. Pashos and colleagues used CRISPR/Cas genome editing to confirm a functional effect of novel risk variants on gene expression (Pashos et al., 2017). Comparing in vitro derived sensory neurons with primary dorsal root ganglia however reveals experimental variation of in vitro derived cells as major limitation of this approach. Given that detecting effects of regulatory variants with moderately large effect sizes requires at least 20 to 80 hiPSC lines, this approach depends on simple and robust protocols to differentiate large cohorts of hPSCs (Schwartzentruber et al., 2017). It remains to be seen whether in vitro association studies can reliably detect novel associations not observed in primary tissues or patients.
HPSC models of complex diseases - Functional analysis of disease-associated risk variants

Correlating genomic variants with quantitative traits such as gene expression can isolate disease-relevant genes and pathways but does not identify causal risk variants or provide molecular insights into how sequence variation contributes to the pathogenesis of complex diseases (Ward and Kellis, 2012). Compelling evidence suggests that disease-association of non-coding risk variants is driven by sequence-specific alterations in the function of non-coding regulatory elements that affects transcription and disease susceptibility. Large-scale collaborative initiatives, such as the ENCODE (ENCODE Project Consortium, 2012), the PsychENCODE (PsychENCODE Consortium, 2015) and Roadmap Epigenomics project (Roadmap Epigenomics Consortium, 2015) are accelerating discovery of functional non-coding elements. These resources provide a comprehensive catalogue for cell and tissue-type specific regulatory sequences including the human brain.

Consistent with gene-regulatory variants impacting disease risk, GWAS disease associations are enriched in open chromatin regions with enhancer-associated histone modifications in tissues and cell-types affected by the respective disease. (Ernst et al., 2011; Hnisz et al., 2013; Maurano et al., 2012). Distal enhancer elements bound by transcription factors control gene expression in a tissue and cell-type specific manner and sequence-specific changes in transcription factor-binding in enhancers correlate with changes in chromatin state and cis-regulated gene expression (Karnuta and Scacheri, 2018; Lee and Young, 2013; Wamstad et al., 2014). This suggests a simple molecular mechanism of how individual sequence variants modify transcription factor binding, contribute to alterations in transcription and affect disease susceptibility.

The integration of epigenetic signatures with GWAS datasets may facilitate the discovery of functional disease-associated risk variants in regulatory elements (Figure 2). Mapping disease-associated SNPs to transcription factor binding sites that regulate tissue-specific enhancer function and gene expression identified causal risk variants in metabolic (Claussnitzer et al., 2015) and cardiac diseases (Wang et al., 2016). To refine prioritizing plausible functional variants, current efforts focus on establishing a rigorous experimental and statistical framework integrating novel and more complex epigenetic data (Schaid et al., 2018; Ward and Kellis, 2012). Such datasets include the 3-dimensional organization of the genome to map interactions between enhancers and target genes (Hnisz et al., 2016), allele-specific epigenetic and gene expression information to facilitate the identification of cis-regulatory effects of sequence variants and mapping of RNA binding proteins to identify regulators of alternative splicing. Again, cellular heterogeneity and limited access to brain tissues restrict generation of high-quality datasets with sufficient sequencing coverage. One current strategy complements existing patient tissue-derived datasets with hiPSC-derived somatic cell populations, preferentially from the same donor with identical genetic background (Figure 2).

To advance from association to causal variant, we recently applied this approach to prioritize risk variants based on the epigenetic signature to sporadic Parkinson’s disease. After compiling a candidate list of probable causal risk variants in active brain enhancers by
integrating epigenetic and GWAS datasets, we employed CRISPR/Cas genome editing to
dissect the function of diseases-associated risk variants (Figure 2). We identified the
sequence-specific binding of transcription factors in a non-coding risk variant in an intronic
enhancer element to regulate the expression of SNCA, a key gene implicated in the
pathogenesis of Parkinson’s disease (Soldner et al., 2016). This proof-of-concept illustrates
that the integration of population genetic and genome-scale epigenetic data combined with
hiPSC and genome editing technologies provide a platform to study complex disorders in a
genetically controlled and systematic manner. A similar approach was used to dissect the
genetic risk associated with 5 vascular diseases, including coronary artery disease, migraine,
cervical artery dissection, fibro- muscular dysplasia and hypertension and identified a non-
coding risk variant in an intronic enhancer of PHCTR1 to regulate the expression of
endothelin 1, a gene with physiological function in vasculature (Gupta et al., 2017).

Current efforts focus on advancing from the “one risk variant at a time” approach to
systematically evaluate risk variants on a genome scale. Targeted engineering of existing and
newly identified species of Cas proteins has substantially broadened the versatility of the
CRISPR/Cas toolbox to now target almost any genomic base (Komor et al., 2017). Several
studies performed pooled genetic screens using tiling libraries of gRNAs to target and
disrupt the function of thousands of endogenous non-coding elements for functional
interrogation (reviewed in (Klann et al., 2018). Alternatively, targeting regulatory sequences
with a nuclease deficient Cas9 protein (dCas9) fused to transcriptional activators
(CRISPRa), suppressors (CRISPRi) (Gilbert et al., 2014) or active domains of epigenetic
modifiers such as DNA methylation/demethylation or histone acetylation can change the
activity of non-coding regulatory elements (Amabile et al., 2016; Kearns et al., 2015; Liu et
al., 2016). An example of a potential candidate for therapeutic application of epigenetic
editing is fragile X syndrome, which is caused by a GGC expansion leading to the silencing
of FMR1. Targeting the repeat by a dCas9-Tet fusion protein caused demethylation of the
repeat, restoration of FMR1 expression and reversal of electrophysiological abnormalities in
fragile X patient-derived neurons (Liu et al., 2018). Pooled epigenome editing screens can
systematically analyze the functional consequences of modulating the activity of non-coding
elements and interrogate the function of GWAS identified risk variants in regulatory
elements on a genome scale (Fulco et al., 2016; Hilton et al., 2015; Thakore et al., 2015).
However, these experiments used tumor cell lines and examined large biological effects that
enrich functional gRNAs. Whether small effects on gene expression as a molecular
mechanism suggested for disease susceptibility are sufficient to identify functional gRNA
targets and translate such approaches to human post- mitotic neurons remains an open
question. A promising novel approach termed Pertrub-seq may combine single cell
sequencing with CRISPR/Cas mediated modulation/disruption of non-coding regulatory
elements to dissect the molecular effects on gene expression in individual cells (Adamson et
al., 2016; Dixit et al., 2016; Jaitin et al., 2016).

**HPSC models of complex phenotypes**

A major limitation of hiPSC approaches is the lack of robust differentiation protocols. The
traditional protocols for *in vitro* differentiation recapitulate the order and timing of regional
specification and neuronal commitment during embryonic development by supplying
developmental cues like morphogens or modulating developmentally relevant pathways using small molecules (Keller, 1995). This approach can direct cell fates with remarkable precision to all major cell types of the nervous system including various distinct neuronal sub-types resembling dopaminergic, glutamatergic, GABAergic, motor, hypothalamic and striatal medium spiny neurons (Figure 3) (Tao and Zhang, 2016). However, differentiation outcomes may suffer from substantial clone-to-clone variability, adding a significant source of variability to cellular disease models (Soldner and Jaenisch, 2012). In addition, any disease-related genetic variant could also affect the differentiation process adding another layer of complexity to the interpretation of phenotypic variation between disease and control cultures.

One strategy to robustly control in vitro differentiation is directed conversion of hPSCs to pure and functionally mature neurons driven by forced expression of cell fate specifying neuronal transcription factors such as NGN2 to induce cortical excitatory neurons (iNs) (Zhang et al., 2013) or NGN2, ISL1, LHX3, NEUROD1, BRN2, ASCL1 and MYT1L to induce motor neurons (iMNs) (Shi et al., 2018). This approach allowed dissection of the molecular effects of distinct ApoE isoforms, a strong risk factor for Alzheimer’s disease, amyloid-beta precursor protein (APP) expression and disease susceptibility (Huang et al., 2017). An alternate strategy selects specific cell types within a heterogeneous population by genetically inserting cell-type specific selection or fluorescent marker proteins. Using longitudinal tracking of the fluorescent motor neuron specific reporter Hb9-RFP in hiPSCs derived from patients with GGGGCC repeat expansion in C9ORF72, which underlies amyotrophic lateral sclerosis and frontotemporal dementia, Shi and colleagues observed a robust degeneration of patient-derived cells and identified disruption of vesicle trafficking as the pathogenic molecular mechanism (Shi et al., 2018).

**HPSC models of complex phenotypes - Programming a physiological niche**

Many in vitro differentiation protocols are thought to produce rather immature cells with fetal-stage like properties independent of the age of the initial donor (Studer et al., 2015). The failure to generate mature functional cells in vitro may result from the lack of a physiological environment to direct proper maturation. Using concepts from reprogramming of fibroblasts to hiPSCs or direct lineage conversion of fibroblast to induced neurons, it may be possible to convert cell identity from an immature to a more mature state by recreating the transcriptional program of mature cells. Access to transcriptional and epigenomic information from in vitro differentiated neuronal cell types and the in vivo counterparts derived from mature primary tissue will permit mapping epigenetic and transcriptional differences between immature and mature cell states. These differences may identify key transcriptional regulators that may circumvent the requirement for an appropriate exogenous niche and program the conversion of an immature to a mature cell identity.

Since aging is an important risk factor for many neurodegenerative diseases, current efforts attempt to mimic age-related processes in cell culture. In vitro ectopic overexpression of Progerin, a truncated form of the mutated Lamin A gene associated with the premature aging syndrome Hutchinson-Gilford Progeria, in hPSC-derived neurons induce aging associated changes, such as genomic instability, epigenetic dysregulation and mitochondrial...
dysfunction. Overexpression of Progerin was sufficient to accelerate the appearance of disease-associated phenotypes in a hiPSC model of Parkinson’s disease (Miller et al., 2013). It remains to be seen whether this approach can be applied more broadly to age-related diseases and whether phenotypes observed in a premature aging context correlate with disease pathologies.

**HPSC models of complex phenotypes - neuronal co-culture systems**

Because neuronal cell loss is central to many neurological disorders, much current research focuses mainly on generating homogenous neuronal cultures. A pure cell population remains the gold standard to study cell-autonomous molecular and cellular phenotypes due to reduced confounding effects from cellular heterogeneity. However, non-neuronal cell types, including astrocytes and microglia, and non-cell autonomous interactions among these cells are essential for brain development and function, the proper maturation of in vitro cultured neuronal cells and play a key role in the pathogenesis of neurological diseases (Clarke and Barres, 2013; Salter and Stevens, 2017). Different strategies exist to recreate a physiological in vivo niche to allow for adequate differentiation, aging and interaction between in vitro derived cell types (Figure 3).

Initial hPSC-based disease modeling studies, using simple co-culture systems between astrocytes and neurons, successfully dissected the effects of SOD1 mutations in glial cells on motor neuron survival in amyotrophic lateral sclerosis (Di Giorgio et al., 2008). These advances spurred efforts to produce more sophisticated in vitro systems by combining hPSC technology with basic engineering principles using novel synthetic biomaterials, hydrogels, microfluidics and 3D printing with the goal to recreate some of the complexity of the human brain. These approaches allow tight control of physiological parameters such as cell-type composition, liquid and oxygen exchange or nutrients and growth factor delivery (Figure 3). Now, novel microdevices, commonly referred to “organs-on-a-chip”, are being developed to recapitulate the complex structure and functionality of a human organs (Ronaldson-Bouchard and Vunjak-Novakovic, 2018).

**HPSC models of complex phenotypes - cerebral organoids**

Although conventional 2-dimensional (2D) monolayer co-cultures can recapitulate many non-cell-autonomous interactions, such models are still limited to provide the complexity of 3-dimensional (3D) brain structures to study complex features, such as spatial organization of neuronal structures (e.g. neocortical layer formation), migration, axon guidance or establishment and response to morphogen gradients. These dysfunctional processes likely contribute to disease pathogenesis especially in neurodevelopmental disorders. Recently, 3D-culture models resembling human brain structures, commonly referred to as cerebral organoids, have facilitated the unprecedented opportunity to study complex aspects of human brain development and disease pathology in a cell culture dish (Kadoshima et al., 2013; Lancaster et al., 2013).

Exploiting the self-organizing capacity of neuroectodermal precursor cells, cerebral organoids recapitulate many features of human brain development and morphology,
including ventricle and neural tube-like structures with apical-basal polarity, progenitor zone organization with outer radial glial stem cells, cortical plate development and electrophysiological mature neurons that participate in neuronal network activity (Kadoshima et al., 2013; Karzbrun et al., 2018; Lancaster et al., 2013; 2017; Li et al., 2017b; Paşca et al., 2015; Quadrato et al., 2017). Cerebral organoids have already provided novel insights into human-specific processes, which could not be studied in conventional 2D culture systems, such as brain gyrification (Karzbrun et al., 2018; Li et al., 2017b), cell migration and outer radial glial mitosis defects in Miller-Dieker syndrome (Bershteyn et al., 2017).

Organoid models are now a common tool to study neurodevelopmental disorders, including microcephaly (Lancaster et al., 2013; Li et al., 2017a), autism spectrum disease associated with macrocephaly (Mariani et al., 2017; Wang et al., 2017), lissencephaly (Bershteyn et al., 2017; Karzbrun et al., 2018), Rett Syndrome (Mellios et al., 2017), Thymothy syndrome (Birey et al., 2017), lysosomal storage diseases such as Sandhoff disease (Allende et al., 2018), schizophrenia (Srikanth et al., 2018; Stachowiak et al., 2017) and Zika infection on neonatal brain development (reviewed in (Qian et al., 2017). Despite the early success with this approach for disease-modelling, challenges remain. The current organoid approach produces substantial variability between individual cerebral organoids regarding organoid size and cellular composition (Quadrato et al., 2017). More sophisticated bioengineering approaches such as the use of mini-bioreactors to grow individual organoids under controlled conditions (Qian et al., 2016) may reduce variability, which should allow studying a broader range of complex diseases.

Although neuronal cell types are abundant in cerebral organoids, the detailed analysis of cellular diversity reveals the lack of non-neuronal cell types such as endothelial cells and microglia. Since these cell types likely have a central role in brain development and disease pathology, one strategy is to externally provide the missing cell types during organoid formation. We demonstrated in vitro hPSC-derived microglia efficiently invade cerebral organoids, showing the utility of this strategy (Muffat et al., 2016).

One concern is whether neuronal cells in cerebral organoids mature past early events in brain development. In vitro derived organoids recapitulate gene expression and epigenetic programs of fetal cortical development. After extended periods in culture, organoids show a gene expression signature comparable to mid-gestation embryos between post-conception week 17-24 (Camp et al., 2015; Luo et al., 2016; Paşca et al., 2015; Quadrato et al., 2017). This may explain why most reported organoid studies describe the dysfunction of processes that occur during early brain development.

One major barrier for extended in vitro culture with further maturation is the lack of vascularization for nutrient and oxygen penetration, as late stage cerebral organoids commonly develop extensive necrosis and massive cell death. Current attempts to engineer a vascular system into organoids either exploit the self-organizing capacity of stem cells to construct a functional vascular system by providing appropriate vascular precursor populations during organoid development or using sophisticated bio and tissue-engineering tools to construct artificial vasculature like structures. One alternate strategy to overcome
this limitation transplants cerebral organoids into the adult mouse brain. Transplanted organoids survive and mature for extended periods of time in the host brain. Surprisingly, mouse blood vessels invade the organoids to create a functional vascular system, resulting in significantly improved neuronal survival and the establishment of extensive functional-graft-to-host synaptic connection (Mansour et al., 2018).

It remains uncertain whether cerebral organoid models can be applied to study late age onset disorders. Yet, experimental evidence exists modelling Alzheimer’s disease like pathology in a 3D culture system with ectopic overexpression of familial Alzheimer’s diseases-associated mutations in beta-amyloid precursor protein and presenilin 1 (Choi et al., 2014; Park et al., 2018) and in cerebral organoids derived from Alzheimer’s patients (Raja et al., 2016), suggesting the feasibility to replicate disease-associated pathology not observed in Alzheimer’s disease mouse models.

**HPSC models of complex phenotypes - interspecies chimeras**

Generating interspecies chimeras using cells from more than one individual may provide the most physiologically relevant environment to study human disease in an in vivo context. Experimental chimeras derived from cells within the same species are routinely used for biomedical application including generating genetically engineered mouse models. The idea of interspecies chimeras between human cells and a model organism such as the mouse, rat, pig or monkey is particular tantalizing for late age of onset disorders, as it would allow studying patient-derived cells in a genuine in vivo environment for extended periods of time over the lifetime of the animal. Although transplantation of human cells into adult hosts animals provides an important research model to study the feasibility of cell-replacement therapies for regenerative applications, we will only focus here on approaches that may support the functional integration of host cells during normal embryonic development.

One approach generates interspecies chimeras by injecting hPSCs into the preimplantation embryo, so human cells would participate in normal development and functionally integrate throughout the entire host embryo. However, this approach currently seems rather inefficient, as the extent of functional integration into the host animal remains an intense debate (Gafni et al., 2013; Theunissen et al., 2016; Wu et al., 2016) A recent report described the robust contribution of human cells to the pre-implantation embryo but only limited contribution to the post-implantation pig embryo (Wu et al., 2017). Currently it is not clear whether the variable contributions result from differences in evolutionary distance, better matched developmental states or reduced ligand-receptor incompatibility between human donor cells and the host embryo.

An alternative approach generates chimeras by transplanting lineage-restricted multipotent stem cells into the developing post-gastrulation embryo. The donor cells, after transplanting into a developmental-stage matched host, should participate in normal embryonic development to generate functionally integrated human cells of the respective cell lineage. Using this approach, we showed that human neural crest stem cells transplanted into the gastrulating mouse embryo followed neural crest specific migration patterns and contributed to the pigment lineage in postnatal mice (Cohen et al., 2016). Importantly, genetic depletion...
of the host melanoblast lineage significantly increased integration of human cells, suggesting that an “empty host niche” gives donor cells a selective advantage to efficiently contribute to interspecies chimeras. Similarly, transplanted human neural precursor cells into E14.5 embryos or terminally differentiated cortical neurons into neonatal mice robustly functionally integrate into mouse brain (Espuny-Camacho et al., 2013; Nagashima et al., 2014).

One surprising difference between these two transplantation paradigms is that neural crest cells may acquire functional properties at the time of birth of the host animal, while transplanted human neuronal cells may functionally mature substantially slower than host neurons. This indicates, that human transplanted neural cells follow a developmental maturation pattern resembling human embryonic development rather than the host environment (Espuny-Camacho et al., 2013). Similarly, this cell autonomous developmental clock may inhibit or delay the appearance of mature cell types in vitro. In both experimental paradigms, the overall contribution of human cells is rather limited. This could be caused by technical challenges related to transplantation (Cohen et al., 2016; Nagashima et al., 2014), imperfect adjustment of developmental timing between donor and host cells at the time of transplantation (Cohen et al., 2018) or species-specific differences.

Inter-species immune-rejection and the lack of crucial signals from the host environment may contribute to limited survival. Using sophisticated immunodeficient mouse models or humanized hosts carrying fully functional human immune system or genetically engineered human specific growth factors could enhance the chimeric contribution of human cells (Zhang et al., 2004). Recent intraspecies chimera experiments suggest that matched developmental timing is crucial for successful contributions by transplanted cells (Cohen et al., 2018). It is still unclear, whether inhibition of cell death (e.g. overexpression of antiapoptotic factors such as Bcl2) is sufficient to relax this requirement (Cohen et al., 2018; Masaki et al., 2016).

Proof of principle examples demonstrate the unique promise of studying human cells in a physiologically relevant, in vivo environment. Human neurons transplanted into a mouse model of Alzheimer’s disease display signs of neuronal degeneration accompanied by cell loss and some pathological features observed in patients with Alzheimer’s disease, such as dystrophic presynaptic changes and hyper-phosphorylation of tau (Espuny-Camacho et al., 2017). The most stunning demonstration of interspecies chimera approach generated human glia chimeric mice by injecting bipotential astrocyte-oligodendrocyte precursors into the neonatal brain. The transplanted human cells did not only survive and integrate, but also replaced the entire glial population of the host brain. The engraftment of human glia in the mouse brain enhanced synaptic plasticity and learning in the transplanted mice (Han et al., 2013). Transplanting glial precursor cells derived from schizophrenia patients induced schizophrenia-like behaviors like excessive anxiety, antisocial traits and disrupted sleep in mice indicating a causal role of glial pathology in disease development (Windrem et al., 2017).
HPSC-based drug discovery

A main motivation behind hPSC-based disease modeling is to discover novel therapeutic targets and develop new drugs. Since human model system are anticipated to be more predictive for clinical efficacy of novel drugs, this approach may substantially reduce cost and time to translate potential therapies into the clinic. Initial proof-of-concept experiments employed a candidate approach testing a small number of compounds to reverse disease-associated phenotypes in human disease-relevant cell types. Selected candidate small molecules included compounds known to target disease-relevant molecular pathways, compounds expected to be effective based on analysis in non-hiPSC disease models and compounds previously described as effective in related diseases to evaluate for repurposing. This strategy yielded numerous biologically active compounds that can reverse disease-associated phenotypes *in vitro* for a wide variety of neurodevelopmental, psychiatric and neurodegenerative disorders (reviewed in (Avior et al., 2016)).

Initial compounds, which either used hiPSC-based disease models for pre-clinical drug development or were developed based on observations from hiPSC disease models, are already in clinical trials. These includes a novel class of splicing modifying compounds to treat spinal muscular atrophy (Naryshkin et al., 2014) (https://clinicaltrials.gov; NCT02913482), a humanized antibody targeting secreted eTau to treat tauopathies (Bright et al., 2015), including progressive supranuclear palsy (https://clinicaltrials.gov; NCT03068468) and early Alzheimer’s disease (https://clinicaltrials.gov; NCT03352557), and the Kv7.2/3 potassium channel agonist ezogabine, an approved antiepileptic medication to treat amyotrophic lateral sclerosis (Kiskinis et al., 2014; Wainger et al., 2014) (https://clinicaltrials.gov; NCT# 02450552).

A major promise of hiPSC technology is the potential to perform unbiased high-throughput compound (HTS) or genome-scale genetic screens. Drug development in the past predominantly employed target-based drug discovery (TDD) that requires identifying defined molecular targets relevant to disease pathogenesis (Moffat et al., 2017). However, TDD has been remarkably ineffective for neurological diseases given the complexity of pathological phenotypes and lack of reliable model systems to identify essential disease-relevant targets. As alternative, recent progress in hiPSC technology has revived phenotypic drug discovery (PDD). This approach does not rely on a priori knowledge of a specific drug target and its role in disease pathology. Instead, it screens for reversing complex disease-associated phenotypes, even with little insight into the underlying molecular mechanisms.

The critical requirement for phenotypical screens are robust, simple to analyze disease-relevant phenotypes, which can be easily scaled up for high-throughput applications. Despite the limitations, some unbiased medium-scale screens have already identified active disease-phenotype modifying compounds for neurological disorders, including familial dysautonomia (Ramirez et al., 2012), fragile X syndrome (Kaufmann et al., 2015; Kumari et al., 2015; Li et al., 2016), Alzheimer’s diseases (Brownjohn et al., 2017; Kondo et al., 2017; Xu et al., 2013), Parkinson’s disease (Ryan et al., 2013), motor neuron disease (Höing et al., 2012), amyotrophic lateral sclerosis (Burkhardt et al., 2013; Egawa et al., 2012; Shi et al., 2018) and Zika infection on fetal brain development (Zhou et al., 2017). Dictated by the
underlying disease pathophysiology, the screening readouts can range from simple quantitative traits such as re-expression of the FMR1 gene for fragile-X-syndrome to complex higher order disease-associated traits such as cellular degeneration for amyotrophic lateral sclerosis (Shi et al., 2018).

Developing high-content screening platforms, which implement automatic digital microscopy combined with sophisticated computational image analysis to quantify complex traits, will become critical for successful phenotypical screening approaches (Finkbeiner et al., 2015). However, practical and technical challenges associated with the complexity of the hPSC platform currently limit scaling-up, required for a broader application of PDD. Efforts to automatize and miniaturize the hPSC-screening platforms using cutting edge robotics, microfluidics and nanotechnology may overcome this limitation. Current phenotypical screens only screen small molecule collections in the order of several hundred to several thousand compounds, not the million-compound libraries commonly used in the pharmaceutical industry. These screening libraries focus on well-curated compounds with known molecular targets or biological activity, or utilize small-molecule libraries that contain drugs already approved for clinical applications. Drug repurposing should substantially reduce cost and facilitate translation of novel therapeutics with known safety, pharmacodynamics and pharmacokinetics information.

An alternative to chemical compound-based drug discovery are genome-scale genetic screens. Genetic screens in various model systems like yeast, flies, nematodes and mice prove to be invaluable to investigate gene function, dissect complex biological processes and identify drug targets. In contrast to the enormous chemical space of pharmacological active molecules, the number of genes in the human genome constrains the scale of genetic compared to compound screens. Genetic screens resolve some limitations in chemical phenotypic screens like identifying molecular targets of compounds recovered in chemical screens as they directly pinpoint the gene and pathways implicated in the observed phenotypes. Importantly, genetic screens may provide targets for subsequent target-based drug discovery (TDD) efforts. The ease and efficiency of the CRISPR/Cas system to target any locus in the genome provides great flexibility to develop genome-wide knock-out, gene activation (CRISPRa), gene suppressor (CRISPRi) or epigenome modifier screens in human cells (Gilbert et al., 2014). Genome-scale CRISPR screens are already used to identify genes implicated in many biologically relevant phenotypes, such as gene essentiality (Wang et al., 2017), caner progression (Chen et al., 2015; Shalem et al., 2014; Shi et al., 2015), drug resistance (Shalem et al., 2014; Wang et al., 2014), viral infection, immune response and vulnerability to bacterial toxins (Zhou et al., 2014). While published genome-scale screens predominantly utilize fast proliferating tumor cell lines, genome-scale phenotypical genetic screens in post-mitotic neuronal cells may be already feasible and become an invaluable tool for drug discovery in neurological diseases.

Conclusion

Rapid advances in the hPSC approach combined with progress in genome editing technologies and the availability of genetic and epigenetic information now provide a realistic opportunity to study human neurological disorders in a disease-relevant and
systematic manner. Most sporadic diseases arise through a complex interaction of multiple genetic and non-genetic risk factors suggesting that each person’s disease prognosis, including onset, progression and responsivity to treatment, is an individual trait. Personalized medicine will tailor therapies to a patient’s unique clinical, physiological and environmental information.

We described how hiPSCs could bridge this gap by generating disease relevant models from the patient’s own cells. This hPSC approach allows easily alternation between the in vitro disease platform and in vivo patient context. We can now generate novel disease hypotheses based on specific clinical observations and functionally test them with high molecular precision in disease-relevant cellular models from the same individual or reversely use insights gained from cellular models and directly confirm the relevance in a patient’s context. Combined with continuing advances in tissue engineering that can generate 3-dimensional organoid structures resembling entire organs, we envision that complex patient-derived human model systems will become a powerful research tool in our understanding of human physiology and disease development, as seen with other tractable model organisms. Working with human systems will overcome many limitations of non-human model organisms, which will more successfully translate newly identified molecular pathways or disease phenotype modifying chemical compounds into novel targets for therapeutic interventions. These advances remarkably occurred only 20 after the first report deriving a hESC line and just over 10 years after the stunning discovery that reprogrammed human somatic cells into hPSCs.

Now, the first clinical trials are already underway evaluating the efficacy of hPSC-derived therapeutics for transplantation of novel drugs developed in hPSC disease models. This stands in contrast to the initial uncertainty if cell culture models for late age of onset neurological disorders could identify any disease-relevant phenotype. The fundamental impact of hPSCs on disease modeling cannot be overstated. There remains no effective alternative approach to functionally dissect the molecular and cellular effects of non-coding variation and to develop reliable models for many devastating sporadic disorders. However, we acknowledge these initial steps will further the promise of hPSCs technology to transform biomedical research. As for every nascent field, unresolved issues of the hPSC approach remains, as discussed in this review. We are confident that the rapidly evolving field will overcome these barriers, so that hPSCs will increasingly yield basic biological and clinical translational insights for years to come.

Acknowledgements:

Work in our lab is supported by grants from the NIH. We would like to thank members of the Jaenisch lab as well as many other colleagues for stimulating discussions and Brandi Mattson for help with editing the manuscript. We apologize to many whose work we did not adequately describe or cite due to space constraints.

References:


Cell. Author manuscript; available in PMC 2019 October 18.


Figure 1: Comparison of hiPSC-based strategies to model complex diseases in cell culture

A. HiPSC-based strategies to identify diseases-associated phenotypes for monogenetic diseases comparing (i) the conventional approach with (ii) an isogenic approach. Genome editing allows to generate an experimental system with controlled genetic background by either correcting a disease-associated mutation in patient-derived cells or inversely inserting the mutation into wild-type hPSCs.

B. HiPSC-based strategies to model sporadic diseases comparing (i) the conventional approach with (ii) a patient stratification approach based on known disease associated genotypes (risk score or specific risk genes) or clinical phenotypes and (iii) population genetics approaches to identify associations between specific genotypes and QTLs and (iv)
the functional analysis of specific disease-disease associated risk variants. Genome editing allows to generate an allelic series of isogenic cell lines carrying distinct risk variants.
Figure 2: Identification of functional risk variants based on epigenetic signatures

Candidate disease-associated risk variants (e.g. SNP) are prioritized based on integrating GWAS data with a variety of epigenetic datasets. Risk variants overlapping with tissue-specific regulatory elements are more likely to be functional compared to risk variants outside of regulatory elements. Epigenetic datasets can be generated either from primary tissue or from \textit{in vitro} hPSC-derived somatic cell types. Schema at the bottom depicts (i) the identification of a hypothetical GWAS identified diseases-associated risk variant overlapping with a brain-specific distal enhancer (indicated by enrichment for histone H3 lysine 27 acetylation Chip-seq signal), (ii) the CRISPR/Cas9 mediated SNP exchange and (iii) the functional analysis of the SNP exchange indicated by a genotype-dependent cis-regulatory effect on expression of Gene Y (eQTL).
Figure 3: Comparison of *in vitro* and chimeric *in vivo* approaches to model complex disease-associated phenotypes

hPSCs, human pluripotent stem cells; NPs, neural progenitors; TF, transcription factor.