Investigating monogenic and complex diseases with pluripotent stem cells

Hao Zhu[*,‡], M. William Lensch[‡], Patrick Cahan[‡] and George Q. Daley[‡]*

Abstract | Human genetic studies have revealed the molecular basis of countless monogenic diseases but have been less successful in associating phenotype to genotype in complex multigenic conditions. Pluripotent stem cells (PSCs), which can differentiate into any cell type, offer promise for defining the functional effects of genetic variation. Here, we recount the advantages and practical limitations of coupling PSCs to genome-wide analyses to probe complex genetics and discuss the ability to investigate epigenetic contributions to disease states. We also describe new ways of using mice and mouse embryonic stem cells (ESCs) in tandem with human stem cells to further define genotype–phenotype relationships.

Genome-wide association study
A whole-genome examination of genetic variation and how it statistically correlates with traits and diseases. This technique has led to the discovery of associations of particular genes with diseases such as age-related macular degeneration and diabetes.

Synteny
In classical genetics, synteny describes the colocalization of genetic loci on the same chromosome between individuals or species.

**Division of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA.
‡Stem Cell Transplantation Program; Division of Pediatric Haematology/Oncology; Harvard Stem Cell Institute; Howard Hughes Medical Institute; Children's Hospital Boston and Dana-Farber Cancer Institute; 300 Longwood Avenue, Boston, Massachusetts 02115, USA.
Correspondence to G.Q.D.
e-mail: george.daley@childrens.harvard.edu*
doi:10.1038/nrg2951
Published online 9 March 2011

Understanding the mechanisms by which genetic variation contributes to disease is a central goal of human genetics and will facilitate the development of preventive strategies and treatments. Despite the torrent of genetic data flowing from genome-wide association studies (GWA studies), numerous barriers remain to defining the functional genetic contributions to complex traits. Classically, the physiological or developmental function of a gene has been explored *in vivo* through the generation of gain- or loss-of-function mutants in model organisms. Genetically altered mice have been central to the elucidation of mammalian genetic mechanisms. However, numerous human phenotypes fail to be successfully replicated in mice owing to either fundamental biological differences between the two species or a lack of synteny, as in the example of contiguous-gene effects deriving from chromosome-level aberrations. Medical genetics is also performed through human clinical studies; however, access to tissues other than blood (for example, brain, heart or pancreas) is quite limited. Moreover, it is challenging to obtain large volumes of material for experimental studies, and efforts to expand biopsy material (for example, by immortalizing primary cell lines) frequently result in genetic alterations and phenotypic artefacts owing to prolonged growth in cell culture. Available cell lines or tissues are often not the tissue of interest, rendering much of the analysis irrelevant.

Embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) — the main features of which are described in BOX 1 — present potential opportunities (and looming challenges) for human disease modelling. The use of stem cells to address pathophysiological questions has exploded over the past 5 years, with many of these early studies appropriately focusing on monogenic disorders. An increasing command of protocols for directed tissue differentiation, combined with a rich body of human GWA studies that are eager for biological validation, has created the opportunity to interrogate the cellular and biochemical consequences of genetic variation. Although few examples of using stem cells to model complex diseases have been published, the field is poised to make important advances. Here, we discuss these studies and the insights that they offer. We also highlight the areas of caution and potentially provide some inspiration for those ready to tackle these challenges.

Rationale for using PSCs in genetic studies

Advantages of PSCs. ESCs and iPSCs, collectively referred to as pluripotent stem cells (PSCs), offer what may prove to be a cornerstone technology for dissecting human genotype–phenotype relationships. iPSCs can be generated from terminally differentiated cells using a combination of genetic and expression-modulating factors (BOX 1). iPSCs capture and immortalize individual genomes and permit tests of cellular function in a scalable experimental format. Patient-specific cells, annotated with clinical history, allow the investigator to probe genotype–phenotype relationships for conditions with a monogenic and, increasingly, with a complex basis. iPSCs constitute the genetic background of the original patient, which may prove crucial as phenotypic variation may entail interactions between a polymorphism and modifier loci.

© 2011 Macmillan Publishers Limited. All rights reserved

Amyotrophic lateral sclerosis
A neurodegenerative disease associated with a dominant L144F substitution in the superoxide dismutase 1 (*SOD1*) gene.

PSCs are inexhaustible, scalable and physiologically native material for experimentation, which is in contrast with other types of patient-derived cell lines that involve oncogenic transformation to ensure long-term immortal and scalable culture. The capacity of PSCs to self-renew facilitates cell-based genetic or drug screens, which would otherwise be difficult, if not impossible, in primary human cell lines or mouse models. Another crucial advantage of PSCs is their pluripotent nature. In principle, directed *in vitro* and/or *in vivo* differentiation permits the study of essentially any cell or tissue, affording access to difficult-to-obtain cell populations (for example, neurons and fetal cells) and allowing interrogation of multiple cellular phenotypes from a single starting biopsy. For most cell and tissue types, however, differentiation remains a challenge.

The ability to derive iPSCs from patients with genetic disease was first described in two papers in 2008 (REFS 1,2). iPSCs from multiple single-gene disorders were generated, including a case of amyotrophic lateral sclerosis (ALS)[1], as well as complex diseases, such as diabetes mellitus and Parkinson's disease, and a chromosomal trisomy (Down's syndrome). These initial reports have been followed by a series of studies on monogenic diseases (for example, dyskeratosis congenita[3], epidermolysis bullosa[4] and many others[5]) and more complex conditions (such as Rett's syndrome and autism spectrum disorder[6]), demonstrating the attractiveness and interest in using PSC-based platforms to study human development and disease.

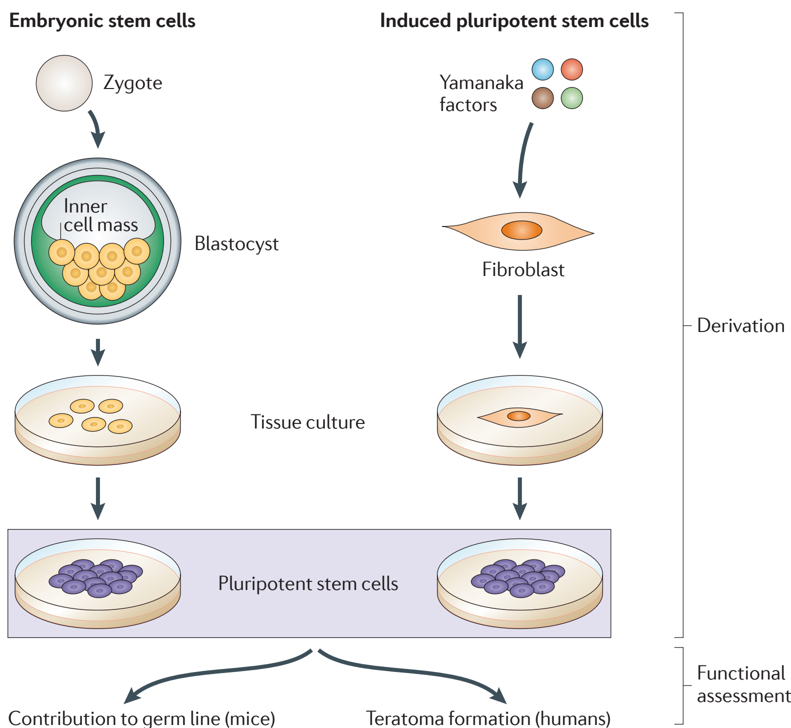
Limitations of PSCs. Not all conditions are readily modelled using iPSCs. Cells from patients with Fanconi's anaemia are defective in DNA repair and are refractory to reprogramming without antecedent gene correction, making it difficult to model this condition *in vitro*[7]. Instead, by using RNAi against multiple Fanconi's anaemia genes in human ESCs, Tulpule *et al.*[8] demonstrated that haematopoietic cells were markedly reduced at the earliest stages of blood lineage specification. There can also be discordance between the same disease modelled by ESCs or iPSCs, as illustrated by fragile X syndrome (FXS; also known as FRAX); the mutant, trinucleotide expanded fragile X mental retardation 1 (*FMR1*) allele is expressed in FXS ESCs but fails to be transcriptionally reactivated in FXS iPSCs following reprogramming[9]. These cases highlight areas in which ESCs instead of iPSCs are more useful and remind us that ESCs themselves are amenable to disease gene modification.

More will need to be learned about the factors that limit cellular reprogramming, as such insights may define what classes of disease will be difficult or impossible to model through reprogramming, although it is likely that most diseases can be 'captured' in one of the two settings.

Box 1 | Embryonic stem cells and induced pluripotent stem cells

Pluripotent stem cells (PSCs) are defined by their ability to differentiate into the hundreds of somatic cell types of an adult animal, and by their ability to continually self-renew. PSCs called embryonic stem cells (ESCs) were first isolated in 1981 by culturing the inner cell mass of pre-implantation mouse blastocysts[59,60] (see the figure, left panel). The derivation of human ESCs, which was made possible by finding amenable cell culture conditions, was achieved in 1998 by Thomson *et al.*[61]. In 2006, Yamanaka and colleagues discovered that introducing four transcription factors that are typically expressed in ESCs (the so-called 'Yamanaka factors': OCT4 (also known as POU5F1), SOX2, Krüppel-like factor 4 (KLF4) and MYC) converted mouse skin fibroblasts to PSCs, a powerful example of the general process termed 'reprogramming'[62] (see the figure, right panel). Similar lines generated from human cells followed soon thereafter[63–65], and now investigators routinely derive and study pluripotent cells from different species, tissues and patients[66–69].

Currently, the best way to functionally assess pluripotency in mice is to generate chimeric mice and subsequently assess the ability of the stem cells to be represented in the germ line. To assess pluripotency from human PSCs, cells are injected subcutaneously into immunodeficient mice and the resulting tumours are assessed by pathologists for contributions to all three germ layers[25]. By these standards, both induced PSCs (iPSCs) and ESCs are pluripotent. However, careful phenotypic, functional and molecular studies of PSCs are revealing subtle differences between ESCs and iPSCs that may be particularly relevant to *in vitro* contexts (discussed in detail in the main text).



Modelling of specific diseases

The following section describes the successes in modelling tissue-specific diseases and elaborates on the shared features of informative PSC models. Rather than broadly focusing on PSC disease modelling, the monogenic examples we have selected represent common diseases that affect tissues into which PSCs can be readily differentiated (heart, liver and β-cells) and offer differentiation and phenotyping assays that could be tailored to the study of more complex diseases. A few examples describe polygenic disorders (such as Down's syndrome) for which PSCs have already been used to explore biological mechanisms. In other cases, we explore areas of opportunity that are primed for the use of PSC tools (for example, diabetes). Where useful, the distinction between the use of ESCs and iPSCs is discussed.

© 2011 Macmillan Publishers Limited. All rights reserved

Box 2 | Challenges of pluripotent stem cell-derived complex disease modelling

This box gives a brief outline of the problems that need to be confronted when creating disease models using pluripotent stem cells (PSCs). We place particular emphasis on the special barriers to complex disease studies. The points in the box are addressed in more detail in the other Reviews in this Focus issue (where noted) and in the subsequent sections of this Review.

Differentiation. Robust and efficient differentiation towards selected cell and tissue types is without question the most daunting barrier to studying diseases in those tissues[19].

Non-cell-autonomous phenotypes. It is challenging to study cellular interactions *in vitro*. As described in the main text, there are several ways in which this problem might be overcome, such as through the use of co-culture, organoids, human–mouse and mouse–mouse chimaeras or transplants.

Presence of viral vectors. Most disease- and patient-specific induced PSCs (iPSCs) are generated using integrating retro- or lentiviruses. There could therefore be a small but significant influence of these virus vectors on gene expression and phenotype, which could hamper efforts to define the weaker, potentially lower-penetrance phenotypes associated with complex diseases and traits (alternative reprogramming methods are reviewed in this issue[54]).

Targeted genetic modification. It is difficult to perform homologous recombination for the purposes of making knock-ins and knockouts, which are the 'gold standard' methods of gene modification in mouse genetics. The inability to do this efficiently makes it difficult to identify proper controls and demands the use of naturally occurring mutations, which can be a challenge to both procure and match with appropriate controls.

Lack of proper controls. Given the genetic and epigenetic variability found among human iPSC lines, among mouse iPSC lines, and even among early-passage and late-passage PSCs of the same cell line, one can easily appreciate the need for appropriate controls. This is a problem that is closely linked to issues around genetic modification, both targeted and viral.

Penetrance. Low-penetrance and modest or undetectable phenotypes are a major challenge when studying complex or polygenic factors. This point is discussed further in BOX 4.

Onset of diseases associated with ageing. Long periods of time may be needed to recapitulate phenotypes of adult- or advanced-age-onset disease in the cell-culture dish. Although it is possible that the stresses of constant proliferation and passaging of diseased cells (or purposefully introduced stress) will elicit late-onset phenotypes, the solutions to this issue will be dependent on the experimental context. This issue is discussed further in BOX 4.

Monogenic diseases as proof-of-principle models. Although the field envisions using PSCs to investigate polygenic conditions, it is imperative to use monogenic disorders to develop proof-of-principle differentiation and phenotyping assays, so that potentially subtle cell-culture phenotypes can be correlated to strong clinical phenotypes. For example, the diabetes field awaits the modelling of a monogenic disease such as maturity onset diabetes of the young (MODY), an autosomal dominant form of diabetes. After such a proof-of-principle disease is shown to recapitulate the human phenotype, more subtle mutations or variants can be assessed for genetic interactions. For example, candidate genes from GWA studies can be overexpressed or knocked down using viral transgenesis or zinc-finger nucleases (discussed later) in the context of these disease models (BOX 2) to test for phenotype enhancement or suppression. It will also be important to explore the impact of environmental perturbations on these well-characterized mutations and their phenotypic outputs, so that the dynamic range of individual variants and assays can be mapped.

Some of the following monogenic examples serve as 'positive controls' and so can be used as templates for follow-on genetic modifier studies. They can also be used for environmental screens through the use of drugs, small molecules, toxins or non-chemical stresses (FIG. 1).

Capturing genetic variation in cardiovascular disease: long QT syndrome. The modelling of cardiovascular biology in mice has been instrumental for understanding the biology of the human heart, but in some cases it has been limiting. The mouse heart beats roughly seven times faster than the human heart, making it difficult to phenocopy human arrhythmia in mice[10]. Moretti *et al.*[11] generated iPSCs from long QT syndrome type I patients with mutations in *KCNQ1* (potassium voltage-gated channel, KQT-like subfamily, member 1), which encodes the pore-forming subunits of the channels that generate potassium currents. These iPSCs were then differentiated into cardiomyocytes, which displayed 'ventricular' or 'atrial' features that could be distinguished from 'nodal' cell types *in vitro*. Whereas nodal cardiomyocytes were similar between patients and control subjects, the action potentials of ventricular and atrial cardiomyocytes derived from iPSCs were significantly longer and had a slower repolarization velocity than did control cells[11]. This electrophysiological profile was distinct from those found in transgenic animal models and more closely resembled what is seen in human long QT syndrome patients. One notable finding of this study was that the mutant form of the KCNQ1 protein subunit interfered with the function of the wild-type subunit, and that varying the relative quantities of these subunits resulted in markedly altered subcellular localization of the potassium channels. This finding illustrates the power of studying mutations within their naturally occurring gene regulatory environment; had a mutant transgene under a ubiquitous promoter been used, there would probably have been a different ratio of mutant to wild-type protein and, as a result, a different experimental outcome.

Long QT syndrome
An inborn heart condition in which delayed repolarization of the heart increases the risk of ventricular arrhythmias.

© 2011 Macmillan Publishers Limited. All rights reserved

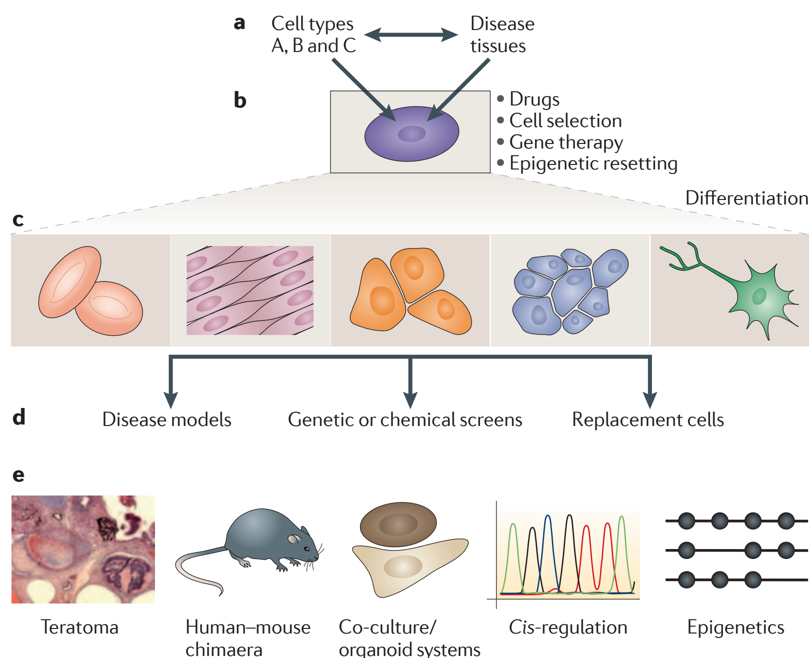


Figure 1 | **Pluripotent stem cells are a hub for genetic interrogation.** **a** | Essentially any source of cells (A, B and C designate distinct lineages), normal or diseased, can be converted into induced pluripotent stem cells (iPSCs). **b** | With iPSCs or embryonic stem cells (ESCs) in hand, chemical or genetic modifications, cell passaging, or epigenetic modulation can be performed in order to study disease, drive differentiation or select for clones with particular phenotypes. **c,d** | Differentiated cell types can be derived (**c**) and used as disease models, unlimited material for cell-based chemical or genetic screens or, potentially, therapeutic tissue sources (**d**). **e** | Disease modelling options include, but are not limited to, teratoma formation assays, human–mouse chimaeras, co-culture/organoid systems, allele-specific expression (ASE) and RNA-seq, and epigenetic assays.

Moretti *et al.* also found that the long QT iPSCs were vulnerable to catecholaminergic stress, which opened up the possibility to screen for compounds that can ameliorate or exacerbate the clinical phenotype. Another study by Braam *et al.*[12] showed that treatment of human ESC-derived cardiomyocytes with 12 cardiac and noncardiac drugs could recapitulate the altered action potentials seen in patients. These studies show that PSC-derived cells are a viable screening platform for investigating environmental, pharmaceutical or genetic modulators of cardiac electrophysiology[12]. Given the robustness of these phenotypes, it is conceivable that phenotypes associated with variation at the loci identified in two recent long QT GWA studies could be tested electrophysiologically in patient-specific iPSC-derived cardiomyocytes[13,14]. Alternatively, genes from these loci could be directly modified in human ESCs for further testing (BOX 2).

Hepatogenesis and liver disease. To better understand conditions that affect the liver, Rashid *et al.* established a protocol to differentiate human iPSCs into hepatocytes that recapitulate the adult phenotypes of three distinct liver diseases *in vitro*[15]. These iPSC-derived hepatocytes shared marker-expression, morphological and physiological features of normal human hepatocytes. The iPSC-derived hepatocytes made from cells taken from patients with α1-antitrypsin deficiency showed the characteristic accumulation of α1-antitrypsin polymers. Of importance, the authors noted little variability in polymer accumulation among multiple iPSC lines derived from the same patient but greater variability among iPSC lines derived from different patients, indicating reproducibility of the phenotype within individual genotypes. In addition, proteosome inhibitor treatment of the cells exacerbated the phenotype, possibly permitting the detection of small but relevant functional differences in genotype; similarly, the assay might detect the impact of environmental modifiers and potentially be used in toxicity studies for the analysis of primary and/or secondary metabolites. This report opens the door to studying more subtle genetic disorders of the liver *in vitro*. More generally, testing drug–genotype interactions in specific cell types will expedite the efforts of personalized medicine and pharmacogenomics, especially for common variants.

***Endocrine disorders: systemic physiology captured* in vitro.** Intense focus has been directed towards *in vitro* and *in vivo* β-cell differentiation, with the long-term view of generating patient-specific, insulin-secreting β-islet cells for therapeutic use in type 1 diabetes, a disease that results directly from β-cell failure or destruction. Although type 2 diabetes is not primarily caused by β-cell destruction, β-cell exhaustion in the face of chronic insulin resistance contributes to the phenotype[16]. Thus, research into both type 1 and 2 diabetes will benefit from a better understanding of human β-cell biology.

Fortunately, multiple groups have made headway in differentiating PSCs into β-cells[17], and one could use these protocols to study genetic variation in β-cell function, survival and susceptibility to autoimmune attack. Currently, it is possible to convert ESCs into definitive endoderm and then drive their development towards becoming pancreatic and endocrine progenitor cells[18,19]. When these progenitor cells are transplanted into mice, they can correct hyperglycaemia[20]. Producing large numbers of definitive β-cells before transplantation is challenging, despite extensive efforts to understand the discrete steps during differentiation. If β-cells could be predictably derived from diabetic patients, drug and genetic screening could be used to phenotypically modulate β-cell number and function using *in vitro* insulin production assays. More specifically, such assays could be used to evaluate iPSCs from patients with type 2 diabetes who do and do not have SNPs that have been identified by GWA studies and to evaluate transgenic ESCs containing gain- and loss-of-function modifications of GWA candidate genes. In addition to *in vitro* insulin assays for β-cells or transplant assays (described below), glucose uptake assays can be performed in muscle, hepatocytes or adipocytes derived from human diabetic iPSCs. The advantage of being able to generate different tissue types from PSCs is that multiple disease-relevant cell types with the same genome can be directly compared.

© 2011 Macmillan Publishers Limited. All rights reserved

Malignancy: assessing host–tumour interactions using iPSCs. PSCs share many features in common with malignant cells (see REF. 21 for a review), so the use of PSCs permits the study of processes that are relevant to tumour growth. One example here relates to tumour angiogenesis, wherein a growing mass recruits/remodels existing vasculature to support new tissue formation.

Down's syndrome is a complex condition resulting from the overexpression of more than 200 genes on the trisomic chromosome 21. A lesser known feature of the Down's syndrome phenotype is a reduced incidence of solid tumour formation; the age-related cancer mortality in patients with Down's syndrome is less than 10% of that in individuals without Down's syndrome[22]. The Down's syndrome candidate region 1 (*DSCR1*; also known as *RCAN1*) gene is both localized to chromosome 21 and a negative regulator of vascular endothelial growth factor (VEGF)-mediated calcineurin signalling, so to define the molecular basis of the antitumour effect seen in Down's syndrome, Baek *et al.*[23] studied the calcineurin pathway. In a mouse model of Down's syndrome[24], *Dscr1* overexpression impaired angiogenesis on its own, although this activity was enhanced by the overexpression of other genes on the same chromosome, including *Dyrk1a* (dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1a), a serine/threonine kinase that impairs the activation of genes distal to VEGF–calcineurin–NFAT (nuclear factor of activated T cells) signalling. This result supports a role for *Dscr1* and other genes in impairing tumour angiogenesis in a mouse model, but to test the applicability of the findings to humans, a teratoma formation assay (see REF. 25 for a review of this assay) was carried out with fibroblast-derived iPSCs from patients with Down's syndrome[23]. Teratomas made from Down's syndrome iPSCs indeed produced fewer $CD31^+$/$CD45^-$ endothelial cells and had significantly lower microvascular density than their non-Down's syndrome counterparts. The teratoma assay thus reflected a phenotypic feature — tumour vascularity — of a human multigenic condition, lending mechanistic insights into this human genotype–phenotype connection.

Non-cell-autonomous phenotypes

Using stem cell differentiation to study cell-autonomous phenotypes is a first step; however, phenotypes often arise from cellular interactions within tissues or across organs. To study non-cell-autonomous phenotypes, it might be necessary to pair the advantage of using native, iPSC-derived disease tissues with the power of co-culture and organoid systems. Where it is possible scientifically and ethically, transplantation of human PSC-derived tissues into mice will be extremely informative. Lastly, using PSC cultures that are derived from mouse models might also help to answer questions about mechanism and cell autonomy.

In vitro *assays: co-culture and organoids.* Genotype–phenotype mapping in ALS has benefited from PSC co-culture modelling using both human iPSCs[1] and ESCs[26,27]. Twenty-five per cent of patients with inherited ALS (10% of the total number of patients with ALS) have superoxide dismutase 1 (*SOD1*) mutations that, based on information gleaned from mouse studies, act in a dominant-negative fashion. Di Giorgio *et al.*[27] showed that motor neurons derived from mouse ESCs harbouring a mutation identical to the human mutant *SOD1* G93A allele showed cell-autonomous neurodegenerative features, including protein inclusions. When either wild-type or mutant motor neurons were co-cultured with *SOD1*-G93A mutant glia, both displayed degenerative features, suggesting that the mutation leads to both cell-autonomous and non-cell-autonomous effects on neurons. Chimeric mice harbouring this mutation also showed non-cell-autonomous influences on motor neurons, although the specific nature of cellular interactions could not be identified[28]. These ESC-based experiments offered an alternative to *in vivo* transplantation assays to explore cell intrinsic versus extrinsic phenomena.

The development of cell culture organoids from PSCs is an alternative approach to assess both cellular differentiation potential and interactions between different cell types on the same genetic background. Organoids derived from vascular, hepatic and chondrocyte tissues have been used with varying degrees of success[29], and in the case of neurospheres and mammospheres, have been integral to understanding stem cell dynamics in development and cancer[30–32]. The generation of crypt–villus organoids derived from leucine-rich-repeat-containing G protein-coupled receptor 5-positive (*Lgr5*$^+$) intestinal stem cells represents a key advance that can be applied to tissue-specific progenitors from PSCs[33–36]. For example, Spence *et al.*[36] have engineered a robust process to differentiate human PSCs into intestinal organoids using a series of growth factor treatments. They also performed monogenic disease modelling by showing that neurogenin 3 (NGN3), a factor mutated in an autosomal recessive disorder characterized by a paucity of enteroendocrine cells, is integral to enteroendocrine cell development in their system[36]. Although strong monogenic phenotypes such as this one can be clearly discerned in organoids, it will certainly be more challenging for non-structural, modest physiological phenotypes to be detected. That said, methods that combine organoid differentiation protocols with PSC genetics might allow the interrogation of the complex genetic basis of human disorders.

In vivo *assays: transplantation and chimaeras.* Although organoids or co-culture systems will be useful, it is necessary to integrate *in vivo* systems with PSC biology. In most cases, such as in the ALS and Down's syndrome examples previously mentioned, investigators will need to continue using both human PSCs and mouse models to dissect polygenic traits. The advantages of using human cells are amplified by exploring interactions with *in vivo* environments. A fundamental way to test these interactions is through the transplantation of differentiated products of patient-derived iPSCs into mouse models, in which the cell-autonomous contribution to disease can be tested by measuring physiologically relevant integration *in vivo*.

Teratoma formation assay
A teratoma is a tumour with tissue components resembling normal derivatives of all three germ layers. Teratoma formation is a key criterion for evaluating the pluripotency of human pluripotent stem cells, which are capable of forming multi-lineage tumours in immunodeficient murine hosts.

Cell-autonomous phenotype
If a gene's activity affects only those cells that express it, its function is cell-autonomous; if it affects cells other than (or in addition to) those expressing it, its function is non-cell-autonomous.

Organoid
A complex three-dimensional cluster of tissues comprising multiple cell types that self-organize into an organ-like structure, as generated by non-adherent culturing of stem or progenitor cells.

© 2011 Macmillan Publishers Limited. All rights reserved

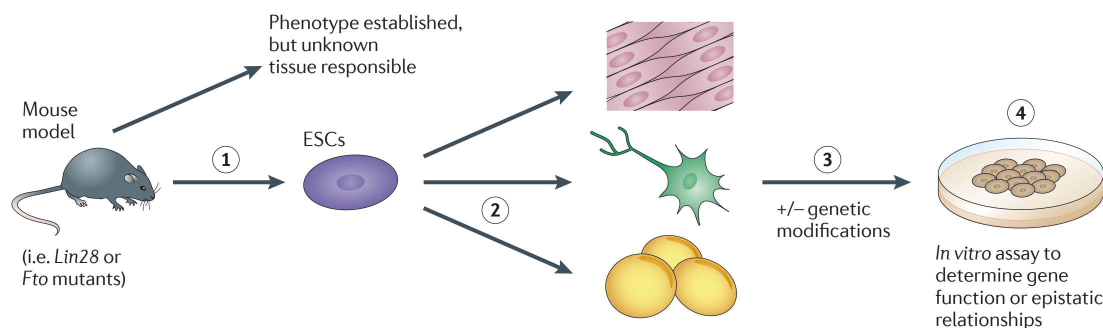


Figure 2 | **Using embryonic stem cells to mechanistically analyse phenotypes from mouse models.** Cell-culture systems can also be used to address mechanisms and cell autonomy hypotheses that might have arisen from compelling *in vivo* observations. In the cases of the fat mass and obesity associated (*Fto*) and *Lin28* transgenic mouse models, the cells or tissues that are responsible for these phenotypes are unclear and could be tested using directed differentiation and subsequent *in vitro* assays. Embryonic stem cells (ESCs) from *Fto* gain- or loss-of-function mice could be isolated (step **1**) and differentiated into adipocytes or other cell types (step **2**). Ultimately, phenotypes within these adipocytes could be assayed to determine cell autonomy (step **4**). If indeed cell-based phenotypes are defined, epistasis studies with candidate modifiers can be undertaken using gain- or loss-of-function short hairpin RNAs introduced at the ESC or differentiated cell stage (step **3**).

Simple transplant models that involve placing human cells into mouse hosts (that is, the generation of human–animal tissue chimaeras) require immunodeficient mice or 'humanized' mice containing human immune systems. Transplants into immunodeficient mice are routinely used to study benign and malignant haematopoietic diseases[37]. Efforts to engineer normal blood cells from iPSCs are also making use of such immunodeficient mice recipients. Human iPSC-derived β-cell function can also be tested after transfer into immunodeficient and diabetic mice[38]. In the case of type 1 diabetes, which results from an interplay among β-cells, the immune system and the environment, it may be necessary to reconstitute the mouse with both the patient-derived immune system and β-cells to recapitulate important aspects of the disease. Transplantation studies in other organ systems such as the brain will depend on the existence of mouse host models that are able to receive human iPSC-derived products in an informative and ethically sound fashion. For example, transplantation of either glia or motor neurons derived from iPSCs taken from patients with ALS could be useful to further define cell autonomy questions in that disease.

Complementing knowledge from mouse models. Cell culture systems can also be used to address mechanisms and cell autonomy hypotheses that might have arisen from compelling *in vivo* observations (FIG. 2). The functions of several genes identified by GWA studies have been successfully probed in mouse models, namely B-cell CLL/lymphoma 11A (*Bcl11a*) for fetal haemoglobin production[39,40], fat mass and obesity associated (*Fto*) for obesity and diabetes[41–43] and *Lin28a* for height/puberty[44]. In the cases of *Fto* and *Lin28a*, the cells or tissues responsible for these phenotypes are unclear and could be interrogated using directed differentiation and subsequent *in vitro* assays. For example, ESCs from *Fto* gain- or loss-of-function mice can be differentiated into adipocytes, which can be assayed (FIG. 2). If these *in vitro* phenotypes recapitulate the *in vivo* mouse phenotypes, then this provides greater confidence that human *FTO* variant alleles could be evaluated in the same way.

Characterizing regulatory variation

Mapping transcriptional events. The recent flood of findings from GWA studies has compelled investigators to design innovative ways to functionally interrogate genotype–phenotype relationships (BOX 3). It has been hypothesized that SNPs identified in GWA studies are associated with *cis*-regulatory variation buried within large regions of perigenic DNA. The first step in confirming such relationships is to accurately measure the transcriptional output of particular alleles relative to putative SNPs that are associated with a given trait. One of the greatest advantages of using patient-derived PSCs is that they contain all of the regulatory complexity underlying the phenotype, allowing SNPs identified by GWA studies to be tested for association with regulatory variation (BOX 3). Lee *et al.* attempted to map SNP-dependent regulatory variants by performing allele-specific sequencing and RNA-seq on iPSCs and differentiated human cells[45,46]. They used 'padlock probes' to specifically sequence 'reporter' SNPs (exonic expression SNPs) to determine whether these cell lines exhibited allele-specific expression (ASE). They were able to detect allelic imbalance ratios as small as 60/40 and map 27% of the expression SNPs; 3–10% of these expression SNPs were tissue-specific, indicating the sensitivity of this method to detect regulatory variation despite high levels of noise.

This work suggests that it is possible to study gene expression using reporter SNPs in relation to GWA-study marker SNPs by generating a cohort of iPSC lines bearing different haplotypes in genomic regions of interest (BOX 3). Under which circumstances would it be necessary to generate iPSCs to perform allele-specific RNA-seq? In cases in which transcriptional and post-transcriptional modulation of relevant loci are only

RNA-seq
High-throughput sequencing of cDNA that aims to obtain information about RNA content. RNA-seq measures transcriptome data, provides information on how different alleles of a gene are expressed, detects post-transcriptional mutations and identifies gene fusions.

Padlock probe
A probe with two target-complementary segments, which on hybridization are brought close to each other so that they can be covalently linked. This results in a circularized probe that is amenable to locus amplification and direct sequencing.

© 2011 Macmillan Publishers Limited. All rights reserved

Box 3 | **Pluripotent stem cells and genome-wide association studies**

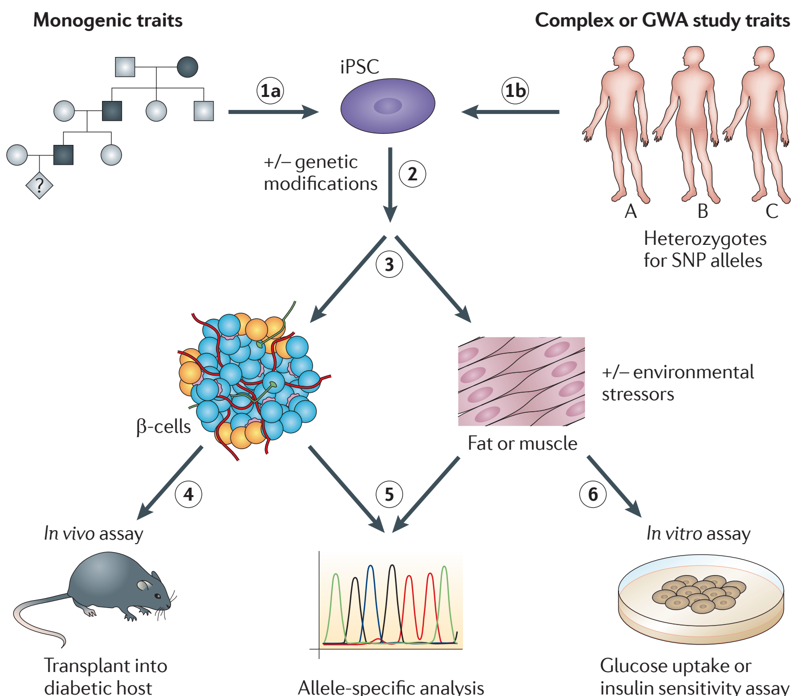
Genome-wide association (GWA) studies have identified hundreds of disease-associated loci, implicating both previously suspected and novel genes and pathways in pathogenesis. In a few cases, the sequencing of associated loci has identified missense and nonsense mutations, generating plausible hypotheses that link the association to disease status[70]. However, most associated loci do not harbour variants with an obvious functional link to the trait.

Often, it is hypothesized that these regions contain variants that affect gene regulation. As a first step to testing this hypothesis, investigators assess whether genes encoded near the variant are differentially expressed between genotypes (for example, between primary prostate and tumour tissues[71]); however, this step is not possible for all studies owing to the limited availability of primary cells. For this reason, and because follow-up functional studies require substantial biological material, pluripotent stem cell (PSC)-differentiated cells represent an attractive complement to primary cells, cell lines and animal models.

We envision a scenario in which fibroblasts, or more accessible cells, from GWA study participants and from patients with monogenic conditions are prospectively banked; a subset with and a subset without candidate alleles would be reprogrammed to pluripotency and subsequently differentiated towards relevant tissue(s) (see the figure, steps **1a**, **1b** and **3**). If the candidate variants are hypothesized to influence expression in *cis*, then a correlation between expression and genotypes could be tested immediately. Follow-up sequencing and complementation studies could demonstrate whether the variant causes the expression variation (see the figure, step **5**).

It would be even more interesting to test the extent to which forced modulation of expression alters a cell activity that is linked to the disease or trait: for example, insulin production in pancreas islet cells or insulin responsiveness in muscle, fat and liver. A straightforward approach would be to generate a GWA study candidate panel of gain- and loss-of-function embryonic stem cells (ESCs) using lentiviral transgenesis (see the figure, step **2**). In a more sophisticated fashion, targeted mutagenesis (in cells from control participants) or correction (in cells from affected individuals) of candidate variants combined with cell-based functional assays would test the contributing role of the variant (see the figure, steps **1b**–**6**). To enhance potentially weak phenotypes, environmental stressors could be used (see also BOX 4).

Functional dissection of suspected regulatory elements can also be performed using ESCs or induced PSCs (iPSCs)[72]. Defining the ways in which tissue-specific promoters or enhancers can fine-tune gene expression levels — as opposed to the binary 'on' or 'off' categories to which our current understanding is limited — will go a long way towards testing dose-appropriate gain, loss or rescue of genes from relevant GWA study loci (for use in step **2**). PSCs from mice or humans may soon be used to study human disease in a similar fashion to established fly, zebrafish and mouse models.



appreciable in an unobtainable tissue, generated iPSCs can be differentiated into tissues (and thus developmental contexts) in which the SNP has been proven to have some predictive power, followed by assaying for ASE of specific genes.

Mapping post-transcriptional events. Although the expression-mapping approach described above would not identify protein level differences, it would capture the transcriptional variation associated with SNPs identified from GWA studies. RNA-seq can extend this characterization to reveal post-transcriptional events, including RNA processing and alternative splicing[47,48]. As the set of alternatively spliced genes and products are highly cell-type-specific, it would be useful to isolate or generate the correct cell type before sequencing efforts are undertaken to determine the effect of SNPs on alternative splicing outputs. In these cases, iPSC generation, subsequent differentiation and genomic or proteomic level analyses could greatly facilitate the biological annotation of variants suggested by GWA studies.

Epigenetics versus genetics

Imperfect clearing of epigenetic memory. Although the genetic context of a cell is preserved with iPSCs, in taking a cell from terminal differentiation in a given developmental lineage back to the pluripotent state and then forward into a different lineage, one is altering the epigenetic landscape in marked and unpredictable ways. This was clearly illustrated recently by studies in our laboratory and the Hochedlinger laboratory, which showed that the epigenetic 'memory' of the original differentiated state may not be perfectly erased during reprogramming[49,50]. Functionally, this means that certain gene expression or expression-permissive chromatin states that are ordinarily restricted to the differentiated, parent cell type (for example, a blood cell or skin fibroblast) persist in the reprogrammed iPSCs. These chromatin states have a functional outcome: they impair differentiation of reprogrammed cells towards lineages different from their cell of origin and favour differentiation towards lineages close to the cell type of origin. Practically, this suggests that using a closely related parent cell or tissue type to make iPSCs is likely to be the most efficient way to achieve the directed differentiation of desired tissues[49,50].

Despite an incomplete understanding of the molecular basis of epigenetic memory, it remains possible to explore its imperfect clearing. One approach is to study the epigenetic memory of disease states, as this would highlight not only those conditions that are difficult to epigenetically reset but also the likelihood that the iPSC-derived cells might harbour deleterious epigenetic marks that will pose barriers to therapy (FIG. 3). It has long been known that somatic cell nuclear transfer (SCNT) in mice can lead to 'large offspring syndrome', an abnormal developmental condition that is similar in some ways to human Beckwith–Wiedemann syndrome. This begs the question of what epigenetic defects arise during SCNT[51] — in some cases, defects in insulin-like growth factor signalling have been implicated[52].

© 2011 Macmillan Publishers Limited. All rights reserved

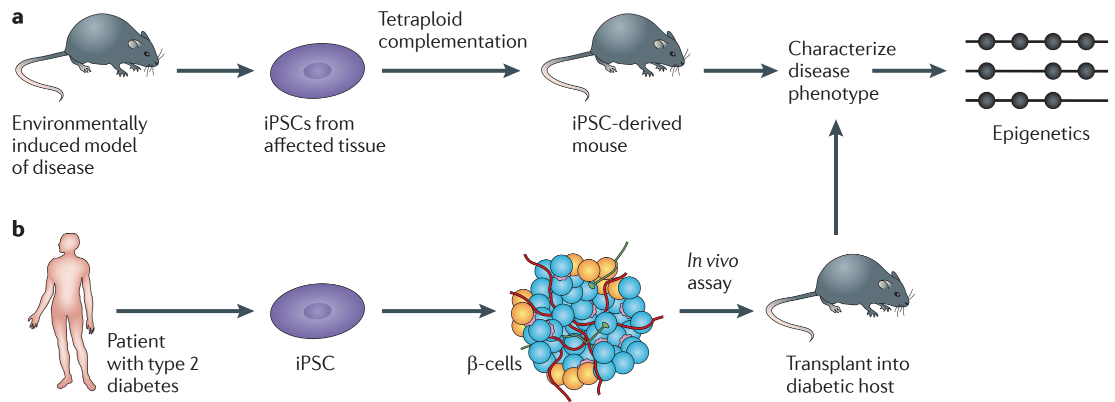


Figure 3 | **Testing for epigenetic contributions to disease. a** | One could first derive induced pluripotent stem cells (iPSCs) from the affected tissues of environmentally induced disease models such as obesity. Next, mice from these iPSCs could be generated with tetraploid complementation. Analysis of these mice for the original disease phenotype would indicate an epigenetic legacy of disease. **b** | Alternatively, patient-specific iPSCs from healthy and diseased individuals could be differentiated and tested in a variety of ways. Shown in panel **b** is an example of a human–mouse transplant assay. Although disease-related loci would be challenging to identify, new tools and data sets are being generated to help identify the landscape of these epigenetic alterations and, potentially, the causative loci.

Somatic cell nuclear transfer
(SCNT). This involves replacing the nucleus of an unfertilized egg cell with the nucleus from a differentiated 'somatic cell' (a skin cell, for example). Stimulating the resulting pseudo-zygote to begin dividing leads to the creation of a cloned blastocyst. Subsequent uterine transfer can lead to the birth of a cloned animal or, in a process sometimes called 'therapeutic cloning', donor-identical cells can be extracted from the blastocyst to generate embryonic stem cells for disease modelling.

Beckwith–Wiedemann syndrome
An overgrowth disorder characterized by an increased risk of childhood cancer. Common features include large birth weight and length, large tongue, abdominal wall defects, ear creases or ear pits, and neonatal hypoglycaemia.

Tetraploid complementation
This technique is used to test the pluripotency of pluripotent stem cells and to generate genetically modified animals. By combining cells from a tetraploid embryo with diploid embryonic stem cells (ESCs), the resultant fetus is entirely derived from ESCs, whereas the extra-embryonic tissues are exclusively derived from tetraploid cells.

Penetrance
The proportion of individuals carrying a particular variation of a gene that also express an associated trait or disease.

A recent report in rats showed that fathers fed a high-fat diet (HFD) gave birth to daughters with a predisposition to diabetes[53]. This would also suggest that islets derived from type 1 or 2 diabetics might also be epigenetically defective. One could imagine an experiment in which one used tetraploid complementation with iPSCs derived from the fat or islets of HFD-induced diabetic mice and then determined whether any resulting iPSC-derived mice were more highly predisposed to diabetes than controls (FIG. 3a). Such an experiment might suggest that making iPSC-derived β-cells using cells from a patient with diabetes might not fully erase any deleterious epigenetic signatures related to the pathogenesis of the patient's diabetes. The function of iPSC-derived β-cells generated from human diabetics could also be tested in immunocompromised mice (FIG. 3b). Such an approach could in theory be performed with any disease in which an epigenetic contribution is suspected. Although disease-related loci influenced by such reprogramming resistance would be difficult to identify, this would provide evidence for diseases in which epigenetics has a prominent role.

Given the incipient stage of this field, it is impossible to predict whether epigenetic memory will have a substantial negative impact on cells derived for therapy or a positive impact on determining the epigenetic basis of disease. Nevertheless, we hope that investigators will be motivated to address both epigenetic and polygenic problems using these PSC-derived assays.

Challenges and prospects

The predictable differentiation of PSCs into specific tissue lineages remains a great challenge that will continue to be met head-on by groups interested in the development and engineering of cellular fates[19]. This task is complicated by the fact that the developmental and disease phenotype of iPSCs and their derivatives may be heavily influenced by the epigenetic memory of their developmental or disease origins. By contrast, if a condition is thought to bear an epigenetic component, the act of cellular reprogramming may eradicate the very genomic influences one wishes to study. In addition, there are technical challenges being faced at the interface of genetic manipulation, gene therapy and PSCs.

Genetic modification of PSCs. A hurdle to generating genetically informative and potentially therapeutic cells is posed by the difficulties of creating mutations in human ESCs or iPSCs, which often obliges investigators to study existing mutations in patient-derived cells. The use of patient-derived iPSCs is complicated by the problem of finding appropriate non-mutated controls and matching genetic backgrounds, which is an important issue when studying low-signal-to-noise and low-penetrance phenotypes. Traditionally, age-matched, unaffected controls from within the same pedigree have been used in such studies, although they are often unavailable and represent an approximate control at best, as they can differ in genetic background, gender, age and exposure history. Introducing targeted genetic modifications at defined loci into human ESCs would ameliorate this problem, although this is currently an immature technology.

It is possible to achieve gain-of-function phenotypes (via transgenes) and loss of-function phenotypes (via short hairpin RNA (shRNA)) by using lentiviral transduction in either ESCs or iPSCs. However, this approach raises a few issues. First, it demands the use of crude *cis*-regulatory elements and does not take advantage of endogenous regulatory motifs at the actual genomic locus. Second, it poses the risk of potentially mutagenic viral insertion. Third, all current iPSC disease-modelling studies have used viral vectors during the reprogramming process, raising concerns about how basal vector expression affects gene expression. RNA- and small-molecule-based methods, which are in rapid development, are reviewed in detail in this issue[54]. By using

© 2011 Macmillan Publishers Limited. All rights reserved

Box 4 | Penetrance in complex disease models using pluripotent stem cells

Here we highlight some of the challenges posed by penetrance for modelling complex traits using pluripotent stem cells (PSCs). Some types of manipulation can increase penetrance and assay optimization and have been mentioned in the main text. For example, differentiating PSCs into the appropriate tissue is a necessary first step in generating the correct context to find phenotypic differences that follow from genetic alterations.

Modelling the correct timing of onset for a phenotype is perhaps as important a variable *in vitro* as it is in the clinical course of disease. Although some disorders are of late-adult onset, we should be aware that induced PSC (iPSC) passaging or differentiation might theoretically accelerate the onset of a phenotype *in vitro*. By contrast, if a strong disease phenotype is modelled early on *in vitro*, it could be selected out during passaging if the mutation is in flux within a heterogeneous epigenetic or genetic environment[3]. Although this *in vitro* system complicates the interpretation of the disease phenotype, it could nevertheless offer a model for the selection of therapeutically advantageous clones from a given PSC lineage and their subsequent enrichment relative to undesired populations in mixed cultures.

If the phenotype is late-onset and mirrors the human condition, environmental stresses could be used to elicit the phenotype, for example by using hypoxic, nutrient, toxic, tonic or thermal stress. The feasibility of this approach, however, remains unproven. As an example, in the cardiomyocyte context, certain long-QT-inducing drugs might exacerbate and make complete an otherwise incompletely penetrant phenotype.

Although the penetrance issue is particularly important for QTLs, it may be that cell-culture systems inherently lead to a reduction in phenotypic thresholds owing to the reduced ability of cells to achieve homeostasis *in vitro*.

these imperfect viral and nonviral strategies, it remains possible to create panels of human PSCs with gain- or loss-of-function modifications, which would be a more difficult task for mouse knockouts or knock-ins.

More modular methods of genetically engineering inducible, tissue-specific promoter transgenics, knock-ins and knockouts in human PSCs are being developed. Currently, these methods are inefficient owing to an impaired ability for human PSCs to grow as single cells, although the use of ROCK inhibitors may help, as they promote single-cell cloning efficiency[55]. Zou *et al.*[56] have used zinc finger nucleases to create double-strand breaks in targeted loci within human ESCs. They reported an efficiency of 0.14–0.24% in human ESCs and iPSCs, although with the caveat that unrelated, small and undetectable DNA changes (for example, microdeletions) were not fully characterized. Another approach would be to use bacterial artificial chromosomes (BACs) containing human gene variants and assaying phenotypes in the context of mouse ESCs, as was done for human *BRCA* variants or human β-globin gene regulation[57,58].

ROCK inhibitors
Y-27,632 is a selective inhibitor of p160-RHO-associated coiled-coil kinase (ROCK). Dissociation-induced apoptosis of human embryonic stem cells is reduced upon inhibitor treatment, and the resulting single-cell cloning efficiency is increased from 1% to 27%.

A key challenge in modelling complex traits is low penetrance or weak phenotypes. The study of less-penetrant phenotypes may be justified by the potentially smaller investment required to generate human genetic variants within iPSCs or ESCs in comparison to creating similar variants in mice. Each mutation and experimental question comes with its own specific options for altering penetrance and optimizing the dynamic range of an assay. Where penetrance is concerned, the field of multigenic, complex and even monogenic disease modelling using PSCs is so immature that to venture into any greater depth would be purely speculative. However, some general points are made in BOX 4.

Conclusion

Although PSCs are currently being used to probe human monogenic disease in cell-culture systems as a complement to mouse models, we anticipate that the next phase of iPSC research will capture human genetic variants in panels of iPSCs for direct assays of allele-specific transcriptional variations in phenotype-relevant tissues. This would be a major step forward in the confirmation and functional characterization of GWA study gene targets, provided that phenotypic signals can be detected in these differentiated tissues.

The potential manipulations used to push the boundaries of the genetic system are only limited by the creativity of each investigator. It may be that although these new-found sources of human materials lend themselves to the study of complex diseases, new assays will be required to adequately test these systems. The old saying "You can't get there from here…" may well describe the case for complicated diseases and predispositions that beg experimental study. Given that this Review is based on a foundational new technology that permits scientists to transform skin cells into insulin-producing β-islets, an ability to someday model *in vitro* a condition as challenging as diabetes may not be far-fetched.

1. Dimos, J. T. *et al.* Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* **321**, 1218–1221 (2008).
2. Park, I. H. *et al.* Disease-specific induced pluripotent stem cells. *Cell* **134**, 877–886 (2008).
This report describes the generation of a panel of iPSC lines for a range of human diseases.
3. Agarwal, S. *et al.* Telomere elongation in induced pluripotent stem cells from dyskeratosis congenita patients. *Nature* **464**, 292–296 (2010).
4. Tolar, J. *et al.* Induced pluripotent stem cells from individuals with recessive dystrophic epidermolysis bullosa. *J. Invest. Dermatol.* 2 Dec 2010 (doi:10.1038/jid.2010.346).
5. Unternaehrer, J. J. & Daley, G. Q. Induced pluripotent stem cells for modelling human diseases. *Phil. Trans. R. Soc. B* (in the press).
6. Marchetto, M. C. *et al.* A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* **143**, 527–539 (2010).
7. Raya, A. *et al.* Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. *Nature* **460**, 53–59 (2009).
8. Tulpule, A. *et al.* Knockdown of Fanconi anemia genes in human embryonic stem cells reveals early developmental defects in the hematopoietic lineage. *Blood* **115**, 3453–3462 (2010).
This study shows that the use of shRNA knockdown in human ESCs is a useful strategy for modelling human genetic diseases.
9. Urbach, A., Bar-Nur, O., Daley, G. Q. & Benvenisty, N. Differential modeling of fragile X syndrome by human embryonic stem cells and induced pluripotent stem cells. *Cell Stem Cell* **6**, 407–411 (2010).
10. Freund, C. & Mummery, C. L. Prospects for pluripotent stem cell-derived cardiomyocytes in cardiac cell therapy and as disease models. *J. Cell. Biochem.* **107**, 592–599 (2009).
11. Moretti, A. *et al.* Patient-specific induced pluripotent stem-cell models for long-QT syndrome. *N. Engl. J. Med.* **363**, 1397–1409 (2010).
This reference describes the use of human iPSCs to model long QT syndrome in the cardiovascular system.
12. Braam, S. R. *et al.* Prediction of drug-induced cardiotoxicity using human embryonic stem cell-derived cardiomyocytes. *Stem Cell Res.* **4**, 107–116 (2010).
13. Pfeufer, A. *et al.* Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nature Genet.* **41**, 407–414 (2009).
14. Newton-Cheh, C. *et al.* Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nature Genet.* **41**, 399–406 (2009).
15. Rashid, S. T. *et al.* Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. *J. Clin. Invest.* **120**, 3127–3136 (2010).
The authors describe the use of human iPSCs to model a series of monogenic liver diseases.
16. Billings, L. K. & Florez, J. C. The genetics of type 2 diabetes: what have we learned from GWAS? *Ann. N. Y. Acad. Sci.* **1212**, 59–77 (2010).
17. Borowiak, M. & Melton, D. A. How to make β cells? *Curr. Opin. Cell Biol.* **21**, 727–732 (2009).
18. Zhou, Q. & Melton, D. A. Pathways to new β cells. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 175–181 (2008).

© 2011 Macmillan Publishers Limited. All rights reserved

19. Cohen, D. E. & Melton, D. Turning straw into gold: directing cell fate for regenerative medicine. *Nature Rev. Genet.* **12**, 243–252 (2011).
20. Alipio, Z. *et al.* Reversal of hyperglycemia in diabetic mouse models using induced-pluripotent stem (iPS)-derived pancreatic β-like cells. *Proc. Natl Acad. Sci. USA* **107**, 13426–13431 (2010).
21. Blum, B. & Benvenisty, N. The tumorigenicity of human embryonic stem cells. *Adv. Cancer Res.* **100**, 133–158 (2008).
22. Yang, Q., Rasmussen, S. A. & Friedman, J. M. Mortality associated with Down's syndrome in the USA from 1983 to 1997: a population-based study. *Lancet* **359**, 1019–1025 (2002).
23. Baek, K. H. *et al.* Down's syndrome suppression of tumour growth and the role of the calcineurin inhibitor DSCR1. *Nature* **459**, 1126–1130 (2009).
This work describes the use of human iPSCs and the teratoma formation assay to explore the role of chromosome 21 genes in the reduced cancer incidence that is seen in Down's syndrome.
24. Reeves, R. H. *et al.* A mouse model for Down syndrome exhibits learning and behaviour deficits. *Nature Genet.* **11**, 177–184 (1995).
25. Lensch, M. W., Schlaeger, T. M., Zon, L. I. & Daley, G. Q. Teratoma formation assays with human embryonic stem cells: a rationale for one type of human–animal chimera. *Cell Stem Cell* **1**, 253–258 (2007).
26. Di Giorgio, F. P., Boulting, G. L., Bobrowicz, S. & Eggan, K. C. Human embryonic stem cell-derived motor neurons are sensitive to the toxic effect of glial cells carrying an ALS-causing mutation. *Cell Stem Cell* **3**, 637–648 (2008).
27. Di Giorgio, F. P., Carrasco, M. A., Siao, M. C., Maniatis, T. & Eggan, K. Non-cell autonomous effect of glia on motor neurons in an embryonic stem cell-based ALS model. *Nature Neurosci.* **10**, 608–614 (2007).
A study that used co-culture to show the non-cell-autonomous effects of glia on motor neurons in an ALS disease model using ESCs.
28. Clement, A. M. *et al.* Wild-type nonneuronal cells extend survival of SOD1 mutant motor neurons in ALS mice. *Science* **302**, 113–117 (2003).
29. Ochoa, E. R. & Vacanti, J. P. An overview of the pathology and approaches to tissue engineering. *Ann. N. Y. Acad. Sci.* **979**, 10–26; discussion 35–38 (2002).
30. Jensen, J. B. & Parmar, M. Strengths and limitations of the neurosphere culture system. *Mol. Neurobiol.* **34**, 153–161 (2006).
31. Ronnov-Jessen, L., Petersen, O. W., Koteliansky, V. E. & Bissell, M. J. The origin of the myofibroblasts in breast cancer. Recapitulation of tumor environment in culture unravels diversity and implicates converted fibroblasts and recruited smooth muscle cells. *J. Clin. Invest.* **95**, 859–873 (1995).
32. Ewald, A. J., Brenot, A., Duong, M., Chan, B. S. & Werb, Z. Collective epithelial migration and cell rearrangements drive mammary branching morphogenesis. *Dev. Cell* **14**, 570–581 (2008).
33. Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).
34. Snippert, H. J. *et al.* Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134–144 (2010).
35. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
36. Spence, J. R. *et al.* Directed differentiation of human pluripotent stem cells into intestinal tissue *in vitro*. *Nature* **470**, 105–109 (2011).
37. Behringer, R. R. Human–animal chimeras in biomedical research. *Cell Stem Cell* **1**, 259–262 (2007).
38. Brehm, M. A. *et al.* Human immune system development and rejection of human islet allografts in spontaneously diabetic *NOD-Rag1*^{null} *IL2rγ*^{null} *Ins2*^{Akita} mice. *Diabetes* **59**, 2265–2270 (2010).
39. Sankaran, V. G. *et al.* Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor *BCL11A*. *Science* **322**, 1839–1842 (2008).
40. Sankaran, V. G. *et al.* Developmental and species-divergent globin switching are driven by *BCL11A*. *Nature* **460**, 1093–1097 (2009).
41. Fischer, J. *et al.* Inactivation of the *Fto* gene protects from obesity. *Nature* **458**, 894–898 (2009).
42. Church, C. *et al.* Overexpression of *Fto* leads to increased food intake and results in obesity. *Nature Genet.* **42**, 1086–1092 (2010).
43. Church, C. *et al.* A mouse model for the metabolic effects of the human fat mass and obesity associated *FTO* gene. *PLoS Genet.* **5**, e1000599 (2009).
44. Zhu, H. *et al.* *Lin28a* transgenic mice manifest size and puberty phenotypes identified in human genetic association studies. *Nature Genet.* **42**, 626–630 (2010).
45. Lee, J. H. *et al.* A robust approach to identifying tissue-specific gene expression regulatory variants using personalized human induced pluripotent stem cells. *PLoS Genet.* **5**, e1000718 (2009).
The authors describe the use of allele-specific sequencing and RNA-seq to determine *cis*-regulatory variation in loci within iPSCs, fibroblasts and lymphoblasts.
46. Zhang, K. *et al.* Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. *Nature Methods* **6**, 613–618 (2009).
47. Cooper, T. A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136**, 777–793 (2009).
48. Majewski, J. & Pastinen, T. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends Genet.* **27**, 72–79 (2010).
49. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467**, 285–290 (2010).
This report shows that the epigenetic legacy of cell types used to generate iPSCs determines the differentiation efficiency and potency of those iPSCs.
50. Stadtfeld, M. *et al.* Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* **465**, 175–181 (2010).
This study shows that epigenetic silencing on chromosome 12qF1 limits the pluripotency of some mouse iPSCs.
51. Bortvin, A. *et al.* Incomplete reactivation of *Oct4*-related genes in mouse embryos cloned from somatic nuclei. *Development* **130**, 1673–1680 (2003).
52. Young, L. E. *et al.* Conservation of *IGF2-H19* and *IGF2R* imprinting in sheep: effects of somatic cell nuclear transfer. *Mech. Dev.* **120**, 1433–1442 (2003).
53. Ng, S. F. *et al.* Chronic high-fat diet in fathers programs β-cell dysfunction in female rat offspring. *Nature* **467**, 963–966 (2010).
54. González, F., Boué, S. & Izpisúa Belmonte, J. C. Methods for making induced pluripotent stem cells: reprogramming à la carte. *Nature Rev. Genet.* **12**, 231–242 (2011).
55. Watanabe, K. *et al.* A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nature Biotech.* **25**, 681–686 (2007).
56. Zou, J. *et al.* Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* **5**, 97–110 (2009).
57. Chang, S., Biswas, K., Martin, B. K., Stauffer, S. & Sharan, S. K. Expression of human *BRCA1* variants in mouse ES cells allows functional analysis of *BRCA1* mutations. *J. Clin. Invest.* **119**, 3160–3171 (2009).
58. Kaufman, R. M., Pham, C. T. & Ley, T. J. Transgenic analysis of a 100-kb human β-globin cluster-containing DNA fragment propagated as a bacterial artificial chromosome. *Blood* **94**, 3178–3184 (1999).
59. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl Acad. Sci. USA* **78**, 7634–7638 (1981).
60. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
61. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
62. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
63. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
64. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
65. Park, I. H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141–146 (2008).
66. Yamanaka, S. & Blau, H. M. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* **465**, 704–712 (2010).
67. Kiskinis, E. & Eggan, K. Progress toward the clinical application of patient-specific pluripotent stem cells. *J. Clin. Invest.* **120**, 51–59 (2010).
68. Hochedlinger, K. & Plath, K. Epigenetic reprogramming and induced pluripotency. *Development* **136**, 509–523 (2009).
69. Lensch, M. W. Cellular reprogramming and pluripotency induction. *Br. Med. Bull.* **90**, 19–35 (2009).
70. Galarneau, G. *et al.* Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nature Genet.* **42**, 1049–1051 (2010).
71. Pomerantz, M. M. *et al.* Analysis of the 10q11 cancer risk locus implicates *MSMB* and *NCOA4* i n human prostate tumorigenesis. *PLoS Genet.* **6**, e1001204 (2010).
72. Zhang, X., Horrell, S. A., Delaney, D. & Gottlieb, D. I. Embryonic stem cells as a platform for analyzing neural gene transcription. *Stem Cells* **26**, 1841–1849 (2008).
This study describes the use of zinc-finger nucleases in human PSC gene targeting.

Acknowledgements

G.Q.D. is an investigator of the Howard Hughes Medical Institute and the Manton Center for Orphan Disease Research. G.Q.D. is supported by grants from the US National Institutes of Health (RO1-DK70055, RO1-DK59279 and UO1-HL100001, and special funds from the American Recovery and Reinvestment Act (ARRA) stimulus package (RC2-HL102815)), the Roche Foundation for Anaemia Research, Alex's Lemonade Stand Foundation and the Harvard Stem Cell Institute.

Competing interests statement

G.Q.D. declares competing financial interests; see web version for details.

FURTHER INFORMATION

George Q. Daley's homepage: http://daley.med.harvard.edu
Nature Reviews Genetics Focus issue on Stem cells: http://www.nature.com/focus/stemcells/index.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2011 Macmillan Publishers Limited. All rights reserved