Review

Stem Cell Engineering and Differentiation for Disease Modeling and Cell-based Therapies

Lauren N. Randolph [1,2,§], **Yuqian Jiang** [1,2,§] **and Xiaojun Lian** [1,2,3,*]

[1] Department of Biomedical Engineering, The Pennsylvania State University, University Park, PA, 16802, USA

[2] The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA, 16802, USA

[3] Department of Biology, The Pennsylvania State University, University Park, PA, 16802, USA

* **Correspondence:** Email: Lian@psu.edu; Tel: 814-865-8093.
[§] These authors contributed equally.

Abstract: The identification and characterization of stem cells, especially human embryonic stem cells, has revolutionized the field of developmental biology by providing an *in vitro* system to study human development. In addition, reprogramming adult cells from patients into an embryonic stem cell-like state using induced pluripotent stem cell (iPSC) technology can potentially generate an unlimited source of human tissue carrying genetic mutations that caused or facilitated disease development, providing unprecedented possibilities to model human disease in the culture dish. To do this, however, efficient differentiation methods to direct iPSCs through multiple progenitor stages to yield homogeneous populations of somatic cells must be established. Furthermore, disease modeling using iPSCs requires proper controls for this "disease-in-a-dish" approach. Therefore, methods to efficiently engineer the genome of iPSCs to correct the mutations become vital in stem cell research. Here we reviewed the

iPSC generation techniques and several genome-editing tools, such as TALENs and CRISPR-Cas9, for performing iPSC gene knock-in and knockout. We also present several efficient stem cell directed differentiation methods for converting iPSCs to neural, hematopoietic, cardiac, and pancreatic lineages. Together, this knowledge will provide insight into design principles for disease modeling using iPSCs and stem cell-based therapies.

1. Introduction

The contemporaneous generation of human induced pluripotent stem cells (iPSCs) by the Yamanaka and Thomson groups has revolutionized the stem cell research field and provided new avenues of research, as well as renewed focus on the development of cell therapies and disease models [1,2]. The emergence of iPSC technology has provided a renewable stem cell source free of the controversy surrounding human embryonic stem cells (hESCs) and introduced the concept of cellular reprogramming, highlighting cell state plasticity. Additionally, this technology has provided an unprecedented opportunity to study human development, model genetic diseases, and develop therapeutics for the treatment of a variety of disease conditions, in particular degenerative disorders.

The loss of cell function or cell death occurring in a cell type unable to self-regenerate presents a challenge in the development of therapeutics for degenerative diseases, such as Parkinson's, heart failure, or type II diabetes. Cellular reprogramming and directed differentiation techniques are crucial to the development of cell-based therapies for degenerative diseases. This is due to the lack of a cell source for many difficult to obtain cells, such as neurons, cardiomyocytes, or beta cells, that are often the cells affected by loss of function or cell death in degenerative disease. Cellular reprogramming strategies carry great potential towards in situ patient application of therapeutic discoveries from disease modeling.

Directed differentiation strategies provide a means to generate these cells for creating disease models and studying healthy and diseased cell function *in vitro* and *in vivo*. A detailed understanding of the factors involved in directing cell fate as well as the identification of reliable markers of a specific cell type are crucial milestones to be met for cell based therapies and clinical translation. Using directed differentiation, it is possible to generate a patient derived somatic cell source to better explore therapeutic options and personalized medicine through disease modeling. Advances in the field of genetic engineering, including zinc finger nucleases, TALENs and CRISPR/Cas9, have aided the elucidation of differentiation pathways and markers of multi-potent and committed progenitors. This

technology has also provided the ability to recapitulate diseased or healthy genotypes in same cellular context lending unprecedented experimental control in the generation of disease models.

In this review, we will discuss the generation of iPSCs and progress on differentiation and reprogramming strategies for cell types from each of the germ layers in the context of developing cell based therapies. We will also review recent advances in genetic engineering strategies and prominent examples of application to disease modeling and the development of cell based therapies.

2.   Differentiation Strategies

Ideal differentiation strategies for clinical applications are fully defined and highly efficient. Successful development of these methods is commonly achieved through careful study and mimicry of human or model animal development. Initially, a strategy often relies on the use of animal serum, co-culture, or conditioned medias and/or growth factors; however, over time the protocol is refined to define all factors used and increase efficiency. The establishment of markers for progenitor cell types, lineage commitment, and terminally differentiated cell types are crucial in evaluating the success and efficiency of a differentiation protocol.

2.1.   Induced Pluripotent Stem Cells (iPSCs)

Derivation and differentiation of hESCs has offered great hope for regenerative medicine, disease modeling, drug discovery, and development biology. Nonetheless, application of hESCs has been faced with several issues. First, hESCs are derived from the inner cell mass of a human blastocyst [3], an early-stage preimplantation embryo, which sparks ethical controversy. In addition, it is technically challenging to derive patient-specific hESCs, and therefore transplantation with hESC-derived somatic cells can lead to immune rejection.

Generation of iPSCs from human somatic cells bypasses the utilization of human embryos; this partially resolves the problems above and holds promise for deriving patient-specific iPSCs. Derived iPSCs have similar morphology to hESCs with round shape, a large nucleolus, and scant cytoplasm, and they form sharp-edged, flat, tightly packed colonies. Moreover, iPSCs express similar genes to hESCs, like Oct-3/4, Sox2, Nanog, GDF3, REX1, FGF4, ESG1, DPPA2, DPPA4, and hTERT, and also have the potential to differentiate into all somatic cell types in the body, such as hepatocytes, pancreatic β cells (endoderm), chondrocyte, muscle cells (mesoderm) and nerve cells (ectoderm).

iPSC generation was first reported by the Shinya Yamanaka group using retroviral transduction of Oct4, Sox2, Klf4, and c-myc into mouse embryonic fibroblasts or tail-tip fibroblasts, which were engineered to survive drug selection determined by the expression of pluripotent marker Fbx15 [4]. Ultimately, 0.02% of cells were successfully reprogramed, and no viable chimeras were produced

after injecting generated cells into developing mouse embryos [4]. The second generation of mouse iPSCs was reported by using the same set of reprograming factors but with Nanog or Oct4 instead of Fbx15 as reporter genes for pluripotency. The researchers generated 0.05–0.08% iPSCs indistinguishable from mESCs as well as viable chimera mice, which is considered to be the 'gold standard' for pluripotency [5–7].

In November 2007, the Yamanaka group and the Thomson group independently reported the first generation of human iPSCs, marking another dramatic advance in the field of stem cell research [1,2]. The Yamanaka group used the same four reprogramming factors with a retroviral system leading to a final reprogramming efficiency of 0.02%. Considering c-myc, a proto-oncogene, caused cell death and differentiation of hESCs [8], the Thomson group used a different set of transcription factors: Oct4, Sox2, Nanog, and Lin28 in a lentiviral system, resulting in an efficiency of 0.03–0.05% for fetal fibroblasts and 0.01% for newborn fibroblasts.

Due to concerns about genomic insertion during iPSC generation, other non-integrating vehicles like adenovirus or sendai virus as well as other nonviral reprogramming methods, such as mRNA or miRNA transfection, minicircle vectors, episomal plasmids, the PiggyBac transposon system, or direct protein delivery, are under consideration for footprint-free reprogramming. However, these non-integrating methods often come with a trade-off with lower efficiency [9]. For instance, even with daily transfections, loss of the episomal vectors at a rate of 5% per generation leads to a reprogramming efficiency as low as 0.03%. In contrast, mRNA transfection has reached a slightly higher efficiency up to 4.4%, but only limited to transfections in human fibroblasts. Also, 2-week continuous transfections of mRNA make this approach rather labor-intensive and expensive.

2.2. *Motor Neurons*

Three general stages can be identified in the differentiation process from human pluripotent stem cells (hPSCs) to motor neurons (MN): neural induction, MN generation, and MN maturation [10]. First, neural induction is a phase when neural progenitors are specified from hPSCs, in the form of a neural rosette (2D) or embryoid body formation (3D). This phase is initiated by inhibiting both bone morphogen protein (BMP) and TGF-β signaling, usually with recombinant protein Noggin or small molecule Dorsomorphin as BMP inhibitors. MN progenitors can be characterized by expression of markers like Olig2, Pax6, Nestin, Islet1/2, and NKX6.1. Du et al. proposed the application of WNT agonist, CHIR99021, during MN progenitor differentiation, which was an efficient way to caudalize neural progenitors and reduced co-expression of Olig2 and Nkx2.2. Combination of CHIR, SB431542 (SB, inhibitor of Activin-Nodal signaling), and DMH1 (inhibitor of BMP signaling) leads to chemically defined generation of homogenous caudal neuroepithelial progenitors from hPSCs [11].

The next stage in MN generation has been primarily manipulated by retinoic acid (RA) signaling,

which induces neutralization and caudalization, and Sonic hedgehog (Shh), which directs ventralization of spinal neural progenitor cells; this strategy derives from work by Wichterle and colleagues who first utilized RA and Shh for MN differentiation with mouse embryonic stem cells (mESCs) in 2002 [12]. Purmorphamine is used to stimulate the Shh pathway, and RA is added for RA signaling. After this stage, MN progenitors will further differentiate with expression of Hb9, ChAT, Islet1 and pan-neural marker βIII-tubulin.

Mature, functional motor neurons share two physiological characteristics: spike frequency adaptation and rebound action potential firing [13]. To gain highly enriched functional motor neuronal populations, RA and Shh are supplemented with B27, cAMP, heparin, brain-derived neurotrophic factor (BDNF), and ascorbic acid. Construction of Olig2 or Hb9 knock-in reporters is also used to target and purify MN progenitors or mature MNs. Furthermore, NOTCH inhibition was reported to promote MN maturation and give rise to homogenous mature MNs with up to 90% efficiency [11].

Despite the advances in exploring the neuron development, failure in clinical translation from preclinical models remains challenging for neurodegenerative diseases. It is crucial to assure authenticity of cell fate, enrichment and scalability for clinical cell therapy, with a billion cells more than 95% enriched as an operating standard. However, according to a systematic comparison of the neural differentiation potential of different hESC and iPSC lines, large variations in conversion efficiency indicates divergent functional properties of the generated neurons, therefore limiting practical feasibility [14]. In addition, genome-wide gene expression of different hPSC-derived neurons reveals their fetal developmental state, leading to query about the ability of these cells to represent their adult counterparts, which also remains a problem in cardiomyocyte differentiation. Recent studies reported progerin-induced aging is available for hiPSC-derived neurons, bringing hope to constructing more adult-like and reliable neurodegenerative models [15]. As another way to further enhance maturation, neuron-glia co-culture paradigms and complex 3D organoids in combination with biomaterials have also been applied to simulate *in vivo* developmental environments and are likely to broaden our understanding of adult-onset diseases [16].

2.3. *Cardiomyocytes*

Several strategies for deriving cardiomyocytes *in vitro* have been developed since Doetschman et al. first observed spontaneous differentiation of cardiomyocytes in mouse embryoid bodies in 1985 [17]. Cardiomyocytes are easily characterized by spontaneous contraction; however, several molecular markers have been established for cells of the cardiac lineage and specifically for myocytes. Cardiomyocytes derive from the mesodermal germ layer requiring mesodermal specification as a first step in differentiation, which is commonly confirmed by Brachyury (gene T) expression. Following mesodermal specification, the emergence of ISL1 expression is known to mark commitment to

cardiovascular cell lineages giving rise to myocytes, smooth muscle, and endothelial cells [18]. NKX2.5 expression also marks the cardiac progenitor state and continued expression is observed upon differentiation to the cardiomyocyte cell type [19]. Cardiac specific contractile proteins troponin T(cTnT) and troponin I (cTnI) can be used to identify all myocytes; however, as the myocytes mature, markers that indicate the subtype of cardiomyocyte are useful. Ventricular cardiomyocytes are marked by expression of transcription factor IRX4 and a ventricular typed myosin light chain (MLC2v) [20,21]. Atrial cardiomyocytes are similarly identified by atrial typed myosin light chain (MLC2a), and nodal cardiomyocytes are often identified by the presence of specific ion channels [22].

Early directed differentiation strategies for deriving cardiomyocytes relied on embryoid body formation and often resulted in very low efficiencies [23,24]. Laflamme et al. developed a monolayer differentiation strategy that used Activin A and BMP4 to modulate cell signaling pathways and achieved about 30% efficiency in deriving cardiomyocytes [25]. This development along with further study help to elucidate the critical role of canonical Wnt signaling in cardiomyocyte differentiation [26]. A fully defined protocol using biphasic chemical manipulation of the Wnt/β-catenin was developed by Lian et al., and proved to be highly efficient with 98% purity [27–29]. Recent advances have further defined culture reagents for cardiomyocyte differentiation to be both chemically defined and albumin free [30].

One hurdle that faces clinical application of stem cell derived cardiomyocytes is the subtype heterogeneity present in the differentiated population. While many protocols preferentially yield more of one myocyte subtype, for example the protocol by Lian et al. favors ventricular myocytes, uniform populations are desirable for successful clinical application [27,29]. A study in a porcine model has shown that heterogeneous populations of cardiomyocytes transplanted into a ventricular myocardial infarction site will cause ectopic pacemaker activity and arrhythmias [31]. Work is being done to understand the key regulators in myocyte subtype specification. Early studies show that retinoic acid signaling may play a role in specifying atrial vs. ventricular myocyte types [30,32]. The ventricular subtype is of particular interest for the regeneration of heart muscle tissue post infarction or restoration of healthy heart tissue in heart failure. In order to see these iPSC derived cell types reach their full potential as therapeutics, consistent homogenous populations must be achieved. Other protocols have been developed to specifically produce sinoatrial node myocytes [22], with potential applications in situations of pacemaker dysfunction.

Once subtype specification is better understood, large-scale production of uniform cardiomyocyte populations remains a challenge facing both clinical application and disease modeling with this cell type. Cost permissive, defined generation of clinically viable quantities of hPSC-derived stem cells will be an important milestone to achieving translation of this technology. Further research looking into enhancing cardiomyocyte and progenitor proliferation could provide an insightful approach to this issue. Additionally, the time associated with the generation of mature cardiomyocytes of the adult phenotype limits clinical usefulness. Strategies to enhance cardiomyocyte maturation would be a significant step in

improving the time required to generate these cells. An alternative approach to this issue that would also aid *in vivo* survival of the regenerated tissue is cellular reprogramming. Cellular reprogramming efforts have emerged showing the successful chemical conversion of fibroblasts to cardiomyocytes; however efficiencies are very low (less than 8%) and native cardiac fibroblast cell sources have not been used with this technique [33].

As a result of these challenges, the clinical use of hPSC-derived cardiomyocytes for regenerative medicine applications is thus far limited. Studies have been conducted in rat and porcine models to evaluate the therapeutic benefits of stem cell derived cardiomyocytes injected into an infarct zone with mixed success [25,31]. Patient derived iPSCs from a patient with a genetically linked tachycardia have been used to demonstrate efficacy in patient specific drug screening [34]. With refinement of techniques to differentiate and reprogram cells to the cardiomyocyte fate, therapeutic application requires further study but remains promising.

2.4.   Hematopoietic Cells

One of the biggest challenges facing directed differentiation of hematopoietic lineages, in particular hematopoietic stem cells (HSCs), is a lack of defining characterizations and maintenance conditions for this cell state. There are many markers that are used to identify cells committed to the hematopoietic lineages such as CD31, CD34, CD43, and CD45; however, these markers are often expressed by both true HSCs and multi-potent progenitor cells, making the identification and isolation of HSCs difficult. The lack of culture conditions for HSC expansion and challenges associated with obtaining these cells represent roadblocks to characterizing the human HSC state and identifying reliable human HSC markers. The development of culture conditions for HSC expansion would not only propel investigation into the identity of HSCs, but it would dramatically impact clinical implementation of HSC transplant, increasing the harvested population of an already rare cell type. This could potentially allow the treatment of more patients from a single donor harvest.

Directed differentiation of HSCs is further complicated by two developmentally distinct stages of hematopoiesis, primitive and definitive, of which only definitive hematopoiesis gives rise to HSCs with long-term engraftment potential and the ability to produce both myeloid and lymphoid lineages [35]. When mimicking embryonic development, the two occurrences of hematopoiesis make it difficult to isolate the true HSC state, especially without specific markers. The HSC state is currently best identified by its long-term engraftment potential and the ability to regenerate all components of the myeloid and lymphoid lineages, and as a result, many studies focus on murine models where this is more easily studied. However the developmental events of hematopoiesis are highly conserved among mammals, so information gleaned from murine studies may translate to human studies.

Many hematopoietic differentiations rely heavily on growth factors. The first protocol using primarily growth factors to generate a CD45+ population of cells was developed by Chadwick et al. and used a combination of stem cell factor (SCF), Flt-3 ligand, interleukin-3 (IL-3), interleukin-6 (IL-6), granulocyte colony stimulating factor, and BMP4 [36]. More recently, Woods et al. developed a defined media for hematopoietic induction, MesoTotal, which relies on a cocktail of growth factors and serum for embryoid body culture and differentiation to CD34+/CD45+ cells [37]. However this strategy did not produce cells with long-term engraftment potential, as tested in an NSG mouse model [37]. Some studies have looked more closely at mimicking the development of the hemogenic endothelium and aorta-gonad-mesonephros (AGM) as these tissues give rise to HSCs capable of long-term engraftment in a mouse model. Ng et al. used a combination of chemical and growth factor signals to generate hematopoietic progenitor cells with similar characteristics to human AGM; however these cells still did not show long-term engraftment potential demonstrating the potential need for a maturation stage of differentiation [38].

Another study that mimicked the development of the hemogenic endothelium, produced cells with lymphoid potential for both T and B lymphocytes [39]. This strategy used a small molecule GSK3 inhibitor CHIR99021 to produce the hemogenic endothelium followed by growth factor treatment and co-culture with OP9 cells to differentiate both T and B lymphoid cells [39,40]. A recent publication from the Daley group showed the generation of HSCs using growth factors to generate the hemogenic endothelium and a lentiviral vector to express seven transcription factors identified by screening. The progenitor cells obtained were successfully engrafted into both primary and secondary mouse recipients using a humanized mouse model with myeloid, T-cell, and B-cell production [41]. Further work remains to better characterize the HSC state and efficiently generate HSCs with both myeloid and lymphoid potential as well as long-term engraftment ability. While the transplantation of HSCs is a standard of care for many hematological conditions and malignancies, hPSC-derived HSCs have yet to reach this potential due to the inability to generate this cell type as of yet. Once these cells are successfully derived, extensive characterization and significantly improved efficiency will be required to move into translational studies and disease modeling. Furthermore, the successful derivation of these cell types will provide an interesting convergence point for the fields of stem cell technology, genome editing, and immunotherapy.

2.5.   *Pancreatic β cells*

Differentiation of pancreatic β cells from hPSCs is composed of several stages: definitive endoderm (marked by SOX17), foregut/midgut, pancreatic endoderm (marked by PDX1), pancreatic progenitor (marked by PDX1 and NKX6.1), endocrine progenitor, and insulin secreting pancreatic β cells.

Early studies of pancreatic development identified important genes and signaling pathways for pancreatic lineages. In 2001, the Mckay group generated about 31.5% insulin+ cells with mESCs in 24–26 days [42], and in 2005, the Deng group shortened the differentiation to 2 weeks to derive glucose-regulated, insulin-producing cells from mESCs [43]. D'Amour et al. transplanted hESC-derived definitive endoderm cells *in vivo*, which further progress to pancreatic cells [44]. Later, D'Amour et al. generated endocrine cells from hESCs that were polyhormonal and secreted 7% insulin using an 11–18 day protocol. The generated cells respond to multiple secretory stimuli, but minimally to glucose [45]. In 2007, Jiang et al. derived insulin-producing islet-like clusters in a serum-free system containing 15% human C-peptide positive cells with glucose sensing machinery and excitation properties. These cells were still polyhormonal with glucagon and somatostatin secretion but could survive in nude mice and maintain marker expression. After transplantation of these cells *in vivo*, 30% of the mice exhibited restoration of stable euglycemia [46]. In 2011, Nostro et al. generated up to 25% c-peptide+ cells from hPSCs in 20 days; they reported that inhibition of TGFβ signaling is essential for endocrine lineage commitment, and Wnt signaling enhances endocrine development [47]. In 2014, based on previous protocols, the Melton group extended the time in culture with the FGF family member KGF, hedgehog inhibitor SANT1, and a low concentration of retinoid acid (RA), increasing the efficiency to 55% NKX6-1/PDX-1 pancreatic progenitors. They then selected 11 different factors out of >150 combinations with >70 compounds, involving Wnt, Activin, hedgehog, EGF, TGFβ, thyroid hormone, RA, and γ-secretase inhibition, generating about 33% NKX6-1+/C-peptide+ cells that can secrete insulin in a glucose-regulated manner *in vitro* and *in vivo*. In 2015, the Keller group increased the efficiency of NKX6.1+ cells to 83.1% [48].

For further clinical applications, the Melton group extended the approach to generate the stem cell derived pancreatic β cells from type 1 diabetic patients (T1D) in 2016. Those constructed T1D stem cells derived β (SC-β) cells hold great promise for treatment due to expression of β-cell markers, response to glucose both *in vitro* and *in vivo*, ability to prevent alloxan-induced diabetes in mice, and response to anti-diabetic drugs. Moreover, no major difference was observed between T1D SC-β and SC-β cells, which doesn't exclude the genetic causes of T1D but provides alternative approaches to address the problem of allo- or xenogeneic rejection [49].

So far, one of the biggest obstacles in pancreatic differentiation remains that this process is time-consuming and laborious, involving different growth factors or chemicals in each stage. Another problem is low production efficiency for mature pancreatic β cells that can secrete c-peptide in response to glucose stimulus. Therefore, much work remains to be done to fully understand the pivotal regulating pathways and generate a simplified, chemically defined protocol to increase the efficiency of pancreatic differentiation, especially from patient-specific iPSCs that hold promise for immune safety.

3. Genetic Engineering

3.1. Zinc Finger Nucleases

Zinc finger nucleases (ZFNs) are the chronological first in a series of proteins used to manipulate the human genome. ZFNs were first developed by Kim et al. in 1996 by fusing the DNA binding domain of a restriction enzyme FokI to alternative DNA binding zinc fingers [50]. These zinc fingers could be assembled in different combinations to generate unique targeting sites [51]. ZFNs could generate double stranded breaks when dimerized and facilitate genetic manipulations via non-homologous end joining (NHEJ) to generate a mutation or knock out or homology directed recombination (HDR) to generate a knock-in or specific modification [51]. One of the primary challenges with genetic engineering using ZFNs is a high probability of off target binding leading to off target modifications and toxicity. Additionally, the cost associated with implementing this technology presented a challenge to both research and clinical application, necessitating the development of more cost-effective techniques. The increased cost associated with the technology limits patient accessibility to any therapy that is developed using ZFNs. This has spurred further research into genome editing technologies with increased ease of use and lower cost.

3.2. TALENs

Transcription activator-like effector nucleases (TALEN) are assembled with a TAL effector DNA-binding domain found in *Xanthomonas* bacteria and a FokI DNA cleavage domain. Theoretically, they can be engineered to target any desired DNA sequence and induce double-strand breaks (DSB). In contrast to zinc-finger proteins, there was no re-engineering of the linkage between repeats necessary for TALEs construction. Additionally, most of the intellectual property surrounding the ZFN technology platform is proprietary and expensive, making TALEN technology more promising [52].

There are a few key limitations with the TALEN technology. One problem is that it remains unknown why different TALEs designed according to the same cipher act on their targets with different levels of activity, which could be due to yet-unknown sequence dependencies for binding efficiency. This limitation is further compounded by the low transfection efficiencies inherent to pluripotent stem cells, prompting the continuation of future research to improve both transfection and genome editing technologies. Another targeting limitation is that TALE binding sites should start with a tyrosine base. While there is a 25% chance that a desired binding site will have tyrosine as its starting base, this could limit some clinical applications aiming to edit a disease genotype limited to a single loci. Perhaps the most detrimental issue to clinical application is unexpected genome binding can lead to off-target effects, which are difficult to detect without functional assays at that loci. The use of stem cells with off target

mutations for regenerative medicine applications could limit or negate therapeutic efficacy as a result of unforeseen genetic mutations. Side effects caused by off target mutations could complicate clinical trials by muddling patient outcomes as well as create increased challenges with FDA approval of treatments involving TALEN technology. Future development of strategies to minimize and screen for off target mutations will be an important arm of research for translating genome editing techniques to the clinic.

3.3. *CRISPR/Cas9*

Clustered regularly interspaced short palindromic repeats (CRISPR) are DNA segments containing short repetitive sequences firstly described by Yoshizumi et al. [53]. Jasen et al. reported the existence of CRISPR-associated systems (Cas), a set of homologous genes located upstream of repeats showing helicase and nuclease motifs [54]. Later, the CRISPR-Cas system was theoretically and experimentally shown to play a key role in the adaptive immune system of bacteria by spacer acquisition from viral DNA and RNA-guided degradation of foreign DNA [55,56]. Engineering and application of the type II CRISPR-Cas9 system, as first proposed by Zhang et al., has emerged as a breakthrough for highly specific genome editing [57]. With the guidance of customized crRNA/tracrRNA or combined single-guide RNA (sgRNA), a Cas9 nuclease can target and cause DSBs in any desired site near a protospacer adjacent motif (PAM) sequence making CRISPR-Cas9 a genome editor with high fidelity and relatively easy construction [57].

While still relying on transfection, which has low efficiency in hPSCs and other difficult to transfect cells such as neurons, one of the advantages of using CRISPR-Cas9 as opposed to other genome editing techniques is the increased editing efficiency over existing techniques. Optimization of donor sequence design, including circular or linearized delivery, homology arm length, and sgRNA selection, is the subject of ongoing research to provide further improvements of efficacy for HDR mediated editing [58]. This will only continue to increase the ease with which researchers can make precise changes to the genome, leading to easier dissection of the roles genetic elements play in development or disease states aimed at the generation of therapeutics. One of the limitations of CRISPR-Cas9 technology is that potential off-target cleavage activity could still occur on a DNA sequence with even three to five base pair mismatches in the PAM-distal part of the sgRNA-guiding sequence [59]. There are several ways to improve specificity one of which is sgRNA sequence modification by shortening the region complementary to the target site at the 5' end of the sgRNA or addition of two guanine nucleotides to the 5' end of the sgRNA. Another potential strategy is to control the Cas9-sgRNA complex by titrating the amount of Cas9 and sgRNA delivered. Guilinger et al. also generated fusion of catalytically inactive Cas9 with FokI nuclease domain to further improve cleavage specificity [60]. In spite of this limitation, the ease of use, improved cost efficiency, and increased precision of CRISPR-Cas9 provides an advantage over existing technology and holds great promise both

as a tool in stem cell research, but also for the development of disease models using relevant hPSC derived somatic cell types.

3.4.   Applications

One of the many applications of genetic engineering to stem cell differentiation, and eventually stem cell therapy, is the identification of differentiation state markers and crucial factors in directing cell fate. Knowledge of the markers regulating cell fate decisions can provide novel insight to human development, inform the development of directed differentiation and cellular reprogramming studies, and provide a stepwise measure by which to improve the efficiency of these methods. As the hPSCs themselves are not the target cell type for therapeutic application, efficient generation of therapeutically applicable somatic cell types and a thorough understanding of the developmental processes underlying their differentiation are crucial to successful clinical translation. Furthermore, dissecting the role of specific genetic elements in disease states can lead better understanding of the disease, prompting the development of more effective therapies.

This goal is commonly achieved through the construction of reporter cell lines by knocking in fluorescent proteins, for example green fluorescent protein (GFP), targeting genes of interest. For example, targeting GFP to the Oct4 (also known as POUF1) gene identifies the pluripotency of cells, like commercially-available Oct4-GFP H1 cell line from the Thomson group. This could be a useful tool for quickly assaying pluripotency while refining routine *in vitro* culture techniques for these cells. This would not only benefit stem cell research, but also clinical translation. One of the promises of iPSC technology is the realization of personalized regenerative medicine. In order to deliver on this potential, large-scale cost effective culture techniques will be required, and cells lines engineered to indicate maintenance of pluripotency would aide in these efforts.

Construction of reporter cell lines can play a key part in enhancing the differentiation efficiency for cell therapy. For example, an Mixl1-GFP knock in identifies human primitive streak-like cells and has enabled isolation of primitive hematopoietic precursors [61]. An NKX2.5 knock in reporter cell line was generated and used to successfully isolate cardiac progenitors and cardiomyocytes, identifying NKX2.5 as an important marker of the cardiac cell fate [19]. Recently, CRISPR-Cas9 genetic engineering techniques were used to create a homozygous knock out of β-catenin to demonstrate its critical role in mesendoderm differentiations [62]. Several knock-in reporter cell lines have been generated to improve differentiations to cell states marked by Sox17 expression, such as definitive endoderm differentiation and hemogenic endothelium [38, 63-64]. Sox17 reporter hPSCs help to characterize endoderm differentiation and are beneficial to exploring the role of Sox17 gene in organogenesis [65]. The use of these engineered cell lines provides valuable insight to developmental regulation as well as providing a metric for stepwise improvement of directed differentiation or cellular reprogramming efficiencies. To

see clinical translation of stem cell derived cell types, large quantities are needed in a rapid and cost effective manner, demanding efficient methodology.

Genetic engineering is also very useful for generating disease or healthy genotypes to model disease and establish efficacy for therapeutic development. Hinson et al. used a combination of patient derived iPSCs from patients with Titin mutations and CRISPR/Cas9 generated mutations to model dilated cardiomyopathy [66]. By using CRIPSR/Cas9 to recapitulate mutations in an otherwise genotypically normal stem cell, they were able to determine that the Titin mutations were the primary cause of sarcomere insufficiency as it leads to genetically linked dilated cardiomyopathy [66]. This deeper understanding of genetically linked dilated cardiomyopathy could lead to the development of therapeutics to definitively treat this condition, which is currently managed through drug regimens lifestyle changes. Another study by Nelson et al. explored the delivery of CRISPR-Cas9 to remove a mutant exon occurring in the dystrophin gene in Duchenne muscular dystrophy [67]. They were able to see partial recovery of dystrophin function with the removal of the mutant exon via systemic viral delivery of CRISPR-Cas9 in a murine model of Duchenne muscular dystrophy [67]. The successful delivery of CRISPR-Cas9 using in an *in vivo* system is a step towards direct application of this technology in humans for genome editing in adults. Non-human primate studies have used TALEN genome editing techniques to perform genome editing at the embryo stage to generate a disease model of Rett Syndrome [68]. The model recapitulated human genotype and phenotype associated with the disease. This demonstrates the vast potential of genome editing technology to provide accurate disease models for the development of novel therapeutics. While challenges in translating these discoveries to clinical application remain, the increasing precision and ease of genetic engineering techniques is opening the door for more complex disease modeling to more effectively recapitulate disease states and hopefully generate improved therapeutic options. With the relatively rapid ubiquitination of genome editing technology, some of the biggest hurdles to clinical translation are ethical concerns, which could slow the emergence of this technology in the clinic. However, the use of genome editing in stem cell derived somatic cell types for disease modeling or regenerative medicine applications could see a more immediate impact in therapeutic options.

4.   Conclusion

The convergence of major advances in genome-editing technologies, stem cell technology, and efficient directed differentiation protocols to generate functional adult cell lineages is ushering in a new era of using stem cells to study disease and for therapy. CRISPR-Cas9 technology together with stem cell technology have now supported the feasibility of generating authentic knockin reporter lines for screening new chemicals and growth factors critical for directed differentiation, for performing genome-wide genetic interrogation of the important genes for human development, and for designing

gene knockout approaches to study inherited genetic disorders. Gene knockout stem cell model systems, along with other large animal model systems, have already been generated and will ultimately allow direct *in vivo* interrogation of the physiological phenotype. In short, a new renaissance in stem cell disease modeling and cell-based therapies appears to be on the horizon.

Conflict of Interest

All authors declare no conflicts of interest in this paper.

References

1. Takahashi K, Tanabe K, Ohnuki M, et al. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131: 861-872.
2. Yu J, Vodyanik MA, Smugaotto K, et al. (2007) Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Stem Cell Rev* 8: 693-702.
3. Thomson JA, Itskovitz-Eldor J, Shapiro SS, et al. (1998) Embryonic stem cell lines derived from human blastocysts. *Science* 282: 1145-1147.
4. Takahashi K, Yamanaka S (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* 126: 663-676.
5. Okita K, Ichisaka T, Yamanaka S (2007) Generation of germline-competent induced pluripotent stem cells. *Nature* 448: 313-317.
6. Wernig M, Meissner A, Foreman R, et al. (2007) In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* 448: 318-324.
7. Maherali N, Sridharan R, Xie W, et al. (2017) Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and Widespread Tissue Contribution. *Cell Stem Cell* 1: 55-70.
8. Sumi T, Tsuneyoshi N, Nakatsuji N, et al. (2007) Apoptosis and differentiation of human embryonic stem cells induced by sustained activation of c-Myc. *Oncogene* 26: 5564-5576.
9. Baharvand H, Hassani S (2013) Pluripotent Stem Cells. 997: 13-23.
10. Faravelli I, Bucchia M, Rinchetti, et al. (2014) Motor neuron derivation from human embryonic and induced pluripotent stem cells: experimental approaches and clinical perspectives. *Stem Cell Res Ther* 5: 87.
11. Du ZW, Chen H, Liu H, et al. (2015) Generation and expansion of highly pure motor neuron progenitors from human pluripotent stem cells. *Nat Commun* 6: 6626.
12. Wichterle H, Lieberam I, Porter JA, et al. (2002) Directed differentiation of embryonic stem cells into motor neurons. *Cell* 110: 385-397.
13. Takazawa T, Croft GF, Amoroso MW, et al. (2012) Maturation of spinal motor neurons derived

from human embryonic stem cells. *PLoS One* 7: 1-9.

14. Broccoli V, Rubio A, Taverna S, et al. (2015) Overcoming the hurdles for a reproducible generation of human functionally mature reprogrammed neurons. *Exp Biol Med* 240: 787-794.

15. Miller JD, Ganat YM, Kishinevsky S, et al. (2013) Human iPSC-Based Modeling of Late-Onset Disease via Progerin-Induced Aging. *Cell Stem Cell* 13: 691-705.

16. Clevers H (2016) Modeling Development and Disease with Organoids. *Cell* 165: 1586-1597.

17. Doetschman TC, Eistetter H, Katz M, et al. (1985) The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol* 87: 27-45.

18. Bu L, Jiang X, Martinpuig S, et al. (2009) Human ISL1 heart progenitors generate diverse multipotent cardiovascular cell lineages. *Nature* 460: 113-117.

19. Elliott DA, Braam SR, Koutsis K, et al. (2011) NKX2-5(eGFP/w) hESCs for isolation of human cardiac progenitors and cardiomyocytes. *Nat Methods* 8: 1037-1040.

20. Bao ZZ, Bruneau BG, Seidman JG, et al. (1999) Regulation of chamber-specific gene expression in the developing heart by Irx4. *Science* 283: 1161-1164.

21. O'Brien TX, Lee KJ, Chien KR (1993) Positional specification of ventricular myosin light chain 2 expression in the primitive murine heart tube. *Proc Natl Acad Sci USA* 90: 5157-5161.

22. Protze SI, Liu J, Nussinovitch U, et al. (2017) Sinoatrial node cardiomyocytes derived from human pluripotent cells function as a biological pacemaker. *Nat Biotechnol* 35: 56-68.

23. Xu C, Police S, Rao N, et al. (2002) Characterization and enrichment of cardiomyocytes derived from human embryonic stem cells. *Circ Res* 91: 501-508.

24. Laflamme MA, Gold J, Xu C, et al. (2005) Formation of human myocardium in the rat heart from human embryonic stem cells. *Am J Pathol* 167: 663-671.

25. Laflamme MA, Chen KY, Naumova AV, et al. (2007) Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nat Biotechnol* 25: 1015-1024.

26. Paige SL, Osugi T, Afanasiev OK, et al. (2010) Endogenous Wnt/Beta-Catenin signaling is required for cardiac differentiation in human embryonic stem cells. *PLoS One* 5: e11134.

27. Lian X, Palecek SP (2012) Cozzarelli Prize Winner: Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc Natl Acad Sci* 109: E1848-E1857.

28. Lian X, Zhang J, Azarin SM, et al. (2013) Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/β-catenin signaling under fully defined conditions. *Nat Protoc* 8: 162-175.

29. Lian X, Bao X, Zilberter M, et al. (2015) Chemically defined, albumin-free human cardiomyocyte generation. *Nat Methods* 12: 595-596.

30. Pei F, Jiang J, Bai S, et al. (2017) Chemical-de fi ned and albumin-free generation of human atrial and ventricular myocytes from human pluripotent stem cells. *Stem Cell Res* 19: 94-103.

31. Kehat I, Khimovich L, Caspi O, et al. (2004) Electromechanical integration of cardiomyocytes derived from human embryonic stem cells. *Nat Biotechnol* 22: 1282-1289.

32. Zhang Q, Jiang J, Han P, et al. (2011) Direct differentiation of atrial and ventricular myocytes from human embryonic stem cells by alternating retinoid signals. *Cell Res* 21: 579-587.

33. Cao N, Huang Y, Zheng J, et al. (2016) Conversion of human fibroblasts into functional cardiomyocytes by small molecules. *Science* 352: 1216-1220.

34. Maizels L, Huber I, Arbel G, et al. (2017) Patient-Specific Drug Screening Using a Human Induced Pluripotent Stem Cell Model of Catecholaminergic Polymorphic Ventricular Tachycardia Type 2. *Circ Arrhythmia Electrophysiol* 10: 1-15.

35. Ackermann M, Liebhaber S, Klusmann JH, et al. (2015) Lost in translation: pluripotent stem cell-derived hematopoiesis. *EMBO Mol Med* 7: 1388-1402.

36. Chadwick K, Wang L, Li L, et al. (2003) Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells. *Blood* 102: 906-915.

37. Woods NB, Parker AS, Moraghebi R, et al. (2011) Brief report: efficient generation of hematopoietic precursors and progenitors from human pluripotent stem cell lines. *Stem Cells* 29: 1158-1164.

38. Ng ES, Azzola L, Bruveris FF, et al. (2016) Differentiation of human embryonic stem cells to HOXA+ hemogenic vasculature that resembles the aorta-gonad-mesonephros. *Nat Biotechnol* 34: 1168-1179.

39. Galat Y, Dambaeva S, Elcheva I, et al. (2017) Cytokine-free directed differentiation of human pluripotent stem cells efficiently produces hemogenic endothelium with lymphoid potential. *Stem Cell Res Ther* 8: 67.

40. Lian X, Bao X, Ai-Ahmad A, et al. (2014) Efficient Differentiation of Human Pluripotent Stem Cells to Endothelial Progenitors via Small-Molecule Activation of WNT Signalling. *Stem Cell Rep* 3: 804-816.

41. Sugimura R, Jha DK, Han A, et al. (2017) Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature* 545: 432-438.

42. Lumelsky N, Blondel O, Laeng P, et al. (2001) Differentiation of Embryonic Stem Cells to Insulin-Secreting Structures Similar to Pancreatic Islets. *Science* 292: 1389-1394.

43. Shi Y, Hou L, Tang F, et al. (2005) Inducing Embryonic Stem Cells to Differentiate into Pancreatic β Cells by a Novel Three-Step Approach with Activin A and All- *Trans* Retinoic Acid. *Stem Cells* 23: 656-662.

44. D'Amour KA, Agulnick AD, Eliazer S, et al. (2005) Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat Biotechnol* 23: 1534-1541.

45. D'Amour KA, Bang AG, Eliazer S, et al. (2006) Production of pancreatic hormone–expressing endocrine cells from human embryonic stem cells. *Nat Biotechnol* 24: 1392-1401.

46. Jiang J, Au M, Lu K, et al. (2007) Generation of Insulin-Producing Islet-Like Clusters from Human Embryonic Stem Cells. *Stem Cells* 25: 1940-1953.

47. Nostro MC, Sarangi F, Ogawa S, et al. (2011) Stage-specific signaling through TGFβ family members and WNT regulates patterning and pancreatic specification of human pluripotent stem cells. *Development* 138: 861-871.

48. Pagliuca FW, Millman JR, Gurtler M, et al. (2014) Generation of functional human pancreatic β cells in vitro. *Cell* 159: 428-439.

49. Millman JR, Xie C, Van DA, et al. (2016) Generation of stem cell-derived β-cells from patients with type 1 diabetes. *Nat Commun* 7: 11463.

50. Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci USA* 93: 1156-1160.

51. Porteus MH, Carroll D (2005) Gene targeting using zinc finger nucleases. *Nat Biotechnol* 23: 967-973.

52. Sanjana NE, Cong L, Zhou Y, et al. (2012) A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* 7: 171-192.

53. Ishino Y, Shinagawa H, Makino K, et al. (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* 169: 5429-5433.

54. Jansen R, Embden JDA, Gaastra W, at al. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 43: 1565-1575.

55. Mojica FJM, Díez-Villaseñor C, García-Martínez J, et al. (2005) Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J Mol Evol* 60: 174-182.

56. Hsu PD, Lander ES, Zhang, F (2014) Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* 157: 1262-1278.

57. Cong L, Ran FA, Cox D, et al. (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* 339: 819-823.

58. Zhang JP, Li XL, Li GH, et al. (2017) Efficient precise knockin with a double cut HDR donor after CRISPR / Cas9-mediated double-stranded DNA cleavage. *Genome Biol* 18: 35.

59. Zhang XH, Tee LY, Wang XG, et al. (2015) Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther Nucleic Acids* 4: e264.

60. Tsai SQ, Wyvekens N, Khayter C, et al. (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol* 32: 569-576.

61. Davis RP, Ng ES, Costa M, et al. (2008) Targeting a GFP reporter gene to the MIXL1 locus of human embryonic stem cells identifies human primitive streak – like cells and enables isolation of primitive hematopoietic precursors. *Blood* 111: 1876-1884.

62. Lian X, Xu J, Bao X, et al. (2016) Interrogating Canonical Wnt Signaling Pathway in Human Pluripotent Stem Cell Fate Decisions Using CRISPR-Cas9. *Cell Mol Bioeng* 9: 325-334.

63. Wang H, Luo X, Yao L, et al. (2015) Improvement of Cell Survival During Human Pluripotent Stem Cell Definitive Endoderm Differentiation. *Stem Cells Dev* 24: 2536-2546.

64. Clarke RL, Yzaguirre AD, Yashiro-Ohtani Y, et al. (2013) The expression of Sox17 identifies and regulates haemogenic endothelium. *Nat Cell Biol* 15: 502-510.

65. Wang P, Rodriguez RT, Wang J, et al. (2011) Targeting SOX17 in human embryonic stem cells creates unique strategies for isolating and analyzing developing endoderm. *Cell Stem Cell* 8: 335-346.

66. Hinson JT, Chopra A, Nafissi N, et al. (2015) Titin mutations in iPS cells define sarcomere insufficiency as a cause of dilated cardiomyopathy. *Science* 349: 982-987.

67. Nelson CE, Hakim CH, Ousterout DG, et al. (2016) In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystropy. *Science* 351: 403-408.

68. Chen Y, Yu J, Niu Y, et al. (2017) Modeling Rett Syndrome Using TALEN-Edited MECP2 Mutant Cynomolgus Monkeys. *Cell* 169: 945-955.