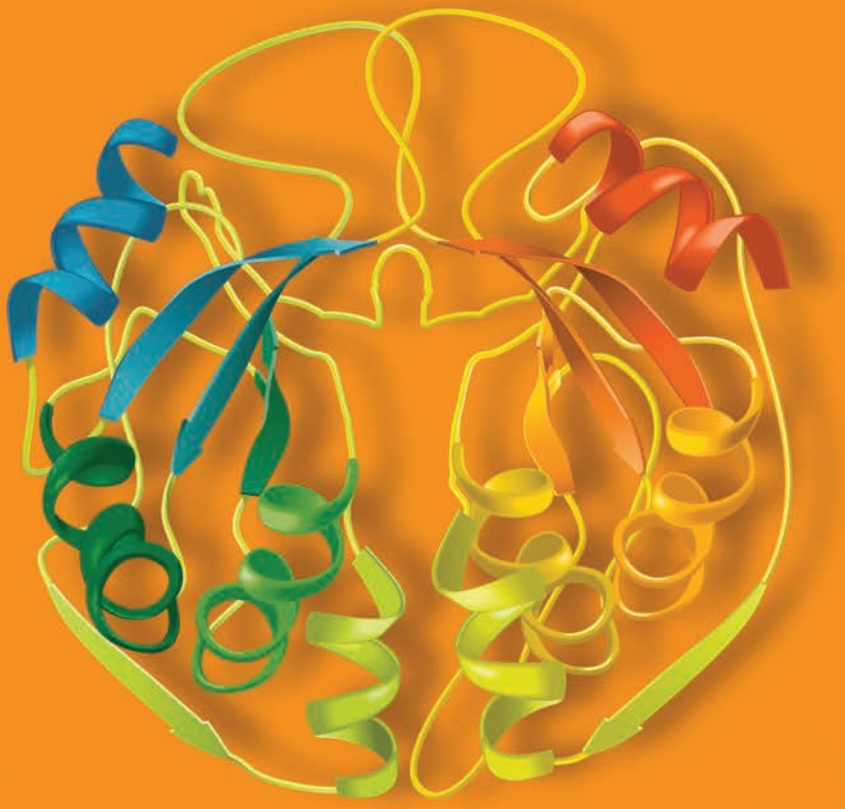
Over 50% of the Human Proteome Made From Human Cells



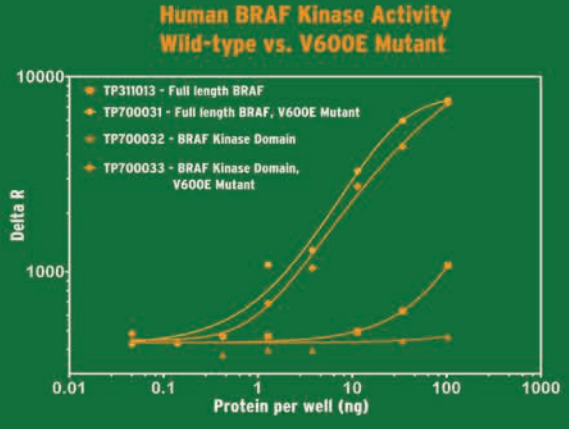
10,000 Proteins Mammalian Cell Produced

Protein functional studies

Compound screening

Immunogens for antibody production

Assay standards: ELISA, Mass Spec, and more



origene.com/protein

ORIGENE

Knock Out Any Gene!

CRISPR/CAS 9 Genome Editing Kits

OriGene's system for genome disruption and gene replacement delivers pre-designed plasmids and all the vectors needed to knock out any human or mouse gene. Knockout is as simple as 1-2-3:

1. Search the gene symbol on origene.com and order
2. Follow the simple protocol for transfection and Puro selection
3. Validate the knockout

Scan to view CRISPR/Cas-9 Video

Kit Components Include:

- 2 guide RNA vectors to ensure efficient cleavage
- Donor vector with predesigned homologous arms
- Knockin GFP-Puro for selection
- Scramble gRNA as negative control included

Come to OriGene, the trusted molecular biology expert, for your CRISPR needs.

ORIGENE

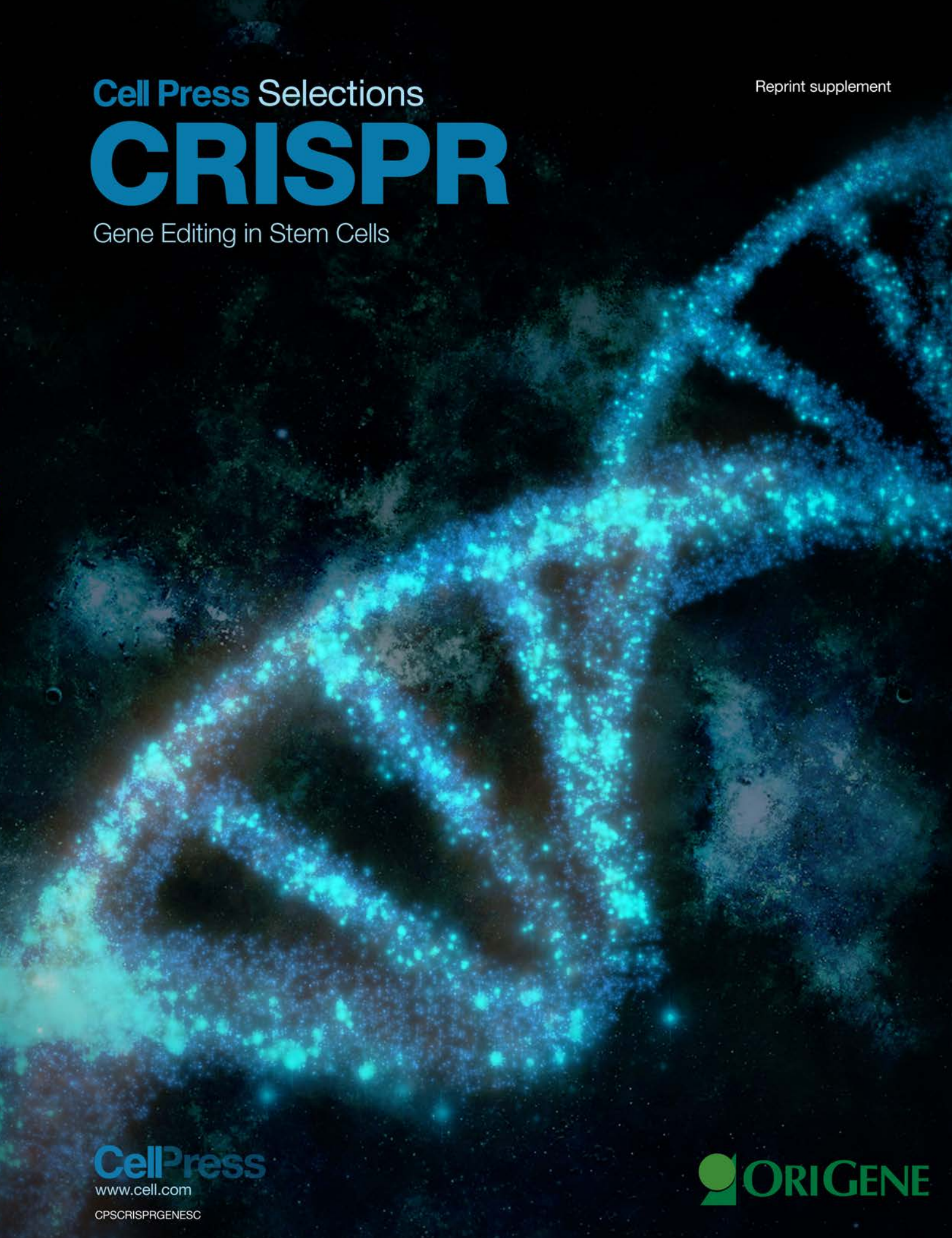
www.origene.com/CRISPR-CAS9

Cell Press Selections

CRISPR

Gene Editing in Stem Cells

Reprint supplement



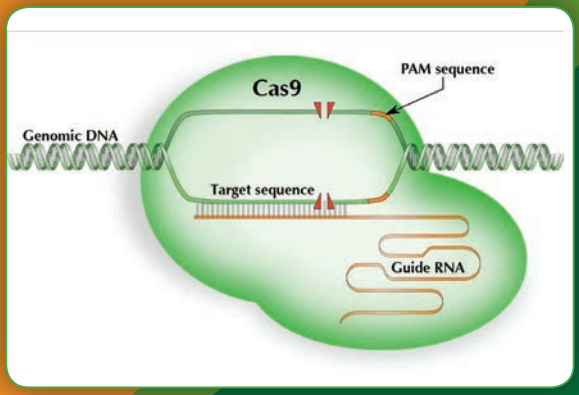
CellPress
www.cell.com
CPSCRISPRGENESC

Make The Right Cut With OriGene's CRISPR/Cas9 Tools

CRISPR is revolutionizing genetics, particularly mammalian genetics.

With unprecedented efficiency, CRISPR enables precise genome disruption in almost all model organisms. RNA-guided cleavage paired with donor-guided repair allows easy introduction of any desired modification in a living cell.

OriGene offers a wide selection of ready-to-use CRISPR tools. Blue Heron, OriGene's gene synthesis division, provides quick, reliable, and confidential service for any CRISPR-related constructs.



All-in-one gRNA/CAS9 vectors

gRNA cloning service

Donor cloning service

Gene knockout kits
(pre-designed gRNAs and donor)

Scan this QR code to view our 4-min video on CRISPR

ORIGENE

www.origene.com

Foreword

Since their discovery, CRISPR-based systems have fundamentally transformed our ability to manipulate genomes. When combined with stem cells, these gene-editing tools also have the power to reshape our understanding of human genetics, developmental biology, and regenerative medicine. The reviews and research in this edition of *Cell Press Selections*, on gene editing in stem cells, offer a snapshot of the latest advances in this rapidly evolving field.

As CRISPR/Cas9 systems, next-generation genome sequencing, and stem cell technologies have matured, so too have the possibilities for their combined use. CRISPR-based platforms have already been successfully applied in stem cells to address basic questions about human biology and produce experimental tools for disease modeling and drug discovery. Looking forward, it seems inevitable that applications of CRISPR technologies will continue to expand their reach, with the first CRISPR-based stem cell therapies already entering clinical testing. The articles compiled in this reprint collection showcase the current progress in these areas, from new CRISPR-based stem cell platforms for disease modeling and therapeutic development to alternative approaches for using CRISPR as a cell biology tool.

These articles represent only a small portion of the exciting research Cell Press has published and will publish on gene editing in stem cells, and we hope you'll visit www.cell.com on a regular basis to keep up with the latest stem cell and genome-editing news.

Finally, we are grateful for the support of OriGene, who helped to make the publication of this collection possible.

Sheila Chari

Senior Scientific Editor, *Cell Stem Cell*

For more information about Cell Press Selections:
Jonathan Christison
Program Director
jchristison@cell.com
617-397-2893

Lenti – Experience the Beauty!

Simple • Convenient • Efficient • Safe

Focus on discovery, not DNA delivery

Lenti Particles and Plasmids are Both Available

- **Lenti-shRNA**
- **Lenti-ORF**
- **Lenti Packaging Kit**

Use code **Lenti10** to get 10% off our lenti products.

Cannot be combined with any other discounts. Valid until Dec 31st, 2016.

20 Years ORIGENE

origene.com/lentiviral

CRISPR

Gene Editing in Stem Cells

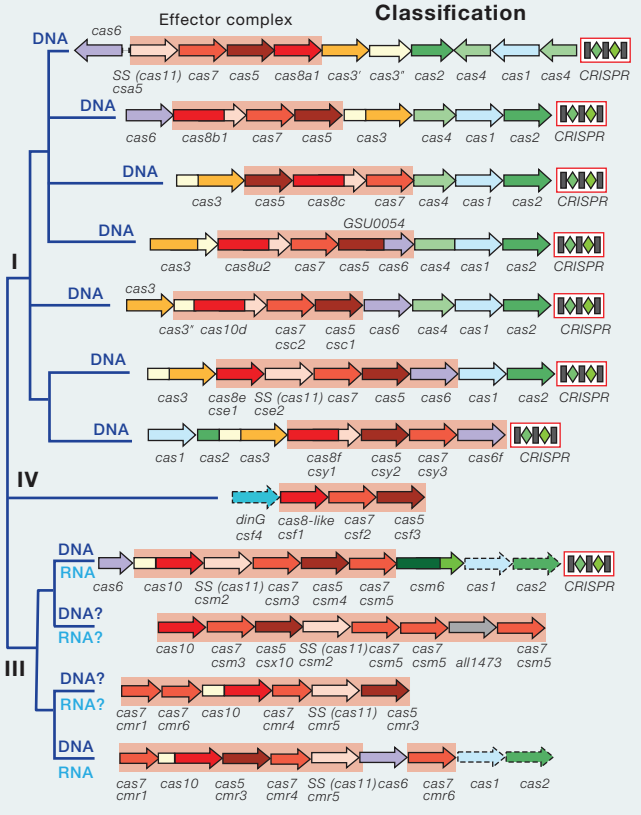
(continued)

SnapShot: Class 1 CRISPR-Cas Systems

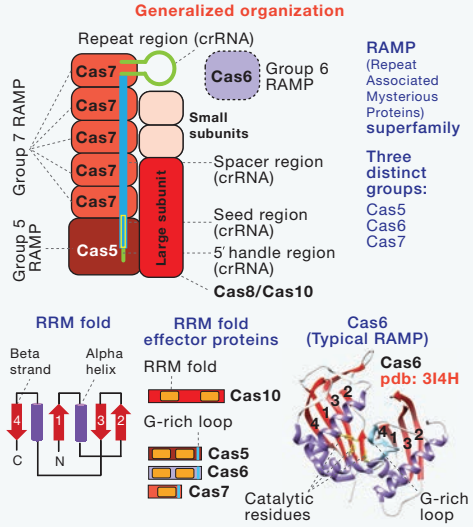
Kira S. Makarova,[1] Feng Zhang,[2,3] and Eugene V. Koonin[1]
[1]National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20894, USA;
[2]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; [3]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Cell

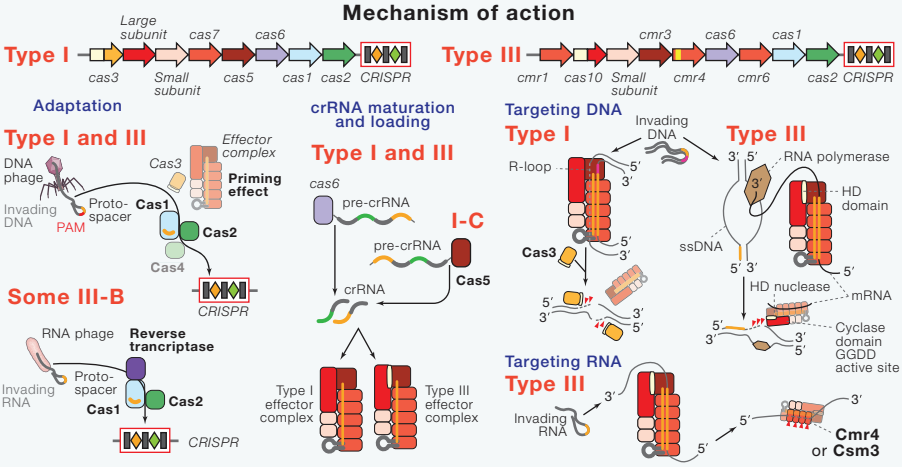
Classification



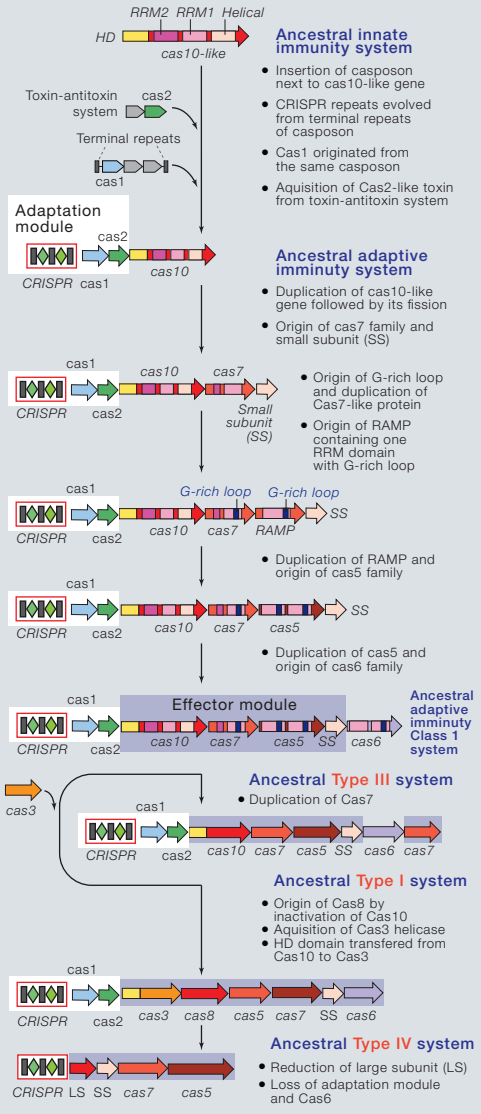
Effector complexes



Mechanism of action



Origin



Ancillary proteins

Cas2, RNA nuclease

- SSO1404 *Sulfolobus solfataricus*
- y1723 *Yersinia pestis*
- LSEI_0356 *Lactobacillus casei*

Cas4, DNA nuclease

- COG1468 PAE0079 *Pyrobaculum aerophilum*
- COG4343 PAE0082 *Pyrobaculum aerophilum*
- GSU0057 *Geobacter sulfurreducens*

CARF (CRISPR-associated Rossmann fold)

- Csx1 pfam09455 — CARF HTH HEPN — PDB: 4EOG/2I71, PF1127 *Pyrococcus furiosus*
- Csm6 pfam09659 — CARF 6H — PDB: 5FSH, Rv2818c *Mycobacterium tuberculosis*
- pfam09670 — TTHB152 *Thermus thermophilus* aq_376 *Aquifex aeolicus*
- pfam09651 — TM RelE PIN — sll7062 *Synechocystis sp.*; ST0035 *Sulfolobus tokodaii*; PDB: 3QYF, Sso1393 *Sulfolobus solfataricus*
- pfam09002 — PDB: 1XMX, Vc1899 *Vibrio Cholerae*
- pfam09623 — NE0113 *Nitrosomonas europaea*
- Csa3 — PDB: 2WTE, SSO1445 *Sulfolobus solfataricus*

Reverse transcriptase (RT)

- Cas1 RT — alr1468 *Nostoc sp.*
- Cas6 Cas1 — Psta_1137 *Pirellula staleyi*

Other families

- WYL domain and HTH — sll7009 *Synecho-cystis sp.*
- Csx3, RNA nuclease — aq_377 *Aquifex aeolicus*
- Csx16 — VVA1548 *Vibrio vulnificus*
- PRIMPOL family protein — Cas1 — Sfum_1354,Sfum_1355 *Syntrophobacter fumaroxidans*
- Csx15 — TTE2665 *Thermoan-aerobacter tengcongensis*
- Csx18 — Ple7327_2251 *Pleurocapsa sp.*
- Csx20 — MJ_1673 *Methano-caldococcus jannaschii*

 DOI http://dx.doi.org/10.1016/j.cell.2017.02.018

See online version for legends and references

With Cell Press Webinars, our network is your network!

Cell Press Webinars give you access to hot topics in emerging research and the application of new technology.

Watch essential, need-to-know webcasts via live streaming or on demand from the comfort and convenience of your office, lab, or home.

- Need-to-know topics, editorially curated
- World-class presenters, experts in their field
- Moderated by Cell Press editors

Tap into our network today!
Visit www.cell.com/webinars

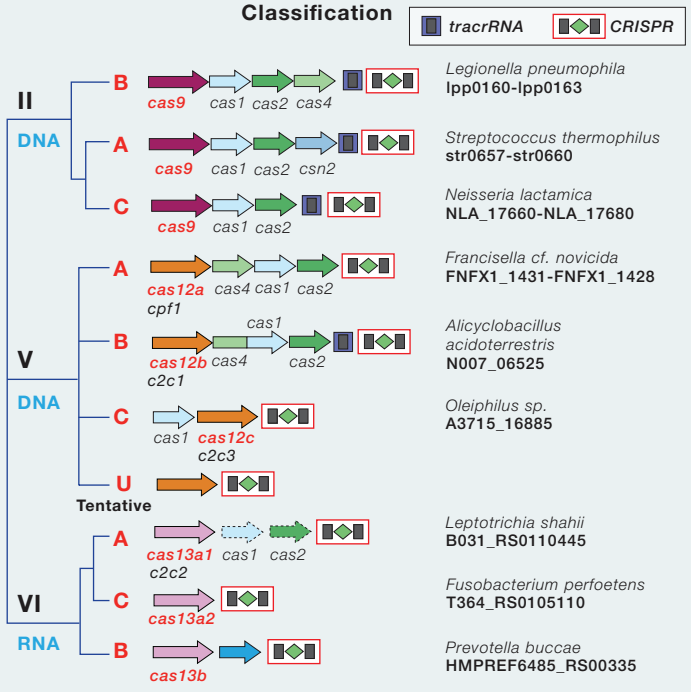
CellPress
Webinars

SnapShot: Class 2 CRISPR-Cas Systems

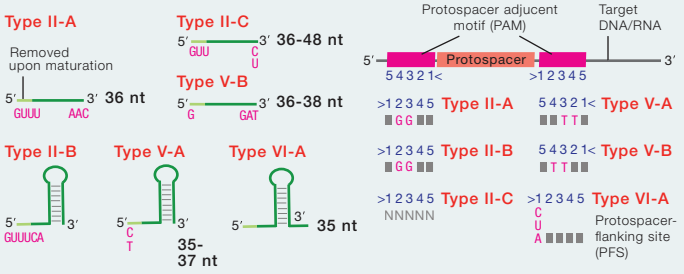
Kira S. Makarova,[1] Feng Zhang,[2,3] and Eugene V. Koonin[1]
[1]National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20894, USA;
[2]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; [3]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Cell

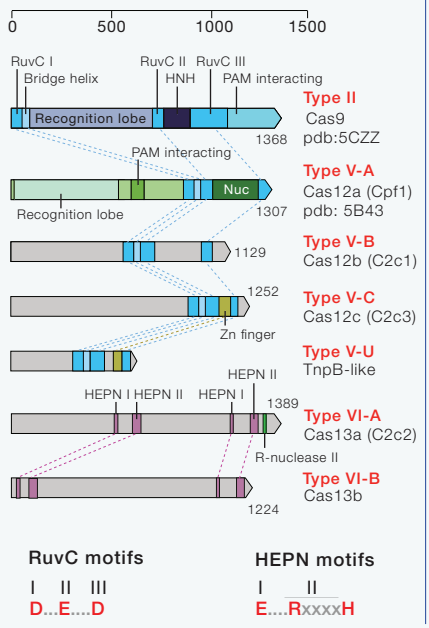
Classification



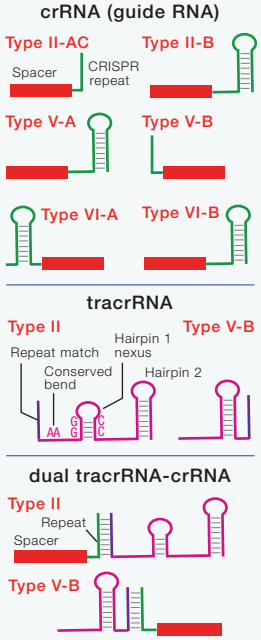
CRISPR and PAM features



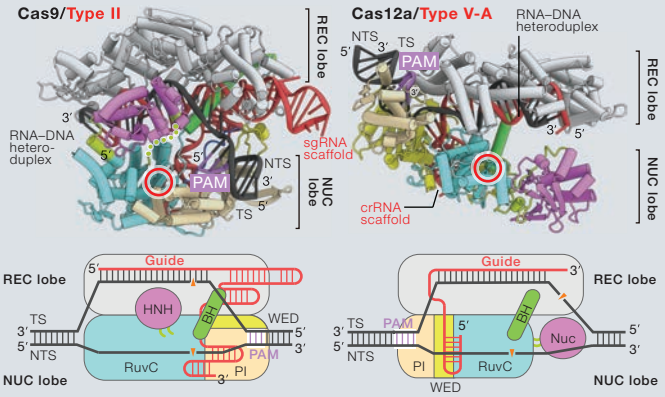
Effector protein domain organization



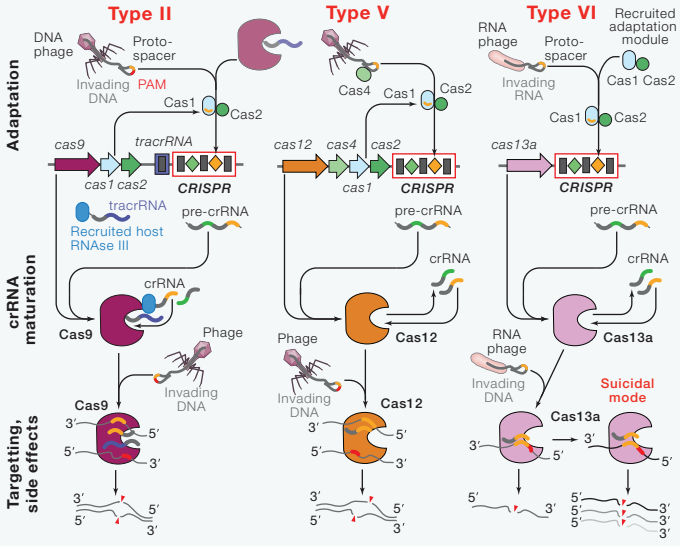
RNA components



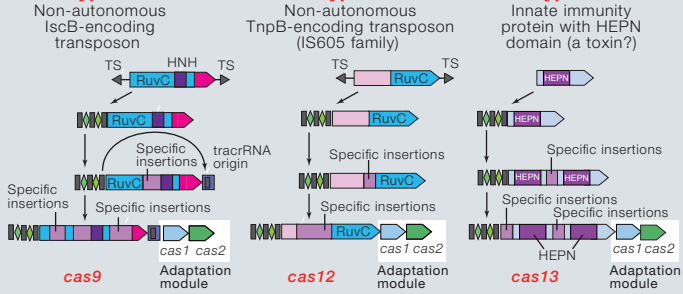
Effector protein structures



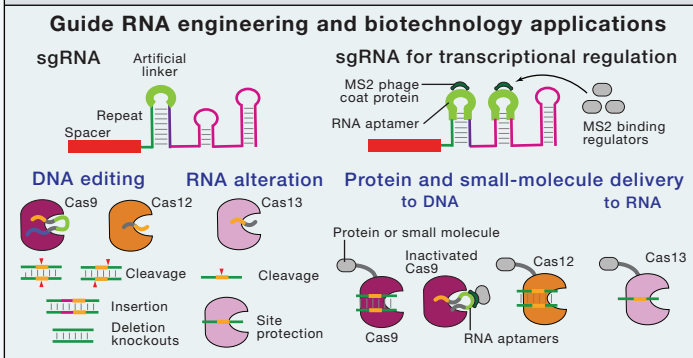
Mechanisms of action



Origin of class 2 systems



Guide RNA engineering and biotechnology applications



 DOI http://dx.doi.org/10.1016/j.cell.2016.12.038

See online version for legends and references

V is for Validated

Our Lenti-Viral clones deliver.

OriGene's TrueORF® Gold expression-validated ORF clones are now available in Lenti-Viral vectors. More cost efficient than gene synthesis or self-cloning, our 20,000 human cDNA clones are sequence verified, validated for protein expression by Western Blot, and are easily shuttled into 70 destination vectors.

Join the crowd.
Choose TrueORF Gold, the validated choice.

TrueORF Gold

- Expression Validated
- Sequence Verified
- Next-Day Shipment

NEW!
Lenti-ORF Human cDNA Clones

- 2 Lenti vector options with Myc-DDK or GFP tags
- Suitable for both transfection and transduction studies
- Ideal for efficient DNA delivery into virtually all cell types

ORIGENE

www.origene.com

Cell Stem Cell
A Decade of iPS Cells
Review

Induced Pluripotent Stem Cells Meet Genome Editing

Dirk Hockemeyer[1,*] and Rudolf Jaenisch[2]
[1]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[2]The Whitehead Institute for Biomedical Research and Department of Biology, MIT, Cambridge, MA 02142, USA
*Correspondence: hockemeyer@berkeley.edu
http://dx.doi.org/10.1016/j.stem.2016.04.013

It is extremely rare for a single experiment to be so impactful and timely that it shapes and forecasts the experiments of the next decade. Here, we review how two such experiments—the generation of human induced pluripotent stem cells (iPSCs) and the development of CRISPR/Cas9 technology—have fundamentally reshaped our approach to biomedical research, stem cell biology, and human genetics. We will also highlight the previous knowledge that iPSC and CRISPR/Cas9 technologies were built on as this groundwork demonstrated the need for solutions and the benefits that these technologies provided and set the stage for their success.

Reprogramming: "The Yamanaka Experiment"

Ten years ago Takahashi and Yamanaka reported on the "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors" (Takahashi and Yamanaka, 2006). The hypothesis of this work was daring and stated that a small set of transcription factors, when ectopically expressed in a somatic cell, can reprogram it back into a pluripotent state. Retrospectively, the simplicity of the experiments that Yamanaka and colleagues used to test this hypothesis was beautiful: take a set of 24 candidate genes, selected mostly for their high and specific expression in pluripotent cells, and simultaneously express them in differentiated cells using integrating retroviruses. Identify cells that induced pluripotency via a selectable marker gene that is not expressed in somatic cells, but is preferentially activated in pluripotent cells. Next, narrow down the cocktail of genes to the minimal set of reprogramming factors (Klf4, Sox2, Oct4, and Myc, a.k.a. KSOM) by the process of elimination. Lastly, demonstrate that the resulting induced pluripotent cells have all the key features of their embryonic stem cell (ECS) counterparts, such as a stem cell-like expression profile, the ability to give rise to differentiated cells in teratoma formation assays, and their contribution to tissues in chimeric mice after blastocyst injections (Takahashi and Yamanaka, 2006).

These experiments had an immediate impact. They came at a time when the potential of pluripotent stem cells (PSCs) in research applications and regenerative medicine had widely been appreciated (Rideout et al., 2002) (Figure 1), but technical and ethical limitations presented a challenge that severely impeded major progress toward realizing their full potential. Decades before the study by Yamanaka, John Gurdon (Gurdon, 1962, 1963) had demonstrated that the epigenetic profile of a fully differentiated cell can be reprogrammed to a pluripotent state. From a set of key experiments Gurdon demonstrated that a nucleus taken from a differentiated frog cell and injected into an enucleated oocyte can give rise to a fully developed frog. This experiment illustrated that during differentiation no essential genetic material is lost and that the epigenetic changes that drive cellular differentiation can be reprogrammed to totipotency. Decades later, the cloning of the sheep "Dolly" also by somatic cell nuclear transfer (SCNT) demonstrated that Gurdon's finding extended to mammals as well (Campbell et al., 1996). SCNT and cell fusion experiments gave two additional insights that set the stage for the Yamanaka experiment. First, they demonstrated that the cytoplasm of an oocyte or an ESC contained diffusible transacting factors capable of reprogramming a somatic nucleus (reviewed in Ambrosi and Rasmussen, 2005). Second, successful derivation of mice by SCNT with nuclei of B cells as a donor, which had undergone VDJ-recombination, provided genetic evidence that terminally differentiated cells can be reprogrammed (Hochedlinger and Jaenisch, 2002). Though more challenging, SCNT was eventually successful in reprogramming human cells into human ESCs (hESCs) in 2014 (Yamada et al., 2014). While these experiments spoke for the possibility of cellular reprogramming, they also suggested highly sophisticated machinery and a complex biological process, making the success of the basic Yamanaka experimental approach even more astounding. Even today, the gradual pace of transcription-factor-mediated reprogramming remains one of the most fascinating facets of the Yamanaka experiment: epigenetic changes after fertilization as well as reprogramming by SCNT occur within a few hours, while reprogramming by the Yamanaka experiment requires significantly more time, generally several days and multiple cell divisions. Yet, both processes result in a functionally equivalent cellular pluripotent state in in vitro cultures that is capable of forming an entirely new organism.

Around the same time as the first mammalian SCNT efforts, James Thomson derived the first hESC lines (Thomson et al., 1998). He used a very similar strategy that had proven successful for Evans and Martin (Evans and Kaufman, 1981; Martin, 1981), culturing the inner cell mass outgrowth of explanted blastocysts. However, it is interesting to note that human and mouse ESC (mESC) maintenance require distinct signaling networks and culture conditions. LIF/Stat3 is required for maintaining the undifferentiated state in mESCs and BMP4 can inhibit the MEK/ERK differentiation pathway resulting in mESC self-renewal. In contrast hESCs and hiPSCs do not require hLIF, and maintenance of pluripotency seems to rely mostly on FGF and MEK/ERK signaling, indicating species-specific requirements for culturing pluripotent cells. It seems likely that this difference can be attributed to a difference in the developmental stage

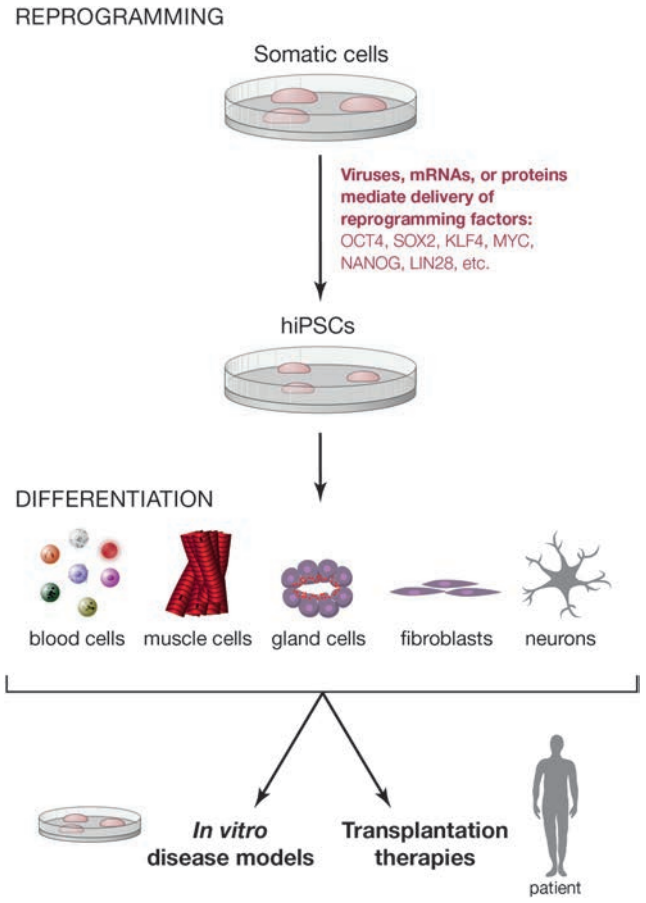


Figure 1. Overview of the iPSC Technology
Patient cells can be reprogrammed into iPSCs using optimized reprogramming protocols that involve small molecules, microRNAs, and combinations of reprogramming factors. iPSCs can be differentiated into somatic cells that could be used either in transplantation therapies or alternatively to model human diseases.

that is captured in vitro from the outgrowth of the inner cell mass, where hPSCs cultured under standard conditions represent a later epiblast-like pluripotent state (Brons et al., 2007; Tesar et al., 2007; Theunissen et al., 2014) (reviewed in Nichols and Smith, 2009).

Proof-of-concept experiments with cells differentiated from hESCs suggested that PSCs could be a source for cell replacement transplantation therapies and could provide a model system to understand early human development and cellular differentiation. However, ethical concerns, limited access to embryos, and the possibility of immune rejection were roadblocks that impeded the promise of hESCs.

In 2006 the ''Yamanaka experiments'' made the ethical debate about PSC research largely obsolete, as they established a robust method to derive human pluripotent cells without the use of human embryos. Furthermore iPSC technology promised to solve complications that were anticipated from immune rejections of heterologous hESC-derived tissues, as it would allow the generation of patient-specific autologous pluripotent cells and derived tissue. The race to perform the key functional follow-up experiments began immediately. For the mouse system it was essential to establish that iPSCs could pass the most stringent test for pluripotency: germline transmission (Maherali et al., 2007; Okita et al., 2007; Wernig et al., 2007) and tetraploid complementation (Kang et al., 2009; Zhao et al., 2009).

For the human system the initial question was whether the same set of factors capable of reprogramming mouse cells would also work for human cells (Takahashi et al., 2007). Yamanaka and Takahashi quickly demonstrated that their factors also worked in human cells (Takahashi et al., 2007). However, additional experiments over that last 10 years in mouse and human cells also revealed that other sets of transcription factor combinations can be equally potent in reprogramming cells to a pluripotent state, providing valuable insights into the transcriptional pluripotency networks and how cells establish pluripotency (Buganim et al., 2012; Apostolou and Hochedlinger, 2013; Park et al., 2008; Takahashi and Yamanaka, 2015; Yu et al., 2007).

For the anticipated clinical application of iPSCs, it was important to demonstrate that reprogramming could be achieved without stably integrating the KSOM factors into the genome of the somatic cell. Such factor-free iPSCs were generated by independent methods such as the excision of reprogramming factors using the Cre/LoxP (Soldner et al., 2009) or the piggyBac system (Kaji et al., 2009; Woltjen et al., 2009) by avoiding integration of the reprogramming factors altogether by using non-integrating viruses (Fusaki et al., 2009), episomal vectors (Yu et al., 2009), or direct transfection of the reprogramming factors as either mRNA (Warren et al., 2010) or protein (Kim et al., 2009). Initially, human cell reprogramming was quite inefficient compared to mouse cells, and thus several technical improvements were made to optimize hiPSC reprogramming protocols, culture conditions, and iPSC characterization procedures to test for the pluripotency of newly isolated iPSCs. Eventually, these optimizations made iPSC technology increasingly more accessible to laboratories without previous stem cell experience and are now so streamlined that iPSC derivation, maintenance, and differentiation are widely used research tools in all aspects of biomedical research. In addition, efficient and robust reprogramming techniques provided insight into the mechanistic steps of reprogramming and the order of events involved in reverting the epigenome from a differentiated to a pluripotent state. A detailed understanding of the forces at work is necessary to answer the key questions of whether the reprogramming of human cells results in a cell state that is equivalent to hESCs or whether iPSCs retain to some extent an epigenetic memory (Kim et al., 2010; Polo et al., 2010). For example, do iPSCs derived from liver cells retain some characteristics of liver cells and do they preferentially differentiate into liver tissue relative to other cell types? Tetraploid complementation and germline transmission experiments gave the clear answers that mouse iPSCs were fully reprogrammed to pluripotency. However these tests are not available for hiPSCs. Moreover, it is not clear to what extent the mouse iPSC's epigenome is reset during the reprogramming process and how much of the resetting occurs in vivo or when the cells pass through the germline. It is not surprising that early cellular stages of the reprogramming process will show epigenetic differences, yet all these differences eventually will converge on the same pluripotent cell state as ESCs. Thus, it is interesting to further examine the level and functional relevance of epigenetic memory; yet it seems that such

epigenetic differences in cellular state are at best small and overshadowed by differences caused by the reprograming method of choice, cell selection during propagation, culture conditions, and more importantly genetic background of the parental somatic cell (Guenther et al., 2010; Kyttälä et al., 2016; Rouhani et al., 2014). For example, it has been demonstrated that the epigenetic memory, i.e. epigenetic characteristics reflecting the state of the donor cells seen initially in the iPSCs, is lost upon prolonged cell passages, suggesting that this donor-cell-specific memory may be of little functional relevance (Polo et al., 2010).

The most attractive application of the iPSC technology is that it allows the isolation of patient-derived cells that carry all genetic alterations that cause the particular disease. Thus, these cells provide an experimental system to study pathogenesis of the disease in an in vitro system and to possibly devise therapeutic strategies (Robinton and Daley, 2012). Importantly, the iPSC technology allows comparison of the neuroanatomical features and physiology of the iPSCs to the clinical features of the donor patient.

The Power and Limitations of iPSCs

In addition to the prospect of future iPSC-based cell replacement therapies, the ability to derive iPSCs from patients' cells had a striking effect on human disease modeling. Some of the most remarkable advances were made in diseases such as neurodegenerative diseases that are only partially recapitulated in animal models. Here, iPSC technology was particularly transformative, as it made it possible to study the effects of familial monoallelic diseases as well as complex idiopathic diseases in the context of patient-derived neurons and tissue, systems that were previously not readily available for experimental investigation. For example, studying dopaminergic neurons differentiated from Parkinson patient-derived iPSCs yielded insights into the molecular causes of the disease and the identification of cellular stressors that might exacerbate the phenotype (Soldner et al., 2011; Soldner and Jaenisch, 2012);. As a result of such advances, iPSC-based and primary tissue culture systems have largely replaced previous experimental systems that studied human genetic diseases using overexpression studies in cancer cell lines. Indeed, the number of human diseases modeled in culture using patient-derived iPSCs (''disease in the dish'') is growing rapidly (summarized in Avior et al., 2016; Sterneckert et al., 2014).

While the approach of studying human disease in the disease-relevant cell type resulted in many success stories and insights, several challenges of iPSC disease modeling quickly became apparent. For one, it became evident that many protocols that were developed for the differentiation of hPSCs into functional tissue resulted in embryonic rather than adult human cell types (Bedada et al., 2015; Forster et al., 2014; Hrvatin et al., 2014; Spence et al., 2011; Takebe et al., 2013). This observation might not pose a problem for studies that aim to recapitulate cell-autonomous defects of developmental diseases that likely will become apparent after a few weeks of in vitro differentiation. However, iPSC differentiation experiments that aim to understand human disease and pathologies within the context of the adult or as a function of human aging suffer from a lack of cellular maturity as well as a relatively short time span limited by culture conditions. One approach to increase the maturity of in vitro cell systems and to mimic cellular aging is to expose these cells to stressors that are associated with aging (Miller et al., 2013; Studer et al., 2015). Significant progress has also been made to current strategies of investigating cell non-autonomous biological problems, including the development of co-culture experiments and protocols to differentiate hPSCs into tissue stem cells and organoid cultures. Organoid cultures are small functional tissue units composed of several distinct cell types that can be maintained and used to recapitulate features of tissues rather than that of individual cell types in vitro (Lancaster et al., 2013; Sato et al., 2011; Sato et al., 2009) (reviewed in Lancaster and Knoblich, 2014; Sato and Clevers, 2013).

An important and often ignored challenge of iPSC technology is the variability between individual iPSC lines in their potential to differentiate into functional cells of a given lineage. This variation between cell lines is unpredictable and mostly caused by genetic background differences as well as the reprogramming history of a given cell line. Thus, in efforts to model a disease, detection of small phenotypic differences between cells differentiated from a patient or control iPSCs may not reveal a disease-relevant phenotypic difference but rather reflect the system's immanent variation between individual iPSC lines (Soldner and Jaenisch, 2012). The generation of isogenic pairs of disease-specific and control iPSCs that differ exclusively at the disease-causing mutation has been used to control for the variation and has led to defining subtle disease-relevant differences in monogenic diseases (Soldner et al., 2011). The problem is, however, exacerbated when studying more clinically important sporadic or polygenic diseases where low-effect-size, disease-causing loci are defined by genome-wide association studies (GWASs). Since phenotypic differences would be expected to be small, the use of isogenic pairs of disease-specific and control cells would be even more important. Finally, ongoing efforts to learn about human genetic variation by studying dozens or even hundreds of iPSC lines derived from healthy donors may give little interpretable information because of the unpredictable system-inherent phenotypic variability between individual iPSC lines (differing in millions of SNPs within each genome) and experimental variations in their differentiation. Making isogenic iPSC controls by genome editing that differ only in a single or a few SNPs could reduce variations due to genomic variability.

The challenge associated with the genetic variability of hPSCs is compounded by another remarkable difference between mouse and human PSCs: the striking resilience of hPSCs to conventional gene targeting approaches. This dearth of genetic control in hPSCs prevented genetic experiments that were considered standard in mESCs. Nevertheless conventional gene targeting has been accomplished in hPSCs (Zwaka and Thomson, 2003). Protocols for conventional gene targeting have been optimized to modify hPSCs (Costa et al., 2007; Davis et al., 2008a; Irion et al., 2007; Ruby and Zheng, 2009) and have been successfully used to establish hPSC models for human disease such as Lesch-Nyhan syndrome (Urbach et al., 2004). Moreover, this approach has been used to correct the disease-causing mutation with ornithine-d-aminotranferase that is mutated in patients with gyrate atrophy (Howden et al., 2011) or to alter the amount of disease-causing CAG repeat expansions in the huntingtin gene of patient-specific iPSCs (An et al., 2012). Furthermore, these protocols have been used to generate linage reporters for genes such as MIXL and Olig2 to study cell

fate decision of differentiating human stem cells (Davis et al., 2008b; Xue et al., 2009). Overall, however, these approaches are very time consuming, as they generally require the generation of large targeting constructs and even then are very inefficient and in many cases not successful. It appears that cell-intrinsic features such as low homologous recombination and single-cell survival rates make conventional genome modification as described by Capecchi and Smithies for mESCs (Doetschman et al., 1987; Thomas and Capecchi, 1987) very inefficient in hPSCs.

Both of these challenges have been overcome: the development of the Rho-kinase inhibitor Y-27632 to suppress anoikis during the disaggregation of hPSC colonies dramatically increased single-cell survival of hPSCs (Watanabe et al., 2007), and the low frequency of spontaneous homology-mediated gene targeting in hPSCs was dramatically increased through the development of site-specific nucleases (SSNs) as a tool for their genetic engineering (reviewed in Carroll, 2014; Hsu et al., 2014; Urnov et al., 2010).

Genome Editing BC (Before CRISPR/Cas9)

The development of SSNs as research tools parallels the development of iPSCs: key experiments uncovered the biological principles and highlight how a generalized platform for genome editing would advance basic and biomedical research. Repurposing of the CRISPR/Cas9 system as an engineered SSN removed the impediments that limited the full potential of genome editing by providing this general platform.

Key experiments more than 15 years ago in mammalian cells demonstrated that a double-strand break (DSB) generated by an SSN at a defined genomic site can be repaired either by the endogenous homology-mediated repair machinery using an exogenous provided repair template or by the error-prone non-homologous end joining (NHEJ)-DNA repair pathway (Rouet et al., 1994a, 1994b). The crucial observation made during these experiments was that a DSB increased the rate of homology-mediated genomic changes at the break site by several orders of magnitude compared to conditions in which only an exogenous repair template was provided without the induction of a DSB. Importantly, this principle of employing a DSB to facilitate DNA-repair mediated editing of genomes proved to be almost universal and applies to hPSCs as well as other systems such as *Caenorhabditis elegans* (Morton et al., 2006; Wood et al., 2011) and *Drosophila melanogaster* (Beumer et al., 2008; Bibikova et al., 2002, 2003), which are similarly resilient to conventional gene-targeting strategies as hPSCs are.

Already in 2005, Urnov et al. demonstrated that engineered zinc finger nucleases (ZFNs) can serve as a designer SSN to correct X-linked SCID disease-relevant mutations in patient-specific cells (Urnov et al., 2005). It was this study that coined the term ''genome editing.'' Ten years later the first clinical trials based on this ZFN technological platform are underway to disrupt CCR5 in T cells to treat HIV patients (Tebas et al., 2014).

Based on these pioneering experiments, we and others implemented the use of SSNs such as ZFNs and transcription activator-like effector nucleases (TALENs) to engineer hPSCs (DeKelver et al., 2010; Hockemeyer and Jaenisch, 2010; Hockemeyer et al., 2009, 2011; Lombardo et al., 2007; Sexton et al., 2014; Soldner et al., 2011; Zou et al., 2009). These experiments provided proof of principle for SSN-mediated gene knockouts, for the insertion of transgenes into expressed and non-expressed genes to generate cell-type-specific lineage reporters, for the overexpression of transgenes from genetically defined loci, and for the insertion or repair of disease-relevant point mutations in hPSCs (Figure 2).

The technical advances that established genetic control in hPSCs proved to be highly synergistic with the development of iPSC technology. Genome editing in hPSCs overcame the issue of enormous genetic background variability inherent in iPSC-based disease models. Independent proof-of-concept studies demonstrated that SSNs can be used to repair or introduce disease-relevant mutations in hPSCs (Soldner et al., 2011; Yusa et al., 2011). The resulting pairs of PSC lines are isogenic, except for the disease-relevant mutation. Parallel differentiation of such isogenic sets of cells into disease-relevant cells and tissues can be used to directly assess the contribution of a mutation to cellular pathology (Chung et al., 2013; Ryan et al., 2013; Wang et al., 2014b; Yusa et al., 2011).

The initial ZFN and TALEN platforms to generate SSNs for genome editing in stem cells were costly and labor intensive and their implementation as research tools therefore developed comparatively slowly. However, extensive work with ZFNs and TALENs has demonstrated the power of genome editing and highlighted the impact that a universal, cheaper, and simpler platform to make SSNs would have.

CRISPR/Cas9: Everyone Can Edit Anything

The need for a simple and unified platform to generate SSNs was met and resolved, similarly to the need for an easy way to make iPSCs, through a single experiment: by repurposing the bacterial Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) adaptive immune systems (reviewed in Marraffini, 2015) as an SSN. In 2012 the collaborative work of the Jennifer Doudna and Emmanuelle Charpentier laboratories demonstrated that in CRISPR type-2 systems a single protein, Cas9, can function as a designer SSN by associating with an engineered single guide RNA (sgRNA) that bears homology to a genetic locus of interest (Jinek et al., 2012). In this process, the sgRNA substitutes the natural Cas9-associated bacterial RNAs that normally confer target specificity for the bacterial pathogen DNA and instead directs Cas9 to introduce a blunt DSB in any target DNA with complementarity to a 20-nucleotide (nt)-long sequence in the sgRNA. Doudna and colleagues predicted that this simple way of engineering SSNs could be exploited to streamline genome editing (Jinek et al., 2012). In less than 4 years this prediction became reality and Cas9-mediated genome engineering was developed into the platform of choice to generate SSNs and to genetically modify hPSCs (Chen et al., 2015; Cong et al., 2013; Fu et al., 2014; González et al., 2014; Kleinstiver et al., 2015; Hou et al., 2013; Hsu et al., 2014; Kleinstiver et al., 2016; Liao and Karnik, 2015; Lin et al., 2014; Mali et al., 2013; Ran et al., 2015; Slaymaker et al., 2016; Tsai et al., 2014; Wu et al., 2014b). Some important adaptations and improvements to increase the ease and scope of Cas9-mediated genome engineering in hPSCs were the establishment of CRISPR/CAS-systems from different organisms (Hou et al., 2013; Zetsche et al., 2015) that respond to different PAM sequences and the engineering of spCas9 to associate with

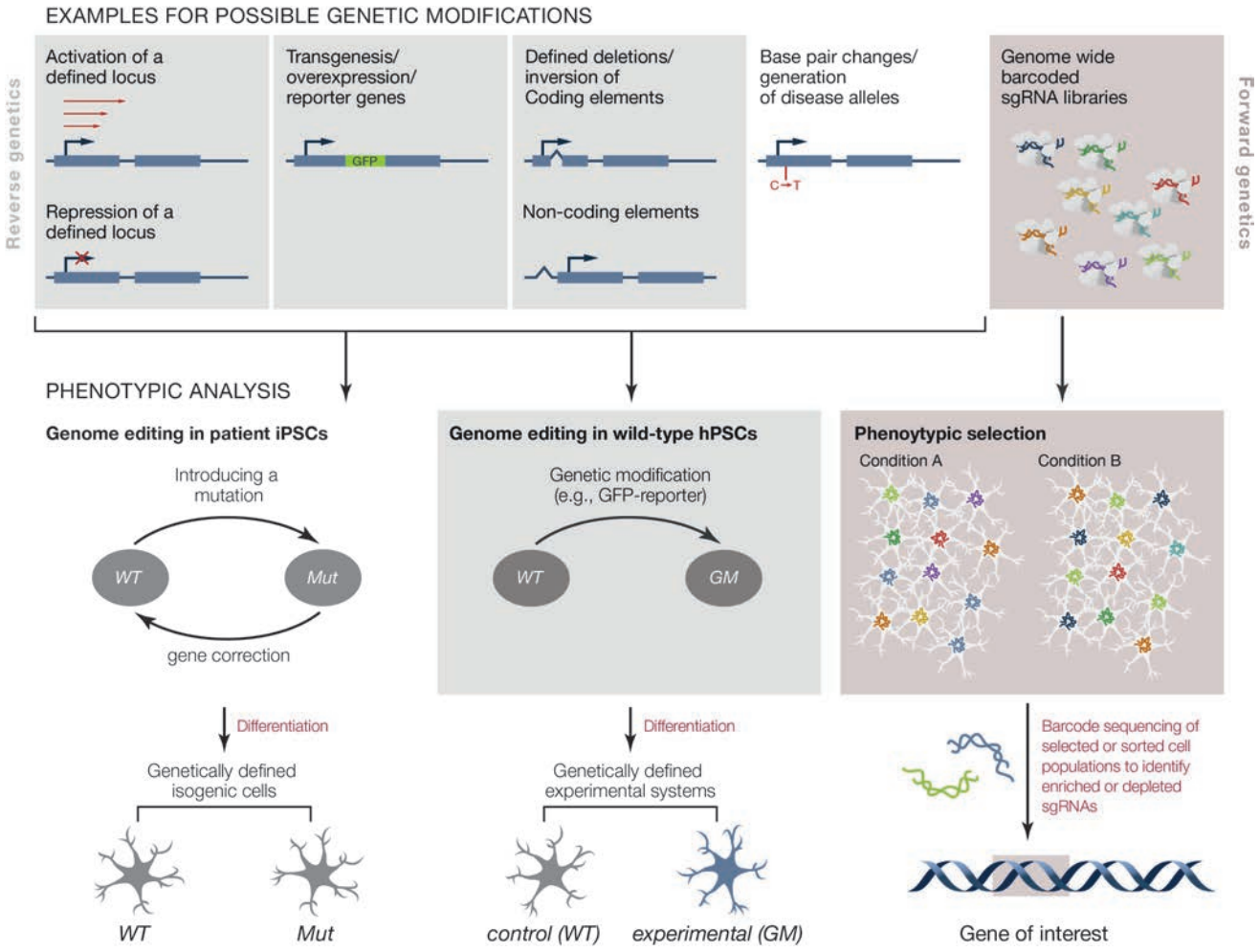


Figure 2. Genome Editing Applications in hiPSCs

Genome editing allows the genetic modification of hiPSCs. The top panel (left side) depicts examples of reverse genetic approaches to study hPSCs using genome editing. Gene expression can be modulated (activated or repressed: CRISPRi and CRISPRa) by reversibly targeting their endogenous promoter. Genes can be inserted to generate reporter genes or to achieve ectopic expression. Genetic information can be deleted or inverted and modifications as small as single base pair changes can be generated to introduce mutations or polymorphisms or to repair disease-relevant mutations. The resulting genetically engineered hPSCs differ from wild-type cells exclusively at the edited locus and are otherwise isogenic (bottom left). Parallel differentiation of these isogenic cell lines into disease-relevant cell-types can provide the basis for the phenotypic analysis of disease-specific cellular pathologies. Phenotypes found in these cells can be directly attributed to genetic manipulation. In addition, forward genetic approaches to study hPSCs (top right panel) became available with the development of genome editing as a screening tool. Bulk transduction of hPSCs with either Cas9 or dCas9 in combination with genome-wide barcoded sgRNA libraries—"CRISPR cutting, CRISPRi, and CRISPRa"—can be used to identify genes whose loss or gain of function changes the cellular representation within the infected cell pool. Enrichment or depletion of sgRNAs can be determined by sequencing the sgRNAs, yielding candidate genes of interest (bottom right panel).

alternative PAMs by structure-based engineering of Cas9, thereby extending genomic target range and specificity of spCas9 (Kleinstiver et al., 2015a; Kleinstiver et al., 2015b). Furthermore, several detailed protocols that describe the implementation of genome editing techniques in PSC systems have been optimized and published (Blair et al., 2016; Byrne and Church, 2015; Chiba and Hockemeyer, 2015; Yusa, 2013).

The key advantage of the CRISPR/Cas9 system over previous systems lies in the fact that DNA-binding specificity is encoded solely by the sgRNA and so unlike previous platforms does not require laborious engineering of DNA binding proteins. Thus CRISPR/Cas9-based editing has largely replaced previous SSN technologies. Combining the cellular versatility of iPSC differentiation with the ease of CRISPR/Cas9-mediated genome editing proved to be a very powerful experimental approach, and by now genome editing in hPSCs has become a standard tool in stem cell research and human disease modeling (Johnson and Hockemeyer, 2015; Matano et al., 2015; Schwank et al., 2013).

One of the most exciting experiments that became possible since the development of robust and highly efficient editing technologies in hPSCs is the genetic and functional testing of the onslaught of empirical data generated by GWASs. Similar to the disease-modeling approach, genome editing allows us to engineer variant alleles observed in these studies found to be associated with a specific disease in an otherwise isogeneic cellular setting. Phenotypic comparison of such cells can reveal how non-coding mutations, enhancer polymorphisms, and balancer

mutations can impact tissue-type-specific cellular behaviors that are relevant to the particular condition.

For example, this approach has been used successfully to identify the molecular principles underlying the most frequent non-coding mutations associated with human cancer (Bojesen et al., 2013; Fredriksson et al., 2014; Horn et al., 2013; Huang et al., 2013; Killela et al., 2013). Genetic engineering of these mutations, which occur in the promoter of the catalytic subunit of human telomerase or TERT, revealed that the mutations result in the failure of cells to silence TERT transcription upon cellular differentiation and explains how these mutations function in tumorigenesis (Chiba et al., 2015).

Gene-correction frequencies in hPSCs are generally much lower than in tumor cell lines such as K578 or HCT116 cells that are commonly used for gene editing in cancer cells. A very elegant approach to overcome this challenge and to increase the efficiency of homology-mediated events in iPSCs was used in experiments that employed zinc finger nucleases to correct mutations in iPSCs derived from patients with alpha trypsin deficiency. In these experiments gene targeting efficiencies were increased by the use of a positive selection marker that allowed the efficient isolation of the edited clones and that could subsequently be removed without leaving residual genetic material using piggyBac transposition. This editing strategy allowed the generation of bi-allelic editing events in patient-derived iPSCs to restore alpha trypsin enzymatic function in disease-relevant iPSC-derived hepatocytes in vitro and after xenotransplantation (Yusa et al., 2011).

A similar approach to overcome the challenges associated with the low frequency of gene-correction events in hPSCs was used to correct point mutations in the beta-globin gene of iPSCs derived from patients with sickle cell disease (Zou et al., 2011). In this case a LoxP-site flanked selection cassette was used to increase the genome editing efficiency initially, but was then subsequently removed using Cre-recombinase. This approach results in a single residual LoxP site in an intron of the beta-globin gene. Similarly, two independent studies demonstrated that the SSN can be used to directly correct β-thalassemia mutations in patient-derived iPSCs and restore hematopoietic differentiation (Ma et al., 2013; Xie et al., 2014).

Alternative strategies for increasing editing efficiencies include methods to more efficiently detect and subclone cells that have undergone rare editing events (Miyaoka et al., 2014) as well as to enhance delivery methods for the nuclease and donor template (Lin et al., 2014). An orthogonal approach to simplify the generation of isogeneic hPSC lines was the derivation of an inducible Cas9-expressing cell line by editing a Cas9 expression cassette into the AAVS1 locus. In this system Cas9 expression can be induced by doxycycline so that efficient editing afterward only requires the expression or delivery of the sgRNA (González et al., 2014). This system has been used to generate loss-of-function alleles in EZH2 and to demonstrate the effects of haploinsufficiency for EZH2 in hematopoietic differentiation (Kotini et al., 2015).

Further developments that facilitate the derivation of genome-engineered iPSC cell lines are protocols that directly combine genome editing with reprogramming. Howden et al. demonstrated that human fibroblasts could be simultaneously reprogrammed and edited, resulting in edited iPSCs going through only one single-cell cloning event without the need for drug selection (Howden et al., 2015).

Further implementation of gene-editing in patient-specific iPSCs will have a substantial impact on current disease modeling approaches. An example of the far-reaching effects is illustrated by editing experiments that inserted an inducible Xist lncRNA into chromosome 21 of Down syndrome patient-derived iPSCs. Using this approach Jiang et al. showed that ectopic expression of Xist was sufficient to transcriptionally suppress the targeted third copy of chromosome 21 and to reverse the cellular disease phenotypes in in vitro differentiated cells (Jiang et al., 2013).

Since the implementation of genome editing in hPSCs, several diseases have been modeled using isogenic cell lines that have either corrected a disease-relevant mutation in iPSCs or introduced a disease-relevant allele in wild-type hPSCs. For example, the genetic correction of mutations in Niemann-Pick type C patient-specific iPSCs to rescue metabolic defects in cholesterol metabolism and autophagy, which are responsible for the pathology, represents just one demonstration of how this approach has been successfully implemented (Maetzel et al., 2014). Furthermore, genome editing in hPSCs has been used to establish models for Rett syndrome disrupting MECP2 function in hPSCs (Li et al., 2013), to generate HIV-resistant variants alleles of the CCR5 gene into iPSCs (Ye et al., 2014), to repair MYO15A in iPSCs derived from patients affected by deafness (Chen et al., 2016), and to derive isogeneic cell pairs of COL7A1-corrected iPSCs derived from patients with dystrophic epidermolysis bullosa (Sebastiano et al., 2014).

In a growing number of cases, such approaches have also been used to provide new insight into disease pathology. For example, SSN-mediated correction of disease-causing mutations in LRKK2 that are associated with Parkinson disease (PD) revealed the transcriptional changes caused by disease-associated alleles in patient cells (Reinhardt et al., 2013). Likewise, genome editing of patient-specific iPSCs followed by in vitro differentiation was also used to generate an isogenic disease model for cystic fibrosis by correcting disease-relevant mutations in CFTR followed by differentiation into airway epithelium (Crane et al., 2015; Firth et al., 2015; Suzuki et al., 2016).

The Challenge of Studying Sporadic (Polygenic) Diseases

The application of iPSC technology for the study of sporadic diseases poses particular challenges because disease-specific phenotypic changes are expected to be subtle. The genetic basis of sporadic or idiopathic diseases is thought to be a combination of multiple low-effect-size risk alleles, mostly in regulatory regions such as enhancers, which are identified by GWASs (Gibson, 2011; Merkle and Eggan, 2013). The "common disease-common variant hypothesis" proposes that multiple risk variants with small effect size in combination with additional environmental factors are the drivers of sporadic diseases. Thus, a major challenge of using human-derived cells is that risk variants are not only present in patients but also in unaffected individuals, albeit with lower frequency. Thus, individual risk variants are not sufficient to cause disease-associated phenotypes in carrier individuals or in hiPSCs derived from carriers or patients. While an iPSC isolated from a patient would harbor all risk variants that contribute to the disease, any in vitro study to gain mechanistic insights is complicated by the high system-immanent variability in differentiation into the disease-relevant cells (Soldner and Jaenisch, 2012). Another complicating factor is that the likely effect of

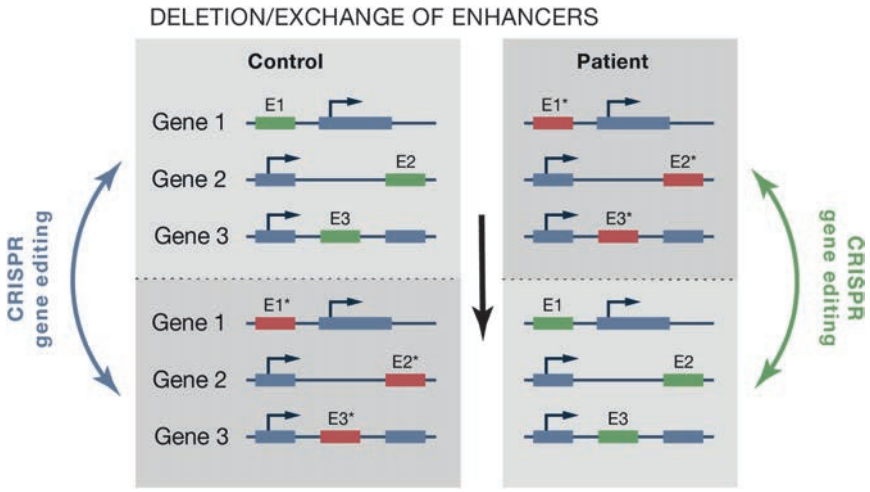


Figure 3. Strategy to Generate Isogenic iPSCs that Differ at Multiple Risk Alleles

GWASs have identified genomic loci that may slightly increase the risk of developing a sporadic disease. The key challenge of using patient-derived iPSCs to get mechanistic insight into risk alleles is to create meaningful control cells. CRISPR/Cas9-mediated gene editing would allow exchanging risk (red squares) and protective (green squares) alleles and generating appropriate control cells that differ exclusively at the risk loci under study.

a GWAS-identified risk regulatory allele on the target gene (or genes) would be predicted to be subtler than would be expected for monogenic diseases as discussed above. Thus, it would be impossible to compare the disease-specific cells to a suitable control cell line because any control cells would have a different genetic background, which will affect the differentiation potential of the cells and thus would prevent a meaningful comparison.

Thus, a major challenge for using iPSCs for the study of sporadic diseases is how to generate pairs of isogenic cells that differ at one or multiple risk alleles. Figure 3 outlines a possible strategy of how the CRISPR/Cas9 gene editing approach could be used to generate isogenic cells that differ at multiple risk loci and thus would enable the mechanistic study of polygenic diseases. This approach was recently used to decipher the impact of PD-associated risk variants. Genetic engineering of a common PD-associated risk variant in a non-coding distal enhancer resulted in deregulation of SNCA expression, a key gene implicated in the pathogenesis of PD, by as little as 10% (Soldner et al., 2016). In order to detect such subtle gene expression differences, an allele-specific assay was developed that allowed the analysis of *cis*-acting effects of candidate variants on allele-specific gene expression as a consequence of deletion or exchange of disease-associated regulatory elements. Detailed analysis of isogenic cells with and without the risk allele further demonstrated that a single base pair change causes loss of transcription factor-binding sites for the transcription factors that otherwise function as a suppressor of SNCA transcription on a non-risk-associated allele.

Epidemiology and population genetics suggest that Sporadic Alzheimer Disease (SAD) results from complex interactions between genetic risk variants and environmental factors. In another approach to study risk alleles, patient-derived hiPSCs were used to dissect the effect of common SAD-associated non-coding genetic variants in the 5′ region of the *SORL1* (sortilin-related receptor, L(DLR class) A repeats containing) gene involved in intracellular vesicular trafficking (Young et al., 2015). While initial experiments did not identify a consistent correlation between SORL1 expression and either disease status or risk haplotype, a small but significant correlation between the SAD-associated *SORL1* haplotype and the BDNF-dependent response of SORL1 expression was found.

Nuclease Specificity and Off-Target Considerations

SSNs are enzymes that are targeted to specific sites in the genome, but their specificity can vary and promiscuous binding to so called off-target sites can lead to unwanted cutting and modifications. Strategies to predict, identify, and reduce these off-target events are largely dependent on the SSN design, organism, and cell type and have already been to some extent implemented in hPSCs. Understanding the frequency and impact of off-targets is highly relevant to the development of SSNs for clinical applications and their reliable use in basic research (Gabriel et al., 2011).

Several studies recently addressed the specificity of Cas9 and its off-target action (reviewed in Wu et al., 2014a). Genome-wide binding studies of dCas9 expressed in mESCs demonstrated that Cas9 can associate with a large number of genomic sites, but off-target cutting of the catalytically active Cas9 at a subset of these sites was infrequent (Wu et al., 2014b). Similarly, single-molecule imaging of Cas9 in living cells has demonstrated that Cas9 searches for target sites by 3D diffusion, and that, in contrast to on-target events, off-target binding events are, on average, short-lived (<1 s) (Knight et al., 2015).

While these data argue for the high specificity of Cas9, data in cancer cells suggest that off-targets can be frequently detected (Frock et al., 2015; Fu et al., 2014; Tsai et al., 2015; Wang et al., 2015b). For example, when using GUIDE-seq (Tsai et al., 2015), a protocol optimized in U2OS and HEK293 to detect off-targets more reliably than other methods such as ChIP-seq, Tsai et al. found many off-targets that computational algorithms had failed to predict. Based on these datasets Tsai et al. proposed that shorter guide sequences that only have about 17-nt homology to the target sequence would improve specificity (Fu et al., 2014). Moreover, the GUIDE-seq protocol was also used to engineer CRISPR-Cas9 nucleases with altered PAM specificities (Kleinstiver et al., 2015a, 2015b) and reduced off-targets (Kleinstiver et al., 2016).

An alternative protocol called BLES-seq, based on directly labeling the DSBs generated by the nuclease in situ followed by enrichment through streptavidin affinity purification and next-generation sequencing (Crosetto et al., 2013), was originally developed to detect DSBs caused by replicative stress by stalled replication in HeLa cells and mouse B lymphocytes. This protocol was further developed to assess Cas9 off-target frequencies of Cas9 and to rationally engineer Cas9 nucleases with improved specificity (Ran et al., 2015; Slaymaker et al., 2016).

Most experiments that have detected significant off-targets have been performed in cancer cells, which may have altered repair pathways that could affect recombination (Fu et al., 2013; Hsu et al., 2013). In contrast, experiments in whole organisms such as mice (Wang et al., 2013), primates (Niu et al., 2014), Zebrafish (Auer et al., 2014), or *C. elegans* (Dickinson et al., 2013) reported off-target frequencies that were low or not detectable, consistent with high specificity of the CRISPR/Cas9-mediated gene targeting. It is also possible that in non-transformed cells off-target cleavages are efficiently counter-selected by the endogenous DNA-damage response. As hPSCs are primary cells with genetically intact check-points it seems possible that off-targets will accumulate less frequently in hPSCs than has been observed in cancer cells. To address this it will be important to determine to what extent off-targets are the result of impaired checkpoint control of cancer cells and whether there are specific cell types and conditions that are predisposed for the accumulation of off-targets. Data from conventional whole-genome sequencing of hPSCs exposed to Cas9 have so far been limited and do not yet fully address the issues due to small sample sizes (Park et al., 2015; Smith et al., 2014).

Understanding how to avoid off-target SSN modification is of particular concern for the eventual clinical application of edited cells. For basic research, however, it seems that the necessary experiments are readily available to control for the effects of eventual off-target action of SSNs. Experiments to adequately address off-target concerns include: (1) the use of several independent guide RNAs to generate a mutant cell line, (2) complementation of loss-of-function phenotypes, and (3) secondary editing of the mutant cell line to revert the mutation to a WT allele followed by confirmation of phenotypic rescue.

Large-Scale Screens, Epigenetic Editing, and Other Applications for Cas9 in iPSCs

In addition to allowing the easy, fast, and inexpensive editing of hPSCs, the advent of Cas9 as a programmable DNA-binding protein allowed the development of forward genetics methodologies that were previously not readily available in hESCs. It is trivial to multiplex guide RNA synthesis, which allows the generation of large barcoded libraries of sgRNAs with several-fold coverage of every gene in the human genome. These libraries can be employed in loss-of-function screens, for example, to identify gene products that are required for drug resistance or the mediation of viral cell death (Gilbert et al., 2014; Hart et al., 2015; Parnas et al., 2015; Shalem et al., 2014; Shi et al., 2015; Wang et al., 2015a; Wang et al., 2014a; Zhou et al., 2014) (Figure 2). Most of the experiments that employ genome-wide screens have been done in cancer cells that can be expanded to accommodate the large numbers of cells that are required to perform these types of genome-wide screens. For the general implementation of these screening approaches in hPSCs or cells differentiated from hPSCs, it will be important to develop protocols that allow the expansion of these cells into large homogenous populations that in turn allow robust selection or enrichment for cellular phenotypes.

It is worth mentioning that the combination of iPSCs and genome editing has not only become a game changer for our approaches to human disease modeling, but it also provides an unprecedented opportunity to study the fundamental principles of cell biology. Previously, cell biologists mostly used aberrant cancer cell lines with often unstable and poorly defined genomes to describe human cellular behavior. This is mainly because human cancer cells presented the only reliable source of human immortal cells that could be expanded sufficiently to facilitate biochemical and genetic experimentation and could be indefinitely propagated, frozen, shipped, and shared between labs. This monopoly of cancer cells as a model system was broken with the advent of hiPSCs and the general availability of hPSCs. Like cancer cells, hPSCs are immortal, but they do not suffer from the disadvantages of the pathologically altered genomes of cancer cells and yet they still retain the capacity to differentiate into any cell type of interest. The combination of hPSCs with the power of genome editing can now be used to study specific aspects of human cell biology. Exploiting this potential will be particularly important in areas of research where fundamental biological processes, such as tumor suppression, cellular immortality, or neuronal biology, diverge between human and other species. Efforts, such as the one launched by the Allen Institute for Cell Science to generate an industrial-scale library of characterized iPSCs that will be used to create a visual, animated model of the cell, suggest that iPSCs will soon replace cancer cells as a model system for basic cell biology (Callaway, 2014).

In the same way that iPSC technology had broad impacts far beyond regenerative medicine and disease modeling, the impact of the discovery of CRISPR/Cas9 on hPSCs is not only its ability to act as an SSN. Catalytically inactive forms of Cas9 (dCas9) have been successfully derived by fusions with functional proteins that bind specific loci, or the activation or repression of gene activity at the target site (Chen et al., 2013; Gilbert et al., 2014; Konermann et al., 2015; Mandegar et al., 2016; Tanenbaum et al., 2014) (CRISPRa and CRISPRi, Figure 2). Some of these platforms have been successfully implemented for genome-wide screens and the manipulation of hPSCs. As demonstrated for TALE proteins and zinc finger DNA binding domains, the range of Cas9 could be extended in the future to also methylate or demethylate DNA or histones/chromatin at precise locations in the genome (Maeder et al., 2013; Meister et al., 2010). Moreover, dCas9 fused to fluorescent reporters has been developed to indicate nuclear organization by visualizing individual genomic loci (Chen et al., 2013; Gilbert et al., 2014; Tanenbaum et al., 2014). It is exciting that more applications are being developed; recently, Cas9 has been programmed to target RNA in vitro and in vivo (O'Connell et al., 2014; Nelles et al., 2016), raising the possibly that it could be used to better understand the transcriptome in addition to the genome.

Arguably the most far-reaching consequence of CRISPR/Cas9 gene targeting is the potential to edit the germline. Because gene editing by homologous recombination is inefficient, cells carrying the desired targeting event need to be selected in culture. Thus, germline modification in the past was restricted to mice as chimera-competent ESCs are not available in other species. Because CRISPR/Cas9 gene editing is so efficient, it requires no selection for the desired targeting events, rendering ESCs superfluous for the generation of mutant animals. CRISPR/Cas9 enabled gene editing in the zygote and was used to efficiently generate animals carrying defined mutations in multiple species including fish, *Drosophila*, mice, and primates (Bassett et al., 2013; Chang et al., 2013; Gratz et al., 2013; Hwang et al.,

2013a, 2013b; Niu et al., 2014; Wang et al., 2013; Yang et al., 2013; Yu et al., 2013).

Challenges and Next Steps

Despite the obvious advances that have been made as a result of iPSC and editing technologies, several challenges remain. A key limitation remains that human cells prefer to choose the imprecise NHEJ pathway to repair a DSB rather than use the more precise homologous DNA repair pathway using an exogenous repair template (Chapman et al., 2012). Due to this pathway choice, editing events often result in NHEJ-mediated insertions and deletions at the DSB rather than the intended homology-mediated modification. NHEJ-mediated gene disruption can be useful when the researcher or clinician intends to generate a loss-of-function event. However, in most clinical treatment settings the generation of a defined allele with high frequency will be essential to devise treatment options that require editing to result in gain of function at endogenous genes. Approaches to shift the balance away from NHEJ and toward homology-mediated repair included inhibiting NHEJ with small molecules or controlling the timing of CRISPR/Cas9 delivery with respect to the cell-cycle stage (Chu et al., 2015; Maruyama et al., 2015; Robert et al., 2015; Yu et al., 2015). These approaches are promising, yet we are currently far away from testing the efficacy of treatment strategies that rely on gene repair or gain-of-function approaches using high-frequency HR repair events of endogenous genes.

Facing this challenge, recent studies used creative ways to take advantage of NHEJ-meditated genome editing and the fact that the simultaneous expression of two nucleases can meditate the excision or inversion of the sequence internal to the two SSNs (Chiba et al.,2015; Chen et al., 2011; Young et al., 2016). In the specific case of Duchenne muscular dystrophy, Cas9 was employed to excise 725 kb of genomic sequences, which removed a premature STOP codon in the disease-causing DMD gene and thereby restored the reading frame and partial protein function (Young et al., 2016).

Similarly, Cas9-mediated genome editing in patient-specific iPSCs was used to genetically correct the disease-causing chromosomal inversions found in patients with Hemophilia A, demonstrating that NHEJ-based approaches can be used to model and correct large-scale genomic alterations underlying human disease (Park et al., 2015).

Elegant work that also takes advantage of the fact that genomic sequences between two SSN cuts can reinsert back into the locus in an inverted manner recently demonstrated that CTCF sites interact with each other in an orientation-dependent manner (Guo et al., 2015). Using this approach Guo et al. elucidate the impact of the directionality of CTCF sites in the mediation of large-scale genome interactions and transcriptional regulation.

Another challenge of genome editing in human cells is that human cells have relatively short conversion tracts (Elliott et al., 1998). This means that even when a DSB is repaired by homology-directed repair (HDR) and not the NHEJ machinery, modifications can only be made with reasonable frequency very close to one side of the DSB. This presents a major obstacle toward the introduction of complex genetic changes in hPSCs. The use of Cpf1, a class 2 CRISPR effector that uses the same basic principles as Cas9, but cleaves DNA further away from the PAM sequence and generates a single-stranded overhang, may help increase the rate of HDR over NHEJ events (Zetsche et al., 2015). Overcoming this challenge will significantly facilitate the engineering of human stem cells, as it will allow us to refine the human genome more efficiently. Eventually this could result in similar resources that have been used in yeast and mESCs, such as a comprehensive collection of conditional human knockout iPSC libraries, with a homozygous iPSC line for each human gene carrying an exon flanked by LoxP sites.

Rethinking the Ethical Debate

It will be important in the near future to navigate the ethical debate that arises from the confluence of genome editing with stem cell technology. This requires a policy framework that supports scientific progress that is independent of special interest groups that would bias a rational risk benefit assessment of this technology. The rampant progress that has been made over the last few years to improve genome editing technologies and to detect and reduce potential off-targets of SSNs has already lead to the first clinical trials for HIV, which are trailblazing through the necessary regulatory hurdles (Tebas et al., 2014). Somatic cell editing and editing in hPSCs in vivo and/or ex vivo coupled with transplantation will progress to become a standard clinical application. These efforts have to be clearly distinguished from editing human germ cells or totipotent cells of the early human embryo. Indeed, the efficiency of altering the genome of mammals by injecting CRISPR/Cas9 RNA or DNA into the fertilized egg (Wang et al., 2013) sparked a debate on whether this technology should be used to modify the human germline (Sheridan, 2015). While technical challenges currently limit the potential application of such modifications, two recent papers describe gene editing of the embryo's genome following injection of gRNAs, CRSPR/Cas9 RNA, and targeting oligos into human zygotes (Kang et al., 2016; Liang et al., 2015). These studies raise a number of scientific issues such as off-target rate, mosaicism, and the likely alteration of the non-targeted wild-type allele when a mutant allele is targeted. More importantly, the technology raises serious ethical issues: do we want to irreversibly alter the human germline? Thus, the clinical application of this gene editing technology for medical purposes raises important ethical issues that will need to be widely discussed and agreed upon as it would affect future generations.

ACKNOWLEDGMENTS

We thank the members of the Hockemeyer laboratory for helpful discussion. D.H. is a New Scholar in Aging of the Ellison Medical Foundation and is supported by the Glenn Foundation as well as the The Shurl and Kay Curci Foundations. The work in the Hockemeyer laboratory is supported by NIH R01 CA196884-01, and in the Jaenisch laboratory, by NIH grants 1R01NS088538-01 and 2R01MH104610-15. R.J. is an advisor to Stemgent and Fate Therapeutics.

REFERENCES

Ambrosi, D.J., and Rasmussen, T.P. (2005). Reprogramming mediated by stem cell fusion. J. Cell. Mol. Med. *9*, 320–330.

An, M.C., Zhang, N., Scott, G., Montoro, D., Wittkop, T., Mooney, S., Melov, S., and Ellerby, L.M. (2012). Genetic correction of Huntington's disease phenotypes in induced pluripotent stem cells. Cell Stem Cell *11*, 253–263.

Apostolou, E., and Hochedlinger, K. (2013). Chromatin dynamics during cellular reprogramming. Nature *502*, 462–471.

Auer, T.O., Duroure, K., De Cian, A., Concordet, J.P., and Del Bene, F. (2014). Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. Genome Res. *24*, 142–153.

Avior, Y., Sagi, I., and Benvenisty, N. (2016). Pluripotent stem cells in disease modelling and drug discovery. Nat. Rev. Mol. Cell Biol. *17*, 170–182.

Bassett, A.R., Tibbit, C., Ponting, C.P., and Liu, J.L. (2013). Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep. *4*, 220–228.

Bedada, F.B., Wheelwright, M., and Metzger, J.M. (2015). Maturation status of sarcomere structure and function in human iPSC-derived cardiac myocytes. Biochim. Biophys. Acta, in press. Published online November 11, 2015. http://dx.doi.org/10.1016/j.bbamcr.2015.11.005.

Beumer, K.J., Trautman, J.K., Bozas, A., Liu, J.L., Rutter, J., Gall, J.G., and Carroll, D. (2008). Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. Proc. Natl. Acad. Sci. USA *105*, 19821–19826.

Bibikova, M., Golic, M., Golic, K.G., and Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics *161*, 1169–1175.

Bibikova, M., Beumer, K., Trautman, J.K., and Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. Science *300*, 764.

Blair, J.D., Bateup, H.S., and Hockemeyer, D.F. (2016). Establishment of Genome-edited Human Pluripotent Stem Cell Lines: From Targeting to Isolation. J. Vis. Exp. *108*.

Bojesen, S.E., Pooley, K.A., Johnatty, S.E., Beesley, J., Michailidou, K., Tyrer, J.P., Edwards, S.L., Pickett, H.A., Shen, H.C., Smart, C.E., et al. (2013). Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat. Genet. *45*, 371–384, 384e371–372.

Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., and Vallier, L. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191–195.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209–1222.

Byrne, S.M., and Church, G.M. (2015). Crispr-mediated Gene Targeting of Human Induced Pluripotent Stem Cells. Curr. Protoc. Stem Cell Biol. *35*, 5a.8.1-22.

Callaway, E. (2014). Microsoft billionaire takes on cell biology. Nature *516*, 157.

Campbell, K.H., McWhir, J., Ritchie, W.A., and Wilmut, I. (1996). Sheep cloned by nuclear transfer from a cultured cell line. Nature *380*, 64–66.

Carroll, D. (2014). Genome engineering with targetable nucleases. Annu. Rev. Biochem. *83*, 409–439.

Chang, N., Sun, C., Gao, L., Zhu, D., Xu, X., Zhu, X., Xiong, J.W., and Xi, J.J. (2013). Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. Cell Res. *23*, 465–472.

Chapman, J.R., Taylor, M.R., and Boulton, S.J. (2012). Playing the end game: DNA double-strand break repair pathway choice. Mol. Cell *47*, 497–510.

Chen, F., Pruett-Miller, S.M., Huang, Y., Gjoka, M., Duda, K., Taunton, J., Collingwood, T.N., Frodin, M., and Davis, G.D. (2011). High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat. Methods *8*, 753–755.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479–1491.

Chen, Y., Cao, J., Xiong, M., Petersen, A.J., Dong, Y., Tao, Y., Huang, C.T., Du, Z., and Zhang, S.C. (2015). Engineering Human Stem Cell Lines with Inducible Gene Knockout using CRISPR/Cas9. Cell Stem Cell *17*, 233–244.

Chen, J.R., Tang, Z.H., Zheng, J., Shi, H.S., Ding, J., Qian, X.D., Zhang, C., Chen, J.L., Wang, C.C., Li, L., et al. (2016). Effects of genetic correction on the differentiation of hair cell-like cells from iPSCs with MYO15A mutation. Cell Death Differ., in press. Published online February 26, 2016. http://dx.doi.org/10.1038/cdd.2016.16.

Chiba, K., and Hockemeyer, D. (2015). Genome editing in human pluripotent stem cells using site-specific nucleases. Methods Mol. Biol. *1239*, 267–280.

Chiba, K., Johnson, J.Z., Vogan, J.M., Wagner, T., Boyle, J.M., and Hockemeyer, D. (2015). Cancer-associated TERT promoter mutations abrogate telomerase silencing. eLife *4*, 4.

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kühn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat. Biotechnol. *33*, 543–548.

Chung, C.Y., Khurana, V., Auluck, P.K., Tardiff, D.F., Mazzulli, J.R., Soldner, F., Baru, V., Lou, Y., Freyzon, Y., Cho, S., et al. (2013). Identification and rescue of α-synuclein toxicity in Parkinson patient-derived neurons. Science *342*, 983–987.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Costa, M., Dottori, M., Sourris, K., Jamshidi, P., Hatzistavrou, T., Davis, R., Azzola, L., Jackson, S., Lim, S.M., Pera, M., et al. (2007). A method for genetic modification of human embryonic stem cells using electroporation. Nat. Protoc. *2*, 792–796.

Crane, A.M., Kramer, P., Bui, J.H., Chung, W.J., Li, X.S., Gonzalez-Garay, M.L., Hawkins, F., Liao, W., Mora, D., Choi, S., et al. (2015). Targeted correction and restored function of the CFTR gene in cystic fibrosis induced pluripotent stem cells. Stem Cell Reports *4*, 569–577.

Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K., et al. (2013). Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. Nat. Methods *10*, 361–365.

Davis, R.P., Costa, M., Grandela, C., Holland, A.M., Hatzistavrou, T., Micallef, S.J., Li, X., Goulburn, A.L., Azzola, L., Elefanty, A.G., and Stanley, E.G. (2008a). A protocol for removal of antibiotic resistance cassettes from human embryonic stem cells genetically modified by homologous recombination or transgenesis. Nat. Protoc. *3*, 1550–1558.

Davis, R.P., Ng, E.S., Costa, M., Mossman, A.K., Sourris, K., Elefanty, A.G., and Stanley, E.G. (2008b). 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.

DeKelver, R.C., Choi, V.M., Moehle, E.A., Paschon, D.E., Hockemeyer, D., Meijsing, S.H., Sancak, Y., Cui, X., Steine, E.J., Miller, J.C., et al. (2010). Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. Genome Res. *20*, 1133–1142.

Dickinson, D.J., Ward, J.D., Reiner, D.J., and Goldstein, B. (2013). Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. Nat. Methods *10*, 1028–1034.

Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature *330*, 576–578.

Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J.A., and Jasin, M. (1998). Gene conversion tracts from double-strand break repair in mammalian cells. Mol. Cell. Biol. *18*, 93–101.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156.

Firth, A.L., Menon, T., Parker, G.S., Qualls, S.J., Lewis, B.M., Ke, E., Dargitz, C.T., Wright, R., Khanna, A., Gage, F.H., and Verma, I.M. (2015). Functional Gene Correction for Cystic Fibrosis in Lung Epithelial Cells Generated from Patient iPSCs. Cell Rep. *12*, 1385–1390.

Forster, R., Chiba, K., Schaeffer, L., Regalado, S.G., Lai, C.S., Gao, Q., Kiani, S., Farin, H.F., Clevers, H., Cost, G.J., et al. (2014). Human intestinal tissue with

adult stem cell properties derived from pluripotent stem cells. Stem Cell Reports *2*, 838–852.

Fredriksson, N.J., Ny, L., Nilsson, J.A., and Larsson, E. (2014). Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. Nat. Genet. *46*, 1258–1263.

Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., and Kii, E. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat. Biotechnol. *33*, 179–186.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. *31*, 822–826.

Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat. Biotechnol. *32*, 279–284.

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., and Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci. *85*, 348–362.

Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaeppel, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G., et al. (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat. Biotechnol. *29*, 816–823.

Gibson, G. (2011). Rare and common variants: twenty arguments. Nat. Rev. Genet. *13*, 135–145.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

González, F., Zhu, Z., Shi, Z.D., Lelli, K., Verma, N., Li, Q.V., and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell Stem Cell *15*, 215–226.

Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J., and O'Connor-Giles, K.M. (2013). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics *194*, 1029–1035.

Guenther, M.G., Frampton, G.M., Soldner, F., Hockemeyer, D., Mitalipova, M., Jaenisch, R., and Young, R.A. (2010). Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell *7*, 249–257.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900–910.

Gurdon, J.B. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. J. Embryol. Exp. Morphol. *10*, 622–640.

Gurdon, J. (1963). Nuclear transplantation in Amphibia and the importance of stable nuclear changes in cellular differentiation. Q. Rev. Biol. *38*, 54–78.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell *163*, 1515–1526.

Hochedlinger, K., and Jaenisch, R. (2002). Monoclonal mice generated by nuclear transfer from mature B and T donor cells. Nature *415*, 1035–1038.

Hockemeyer, D., and Jaenisch, R. (2010). Gene targeting in human pluripotent cells. Cold Spring Harb. Symp. Quant. Biol. *75*, 201–209.

Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B., et al. (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat. Biotechnol. *27*, 851–857.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C., et al. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat. Biotechnol. *29*, 731–734.

Horn, S., Figl, A., Rachakonda, P.S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll, I., Nagore, E., Hemminki, K., et al. (2013). TERT promoter mutations in familial and sporadic melanoma. Science *339*, 959–961.

Hou, Z., Zhang, Y., Propson, N.E., Howden, S.E., Chu, L.F., Sontheimer, E.J., and Thomson, J.A. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc. Natl. Acad. Sci. USA *110*, 15644–15649.

Howden, S.E., Gore, A., Li, Z., Fung, H.L., Nisler, B.S., Nie, J., Chen, G., McIntosh, B.E., Gulbranson, D.R., Diol, N.R., et al. (2011). Genetic correction and analysis of induced pluripotent stem cells from a patient with gyrate atrophy. Proc. Natl. Acad. Sci. USA *108*, 6537–6542.

Howden, S.E., Maufort, J.P., Duffin, B.M., Elefanty, A.G., Stanley, E.G., and Thomson, J.A. (2015). Simultaneous Reprogramming and Gene Correction of Patient Fibroblasts. Stem Cell Reports *5*, 1109–1118.

Hrvatin, S., O'Donnell, C.W., Deng, F., Millman, J.R., Pagliuca, F.W., DiIorio, P., Rezania, A., Gifford, D.K., and Melton, D.A. (2014). Differentiated human stem cells resemble fetal, not adult, β cells. Proc. Natl. Acad. Sci. USA *111*, 3038–3043.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262–1278.

Huang, F.W., Hodis, E., Xu, M.J., Kryukov, G.V., Chin, L., and Garraway, L.A. (2013). Highly recurrent TERT promoter mutations in human melanoma. Science *339*, 957–959.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Kaini, P., Sander, J.D., Joung, J.K., Peterson, R.T., and Yeh, J.R. (2013a). Heritable and precise zebrafish genome editing using a CRISPR-Cas system. PLoS ONE *8*, e68708.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R., and Joung, J.K. (2013b). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. *31*, 227–229.

Irion, S., Luche, H., Gadue, P., Fehling, H.J., Kennedy, M., and Keller, G. (2007). Identification and targeting of the ROSA26 locus in human embryonic stem cells. Nat. Biotechnol. *25*, 1477–1482.

Jiang, J., Jing, Y., Cost, G.J., Chiang, J.C., Kolpa, H.J., Cotton, A.M., Carone, D.M., Carone, B.R., Shivak, D.A., Guschin, D.Y., et al. (2013). Translating dosage compensation to trisomy 21. Nature *500*, 296–300.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Johnson, J.Z., and Hockemeyer, D. (2015). Human stem cell-based disease modeling: prospects and challenges. Curr. Opin. Cell Biol. *37*, 84–90.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., and Woltjen, K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature *458*, 771–775.

Kang, L., Wang, J., Zhang, Y., Kou, Z., and Gao, S. (2009). iPS cells can support full-term development of tetraploid blastocyst-complemented embryos. Cell Stem Cell *5*, 135–138.

Kang, X., He, W., Huang, Y., Yu, Q., Chen, Y., Gao, X., Sun, X., and Fan, Y. (2016). Introducing precise genetic modifications into human 3PN embryos by CRISPR/Cas-mediated genome editing. J. Assist. Reprod. Genet., in press April 6, 2016. http://dx.doi.org/10.1007/s10815-016-0710-8.

Killela, P.J., Reitman, Z.J., Jiao, Y., Bettegowda, C., Agrawal, N., Diaz, L.A., Jr., Friedman, A.H., Friedman, H., Gallia, G.L., Giovanella, B.C., et al. (2013). TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. Proc. Natl. Acad. Sci. USA *110*, 6021–6026.

Kim, D., Kim, C.H., Moon, J.I., Chung, Y.G., Chang, M.Y., Han, B.S., Ko, S., Yang, E., Cha, K.Y., Lanza, R., and Kim, K.S. (2009). Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. Cell Stem Cell *4*, 472–476.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I.R., et al. (2010). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285–290.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., Yeh, J.R., et al. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature *523*, 481–485.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Topkar, V.V., Zheng, Z., and Joung, J.K. (2015a). Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat. Biotechnol. *33*, 1293–1298.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., Yeh, J.R., et al. (2015b). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature *523*, 481–485.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495.

Knight, S.C., Xie, L., Deng, W., Guglielmi, B., Witkowsky, L.B., Bosanac, L., Zhang, E.T., El Beheiry, M., Masson, J.B., Dahan, M., et al. (2015). Dynamics of CRISPR-Cas9 genome interrogation in living cells. Science *350*, 823–826.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

Kotini, A.G., Chang, C.J., Boussaad, I., Delrow, J.J., Dolezal, E.K., Nagulapally, A.B., Perna, F., Fishbein, G.A., Klimek, V.M., Hawkins, R.D., et al. (2015). Functional analysis of a chromosomal deletion associated with myelodysplastic syndromes using isogenic human induced pluripotent stem cells. Nat. Biotechnol. *33*, 646–655.

Kyttälä, A., Moraghebi, R., Valensisi, C., Kettunen, J., Andrus, C., Pasumarthy, K.K., Nakanishi, M., Nishimura, K., Ohtaka, M., Weltner, J., et al. (2016). Genetic Variability Overrides the Impact of Parental Cell Type and Determines iPSC Differentiation Potential. Stem Cell Reports *6*, 200–212.

Lancaster, M.A., and Knoblich, J.A. (2014). Organogenesis in a dish: modeling development and disease using organoid technologies. Science *345*, 1247125.

Lancaster, M.A., Renner, M., Martin, C.A., Wenzel, D., Bicknell, L.S., Hurles, M.E., Homfray, T., Penninger, J.M., Jackson, A.P., and Knoblich, J.A. (2013). Cerebral organoids model human brain development and microcephaly. Nature *501*, 373–379.

Li, Y., Wang, H., Muffat, J., Cheng, A.W., Orlando, D.A., Lovén, J., Kwok, S.M., Feldman, D.A., Bateup, H.S., Gao, Q., et al. (2013). Global transcriptional and translational repression in human-embryonic-stem-cell-derived Rett syndrome neurons. Cell Stem Cell *13*, 446–458.

Liang, P., Xu, Y., Zhang, X., Ding, C., Huang, R., Zhang, Z., Lv, J., Xie, X., Chen, Y., Li, Y., et al. (2015). CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. Protein Cell *6*, 363–372.

Liao, J., and Karnik, R. (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat. Genet. *47*, 469–478.

Lin, S., Staahl, B.T., Alla, R.K., and Doudna, J.A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. eLife *3*, e04766.

Lombardo, A., Genovese, P., Beausejour, C.M., Colleoni, S., Lee, Y.L., Kim, K.A., Ando, D., Urnov, F.D., Galli, C., Gregory, P.D., et al. (2007). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat. Biotechnol. *25*, 1298–1306.

Ma, N., Liao, B., Zhang, H., Wang, L., Shan, Y., Xue, Y., Huang, K., Chen, S., Zhou, X., Chen, Y., et al. (2013). Transcription activator-like effector nuclease (TALEN)-mediated gene correction in integration-free β-thalassemia induced pluripotent stem cells. J. Biol. Chem. *288*, 34671–34679.

Maeder, M.L., Angstman, J.F., Richardson, M.E., Linder, S.J., Cascio, V.M., Tsai, S.Q., Ho, Q.H., Sander, J.D., Reyon, D., Bernstein, B.E., et al. (2013). Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat. Biotechnol. *31*, 1137–1142.

Maetzel, D., Sarkar, S., Wang, H., Abi-Mosleh, L., Xu, P., Cheng, A.W., Gao, Q., Mitalipova, M., and Jaenisch, R. (2014). Genetic and chemical correction of cholesterol accumulation and impaired autophagy in hepatic and neural cells derived from Niemann-Pick Type C patient-specific iPS cells. Stem Cell Reports *2*, 866–880.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell *1*, 55–70.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Mandegar, M.A., Huebsch, N., Frolov, E.B., Shin, E., Truong, A., Olvera, M.P., Chan, A.H., Miyaoka, Y., Holmes, K., Spencer, C.I., et al. (2016). CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. Cell Stem Cell *18*, 541–553.

Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature *526*, 55–61.

Martin, G.R. (1981). 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.

Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., and Ploegh, H.L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat. Biotechnol. *33*, 538–542.

Matano, M., Date, S., Shimokawa, M., Takano, A., Fujii, M., Ohta, Y., Watanabe, T., Kanai, T., and Sato, T. (2015). Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids. Nat. Med. *21*, 256–262.

Meister, G.E., Chandrasegaran, S., and Ostermeier, M. (2010). Heterodimeric DNA methyltransferases as a platform for creating designer zinc finger methyltransferases for targeted DNA methylation in cells. Nucleic Acids Res. *38*, 1749–1759.

Merkle, F.T., and Eggan, K. (2013). Modeling human disease with pluripotent stem cells: from genome association to function. Cell Stem Cell *12*, 656–668.

Miller, J.D., Ganat, Y.M., Kishinevsky, S., Bowman, R.L., Liu, B., Tu, E.Y., Mandal, P.K., Vera, E., Shim, J.W., Kriks, S., et al. (2013). Human iPSC-based modeling of late-onset disease via progerin-induced aging. Cell Stem Cell *13*, 691–705.

Miyaoka, Y., Chan, A.H., Judge, L.M., Yoo, J., Huang, M., Nguyen, T.D., Lizarraga, P.P., So, P.L., and Conklin, B.R. (2014). Isolation of single-base genome-edited human iPS cells without antibiotic selection. Nat. Methods *11*, 291–293.

Morton, J., Davis, M.W., Jorgensen, E.M., and Carroll, D. (2006). Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells. Proc. Natl. Acad. Sci. USA *103*, 16370–16375.

Nelles, D.A., Fang, M.Y., O'Connell, M.R., Xu, J.L., Markmiller, S.J., Doudna, J.A., and Yeo, G.W. (2016). Programmable RNA Tracking in Live Cells with CRISPR/Cas9. Cell *165*, 488–496.

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell Stem Cell *4*, 487–492.

Niu, Y., Shen, B., Cui, Y., Chen, Y., Wang, J., Wang, L., Kang, Y., Zhao, X., Si, W., Li, W., et al. (2014). Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. Cell *156*, 836–843.

O'Connell, M.R., Oakes, B.L., Sternberg, S.H., East-Seletsky, A., Kaplan, M., and Doudna, J.A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. Nature *516*, 263–266.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313–317.

Park, I.H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141–146.

Park, C.Y., Kim, D.H., Son, J.S., Sung, J.J., Lee, J., Bae, S., Kim, J.H., Kim, D.W., and Kim, J.S. (2015). Functional Correction of Large Factor VIII Gene Chromosomal Inversions in Hemophilia A Patient-Derived iPSCs Using CRISPR-Cas9. Cell Stem Cell *17*, 213–220.

Parnas, O., Jovanovic, M., Eisenhaure, T.M., Herbst, R.H., Dixit, A., Ye, C.J., Przybylski, D., Platt, R.J., Tirosh, I., Sanjana, N.E., et al. (2015). A Genome-

wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. Cell *162*, 675–686.

Polo, J.M., Liu, S., Figueroa, M.E., Kulalert, W., Eminli, S., Tan, K.Y., Apostolou, E., Stadtfeld, M., Li, Y., Shioda, T., et al. (2010). Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat. Biotechnol. *28*, 848–855.

Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature *520*, 186–191.

Reinhardt, P., Schmid, B., Burbulla, L.F., Schöndorf, D.C., Wagner, L., Glatza, M., Höing, S., Hargus, G., Heck, S.A., Dhingra, A., et al. (2013). Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression. Cell Stem Cell *12*, 354–367.

Rideout, W.M., 3rd, Hochedlinger, K., Kyba, M., Daley, G.Q., and Jaenisch, R. (2002). Correction of a genetic defect by nuclear transplantation and combined cell and gene therapy. Cell *109*, 17–27.

Robert, F., Barbeau, M., Éthier, S., Dostie, J., and Pelletier, J. (2015). Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. Genome Med. *7*, 93.

Robinton, D.A., and Daley, G.Q. (2012). The promise of induced pluripotent stem cells in research and therapy. Nature *481*, 295–305.

Rouet, P., Smih, F., and Jasin, M. (1994a). Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc. Natl. Acad. Sci. USA *91*, 6064–6068.

Rouet, P., Smih, F., and Jasin, M. (1994b). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol. Cell. Biol. *14*, 8096–8106.

Rouhani, F., Kumasaka, N., de Brito, M.C., Bradley, A., Vallier, L., and Gaffney, D. (2014). Genetic background drives transcriptional variation in human induced pluripotent stem cells. PLoS Genet. *10*, e1004432.

Ruby, K.M., and Zheng, B. (2009). Gene targeting in a HUES line of human embryonic stem cells via electroporation. Stem Cells *27*, 1496–1506.

Ryan, S.D., Dolatabadi, N., Chan, S.F., Zhang, X., Akhtar, M.W., Parker, J., Soldner, F., Sunico, C.R., Nagar, S., Talantova, M., et al. (2013). Isogenic human iPSC Parkinson's model shows nitrosative stress-induced dysfunction in MEF2-PGC1α transcription. Cell *155*, 1351–1364.

Sato, T., and Clevers, H. (2013). Growing self-organizing mini-guts from a single intestinal stem cell: mechanism and applications. Science *340*, 1190–1194.

Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J., and Clevers, H. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature *459*, 262–265.

Sato, T., Stange, D.E., Ferrante, M., Vries, R.G., Van Es, J.H., Van den Brink, S., Van Houdt, W.J., Pronk, A., Van Gorp, J., Siersema, P.D., and Clevers, H. (2011). Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology *141*, 1762–1772.

Schwank, G., Koo, B.K., Sasselli, V., Dekkers, J.F., Heo, I., Demircan, T., Sasaki, N., Boymans, S., Cuppen, E., van der Ent, C.K., et al. (2013). Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. Cell Stem Cell *13*, 653–658.

Sebastiano, V., Zhen, H.H., Haddad, B., Bashkirova, E., Melo, S.P., Wang, P., Leung, T.L., Siprashvili, Z., Tichy, A., Li, J., et al. (2014). Human COL7A1-corrected induced pluripotent stem cells for the treatment of recessive dystrophic epidermolysis bullosa. Sci. Transl. Med. *6*, 264ra163.

Sexton, A.N., Regalado, S.G., Lai, C.S., Cost, G.J., O'Neil, C.M., Urnov, F.D., Gregory, P.D., Jaenisch, R., Collins, K., and Hockemeyer, D. (2014). Genetic and molecular identification of three human TPP1 functions in telomerase action: recruitment, activation, and homeostasis set point regulation. Genes Dev. *28*, 1885–1899.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87.

Sheridan, C. (2015). CRISPR germline editing reverberates through biotech industry. Nat. Biotechnol. *33*, 431–432.

Shi, J., Zhao, Y., Wang, K., Shi, X., Wang, Y., Huang, H., Zhuang, Y., Cai, T., Wang, F., and Shao, F. (2015). Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. Nature *526*, 660–665.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84–88.

Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., Wang, Y., Brodsky, R.A., Zhang, K., Cheng, L., and Ye, Z. (2014). Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. Cell Stem Cell *15*, 12–13.

Soldner, F., and Jaenisch, R. (2012). Medicine. iPSC disease modeling. Science *338*, 1155–1156.

Soldner, F., Hockemeyer, D., Beard, C., Gao, Q., Bell, G.W., Cook, E.G., Hargus, G., Blak, A., Cooper, O., Mitalipova, M., et al. (2009). Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. Cell *136*, 964–977.

Soldner, F., Laganière, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S., et al. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell *146*, 318–331.

Soldner, F., Stelzer, Y., Shivalila, C., Abraham, B., Latourelle, J., Barrasa, I., Goldmann, J.J., Myers, R., Young, R., and Jaenisch, R. (2016). Parkinson-associated risk variant in enhancer element produces subtle effect on target gene expression. Nature, in press. Published online April 20, 2016. http://dx.doi.org/10.1038/nature17939.

Spence, J.R., Mayhew, C.N., Rankin, S.A., Kuhar, M.F., Vallance, J.E., Tolle, K., Hoskins, E.E., Kalinichenko, V.V., Wells, S.I., Zorn, A.M., et al. (2011). Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. Nature *470*, 105–109.

Sterneckert, J.L., Reinhardt, P., and Schöler, H.R. (2014). Investigating human disease using stem cell models. Nat. Rev. Genet. *15*, 625–639.

Studer, L., Vera, E., and Cornacchia, D. (2015). Programming and Reprogramming Cellular Age in the Era of Induced Pluripotency. Cell Stem Cell *16*, 591–600.

Suzuki, S., Sargent, R.G., Illek, B., Fischer, H., Esmaeili-Shandiz, A., Yezzi, M.J., Lee, A., Yang, Y., Kim, S., Renz, P., et al. (2016). TALENs Facilitate Single-step Seamless SDF Correction of F508del CFTR in Airway Epithelial Submucosal Gland Cell-derived CF-iPSCs. Mol. Ther. Nucleic Acids *5*, e273.

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

Takahashi, K., and Yamanaka, S. (2015). A developmental framework for induced pluripotency. Development *142*, 3274–3285.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Takebe, T., Sekine, K., Enomura, M., Koike, H., Kimura, M., Ogaeri, T., Zhang, R.R., Ueno, Y., Zheng, Y.W., Koike, N., et al. (2013). Vascularized and functional human liver from an iPSC-derived organ bud transplant. Nature *499*, 481–484.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell *159*, 635–646.

Tebas, P., Stein, D., Tang, W.W., Frank, I., Wang, S.Q., Lee, G., Spratt, S.K., Surosky, R.T., Giedlin, M.A., Nichol, G., et al. (2014). Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. N. Engl. J. Med. *370*, 901–910.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196–199.

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell *15*, 471–487.

Thomas, K.R., and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell *51*, 503–512.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–1147.

Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J., and Joung, J.K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat. Biotechnol. *32*, 569–576.

Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. *33*, 187–197.

Urbach, A., Schuldiner, M., and Benvenisty, N. (2004). Modeling for Lesch-Nyhan disease by gene targeting in human embryonic stem cells. Stem Cells *22*, 635–641.

Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., and Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature *435*, 646–651.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nat. Rev. Genet. *11*, 636–646.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910–918.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014a). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80–84.

Wang, Y., Liang, P., Lan, F., Wu, H., Lisowski, L., Gu, M., Hu, S., Kay, M.A., Urnov, F.D., Shinnawi, R., et al. (2014b). Genome editing of isogenic human induced pluripotent stem cells recapitulates long QT phenotype for drug testing. J. Am. Coll. Cardiol. *64*, 451–459.

Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015a). Identification and characterization of essential genes in the human genome. Science *350*, 1096–1101.

Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.J., and Yee, J.K. (2015b). Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. Nat. Biotechnol. *33*, 175–178.

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A., et al. (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell *7*, 618–630.

Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J.B., Nishikawa, S., Nishikawa, S., Muguruma, K., and Sasai, Y. (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. Nat. Biotechnol. *25*, 681–686.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318–324.

Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hämäläinen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., et al. (2009). piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature *458*, 766–770.

Wood, A.J., Lo, T.W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Meng, X., et al. (2011). Targeted genome editing across species using ZFNs and TALENs. Science *333*, 307.

Wu, X., Kriz, A.J., and Sharp, P.A. (2014a). Target specificity of the CRISPR-Cas9 system. Quant. Biol. *2*, 59–70.

Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S., et al. (2014b). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat. Biotechnol. *32*, 670–676.

Xie, F., Ye, L., Chang, J.C., Beyer, A.I., Wang, J., Muench, M.O., and Kan, Y.W. (2014). Seamless gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. Genome Res. *24*, 1526–1533.

Xue, H., Wu, S., Papadeas, S.T., Spusta, S., Swistowska, A.M., MacArthur, C.C., Mattson, M.P., Maragakis, N.J., Capecchi, M.R., Rao, M.S., et al. (2009). A targeted neuroglial reporter line generated by homologous recombination in human embryonic stem cells. Stem Cells *27*, 1836–1846.

Yamada, M., Johannesson, B., Sagi, I., Burnett, L.C., Kort, D.H., Prosser, R.W., Paull, D., Nestor, M.W., Freeby, M., Greenberg, E., et al. (2014). Human oocytes reprogram adult somatic nuclei of a type 1 diabetic to diploid pluripotent stem cells. Nature *510*, 533–536.

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., and Jaenisch, R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell *154*, 1370–1379.

Ye, L., Wang, J., Beyer, A.I., Teque, F., Cradick, T.J., Qi, Z., Chang, J.C., Bao, G., Muench, M.O., Yu, J., et al. (2014). Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Δ32 mutation confers resistance to HIV infection. Proc. Natl. Acad. Sci. USA *111*, 9591–9596.

Young, J.E., Boulanger-Weill, J., Williams, D.A., Woodruff, G., Buen, F., Revilla, A.C., Herrera, C., Israel, M.A., Yuan, S.H., Edland, S.D., and Goldstein, L.S. (2015). Elucidating molecular phenotypes caused by the SORL1 Alzheimer's disease genetic risk factor using human induced pluripotent stem cells. Cell Stem Cell *16*, 373–385.

Young, C.S., Hicks, M.R., Ermolova, N.V., Nakano, H., Jan, M., Younesi, S., Karumbayaram, S., Kumagai-Cresse, C., Wang, D., Zack, J.A., et al. (2016). A Single CRISPR-Cas9 Deletion Strategy that Targets the Majority of DMD Patients Restores Dystrophin Function in hiPSC-Derived Muscle Cells. Cell Stem Cell *18*, 533–540.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917–1920.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I.I., and Thomson, J.A. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. Science *324*, 797–801.

Yu, Z., Ren, M., Wang, Z., Zhang, B., Rong, Y.S., Jiao, R., and Gao, G. (2013). Highly efficient genome modifications mediated by CRISPR/Cas9 in Drosophila. Genetics *195*, 289–291.

Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., La Russa, M., Xie, M., Ding, S., and Qi, L.S. (2015). Small molecules enhance CRISPR genome editing in pluripotent stem cells. Cell Stem Cell *16*, 142–147.

Yusa, K. (2013). Seamless genome editing in human pluripotent stem cells using custom endonuclease-based gene targeting and the piggyBac transposon. Nat. Protoc. *8*, 2061–2078.

Yusa, K., Rashid, S.T., Strick-Marchand, H., Varela, I., Liu, P.Q., Paschon, D.E., Miranda, E., Ordóñez, A., Hannan, N.R., Rouhani, F.J., et al. (2011). Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. Nature *478*, 391–394.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759–771.

Zhao, X.Y., Li, W., Lv, Z., Liu, L., Tong, M., Hai, T., Hao, J., Guo, C.L., Ma, Q.W., Wang, L., et al. (2009). iPS cells produce viable mice through tetraploid complementation. Nature *461*, 86–90.

Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., and Wei, W. (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. Nature *509*, 487–491.

Zou, J., Maeder, M.L., Mali, P., Pruett-Miller, S.M., Thibodeau-Beganny, S., Chou, B.K., Chen, G., Ye, Z., Park, I.H., Daley, G.Q., et al. (2009). Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. Cell Stem Cell *5*, 97–110.

Zou, J., Mali, P., Huang, X., Dowey, S.N., and Cheng, L. (2011). Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. Blood *118*, 4599–4608.

Zwaka, T.P., and Thomson, J.A. (2003). Homologous recombination in human embryonic stem cells. Nat. Biotechnol. *21*, 319–321.

Trends in Cell Biology

Special Issue: Future of Cell Biology

Review

Applications of CRISPR Genome Engineering in Cell Biology

Fangyuan Wang[1] and Lei S. Qi[2,3,4,*]

Recent advances in genome engineering are starting a revolution in biological research and translational applications. The clustered regularly interspaced short palindromic repeats (CRISPR)-associated RNA-guided endonuclease CRISPR associated protein 9 (Cas9) and its variants enable diverse manipulations of genome function. In this review, we describe the development of Cas9 tools for a variety of applications in cell biology research, including the study of functional genomics, the creation of transgenic animal models, and genomic imaging. Novel genome engineering methods offer a new avenue to understand the causality between the genome and phenotype, thus promising a fuller understanding of cell biology.

Trends

The RNA-guided CRISPR/Cas9 endonuclease and the endonuclease-dead dCas9 protein are powerful genomic manipulation tools for gene editing, transcriptional regulation, and epigenetic modifications.

Both Cas9 and dCas9 enable diverse types of high-throughput screening of gene functions in cell lines and *in vivo*.

The CRISPR/Cas9 accelerates the establishment of many useful transgenic animal models for biomedical research.

The CRISPR/Cas9 is repurposed for genomic imaging and lineage tracing in living cells and tissues.

From DNA Repair Pathways to CRISPR/Cas9-Mediated Genome Editing

Eukaryotic cells use a sophisticated network of genes and genomic regulatory elements to carry out functions related to cell growth and death, organelle formation and organization, metabolite production, and microenvironment sensing. The ability to precisely manipulate the genome is essential to understanding complex and dynamic cellular processes. Broadly speaking, genome engineering defines methodological approaches to alter genomic DNA sequence (gene editing), modify epigenetic marks (epigenetic editing), modulate functional output (transcriptional regulation), and reorganize chromosomal structure (structural manipulation) (Figure 1). These goals require a toolkit of designer molecules that can be conveniently constructed and delivered into cells to perform one of the above functions.

Naturally occurring systems and pathways have provided a rich resource for tool building. The discovery of the homology-directed repair (HDR) pathway inspired a method to modify the DNA sequence at a precise genomic locus in a targeted manner. Using the HDR pathway, a designed DNA template with flanking homologous sequences could be used to precisely recombine at the target genomic locus [1]. However, this application is usually a highly inefficient process in mammalian cells and tissues. By contrast, the presence of a double-stranded DNA break (DSB) can enhance efficiency [2,3]. Furthermore, it has been shown that, in the absence of a DNA template, eukaryotic cells may generate almost random deletion or insertion indels at the site of a DSB via the alternative nonhomology end joining (NHEJ) pathway, offering another approach for targeted gene knockout [4].

Following the developments described above, a major question in the field of gene editing was how to introduce site-specific DSBs to initiate the DNA repair process. Molecules that allow sequence-specific DNA binding were of primary interest. These included programmable

[1]Sino-U.S. Center of Synthetic Biology, Shanghai Institute of Rheumatology, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China
[2]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
[3]Department of Chemical and Systems Biology, Stanford University, Stanford, CA 94305, USA
[4]ChEM-H, Stanford University, Stanford, CA 94305, USA

*Correspondence: stanley.qi@stanford.edu (L.S. Qi).

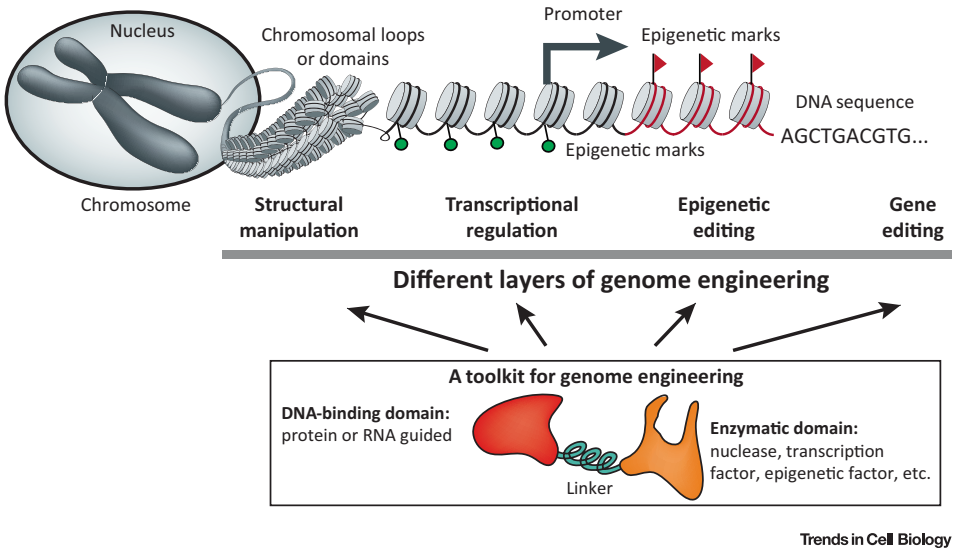


Figure 1. A Schematic View of the Diverse Goals of Genome Engineering. Genome engineering defines methodological approaches to alter the DNA sequence (gene editing), modify the epigenetic marks (epigenetic editing), modulate the functional output (transcriptional regulation), and reorganize the chromosomal structure (structural manipulation).

endonucleases engineered from zinc finger proteins (ZFNs) or transcription activator-like effectors (TALENs) [5,6]. The peptide domains of these proteins could be designed following a simple set of rules for protein–DNA recognition. However, their utility was hindered by an often costly and tedious construction process and by a context-dependency issue in the protein design [7,8]. Nevertheless, previous work showed that these programmable DNA-binding proteins could be coupled to nuclease domains, transcriptional repressors or activators, and epigenetic modifiers to enable diverse types of genomic manipulation [9–12]. However, it remained to be understood how to precisely target a specific DNA sequence of interest via an even simpler mechanism, such as Watson-Crick base pairing.

The CRISPR/Cas system performs such a function. Truly a gift from Nature [13,14], the CRISPR/Cas system was discovered initially in *Escherichia coli* during the 1980s [15], but its function remained elusive until 2007. Working in the yogurt production bacterium *Streptococcus thermophilus*, earlier work demonstrated that encoding the bacteriophage sequence from the host CRISPR locus conferred acquired resistance against the same bacteriophage [16]. Later work showed that CRISPR utilized small CRISPR-associated RNAs (crRNAs) to guide the nuclease activity of Cas proteins in *E. coli* [17]. Together, these studies uncovered a RNA-guided nuclease mechanism for the CRISPR system, which also suggested a genetic system with high specificity and efficiency for DNA binding and cleavage.

The practical use of CRISPR for gene editing began with the elucidation of the mechanism of the type II CRISPR system [18]. The type II CRISPR from *Streptococcus pyogenes* encodes a RNA-guided endonuclease protein, Cas9, which was shown to use only two small RNAs (a mature crRNA and a *trans*-acting tracrRNA) for sequence-specific DNA cleavage [18–20]. Furthermore, a chimeric single guide RNA (sgRNA) fused between crRNA and tracrRNA recapitulated the structure and function of the tracrRNA–crRNA complex, which could efficiently direct Cas9 to induce DSBs *in vitro* [18]. The rules used by Cas9 to search for a DNA target are elegant and simple, requiring only a 20-nucleotide (nt) sequence on the sgRNA that base pairs with the target DNA and the presence of a DNA protospacer adjacent motif (PAM) adjacent to the complimentary region [18,21].

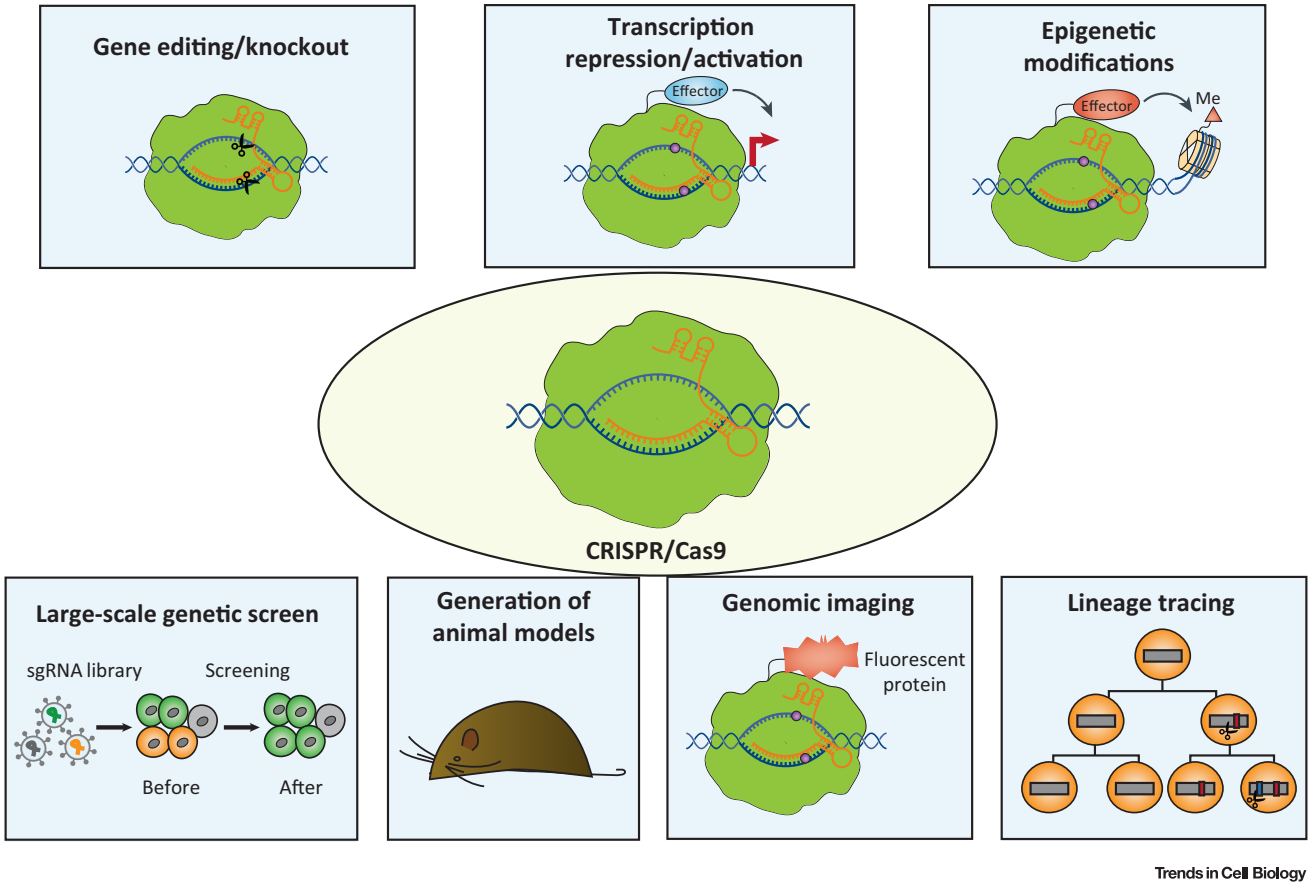


Figure 2. Applications of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-Associated Protein 9 (Cas9) to Cell Biology Research. CRISPR/Cas9 technology has been used for gene editing, transcriptional regulation, epigenetic regulation, large-scale genetic screens, generation of animal models, and genomic imaging. Abbreviation: sgRNA, single guide RNA.

The Cas9 complex has since been developed as a remarkably useful tool for genome editing. As demonstrated by the pioneering work in several cell types and organisms [22–26], the Cas9/sgRNA complex can efficiently generate DSBs, which then facilitates NHEJ-mediated gene knockout or HDR-mediated recombination. This system has since gained rapid acceptance and has been used for genome editing in essentially all organisms that can be cultured in the laboratory. In this review, we focus on recent applications of CRISPR/Cas9 in cell biology research using mammalian cell cultures and animal models (Figure 2).

An Expanding CRISPR Toolkit for RNA-Guided Genome Editing

The different types of natural CRISPR system encode a toolkit for genome editing. Six major types of CRISPR system have been identified from different organisms (types I–VI), with various subtypes in each major type [27,28]. Within the type II CRISPR system, several species of Cas9 have been characterized from *S. pyogenes*, *Streptococcus thermophilus*, *Neisseria meningitidis*, *Staphylococcus aureus*, and *Francisella novicida* [18,29–34]. While these Cas9s have a similar RNA-guided DNA-binding mechanism, they often have distinct PAM recognition sequences. Similar to the toolkit of restriction enzymes for molecular cloning, a large toolkit of Cas9s expands the targetable genome sequence for gene editing and genome manipulation.

Other types of CRISPR system may exhibit different mechanisms. For example, the Type III-B CRISPR system from *Pyrococcus furiosus* uses a Cas complex for RNA-directed RNA cleavage

[35,36], which is indicative of a mechanism for targeting and modulating RNAs in cells. The recent discovery of the protein Cpf1 from the *Prevotella* and *Francisella*-1 type V CRISPR showed that Cpf1 uses a short crRNA without a tracrRNA for RNA-guided DNA cleavage [37–40]. Both biochemical and cell culture work showed that Cpf1-mediated genome targeting is effective and specific, comparable with the *S. pyogenes* Cas9. The type VI-A CRISPR effector C2c2 from the bacterium *Leptotrichia shahii* is a RNA-guided RNase that can be programmed to knock down specific mRNAs in bacteria [41]. These results broaden our understanding of the diversity of natural CRISPR/Cas systems, which also provide a functionally diverse set of tools.

Other enzymatic domains can also be harnessed for genome editing. For example, instead of using the endonuclease activity of Cas9, a mutation in one nuclease domain of Cas9 can create a nickase Cas9 (nCas9) that can cleave one strand of DNA [42]. With a pair of sgRNAs, the specificity of genome editing could be enhanced by using a pair of nCas9s that target each strand of DNA at adjacent sites. Furthermore, recent work demonstrated that a Cas9-fused cytidine deaminase enzyme allowed for direct conversion of a C to T (or G to A) substitution [43]. In this work, fusing the nuclease-deactivated dCas9 or the nCas9 with a cytidine deaminase domain corrected point mutations relevant to human disease without DSBs; therefore, avoiding NHEJ-mediated indel formation.

Applications of CRISPR/Cas9 for Cell Biological Studies

The CRISPR/Cas9 technology has accelerated the discovery and mechanistic interrogation of the genome and organelles in diverse types of cell and organism. Some examples of utilizing CRISPR/Cas9 for studying cellular organelles are summarized in Table 1 and Figure 3. Beyond using CRISPR/Cas9 as a gene-editing tool, we describe the development of CRISPR/Cas9 as a versatile toolkit for transcriptional control and epigenetic regulation, and highlight its utilities for large-scale genetic screens, generation of animal models, genomic imaging, and lineage tracing (Figure 2).

Transcriptional Regulation of the Genome with CRISPR/dCas9

The nuclease-dead dCas9 has provided a broad platform for programming diverse types of transcriptional or epigenetic manipulation of the genome, without altering the genome sequence. In brief, dCas9 was created by introducing point mutations into the HNH and RuvC domains to eliminate endonuclease activity [44]. This repurposed protein became a RNA-guided DNA-binding protein. In bacteria, the dCas9 protein was sufficient to induce strong sequence-specific gene repression, simply by sterically hindering the transcriptional activity of RNA polymerase [44,45]. In eukaryotic cells, fusing dCas9 to transcriptional effector proteins allowed for more efficient RNA-guided transcriptional modulation for both gene interference (CRISPRi) and activation (CRISPRa) [12,46–48].

By fusing dCas9 to transcriptional repressors, such as the Kruppel-associated box (KRAB) domain, CRISPRi can efficiently repress coding and noncoding genes, such as miRNAs and large intergenic noncoding RNAs (lincRNAs) in mammalian cells [46,47,49,50]. Compared with complete loss-of-function using Cas9, CRISPRi can use different sgRNAs that bind to different genomic loci for tunable and titratable gene repression [47]. While complete knockout is useful for studying gene function in many cases, tunable repression of a gene to different levels offers advantages when knocking out a gene leads to lethality of cells or an organism [45].

Earlier work using dCas9 fused to a peptide containing multiple VP16 domains (VP64 or VP128) could only activate endogenous genes mildly [46,51,52]; therefore, several strategies have been developed to improve CRISPRa efficiency. These include recruiting multiple copies of the VP64

Table 1. Examples of CRISPR/Cas9 Being Used for Cell Biology Research

Organelle	CRISPR/Cas9 Target	Finding	Refs
Microtubule	CRISPR/Cas9 generation of mutant flies by deleting a linker region in the centrosome protein CP190	Identified a centrosome and microtubule-targeting region in CP190 for spindle localization; deletion of linker region altered spindle morphology and led to DNA segregation errors	[114]
Mitochondrion	CRISPR/Cas9 knockout of copper transporting ATPase ATP7A in mouse 3T3-L1 cells and in fibroblasts from patients with Menkes Disease (MD)	ATP7A dysfunction damages mitochondrial redox balance	[115]
	CRISPR/Cas9 knockout of FASTKD2, a RNA-binding protein of the FAS-activated serine/threonine kinase family	Defective processing and expression of mitochondrial RNA; cellular respiration damage with depressed activities of respiratory complexes	[116]
	CRISPR/Cas9-mediated repair of ARID5B motif of rs1421085 in primary adipocytes from a patient carrying the risk allele	IRX3 and IRX5 repression restored; browning expression programs activated; thermogenesis restored	[117]
	CRISPR/Cas9-based genetic screen to study cell proliferation suppression due to inhibition of mitochondrial electron transport chain (ETC)	Identified cytosolic aspartate aminotransferase (*GOT1*) as key gene; *GOT1* loss-of-function kills cells upon ETC inhibition	[118]
Endoplasmic reticulum	CRISPR/Cas9 knockout of ATF4 or NLRP1	NLRP1 upregulated during severe ER stress; ATF4 binds and activates NLRP1 promoter during ER stress	[119]
	CRISPR/Cas9-mediated deletion of transmembrane endoribonuclease Ire1α in HEK293 cells	Ire1α forms a complex with the Sec61 translocon to cleave its mRNA substrates; disruption of Ire1α complex reduced cleavage of ER-targeted mRNA	[120]
Centrosome	CRISPR/Cas9 dual-sgRNA to generate a null abnormal spindle (*asp*) allele by excising a 750-bp fragment that included the promoter, 5′ UTR, and the first exon in *Drosophila* neuroblasts	*Asp* null mutations cause spindle defects in neuroblasts; *Asp* regulated by *Drosophila melanogaster* calmodulin (CaM) to crosslink spindle microtubules	[121]
Lysosome	Generation of Niemann-Pick type C 1 (NPC1)-deficient cell line using CRISPR/Cas9	NPC1 moves cholesterol across lysosomal glycocalyx	[122]
Ribosome	CRISPR/Cas9 knockout of nonessential gene of ribosomal protein eS25 (RPS25) in Hap1 cell line; RPS25-SNAP (mutant O6-alkylguanine DNA alkyl-transferase) transgene was transduced into RPS25-KO Hap1 cells to be the only source of the protein	Demonstrated an approach to create fluorescently labeled 40S ribosomal subunits from human cells; studied kinetics of the 40S subunit recruitment to the hepatitis C virus (HCV) internal ribosome entry site (IRES)	[123]
Golgi apparatus	Genome-wide CRISPR/Cas9 loss-of-function screen to identify host targets required for *Staphylococcus aureus* toxin alpha hemolysin (αHL) susceptibility in human myeloid cells	Identified new proteins (SYS1, ARFRP1, and TSPAN14) in regulating presentation of ADAM10 on the plasma membrane post-translationally; cells lacking sphingomyelin synthase 1 (SGMS1) resist αHL intoxication	[124]

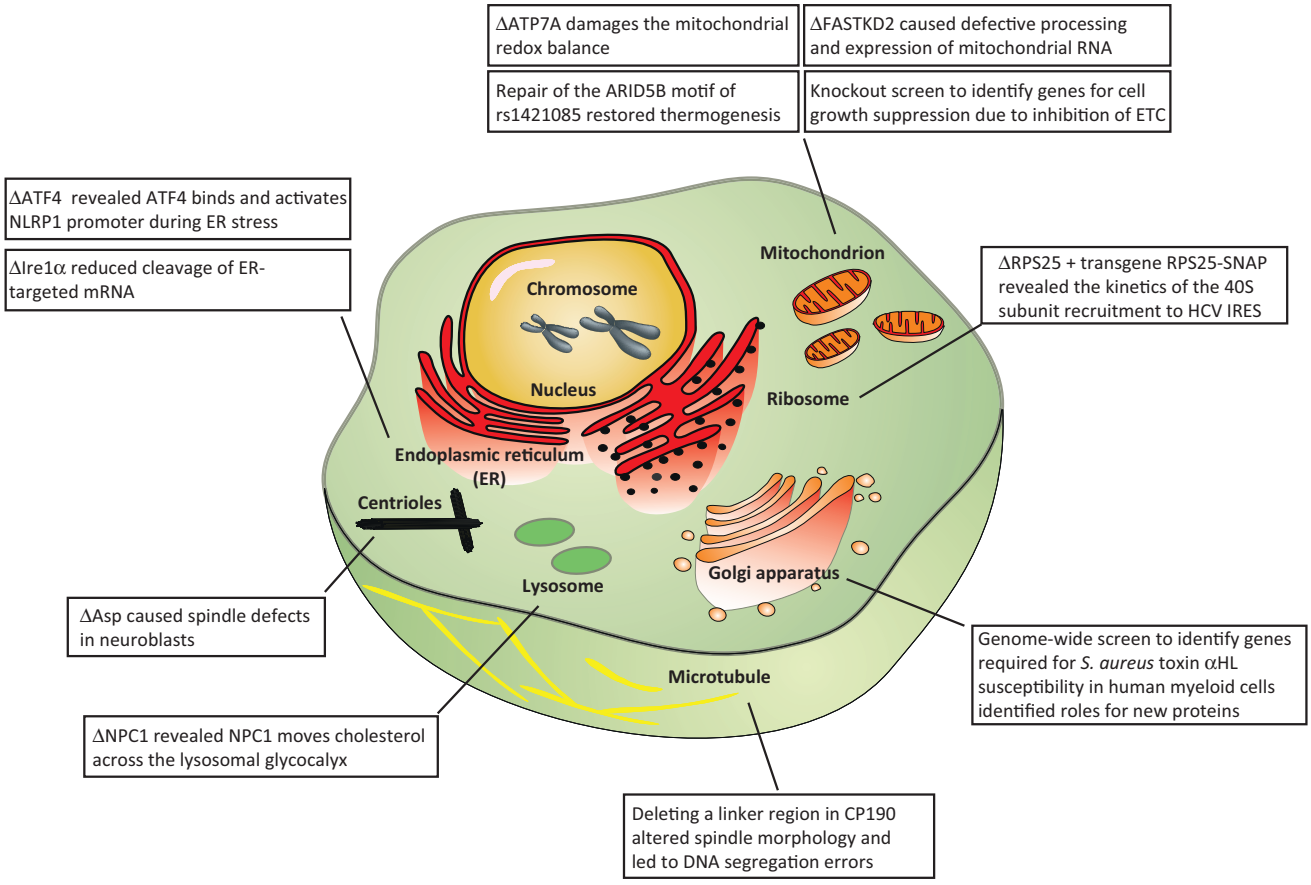


Figure 3. Examples of Applying Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-Associated Protein 9 (Cas9) Technology to Study Cellular Organelles. The figure illustrates exemplar studies in particular organelles, with more details listed in Table 1 (main text).

domain via a multimeric peptide array (SunTag), wherein each peptide domain could bind to a single-chain variable fragment (scFv) fused to VP64 [53]; fusing dCas9 to a synergistic tripartite activator system containing VP64, the activation domain of p65 (p65AD), and Epstein-Barr virus R transactivator (Rta) [54]; and combining dCas9-VP64 with a modified sgRNA engineered with two copies of an MS2 RNA hairpin that could recruit p65AD and the human heat shock factor 1 (HSF1) activation domain via interaction with the MS2-binding protein [48]. A systematic comparison of the efficacy of these methods revealed that these systems perform comparably but are dependent on the genomic and cellular context [55], suggesting that activation efficiency varies for different genes and in different types of cell. In the future, simpler, yet more effective, tools for RNA-guided gene activation should be further developed.

To repurpose more complex gene regulation, sgRNA was engineered as a class of 'scaffold' RNAs (scRNAs) that directly recruit transcription effectors without protein fusion [56]. scRNAs are generated by fusing RNA hairpins to the sgRNA, which interact with the cognate protein to recruit activators or repressors. Using engineered scRNAs, multiple genes can be simultaneously activated and repressed in the same cells. In addition to using scRNAs, multiple orthogonal species of dCas9s could also provide a platform for complex transcription regulation and sophisticated manipulation of the transcriptome.

Epigenetic Regulation with CRISPR/dCas9

dCas9 fused to epigenetic-modifying enzymes has been used to introduce locus-specific epigenetic modifications in the genome. Examples include fusing dCas9 to the core catalytic domain of the human acetyltransferase p300 ($p300^{core}$), which allowed acetylation of histone H3 Lys27 (H3K27) and upregulation of genes when binding to proximal or distal enhancers [57]; fusing dCas9 to lysine demethylase 1 (LSD1) reduced the acetylation level of H3K27 [58]; fusing dCas9 to KRAB increased the H3K9me3 mark near the target site [59]; and fusing dCas9 to the DNA methyltransferase DNMT3A increased CpG methylation near the target site [60]. These studies also demonstrated modified gene expression levels due to Cas9-mediated locus-specific epigenetic modifications. For example, in mouse embryonic stem cells, the enhancers of pluripotency factors, such as Oct4 and Tbx3, could be repressed by dCas9–LSD1 fusion, leading to loss of pluripotency [58,61].

While these examples provide an approach to edit the epigenetic states of essentially any locus in the genome, a largely unexplored question is the fate of the synthetic epigenetic marks, and whether they can be stably inherited when cells proliferate. Furthermore, given the diverse types of epigenetic modification and their mutual interactions, a comprehensive toolkit comprising multiple orthogonally acting dCas9s and their cognate sgRNA that allows the flexible editing of multiple epigenetic (histone or DNA) marks simultaneously is needed. Such a toolkit would be useful for understanding the function of diverse epigenetic marks, their interactions, and their relation to genomic and cellular functions.

Large-Scale Functional Genomic Studies Using CRISPR/Cas9

One of the powerful applications of the CRISPR/Cas9 technology is the high-throughput screening of genomic functions. The oligo libraries encoding hundreds of thousands of sgRNAs can be computationally designed and chemically synthesized to target a broad set of genome sequences. By pairing with Cas9 or dCas9 fusion proteins, this provides an approach to systematically knock out, repress, or activate genes on a large scale. The technique requires a delicate delivery method that ensures that every cell only receives a single sgRNA, usually via lentiviral or retroviral delivery into mammalian cells. The screens are frequently performed in a pooled manner, because cells transduced with the lentiviral library as a mixed population are cultured together. Via deep sequencing and analysis of the sgRNA features in the pooled cells, genes causing changes in cell growth and death can be inferred with bioinformatics. Indeed, CRISPR screens can easily identify genes, their regulatory elements, and protein domains in the mammalian genome responsible for cell growth and drug resistance [62]. A genomic tiling screen using CRISPR/Cas9 precisely mapped functional domains within enhancer elements and found that a p53-bound enhancer of the p53 effector gene *CDKN1A* was required for oncogene-induced senescence in immortalized human cells [63].

Using the endonuclease Cas9, loss-of-function genome-wide knockout screens have been performed in cultured or primary mammalian cells with sgRNA libraries (usually three–ten sgRNAs per gene) to investigate a range of phenotypes, including cell growth, cancer cell drug resistance, and viral susceptibility [64–66]. A genome-scale sgRNA library can also be used to manipulate cultured cells that are later introduced *in vivo*. Indeed, a genome-scale sgRNA library was created to mutagenize a non-metastatic mouse cancer cell line for the study of metastasis in a mouse model [67]. The mutant cell pool rapidly generated metastases when transplanted into immunocompromised mice *in vivo*. Sequencing of the metastatic cells suggested genes that accelerate lung cancer metastases and development of late-stage primary tumors. Moreover, this screening method can be extended to use in primary cells, which can lead to novel findings that are often overlooked using cell lines. Indeed, introducing a genome-wide sgRNA library into primary dendritic cells (DCs) allowed for the identification of genes related to cell growth that induce tumor necrosis factor (TNF) in response to bacterial lipopolysaccharide (LPS), an

essential host response to pathogens [68], which would otherwise be technically challenging with other genome-editing tools.

Cas9-mediated loss-of-function screens have also performed to knock out pairs of genes in combination [69]. A library of 23 409 barcoded dual sgRNA combinations was created and a pooled screen was performed to identify gene pairs in human cells that inhibit ovarian cancer cell growth in the presence of small-molecule drugs. While further work is needed to characterize the efficacy and accuracy of multiplex genetic screening, this work highlights the potential of more sophisticated functional screening studies using CRISPR.

Beyond Cas9-based complete loss-of-function screens, the invention of CRISPRi and CRISPRa further enables both partial loss-of-function and gain-of-function genetic screens [47,48]. Growth-based screens using CRISPRi/a have been used to identify essential genes, tumor suppressor genes, and potential mechanisms that confer cytotoxicity induced by a cholera-diphtheria toxin [47]. Using a library comprising approximately 70 000 guides targeting the human RefSeq coding isoforms, a CRISPRa-based screen identified genes that, upon activation, conferred resistance to a BRAF inhibitor [48].

In addition to the use of pooled screens, multi-well plates have been used in combination with the partial repression feature of CRISPRi to study the function of the full set of essential genes in the Gram-positive bacterium *Bacillus subtilis* [45]. Given that knocking out essential genes results in lethality that prevents further assay of the phenotype, partial knockdown of essential genes becomes a powerful approach. A mutant *B. subtilis* library was created to include gene partial knockdowns (approximately threefold) of all essential genes using CRISPRi, which was tested for the growth phenotype under 35 unique compounds. Using this chemical genomic approach, a comprehensive interconnecting essential gene network was identified, as well as targeted genes that interact with uncharacterized antibiotics. Inducible knockdown of essential genes also allowed for systematic characterization of cell morphology and terminal death phenotypes.

An important question is how these screens compare with each other and with other existing approaches. Several works compared different screens based on CRISPR, CRISPRi, and RNAi. One work performed comparative screens of 46 essential and 47 nonessential genes, and concluded that the CRISPR/Cas9 nuclease system outperformed the shRNA- and CRISPRi/dCas9-based gene regulation systems for the sets of essential and nonessential genes [70]. From the CRISPR screening data, the authors observed less variation across the data, and detected more functional constructs with fewer off-target effects. Another study concluded that CRISPR could identify more essential gene targets compared with RNAi [71]. Since similar precision was observed between the two approaches, it was suggested that combining data from both screens would improve the predictive accuracy. The systematic comparison of different approaches suggests that a comparative screening approach will be more powerful for studying complex cell biology phenotypes.

In addition, new methods to generate CRISPR libraries may help reduce the overall cost associated with this technique and extend its uses to screen a larger chromosomal region (e.g., the tiling along a whole chromosome). While most CRISPR libraries are generated via chemical synthesis of large pools of oligos, a new method, termed CRISPR EATING (Everything Available Turned Into New Guides), can inexpensively generate large quantities of sgRNAs for whole-genome targeting [72]. In this approach, PAM-proximal sequences are extracted by digesting input DNA with restriction enzymes that target immediately 5′ to an NGG or NAG (the PAM sequences for *S. pyogenes* Cas9, N = any nucleic acid). In this study, one library was generated and used to label the whole 3.4-mb region on *Xenopus laevis* chromosome 4 in the

egg extracts. The method allows for the generation of complex and customized libraries from any source of DNA via routine molecular biology methods.

CRISPR/Cas9 for Generating Animal Models

Genetically engineered animal models are crucial for the study of complex cellular and physiological processes. While mouse models have been widely used, the CRISPR/Cas9 gene-editing approach has been established in many other animal models, including worm [73], fly [74], fish [75,76], rat [77], rabbit [78,79], goat [80], sheep [81], dog [82], pig [83], and monkey [84]. The expansion of transgenic animal models beyond mouse is advantageous to biomedical research because it can accelerate the development of new therapeutic strategies.

CRISPR provides an easier approach to establish these transgenic animal models compared with previous gene-editing tools. Traditional approaches to construct transgenic mice via insertional mutagenesis or TALEN-mediated gene editing are time consuming, costly, and inefficient. The robustness and high efficiency of CRISPR/Cas9 simplify the process for creating model systems [85,86]. Moreover, nucleic acids encoding the Cas9 protein and target-specific sgRNAs can be conveniently injected into embryos to generate gene-modified mice with deletions of multiple genes, mutations in defined genes, or insertions of fluorescence reporters or other peptide tags to endogenous genes. For example, co-injection of Cas9 mRNA and sgRNAs targeting *Tet1* and *Tet2* into zygotes generated mice with biallelic mutations in both genes with an efficiency of 80% [85]. Furthermore, co-injection of Cas9 mRNA and sgRNAs with mutant oligos generated precise point mutations simultaneously in two target genes, while co-injecting Cas9 mRNA and sgRNAs into one-cell-stage cynomolgus monkey embryos generated founder animals harboring two gene modifications [84].

The establishment of a Cre-conditional Cas9 knock-in mouse has broadened the applications of Cas9 *in vivo* [87]. The Cas9 knock-in mouse is a great resource to rapidly generate mutations in a subpopulation of cells *in vivo*, and test how mutations cause disease phenotypes. Different methods based on adeno-associated virus (AAV), lentivirus, or nanoparticles can be used to deliver sgRNAs into multiple cell types, such as neurons, immune cells, and endothelial cells, in a Cas9 knock-in mouse to model the dynamics of significantly mutated genes in lung adenocarcinoma [87]. Another work demonstrated that the Cre-conditional Cas9 knock-in mouse phenocopied Cre-mediated genetic deletion of genes in Cre/LoxP mouse models in studying pancreatic ductal adenocarcinoma [88]. Via retrograde pancreatic ductal injection of lentiviral vectors expressing Cre and an sgRNA into Cre-conditional Cas9 knock-in mice, the authors showed knockout of *Lkb1* together with manipulated expression of oncogenic *Kras*. However, due to the heterogeneity of delivery and Cas9-mediated gene editing, caution is required when interpreting results.

In addition to using a Cas9 knock-in mouse model, viral vectors encoding Cas9 and an sgRNA can be directly delivered into wild-type mice or Cre/loxP mouse models to probe gene function. One study used AAV vectors encoding Cas9 and sgRNAs to target a single gene or multiple genes in the normal adult mouse brain *in vivo* [89]. Characterizing the effects of gene modifications in postmitotic neurons revealed similar phenotypes as observed in gene knockout mice. Another work used a lentiviral system that delivers both the CRISPR system and Cre recombination to examine CRISPR-induced mutation of genes in the context of well-studied conditional Cre/*loxP* mouse models of lung cancer and other cancer types [90]. In other research to study cancer genes in the mouse liver, a hydrodynamic injection was used to deliver a plasmid DNA expressing Cas9 and sgRNAs that directly targeted the tumor suppressor genes (*p53* or *PTEN*) alone and in combination into the liver. The authors demonstrated the feasibility of Cas9-mediated mutation of tumor suppressor genes in the liver as an avenue for the rapid development of liver cancer models [91]. However, similar to the Cas9 knock-in mouse, the virally

delivered Cas9 may only edit genes in a fraction of cells, and the approach may be most effective for studying the effects of loss-of-function mutations on cell autonomous properties.

Genome Imaging Using CRISPR/Cas9

Imaging offers a direct approach for studying the spatial and temporal behavior of the genome in living cells [92]. The ability of Cas9 to target specific sequences in the genome makes it a promising imaging tool for directly observing genomic organization and dynamics in cells. The first proof-of-concept work fused the *S. pyogenes* dCas9 to EGFP and used the fusion protein to visualize the dynamics of coding or noncoding sequences in living human cell lines [93]. The authors tracked the dynamics of telomeres, and the repetitive and nonrepetitive sequences of coding genes (*MUC4* and *MUC1*) in a short time frame (~minutes) and throughout the whole cell cycle. In addition, dCas9 fused to EGFP has been used to label endogenous centromeres and telomeres loci in live mouse embryonic stem cells [94]. The development of the SunTag system, a repeating peptide array that can recruit multiple copies of an antibody-fusion protein, enhanced the sensitivity to amplify the dCas9 fluorescent signal in the genome [53]. Using dCas9 orthologs tagged with different fluorescent proteins, it was shown that the dynamics of multiple repetitive genomic loci could be tracked in living cells [95]. A method termed 'Cas9-mediated fluorescence *in situ* hybridization' (CASFISH) further combined dCas9 with fluorescence *in situ* hybridization (FISH) [96]. Due to the specific DNA targeting and unwinding activity of dCas9, CASFISH is a fast and convenient process for labeling DNA elements while avoiding treatment of heat and disruptive chemicals that distort the natural organization of the nucleus, which is normally seen in FISH. Thus, the approach preserves the spatial relations of the genetic elements that are important for studying gene expression.

Recent work also established a CRISPR approach to facilitate super-resolution imaging in living mammalian cells [97]. Current live cell super-resolution imaging normally relies on the overexpression of a host protein fused to a fluorescent protein, which results in artifacts that may obscure the interpretation of imaging results. Using CRISPR/Cas9 to fluorescently tag the endogenous genes that are expressed from their native genomic loci could allow genes to be expressed at close to endogenous levels, thus avoiding artifacts. Based on this idea, a method termed 'reversible saturable optical fluorescence transitions' (RESOLFT) was developed, wherein heterozygous and homozygous Cas9-edited human knock-in cell lines were generated that expressed the reversibly switchable fluorescent protein rsEGFP2 from their respective native genomes, which prevented the appearance of typical overexpression-induced artifacts in these cells.

To enhance signals for endogenous proteins imaging, one study adapted self-complementing split fluorescent proteins, GFP11 and sfCherry11, derived from the sfGFP and sfCherry [98]. The small sizes of these split fluorescent domains (16–18 amino acids) enable them to be easily inserted into endogenous genomic loci via CRISPR gene editing. Tandem arrays of these domains further amplify fluorescence signals in imaging, such as for tracking intraflagellar transport particles.

In addition to DNA imaging, *S. pyogenes* dCas9 can also allow for endogenous RNA imaging in living cells [99]. In the presence of sgRNAs targeting mRNA and a stabilized PAMmer oligonucleotide that contains the PAM domain for dCas9 binding, specifically targeted RNA can be visualized. Indeed, it was observed that nuclear localized dCas9 could be exported to the cytoplasm. Furthermore, dCas9 allowed for tracking of RNA during induced RNA/protein accumulation in the presence of oxidative stress.

Lineage Tracing Using CRISPR/Cas9

Gene editing has been used as tools for cell lineage tracing. One recent study demonstrated a lineage-tracing method termed 'genome editing of synthetic target arrays for lineage tracing'

(GESTALT) [100]. This method uses CRISPR/Cas9 gene editing to generate a combinatorial diversity of mutations that accumulate over cell divisions within a series of DNA barcodes. Via deep sequencing, lineage relations between many cells can be inferred using patterns of the edited barcodes. The approach was developed in both cell culture and zebrafish, by editing synthetic arrays of approximately a dozen CRISPR/Cas9 target sites. The approach generated thousands of unique edited barcodes in cell lines, which could then be sequenced from either DNA or RNA. By injecting fertilized eggs with editing reagents that targeted a genomic barcode with ten target sites, the authors observed the accumulation of hundreds to thousands of uniquely edited barcodes per animal, and further inferred the lineage relations between ancestral progenitors and organs based on mutation patterns. This proof-of-principle study showed that combinatorial and cumulative genome editing is a powerful approach to record lineage information in multicellular systems.

In another study, the type I-E CRISPR/Cas system of *E. coli* was harnessed to generate records of specific DNA sequences in bacterial genomes [101]. Unlike gene editing, the work was based on the native adaptive immunity acquisition ability of CRISPR, because new spacer sequences can be acquired and integrated stably into the CRISPR crRNA array. Using this feature, it was demonstrated that the Cas1–Cas2 complex enables the recording of defined sequences over many days and in multiple modalities. The work elucidated fundamental aspects of the CRISPR acquisition process. The recording system developed could be useful for applications that require long histories of *in vivo* cellular activity to be traced.

While optimization of these methods is required for more robust performance, genome editing and the unique features (i.e., adaptation) of the CRISPR system provide promising approaches to record biological information and history in living cells and tissues. One can envision that these tools may enable mapping of the complete cell lineage in multicellular organisms as well as linking cell lineage information to molecular profiles (e.g., transcription, epigenetics, and proteomics), such as those in single cells.

Concluding Remarks

The CRISPR/Cas9 technology has revolutionized cell biology research. The system is versatile, enabling diverse types of genome engineering approach. While most of the work has used Cas9-mediated knockout or dCas9-mediated repression and activation to study gene function, we expect expansion of these tools to study the epigenome and 3D chromosomal organization in greater detail in the future. Furthermore, studies have used CRISPR to model complex genomic rearrangements *in vitro* and *in vivo*, which resulted in breakthroughs in studying chromosomal translocations [102,103]. Most research has been performed in cell lines, and future work related to the interrogation of cellular functions should be carried out in primary cells derived from animals or humans or *in vivo* using relevant animal models.

CRISPR/Cas9 is emerging as a major genome-manipulation tool for research and therapeutics, yet there are challenges remaining to improve its specificity, efficiency, and utility (see Outstanding Questions). One major concern is the off-target effects, since Cas9 can tolerate mismatches between sgRNA and target DNA [104–106]. Methods have been developed to profile the off-target effects, such as GUIDE-seq [107]. To improve specificity, several strategies have been developed, including using paired nickase variants of Cas9 [32,42], paired dCas9-FokI nucleases [108,109], truncated sgRNAs (17–18 base pairs) that are more sensitive to mismatches [110], and controlling acting concentration of the Cas9/sgRNA complex [111]. Using structure-guided protein-engineering approaches, two studies recently created *S. pyogenes* Cas9 variants with improved specificity [112,113]. For example, a high-fidelity variant of Cas9 harboring designed alterations showed reduced nonspecific DNA contacts, while retaining robust

Outstanding Questions

How can the off-target effects of CRISPR/Cas9 be avoided in mammalian cells and whole organisms?

Can CRISPR/Cas9 technology be developed to insert a large gene fragment into the mammalian genome for gene knock-in studies with similar efficiency to that of gene knockout studies?

Will CRISPR/Cas9 technology be able to efficiently modulate different types of epigenetic modification? Can it control the fate of synthetic epigenetic marks, and whether they can be stably inherited when cells proliferate?

Can CRISPR/Cas9-mediated genetic screens be performed on nonproliferation-based phenotypes such as differentiation?

Can CRISPR/Cas9 technology enable more robust transgenic animal generation by deleting, mutating, and inserting any gene of interest?

Beyond gene editing, how can CRISPR/Cas9 be used to help advance cell biology research?

on-target activities comparable with wild-type Cas9 [113]. Combinations of these methods could provide a route to its ultimate use for gene therapy.

As a powerful, yet versatile, gene-editing and regulation tool, CRISPR/Cas9 technology is already accelerating both research and therapeutics. We believe that its broad applications in genomics research and cell biology research will greatly advance our knowledge of both basic biology and diseases in the years to come.

Acknowledgments

F.W. acknowledges support from the Center of Synthetic Biology and Shanghai Institute of Rheumatology in Renji Hospital affiliated to Shanghai Jiaotong University School of Medicine, National Natural Science Foundation of China (No. 81502233), the Science and Technology Commission of Shanghai Municipality (No. 134119a8100) and the Young Scientific Research Project of Shanghai Municipal Health Bureau (No. 20134Y176). L.S.Q. acknowledges support from the NIH Office of the Director (OD), National Institute of Dental & Craniofacial Research (NIDCR), and NIH Director's Early Independence Award DP5 OD017887.

References

1. Capecchi, M.R. (1989) Altering the genome by homologous recombination. *Science* 244, 1288–1292
2. Rudin, N. *et al.* (1989) Genetic and physical analysis of double-strand break repair and recombination in *Saccharomyces cerevisiae*. *Genetics* 122, 519–534
3. Bibikova, M. *et al.* (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell Biol.* 21, 289–297
4. Bibikova, M. *et al.* (2002) Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* 161, 1169–1175
5. Urnov, F.D. *et al.* (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 435, 646–651
6. Christian, M. *et al.* (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757–761
7. Gaj, T. *et al.* (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405
8. Wolfe, S.A. *et al.* (2000) DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* 29, 183–212
9. Beerli, R.R. and Barbas, C.F., 3rd (2002) Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol.* 20, 135–141
10. Konermann, S. *et al.* (2013) Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 500, 472–476
11. Zhang, F. *et al.* (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.* 29, 149–153
12. Maeder, M.L. *et al.* (2013) CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* 10, 977–979
13. Wiedenheft, B. *et al.* (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331–338
14. Marraffini, L.A. and Sontheimer, E.J. (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* 11, 181–190
15. Nakata, A. *et al.* (1982) Cloning of alkaline phosphatase isozyme gene (iap) of *Escherichia coli*. *Gene* 19, 313–319
16. Barrangou, R. *et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712
17. Brouns, S.J. *et al.* (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960–964
18. Jinek, M. *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821
19. Deltcheva, E. *et al.* (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602–607
20. Sapranauskas, R. *et al.* (2011) The Streptococcus thermophilus CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic. Acids Res.* 39, 9275–9282
21. Marraffini, L.A. and Sontheimer, E.J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 463, 568–571
22. Cong, L. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823
23. Mali, P. *et al.* (2013) RNA-guided human genome engineering via Cas9. *Science* 339, 823–826
24. Jinek, M. *et al.* (2013) RNA-programmed genome editing in human cells. *Elife* 2, e00471
25. Hwang, W.Y. *et al.* (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 227–229
26. Cho, S.W. *et al.* (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31, 230–232
27. Makarova, K.S. *et al.* (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736
28. Chylinski, K. *et al.* (2014) Classification and evolution of type II CRISPR-Cas systems. *Nucleic. Acids Res.* 42, 6091–6105
29. Gasiunas, G. *et al.* (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2579–E2586
30. Ran, F.A. *et al.* (2015) In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 520, 186–191
31. Zhang, Y. *et al.* (2013) Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Mol. Cell* 50, 488–503
32. Mali, P. *et al.* (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* 31, 833–838
33. Sampson, T.R. *et al.* (2013) A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* 497, 254–257
34. Hirano, H. *et al.* (2016) Structure and engineering of *Francisella novicida* Cas9. *Cell* 164, 950–961
35. Hale, C.R. *et al.* (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139, 945–956
36. Hale, C.R. *et al.* (2012) Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol. Cell* 45, 292–302
37. Zetsche, B. *et al.* (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771
38. Yamano, T. *et al.* (2016) Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* 165, 949–962
39. Dong, D. *et al.* (2016) The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* 532, 522–526

40. Fonfara, I. *et al.* (2016) The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* 532, 517–521
41. Abudayyeh, O.O. *et al.* (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353, aaf5573
42. Ran, F.A. *et al.* (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154, 1380–1389
43. Komor, A.C. *et al.* (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424
44. Qi, L.S. *et al.* (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183
45. Peters, J.M. *et al.* (2016) A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell* 165, 1493–1506
46. Gilbert, L.A. *et al.* (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451
47. Gilbert, L.A. *et al.* (2014) Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 159, 647–661
48. Konermann, S. *et al.* (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517, 583–588
49. Larson, M.H. *et al.* (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* 8, 2180–2196
50. Zhao, Y. *et al.* (2014) Sequence-specific inhibition of microRNA via CRISPR/CRISPRi system. *Sci. Rep.* 4, 3943
51. Perez-Pinera, P. *et al.* (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* 10, 973–976
52. Cheng, A.W. *et al.* (2013) Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* 23, 1163–1171
53. Tanenbaum, M.E. *et al.* (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159, 635–646
54. Chavez, A. *et al.* (2015) Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* 12, 326–328
55. Chavez, A. *et al.* (2016) Comparison of Cas9 activators in multiple species. *Nat. Methods* 13, 563–567
56. Zalatan, J.G. *et al.* (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* 160, 339–350
57. Hilton, I.B. *et al.* (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* 33, 510–517
58. Kearns, N.A. *et al.* (2015) Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat. Methods* 12, 401–403
59. Thakore, P.I. *et al.* (2015) Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* 12, 1143–1149
60. Vojta, A. *et al.* (2016) Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic. Acids Res.* 44, 5615–5628
61. Kearns, N.A. *et al.* (2014) Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development* 141, 219–223
62. Shi, J. *et al.* (2015) Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* 33, 661–667
63. Korkmaz, G. *et al.* (2016) Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.* 34, 192–198
64. Shalem, O. *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84–87
65. Bernstein, B.E. *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315–326
66. Zhou, Y. *et al.* (2014) High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 509, 487–491
67. Chen, S. *et al.* (2015) Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* 160, 1246–1260
68. Parnas, O. *et al.* (2015) A genome-wide CRISPR screen in primary immune cells to dissect regulatory networks. *Cell* 162, 675–686
69. Wong, A.S. *et al.* (2016) Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proc. Natl. Acad. Sci. U.S.A.* 113, 2544–2549
70. Evers, B. *et al.* (2016) CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat. Biotechnol.* 34, 631–633
71. Morgens, D.W. *et al.* (2016) Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat. Biotechnol.* 34, 634–636
72. Lane, A.B. *et al.* (2015) Enzymatically generated CRISPR libraries for genome labeling and screening. *Dev Cell* 34, 373–378
73. Friedland, A.E. *et al.* (2013) Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat. Methods* 10, 741–743
74. Bassett, A.R. *et al.* (2013) Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell Rep.* 4, 220–228
75. Jao, L.E. *et al.* (2013) Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13904–13909
76. Chang, N. *et al.* (2013) Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res.* 23, 465–472
77. Li, W. *et al.* (2013) Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. *Nat. Biotechnol.* 31, 684–686
78. Lv, Q. *et al.* (2016) Efficient generation of myostatin gene mutated rabbit by CRISPR/Cas9. *Sci. Rep.* 6, 25029
79. Yan, Q. *et al.* (2014) Generation of multi-gene knockout rabbits using the Cas9/gRNA system. *Cell Regen. (Lond).* 3, 12
80. Wang, X. *et al.* (2015) Generation of gene-modified goats targeting MSTN and FGF5 via zygote injection of CRISPR/Cas9 system. *Sci. Rep.* 5, 13878
81. Crispo, M. *et al.* (2015) Efficient generation of myostatin knockout sheep using CRISPR/Cas9 technology and microinjection into zygotes. *PLoS ONE* 10, e0136690
82. Zou, Q. *et al.* (2015) Generation of gene-target dogs using CRISPR/Cas9 system. *J. Mol. Cell Biol.* 7, 580–583
83. Wang, K. *et al.* (2015) Efficient generation of myostatin mutations in pigs using the CRISPR/Cas9 System. *Sci. Rep.* 5, 16623
84. Niu, Y. *et al.* (2014) Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* 156, 836–843
85. Wang, H. *et al.* (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153, 910–918
86. Yang, H. *et al.* (2013) One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* 154, 1370–1379
87. Platt, R.J. *et al.* (2014) CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* 159, 440–455
88. Chiou, S.H. *et al.* (2015) Pancreatic cancer modeling using retrograde viral vector delivery and in vivo CRISPR/Cas9-mediated somatic genome editing. *Genes. Dev.* 29, 1576–1585
89. Swiech, L. *et al.* (2015) In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat. Biotechnol.* 33, 102–106
90. Sanchez-Rivera, F.J. *et al.* (2014) Rapid modelling of cooperating genetic events in cancer through somatic genome editing. *Nature* 516, 428–431
91. Xue, W. *et al.* (2014) CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature* 514, 380–384

92. Lanctot, C. *et al.* (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.* 8, 104–115

93. Chen, B. *et al.* (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491

94. Anton, T. *et al.* (2014) Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus* 5, 163–172

95. Ma, H. *et al.* (2015) Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3002–3007

96. Deng, W. *et al.* (2015) CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11870–11875

97. Ratz, M. *et al.* (2015) CRISPR/Cas9-mediated endogenous protein tagging for RESOLFT super–resolution microscopy of living human cells. *Sci. Rep.* 5, 9592

98. Kamiyama, D. *et al.* (2016) Versatile protein tagging in cells with split fluorescent protein. *Nat. Commun.* 7, 11046

99. Nelles, D.A. *et al.* (2016) Programmable RNA tracking in live cells with CRISPR/Cas9. *Cell* 165, 488–496

100. McKenna, A. *et al.* (2016) Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* 353, aaf7907

101. Shipman, S.L. *et al.* (2016) Molecular recordings by directed CRISPR spacer acquisition. *Science* 353, aaf1175

102. Maddalo, D. *et al.* (2014) In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* 516, 423–427

103. Choi, P.S. and Meyerson, M. (2014) Targeted genomic rearrangements using CRISPR/Cas technology. *Nat. Commun.* 5, 3728

104. Hsu, P.D. *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832

105. Fu, Y. *et al.* (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31, 822–826

106. Crosetto, N. *et al.* (2013) Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* 10, 361–365

107. Tsai, S.Q. *et al.* (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33, 187–197

108. Guilinger, J.P. *et al.* (2014) Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 32, 577–582

109. Tsai, S.Q. *et al.* (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* 32, 569–576

110. Fu, Y. *et al.* (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* 32, 279–284

111. Davis, K.M. *et al.* (2015) Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* 11, 316–318

112. Slaymaker, I.M. *et al.* (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science* 351, 84–88

113. Kleinstiver, B.P. *et al.* (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490–495

114. Plevock, K.M. *et al.* (2015) Newly characterized region of CP190 associates with microtubules and mediates proper spindle morphology in *Drosophila* stem cells. *PLoS ONE* 10, e0144174

115. Bhattacharjee, A. *et al.* (2016) Activity of Menkes Disease protein ATP7A is essential for redox balance in mitochondria. *J. Biol. Chem.* 291, 16644–16658

116. Popow, J. *et al.* (2015) FASTKD2 is an RNA-binding protein required for mitochondrial RNA processing and translation. *RNA* 21, 1873–1884

117. Claussnitzer, M. *et al.* (2015) FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* 373, 895–907

118. Birsoy, K. *et al.* (2015) An essential role of the mitochondrial electron transport chain in cell proliferation is to enable aspartate synthesis. *Cell* 162, 540–551

119. D'Osualdo, A. *et al.* (2015) Transcription factor ATF4 induces NLRP1 inflammasome expression during endoplasmic reticulum stress. *PLoS ONE* 10, e0130635

120. Plumb, R. *et al.* (2015) A functional link between the co-translational protein translocation pathway and the UPR. *Elife* 4, e07426

121. Schoborg, T. *et al.* (2015) An Asp-CaM complex is required for centrosome-pole cohesion and centrosome inheritance in neural stem cells. *J. Cell Biol.* 211, 987–998

122. Li, J. *et al.* (2015) Glycosylation inhibition reduces cholesterol accumulation in NPC1 protein-deficient cells. *Proc. Natl. Acad. Sci. U.S.A.* 112, 14876–14881

123. Fuchs, G. *et al.* (2015) Kinetic pathway of 40S ribosomal subunit recruitment to hepatitis C virus internal ribosome entry site. *Proc. Natl. Acad. Sci. U.S.A.* 112, 319–325

124. Virreira Winter, S. *et al.* (2016) Genome-wide CRISPR screen reveals novel host factors required for Staphylococcus aureus alpha-hemolysin-mediated toxicity. *Sci. Rep.* 6, 24242

Cell Stem Cell
Short Article

A Single CRISPR-Cas9 Deletion Strategy that Targets the Majority of DMD Patients Restores Dystrophin Function in hiPSC-Derived Muscle Cells

Courtney S. Young,[1,2,3,4] Michael R. Hicks,[2,3,5] Natalia V. Ermolova,[2,4] Haruko Nakano,[3,7] Majib Jan,[2,5,6] Shahab Younesi,[2,5,6] Saravanan Karumbayaram,[3,5] Chino Kumagai-Cresse,[2,4] Derek Wang,[2,5] Jerome A. Zack,[1,3,5] Donald B. Kohn,[1,2,3,5] Atsushi Nakano,[1,2,3,7] Stanley F. Nelson,[1,2,3,8] M. Carrie Miceli,[1,2,3,5] Melissa J. Spencer,[1,2,3,4,*] and April D. Pyle[1,2,3,5,*]

[1]Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA
[2]Center for Duchenne Muscular Dystrophy, University of California, Los Angeles, CA 90095, USA
[3]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, CA 90095, USA
[4]Department of Neurology, University of California, Los Angeles, CA 90095, USA
[5]Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, CA 90095, USA
[6]CIRM Bridges Program, California State University, Northridge, CA 91330, USA
[7]Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, CA 90095, USA
[8]Department of Human Genetics, University of California, Los Angeles, CA 90095, USA
*Correspondence: mspencer@mednet.ucla.edu (M.J.S.), apyle@mednet.ucla.edu (A.D.P.)
http://dx.doi.org/10.1016/j.stem.2016.01.021

SUMMARY

Mutations in *DMD* disrupt the reading frame, prevent dystrophin translation, and cause Duchenne muscular dystrophy (DMD). Here we describe a CRISPR/Cas9 platform applicable to 60% of DMD patient mutations. We applied the platform to DMD-derived hiPSCs where successful deletion and non-homologous end joining of up to 725 kb reframed the *DMD* gene. This is the largest CRISPR/Cas9-mediated deletion shown to date in *DMD*. Use of hiPSCs allowed evaluation of dystrophin in disease-relevant cell types. Cardiomyocytes and skeletal muscle myotubes derived from reframed hiPSC clonal lines had restored dystrophin protein. The internally deleted dystrophin was functional as demonstrated by improved membrane integrity and restoration of the dystrophin glycoprotein complex in vitro and in vivo. Furthermore, miR31 was reduced upon reframing, similar to observations in Becker muscular dystrophy. This work demonstrates the feasibility of using a single CRISPR pair to correct the reading frame for the majority of DMD patients.

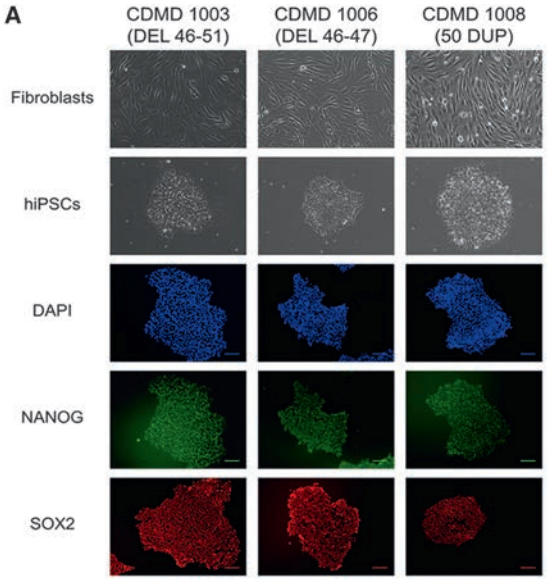
INTRODUCTION

Duchenne muscular dystrophy (DMD) is the most common fatal genetic disease of childhood, affecting ~1 in 3,500–5,000 boys. In DMD, progressive muscle degeneration generally leads to death in the twenties, and there are currently no highly effective therapies. DMD is often caused by frameshifting exonic deletions in *DMD*, which encodes dystrophin. Dystrophin stabilizes the dystrophin glycoprotein complex (DGC) at the sarcolemma; loss of functional dystrophin leads to the degradation of DGC components, which results in muscle membrane fragility and leakage of creatine kinase (CK) (Pearce et al., 1964). Approximately 60% of mutations causing DMD occur between *DMD* exons 45–55 (Béroud et al., 2007). Multiple independent clinical reports in patients and dystrophic mice have revealed that in-frame deletions of exons 45–55 produce an internally deleted dystrophin protein and are associated with a very mild Becker muscular dystrophy (BMD) disease course, with some patients still asymptomatic in their sixties (Béroud et al., 2007; Echigoya et al., 2015; Nakamura et al., 2008; Taglia et al., 2015). Thus, genetic manipulation to create a large deletion of exons 45–55 is a therapeutic strategy to restore the reading frame for 60% of DMD patients with mutations in this region.

One promising approach to induce genetic correction of *DMD* is through the use of the bacterially acquired immune surveillance system known as clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated nuclease (Cas) 9. In this system a short guide RNA (gRNA), which is complimentary to a specific site in the genome, is used to target the Cas9 nuclease and induce double-stranded breaks (DSBs). The DSBs can be repaired through non-homologous end joining (NHEJ) or homology-directed repair.

Previous work has shown that CRISPR/Cas9 components can modify the *DMD* gene (Li et al., 2015; Long et al., 2014, 2016; Nelson et al., 2016; Ousterout et al., 2015; Tabebordbar et al., 2016; Wojtal et al., 2016; Xu et al., 2015). In this investigation, we describe a therapeutically relevant CRISPR/Cas9 platform that we designed to modify *DMD*. Our platform involves excision of exons 45–55 and NHEJ to reframe dystrophin through creation of an internally deleted protein that is stable and functional. The internally deleted protein mimics the naturally occurring exon 45–55 deletion observed in mild BMD patients and encompasses 60% of DMD patient mutations.

For the first time, we demonstrate CRISPR/Cas9-mediated deletion and NHEJ of up to 725 kb of the *DMD* gene in human induced pluripotent stem cell (hiPSC) lines. We show that CRISPR/Cas9 reframed, hiPSC-derived skeletal and cardiac



D

CELL IDENTIFIER	MUTATION	FOR NHEJ # EXONS	FOR NHEJ DISTANCE
CDMD 1003	DEL 46-51	5 EXONS (45 + 52-55)	530 kb
CDMD 1006	DEL 46-47	9 EXONS (45 + 48-55)	670 kb
CDMD 1008	50 DUP	12 EXONS (45-55)	725 kb

Figure 1. CDMD hiPSCs Are Pluripotent and Genetically Stable

(A) CDMD hiPSCs were generated from DMD fibroblasts. Brightfield images depict fibroblasts before and after reprogramming to hiPSCs. Immunocytochemical staining reveals that cells express pluripotency markers NANOG (green) and SOX2 (red). Scale bar, 100 μm.

(B) Karyotyping of all lines is shown.

(C) CDMD hiPSCs were injected into mice to test teratoma formation in vivo. Representative H&E stainings of the three germ layers (endoderm, mesoderm, and ectoderm) are shown.

(D) Patient mutations for each CDMD hiPSC line are shown. In addition, the number of exons and the approximate distance necessary for successful NHEJ is indicated, based on comparative genomic hybridization data for the patient's underlying mutation size.

muscle cells express stable dystrophin that improves membrane stability and restores a DGC member, β-dystroglycan. We also demonstrate reduced microRNA 31 (miR31) levels after the reading frame is restored, consistent with the observations made in BMD patients (Cacchiarelli et al., 2011). Furthermore, we show restoration of dystrophin and β-dystroglycan in vivo after engraftment of reframed hiPSC-derived skeletal muscle cells into a mouse model of DMD. This work sets the stage for use of reframed DMD hiPSC-derived cells or in vivo correction strategies using CRISPR/Cas9 for direct translation to patients with DMD.

RESULTS

DMD hiPSC Lines Are Pluripotent and Genetically Stable

We have developed several xenobiotic-free hiPSC lines derived from wild-type and DMD patient fibroblasts using current good manufacturing practice protocols. Each DMD hiPSC line harbors a unique frameshifting *DMD* mutation within the exon 45–55 hotspot region. All hiPSC lines (Center for Duchenne Muscular Dystrophy [CDMD] 1003, 1006, and 1008) express pluripotency markers (NANOG and SOX2) and are karyotypically normal (Figures 1A and 1B). CDMD hiPSCs maintain pluripotency, as they form teratomas in vivo that represent all three germ layers (Figure 1C), and each harbor unique mutations (Figure 1D).

CRISPR/Cas9-Mediated Deletion and NHEJ of up to 725 kb in the *DMD* Gene

In order to delete exons 45–55 of *DMD*, gRNAs were designed to target introns 44 and 55. gRNA sites were chosen to only retain ~500 bp of the intron next to each of the flanking exons (44 and 56). The rationale for this design is to develop gRNAs applicable to as many patient mutations as possible and to ensure that a small functional chimeric intron is generated. During NHEJ, the 3′ end of intron 44 and the 5′ end of intron 55 join to create a ~1 kb chimeric intron (Figure 2A). We expect that introns generated in this manner are functional and splice correctly to create an in-frame transcript, with exon 44 joined with exon 56.

Since hiPSCs are challenging to genetically manipulate, human embryonic kidney (HEK) 293FT cells were used to screen five gRNAs at each intronic region. All gRNAs demonstrated individual cutting activity on Surveyor assay up to 34% (Figures S1A and S1B). Using multiplex PCR, gRNAs transfected in pairs were shown to effectively delete the entire 708 kb region encompassing exons 45–55 (Figures S1C and S1D).

In order to assess the feasibility of an exon 45–55 deletion across different patient mutations, we applied our gRNAs to three DMD hiPSC lines. The lines (CDMD 1003, 1006, and 1008) require ~530 kb, 670 kb, or 725 kb, respectively, for successful deletion and NHEJ of *DMD*. The gRNAs used were shown to be active in all three lines and effectively deleted exons 45–55 (Figures S2 and S3). Transient puromycin selection of cells nucleofected with the CRISPR plasmids improved the efficiency of deletion in CDMD 1003 and 1006 hiPSCs (Figure S3D).

Clonal Reframed DMD hiPSC Lines Contain No Off-Target Activity at Candidate Sites

Stably deleted DMD hiPSC lines were generated from CDMD 1003 and 1006 by clonal selection after nucleofection with the gRNA pair 44C4 and 55C3 (Figures 2B and 2C) and are pluripotent (Figures 2C and S4B). All reframed lines were karyotypically normal except for one clone (CDMD 1003-81), which was found to contain a 1q32 amplification confirmed via FISH analysis

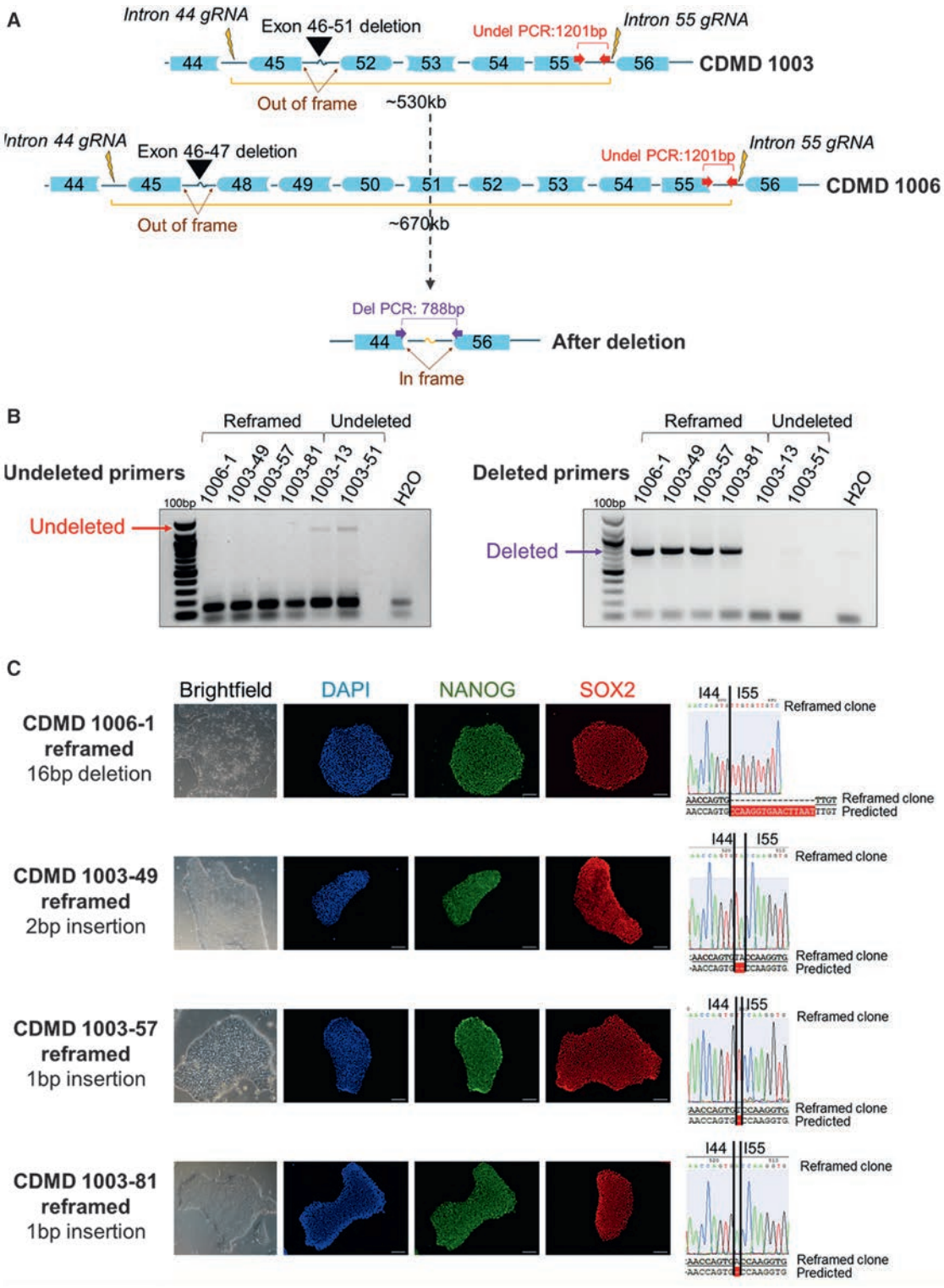


Figure 2. Generation of Stable, Pluripotent CDMD hiPSC Lines with an Exon 45–55 Deletion

(A) Shown is a cartoon (not to scale) of the region of *DMD* targeted for CRISPR/Cas9-mediated deletion using gRNAs specific to introns 44 and 55 (lightning bolts). Successful NHEJ deletes exons 45–55 and restores the reading frame for mutations within this region. Different deletion sizes are required depending on the patient's underlying mutation (black arrow heads).

(B) PCR genotyping of 117 and 109 single-cell clones from parental lines CDMD 1006 and 1003, respectively, was carried out on cells nucleofected with gRNAs 44C4 and 55C3. One clone from CDMD 1006 (CDMD 1006-1) and three from CDMD 1003 (CDMD 1003-49, 1003-57, and 1003-81) were identified as stably

(legend continued on next page)

(Figure S4A), also observed in the original parental line and in all daughter clones after post hoc analysis. The 1q32 amplification is common in hPSCs after extended propagation in culture (Dekel-Naftali et al., 2012), and thus was not a result of CRISPR-mediated off-target activity. To determine off-target activity of our gRNAs, the top ten homologous sites per guide were determined by COSMID (Cradick et al., 2014) and sequenced in all clonal and parental lines. No off-target mutations were observed at any site (Table S2). All variants, besides a heterozygous SNP in chromosome 11, were detected in less than 1% of reads, which is consistent with error in the sequencing method.

Dystrophin (DYS$^{\Delta 45–55}$) Expression Is Restored in Reframed DMD hiPSC-Derived Cardiomyocytes and Skeletal Myotubes

CRISPR/Cas9-mediated deletion of *DMD* should result in an internally deleted dystrophin protein lacking exons 45−55 (hereafter referred to as DYS$^{\Delta 45–55}$). As hiPSCs do not express dystrophin, we differentiated the reframed DMD hiPSC clonal lines to two disease-relevant cell types, cardiomyocytes and skeletal muscle myotubes, using directed differentiation or overexpression (OE) of MyoD to evaluate rescue of DYS$^{\Delta 45–55}$. PCR and sequencing of the exon 44/56 boundary in cDNA from the reframed cardiomyocyte clones demonstrated correct splicing of the dystrophin transcript (Figures S4C and S4D). Additionally, both the reframed cardiac and skeletal muscle cell lines restored dystrophin expression as assayed by immunocytochemistry and western blot (Figures 3A–3C). Compared to wild-type CDMD 1002 or human skeletal muscle myotubes (HSMM), the band was truncated by ~66 kDa as expected.

DYS$^{\Delta 45–55}$ Protein Restores Membrane Functionality to Cardiomyocytes and Skeletal Myotubes In Vitro

Cardiomyocytes or skeletal myotubes lacking dystrophin demonstrate membrane fragility in vitro and respond to osmotic stress by releasing elevated levels of CK (Guan et al., 2014; Menke and Jockusch, 1995), as is seen in human patients (Pearce et al., 1964). To determine whether DYS$^{\Delta 45–55}$ could restore stability to dystrophic plasma membranes, we subjected differentiated cardiomyocytes and skeletal muscle myotubes derived from reframed and out-of-frame hiPSCs to hypo-osmotic conditions. Cells were stressed by incubation in hypo-osmolar solutions (66–240 mosmol) and CK release into the supernatant was measured to show functional improvement after dystrophin restoration. Both the reframed CDMD 1003-49 cardiomyocytes and skeletal muscle cells demonstrated reduced CK release, similar to wild-type (CDMD 1002), versus the out-of-frame CDMD 1003 cells, indicating that DYS$^{\Delta 45–55}$ was capable of reducing membrane fragility (Figure 4A). The same trend was also observed with CDMD 1006/1006-1 cardiomyocytes (Figure S4E). After normalizing and pooling all experiments, we observed that significantly less CK was released at 93, 135, and 240 mosmol in the reframed and wild-type cells compared to out-of-frame (Figure S4F).

CRISPR/Cas9 Reframing Correlates with miR31 Levels in Skeletal Myotubes In Vitro

Elevated levels of miR31 have been observed in DMD patient biopsies compared to wild-type or BMD (Cacchiarelli et al., 2011). We measured levels of miR31 using droplet digital PCR (ddPCR) after differentiation of out-of-frame and reframed CDMD hiPSCs to skeletal myotubes. Reframing *DMD* reduced levels of miR31 (similar to wild-type cells) compared to out-of-frame *DMD,* as is observed in human dystrophinopathies (Figure 4B). Thus, reframing the *DMD* gene normalizes miR31 levels similar to BMD, demonstrating functional rescue of the dystrophic phenotype to a BMD phenotype.

DYS$^{\Delta 45–55}$ Protein Restores the DGC In Vitro and In Vivo

As a third assay of DYS$^{\Delta 45–55}$ functionality, we evaluated its ability to restore the DGC in vitro and in vivo. The DGC member β-dystroglycan was restored and detected at the membrane of reframed hiPSCs, but not out-of-frame hiPSCs, after directed differentiation to skeletal muscle in vitro by immunostaining and western blot (Figures 4C and 4D). Additionally, skeletal muscle cells derived from a wild-type (CDMD 1002), out-of-frame (CDMD 1003), or reframed (CDMD 1003-49) hiPSC line were injected into the tibialis anterior (TA) of NOD *scid* IL2Rgamma (NSG)-mdx mice. Correctly localized dystrophin and β-dystroglycan was only observed in engrafted human cells (demarked by human lamin A/C and spectrin) from the reframed or wild-type lines (Figures 4E and 4F). These studies taken together with the hypo-osmotic stress assays demonstrate the ability of DYS$^{\Delta 45–55}$ to functionally reassemble the DGC and restore membrane stability in vitro and in vivo.

DISCUSSION

Using CRISPR/Cas9 gene editing, we have induced the largest deletion accomplished to date in DMD hiPSCs and restored a functional dystrophin protein. Deletion of *DMD* exons 45–55 has the potential to be therapeutically relevant to 60% of DMD patients. Since this internal deletion has been associated with a very mild disease course in multiple independent patients, a therapy utilizing this approach should create a highly functional dystrophin. We showed successful deletion of exons 45–55 using a single gRNA pair and did not identify any off-target activity at the top ten homologous sites; however, a more comprehensive and unbiased approach should be undertaken such as whole-genome sequencing. Importantly, removal of exons 45–55 resulted in stable dystrophin protein (DYS$^{\Delta 45–55}$) in both cardiomyocytes and skeletal myotubes in vitro. Functionality of

deleted. Deletion PCR genotyping results for six hiPSC clonal lines is shown. One pair of primers (red arrows in A) was located internal to the deletion and only produced a 1,201 bp band in the undeleted clones CDMD 1003-13 and 1003-51. Another primer set (purple arrows in A) flanked the deletion region and produced a 788 bp band only when the deletion and NHEJ occurred successfully, as in the reframed clones CDMD 1006-1, 1003-49, 1003-57, and 1003-81.
(C) Each clonal line maintained normal morphology (brightfield) and expressed NANOG (green) and SOX2 (red) by immunocytochemistry. Scale bar, 100 μm. Shown to the right is the sequence of the gDNA at the rejoining site between introns 44 (I44) and 55 (I55). Sequencing revealed a 16 bp deletion in CDMD 1006-1, a 2 bp insertion in CDMD 1003-49, and 1 bp insertions in CDMD 1003-57 and CDMD 1003-81.
See also Figures S1, S2, S3, S4A, and S4B.

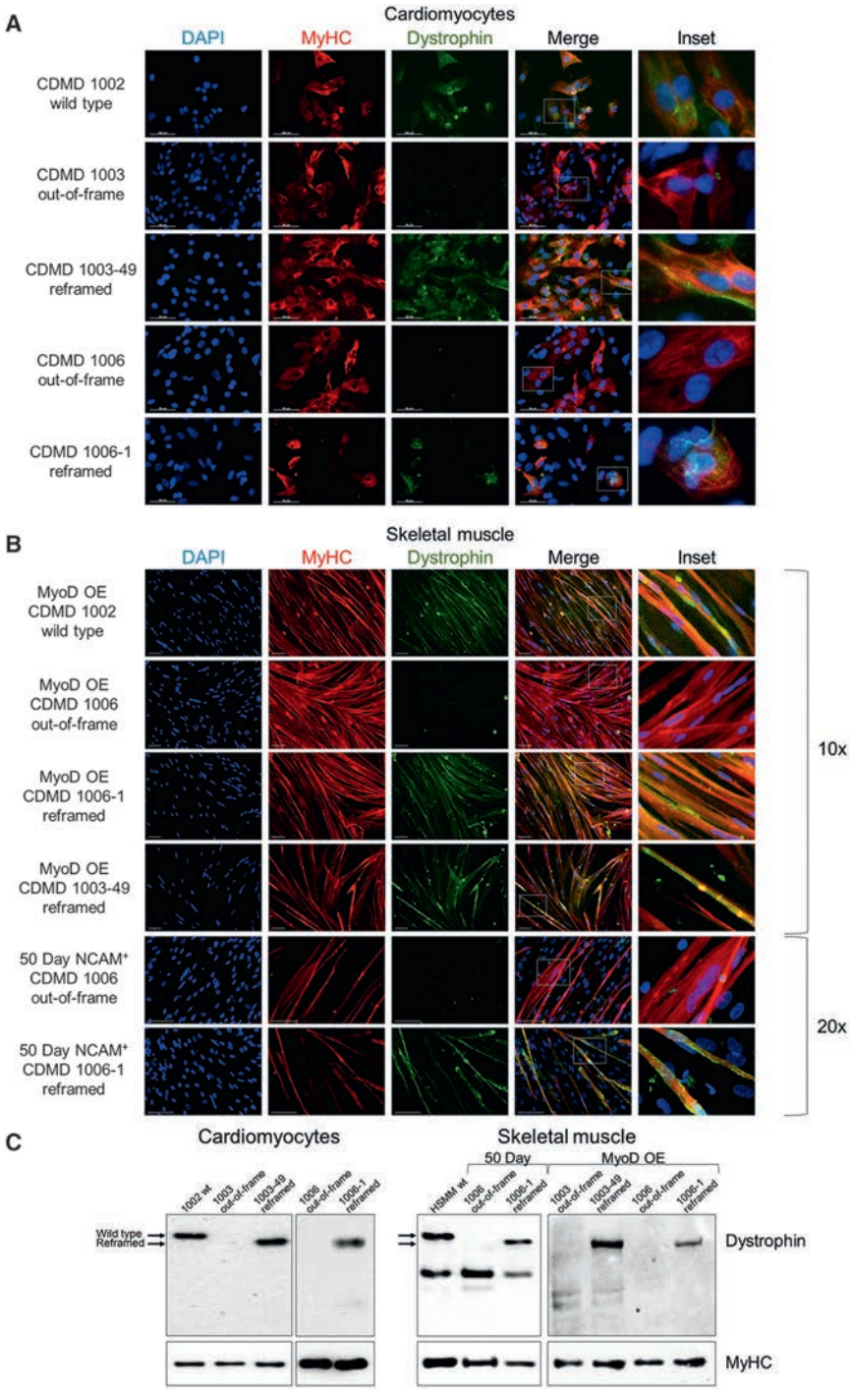


Figure 3. Reframed CDMD hiPSC-Derived Skeletal Muscle and Cardiomyocytes Restore Dystrophin Expression

(A) Immunocytochemical staining of human myosin heavy chain (MyHC, red) and dystrophin (green) of wild-type (CDMD 1002), out-of-frame (CDMD 1003 or 1006) or reframed (CDMD 1003-49 or 1006-1) cardiomyocytes derived from hiPSCs by directed differentiation. Inset depicts zoomed in region defined by the white box. Scale bar, 50 μm.

(B) Immunocytochemical staining of MyHC (red) and dystrophin (green) of wild-type (CDMD 1002), out-of-frame (CDMD 1006) or reframed (CDMD 1006-1 or 1003-49) skeletal muscle myotubes derived from hiPSCs. Myotubes were fused after MyoD OE or from sorted NCAM$^+$ cells after an adapted directed differentiation 50-day protocol was used. Inset depicts zoomed-in region defined by the white box. Scale bar, 100 μm.

(C) Western blots of cell extracts probed with anti-dystrophin. Extracts were from out-of-frame and reframed cardiomyocytes (left) and skeletal muscle myotubes (right), derived from CDMD hiPSCs. Wild-type (wt) hiPSCs (CDMD 1002) or human skeletal muscle myotubes (HSMM) were used as a control for dystrophin. The molecular weight shift caused by the exon 45–55 deletion (1779 bp, ~66 kDa) is evident in reframed versus wild-type dystrophin (arrows). A non-specific band around 220 kDa was seen in some samples. Samples were also probed with anti-MyHC as a loading control (bottom panels).

See also Figures S4C and S4D.

DYS$^{\Delta 45–55}$ was tested in cardiomyocytes and skeletal muscle derived from reframed DMD hiPSCs and demonstrated improved membrane stability by a physiologically relevant measure of CK release, similar to wild-type. The ability to evaluate cardiomyocyte functionality is an advantage of using hiPSCs, as some current preclinical and clinical studies for DMD therapies do not efficiently target the heart (e.g., exon skipping; Arechavala-Gomeza et al., 2012). Additionally, we demonstrated a normalization in miR31 levels, a microRNA that inhibits dystrophin, after reading frame restoration, similar to what is observed in human BMD patients (Cacchiarelli et al., 2011). Finally, we show restored DGC localization in vitro and in vivo, which further validates the functionality of DYS$^{\Delta 45–55}$.

Previous work by Ousterout et al. (2015) demonstrated that multiplexed gRNAs can restore the *DMD* reading frame in primary myoblasts. However, myoblasts do not provide a renewable source of stem cells, which is a requirement for long-term therapeutic efficacy (Partridge, 2002). In contrast, we used hiPSCs, which offer the opportunity to evaluate the internally deleted dystrophin protein in multiple cell types that are affected in DMD, and in future studies, they may provide a renewal source of corrected progenitor cells. Our work is further distinguished from previous studies as we are the only group to show restoration of dystrophin function on membrane integrity, miR31 expression, and the DGC in cardiac and skeletal muscle cells following CRISPR-mediated gene editing.

An advantage of our CRISPR platform is the therapeutic potential of a single pair of gRNAs to treat the majority of DMD patients. By designing gRNAs that accomplish a

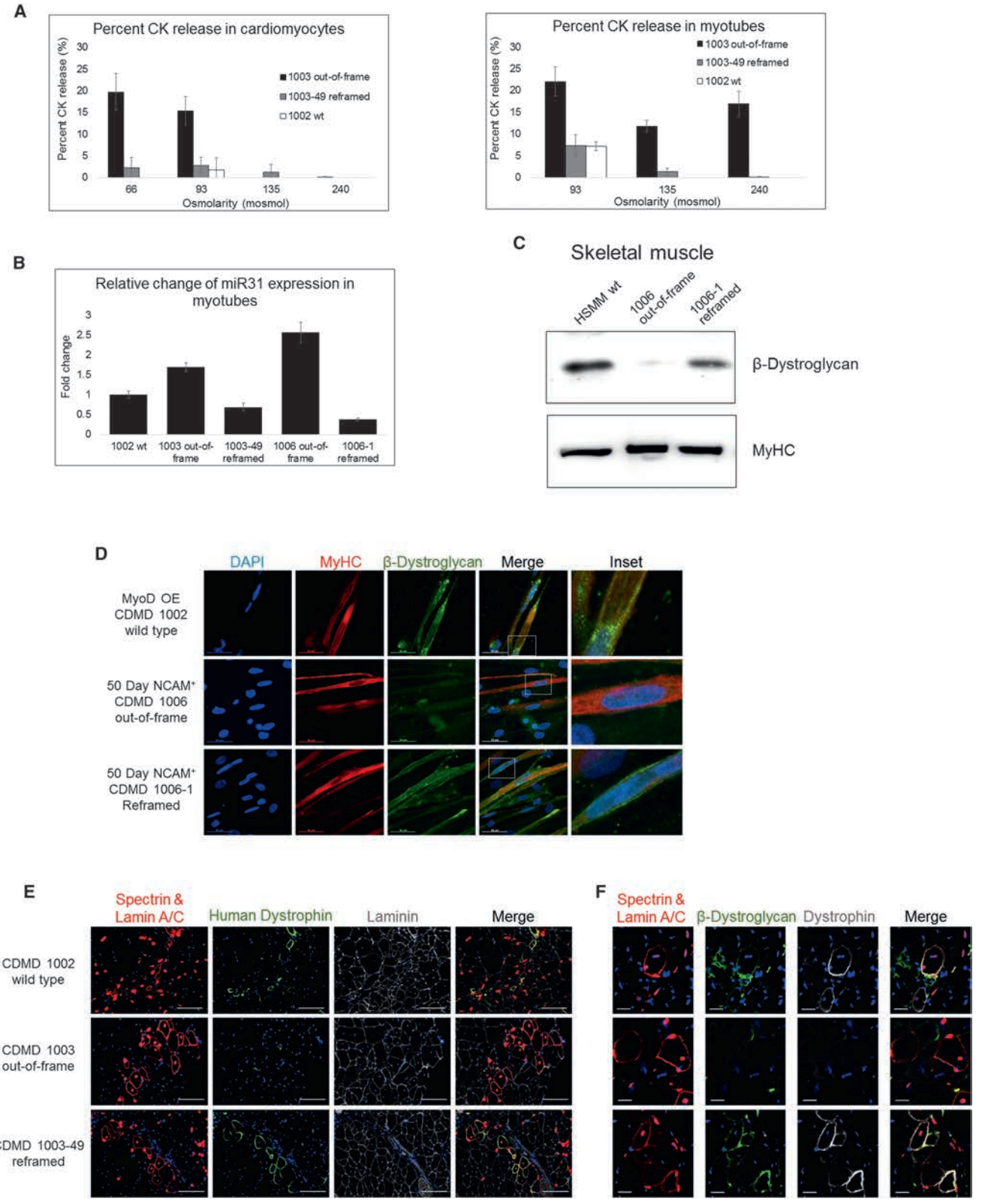


Figure 4. Reframed hiPSC-Derived Cardiomyocytes and Skeletal Muscle Cells Demonstrate Restored Function In Vitro and In Vivo

(A) Representative graphs of CK release assays from cells exposed to hypo-osmotic conditions. Cardiomyocytes and skeletal muscle myotubes derived from hiPSCs were subjected to a range of osmolarities below 240 mosmol, and CK release to the supernatant was measured as an indication of membrane fragility. Data are presented as average ± SE.

(legend continued on next page)

deletion that encompasses the majority of DMD mutations, this approach is optimized for future clinical studies. It would be unreasonable to design, validate, and evaluate off targets for every new CRISPR pair tailored for each individual patient. Additionally, CRISPR/Cas9 is advantageous over exon skipping, as it results in permanent restoration of the reading frame as opposed to transient effects on RNA splicing. Previously, Li et al. (2015) used CRISPR/Cas9 to induce exon skipping, frameshifting, or exon knockin to restore dystrophin in a DMD hiPSC line with an exon 44 deletion; however, their platform is only applicable to 3%–9% of DMD patients (Bladen et al., 2015), and two of their strategies relied on the creation of indels, which would be difficult to apply consistently to each patient. While Ousterout et al. deleted exons 45–55, they removed significantly less of the intervening region (336 kb) and thus their approach would cover fewer patient mutations within the hotspot region. This is because many mutations extend into the intronic region; thus, by designing gRNAs that encompass more of the intron, our platform is applicable to more patients.

Another benefit of using this platform to delete a large portion of *DMD*, as opposed to single exons, is the known correlation of $DYS^{\Delta 45–55}$ with a mild BMD phenotype. Large deletions in the rod domain of dystrophin often produce a more functional (more like wild-type) protein, than even very small deletions (Harper et al., 2002). Larger deletions, which remove hinge III (exons 50–51), are believed to lead to a milder BMD phenotype than smaller deletions, or those that retain hinge III (Carsana et al., 2005). Thus, in many cases larger deletions are more therapeutically beneficial than smaller ones, due to the way they affect the secondary structure of the protein.

In summary, we have developed a potentially therapeutic gene editing platform for DMD to permanently restore the dystrophin reading frame in multiple patient-derived hiPSCs. Our approach using CRISPR/Cas9 and NHEJ deletes up to 725 kb of *DMD* encompassing exons 45–55 and restores dystrophin protein function in both cardiomyocytes and skeletal muscle cells derived from reframed hiPSCs. A current limitation of this platform is that clinical protocols still need to be developed that allow rapid clonal line derivation and the utilization of hiPSC-derived cardiac and skeletal muscle progenitors combined with gene correction. Alternatively, CRISPR/Cas9 to restore the reading frame in DMD mouse models has been delivered directly in vivo (Long et al., 2016; Nelson et al., 2016; Tabebordbar et al., 2016). Thus, applications of this platform in the future will allow for the development of an in situ gene strategy or ex vivo gene correction followed by autologous cell transplantation, either of which offers tremendous potential for DMD.

EXPERIMENTAL PROCEDURES

Differentiation of hiPSCs to Skeletal Muscle Cells and Cardiomyocytes

Skeletal muscle differentiation from hiPSCs was induced using OE of a tamoxifen inducible MyoD-ERT lentivirus or an adapted 50 day directed differentiation protocol where $NCAM^+$ $HNK1^-$ cells underwent fluorescence-activated cell sorting at day 50. Cardiomyocytes were derived through aggregates over 30 days. See Supplemental Experimental Procedures.

Engraftment into Immunodeficient Mice

NSG immunodeficient mice (Jackson Laboratory) were crossed to mdx *scid* mice (Jackson Laboratory) to generate NSG-mdx mice (see Supplemental Experimental Procedures). Five- to seven-week-old NSG-mdx mice were pretreated with 50 μl of 10 μM cardiotoxin (Sigma-Aldrich) injected into the right TA 24 hr prior to engraftment. For MyoD OE cells, 100 μl of 5 mg/ml tamoxifen (Sigma-Aldrich) was i.p. injected for 5 days beginning on the day prior to engraftment. 1×10^6 cells in HBSS were injected intramuscularly and the TA was harvested after 30 days. See Supplemental Experimental Procedures.

Hypo-osmotic Stress CK Release Assay

Terminally differentiated skeletal muscle cells and cardiomyocytes plated in duplicate were stressed by incubation in hypo-osomolar solutions ranging from 66 to 240 mosmol (see Supplemental Experimental Procedures) for 20 min at 37°C. CK was measured in triplicate from the supernatant and cell lysate with the Creatine Kinase-SL kit (Sekisui Diagnostics) according to the manufacturer's instructions.

SUPPLEMENTAL INFORMATION

Supplemental Information for this article includes four figures, two tables and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2016.01.021.

AUTHOR CONTRIBUTIONS

Conceptualization, Methodology, Writing–Original Draft, Visualization, Project Administration, C.S.Y., M.J.S., and A.D.P.; Validation, C.S.Y., M.R.H., and N.V.E.; Formal Analysis, C.S.Y.; Investigation, C.S.Y., M.R.H., N.V.E., H.N., M.J., S.Y., S.K., C.K.-C., and D.W.; Resources, A.N., S.F.N., M.C.M., M.J.S., and A.D.P.; Writing–Review and Editing, C.S.Y., M.R.H., N.E.V., S.K., C.K.-C., D.B.K., A.N., S.F.N., M.C.M., M.J.S., and A.D.P.; Supervision, J.A.Z., D.B.K., A.N., M.C.M., M.J.S., and A.D.P.; Funding Acquisition, M.J.S. and A.D.P.

(B) Fold change in expression of miR31 measured by ddPCR in myotubes derived from out-of-frame or reframed hiPSCs by MyoD OE, normalized to wild-type (CDMD 1002). Data are presented as average ± SD.
(C) Western blots of cell extracts probed with anti-β-dystroglycan. Extracts were from out-of-frame and reframed skeletal muscle myotubes derived by MyoD OE. HSMM was used as a positive control. Samples were also probed with anti-MyHC as a loading control (bottom panel).
(D) Immunocytochemical staining of MyHC (red) and β-dystroglycan (green), a component of the DGC, in wild-type (CDMD 1002), out-of-frame (CDMD 1006), or reframed (CDMD 1006-1) skeletal muscle myotubes. Inset depicts zoomed-in region defined by the white box. Scale bar, 50 μm.
(E) Assessment of human dystrophin restoration in wild-type (CDMD 1002), out-of-frame (CDMD 1003), and reframed (CDMD 1003-49) MyoD OE cells engrafted into the TA of NSG-mdx mice. Engrafted human cells were identified by co-immunostaining for human spectrin and lamin A/C (shown in red). Positive staining for human dystrophin is shown in green and all fibers are shown using laminin (gray). All sections were stained with DAPI (blue) to identify nuclei. Scale bar, 100 μm.
(F) Assessment of β-dystroglycan restoration in human fibers from wild-type (CDMD 1002), out-of-frame (CDMD 1003), and reframed (CDMD 1003-49) MyoD OE cells engrafted into the TA of NSG-mdx mice. Engrafted human cells were identified by co-immunostaining for human spectrin and lamin A/C (shown in red). Positive staining for dystrophin is shown in gray and β-dystroglycan is shown in green. All sections were stained with DAPI (blue) to identify nuclei. Cell order is the same as noted in (E). Scale bar, 20 μm.
See also Figures S4E and S4F.

ACKNOWLEDGMENTS

We would like to thank Doug Black, Richard Wang, and Haibin Xi for suggestions and Diana Becerra and Jane Wen for technical assistance. The following cores were utilized: CDMD Muscle Phenotyping and Imaging Core, HTS and Cell Repository Core, and Bioinformatics and Genomics Core (supported by NIAMS-P30 AR057230); the UCLA Broad Stem Cell Research Center (BSCRC) Flow Cytometry Core Resource, the UCLA Jonsson Comprehensive Cancer Center (JCCC), and Center for AIDS Research (CFAR) Flow Cytometry Core Facility (supported by NIH P30 CA016042, 5P30 AI028697); and the UCLA Integrated Molecular Technologies Core (supported by CURE/P30 DK041301), the UCLA Humanized Mouse Core (supported by the JCCC, BSCRC, CFAR, and NIH/NIAID AI028697), and the UCLA GenoSeq core. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1144087 (C.S.Y.). Funding was provided by NIAMS of the NIH (P30 AR057230 to M.J.S., A.D.P., and A.N. and R01AR064327 to A.D.P.), the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA (M.J.S. and A.D.P.), and a Rose Hills Foundation Research Award (A.D.P.).

Received: September 9, 2015
Revised: December 18, 2015
Accepted: January 22, 2016
Published: February 11, 2016

REFERENCES

Arechavala-Gomeza, V., Anthony, K., Morgan, J., and Muntoni, F. (2012). Antisense oligonucleotide-mediated exon skipping for Duchenne muscular dystrophy: progress and challenges. Curr. Gene Ther. *12*, 152–160.

Béroud, C., Tuffery-Giraud, S., Matsuo, M., Hamroun, D., Humbertclaude, V., Monnier, N., Moizard, M.P., Voelckel, M.A., Calemard, L.M., Boisseau, P., et al. (2007). Multiexon skipping leading to an artificial DMD protein lacking amino acids from exons 45 through 55 could rescue up to 63% of patients with Duchenne muscular dystrophy. Hum. Mutat. *28*, 196–202.

Bladen, C.L., Salgado, D., Monges, S., Foncuberta, M.E., Kekou, K., Kosma, K., Dawkins, H., Lamont, L., Roy, A.J., Chamova, T., et al. (2015). The TREAT-NMD DMD Global Database: analysis of more than 7,000 Duchenne muscular dystrophy mutations. Hum. Mutat. *36*, 395–402.

Cacchiarelli, D., Incitti, T., Martone, J., Cesana, M., Cazzella, V., Santini, T., Sthandier, O., and Bozzoni, I. (2011). miR-31 modulates dystrophin expression: new implications for Duchenne muscular dystrophy therapy. EMBO Rep. *12*, 136–141.

Carsana, A., Frisso, G., Tremolaterra, M.R., Lanzillo, R., Vitale, D.F., Santoro, L., and Salvatore, F. (2005). Analysis of dystrophin gene deletions indicates that the hinge III region of the protein correlates with disease severity. Ann. Hum. Genet. *69*, 253–259.

Cradick, T.J., Qiu, P., Lee, C.M., Fine, E.J., and Bao, G. (2014). COSMID: A web-based tool for identifying and validating CRISPR/Cas off-target sites. Mol. Ther. Nucleic Acids *3*, e214.

Dekel-Naftali, M., Aviram-Goldring, A., Litmanovitch, T., Shamash, J., Reznik-Wolf, H., Laevsky, I., Amit, M., Itskovitz-Eldor, J., Yung, Y., Hourvitz, A., et al. (2012). Screening of human pluripotent stem cells using CGH and FISH reveals low-grade mosaic aneuploidy and a recurrent amplification of chromosome 1q. Eur. J. Hum. Genet. *20*, 1248–1255.

Echigoya, Y., Aoki, Y., Miskew, B., Panesar, D., Touznik, A., Nagata, T., Tanihata, J., Nakamura, A., Nagaraju, K., and Yokota, T. (2015). Long-term efficacy of systemic multiexon skipping targeting dystrophin exons 45-55 with a cocktail of vivo-morpholinos in mdx52 mice. Mol. Ther. Nucleic Acids *4*, e225.

Guan, X., Mack, D.L., Moreno, C.M., Strande, J.L., Mathieu, J., Shi, Y., Markert, C.D., Wang, Z., Liu, G., Lawlor, M.W., et al. (2014). Dystrophin-deficient cardiomyocytes derived from human urine: new biologic reagents for drug discovery. Stem Cell Res. (Amst.) *12*, 467–480.

Harper, S.Q., Hauser, M.A., DelloRusso, C., Duan, D., Crawford, R.W., Phelps, S.F., Harper, H.A., Robinson, A.S., Engelhardt, J.F., Brooks, S.V., and Chamberlain, J.S. (2002). Modular flexibility of dystrophin: implications for gene therapy of Duchenne muscular dystrophy. Nat. Med. *8*, 253–261.

Li, H.L., Fujimoto, N., Sasakawa, N., Shirai, S., Ohkame, T., Sakuma, T., Tanaka, M., Amano, N., Watanabe, A., Sakurai, H., et al. (2015). Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. Stem Cell Reports *4*, 143–154.

Long, C., McAnally, J.R., Shelton, J.M., Mireault, A.A., Bassel-Duby, R., and Olson, E.N. (2014). Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. Science *345*, 1184–1188.

Long, C., Amoasii, L., Mireault, A.A., McAnally, J.R., Li, H., Sanchez-Ortiz, E., Bhattacharyya, S., Shelton, J.M., Bassel-Duby, R., and Olson, E.N. (2016). Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. Science *351*, 400–403.

Menke, A., and Jockusch, H. (1995). Extent of shock-induced membrane leakage in human and mouse myotubes depends on dystrophin. J. Cell Sci. *108*, 727–733.

Nakamura, A., Yoshida, K., Fukushima, K., Ueda, H., Urasawa, N., Koyama, J., Yazaki, Y., Yazaki, M., Sakai, T., Haruta, S., et al. (2008). Follow-up of three patients with a large in-frame deletion of exons 45-55 in the Duchenne muscular dystrophy (DMD) gene. J. Clin. Neurosci. *15*, 757–763.

Nelson, C.E., Hakim, C.H., Ousterout, D.G., Thakore, P.I., Moreb, E.A., Rivera, R.M.C., Madhavan, S., Pan, X., Ran, F.A., Yan, W.X., et al. (2016). In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. Science *351*, 403–407.

Ousterout, D.G., Kabadi, A.M., Thakore, P.I., Majoros, W.H., Reddy, T.E., and Gersbach, C.A. (2015). Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. Nat. Commun. *6*, 6244.

Partridge, T. (2002). Myoblast transplantation. Neuromuscul. Disord. *12* (*Suppl 1*), S3–S6.

Pearce, J.M.S., Pennington, R.J.T., and Walton, J.N. (1964). Serum enzyme studies in muscle disease: Part III Serum creatine kinase activity in relatives of patients with the Duchenne type of muscular dystrophy. J. Neurol. Neurosurg. Psychiatry *27*, 181–185.

Tabebordbar, M., Zhu, K., Cheng, J.K.W., Chew, W.L., Widrick, J.J., Yan, W.X., Maesner, C., Wu, E.Y., Xiao, R., Ran, F.A., et al. (2016). In vivo gene editing in dystrophic mouse muscle and muscle stem cells. Science *351*, 407–411.

Taglia, A., Petillo, R., D'Ambrosio, P., Picillo, E., Torella, A., Orsini, C., Ergoli, M., Scutifero, M., Passamano, L., Palladino, A., et al. (2015). Clinical features of patients with dystrophinopathy sharing the 45-55 exon deletion of DMD gene. Acta Myol. *34*, 9–13.

Wojtal, D., Kemaladewi, D.U., Malam, Z., Abdullah, S., Wong, T.W.Y., Hyatt, E., Baghestani, Z., Pereira, S., Stavropoulos, J., Mouly, V., et al. (2016). Spell Checking Nature: Versatility of CRISPR/Cas9 for Developing Treatments for Inherited Disorders. Am. J. Hum. Genet. *98*, 90–101.

Xu, L., Park, K.H., Zhao, L., Xu, J., El Refaey, M., Gao, Y., Zhu, H., Ma, J., and Han, R. (2015). CRISPR-mediated genome editing restores dystrophin expression and function in mdx mice. Mol. Ther., in press. Published online October 9, 2015. http://dx.doi.org/10.1038/mt.2015.192.

Cell Stem Cell
Resource

CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs

Mohammad A. Mandegar,[1,*] Nathaniel Huebsch,[1,2] Ekaterina B. Frolov,[1] Edward Shin,[1] Annie Truong,[1] Michael P. Olvera,[1] Amanda H. Chan,[1] Yuichiro Miyaoka,[1,12] Kristin Holmes,[1] C. Ian Spencer,[1] Luke M. Judge,[1,2] David E. Gordon,[3,4,5] Tilde V. Eskildsen,[6,7] Jacqueline E. Villalta,[3,4,8,9] Max A. Horlbeck,[3,4,8,9] Luke A. Gilbert,[3,4,8,9] Nevan J. Krogan,[3,4,5] Søren P. Sheikh,[6,7] Jonathan S. Weissman,[3,4,8,9] Lei S. Qi,[10] Po-Lin So,[1] and Bruce R. Conklin[1,3,4,11,*]
[1]Gladstone Institute of Cardiovascular Disease, San Francisco, CA 94158, USA
[2]Department of Pediatrics, University of California, San Francisco, San Francisco, CA 94158, USA
[3]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94158, USA
[4]California Institute for Quantitative Biosciences, QB3, University of California, San Francisco, San Francisco, CA 94158, USA
[5]Gladstone Institute of Virology and Immunology, San Francisco, CA 94158, USA
[6]Department of Cardiovascular and Renal Research, University of Southern Denmark, 5000 Odense C, Denmark
[7]Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, 5000 Odense C, Denmark
[8]Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, CA 94158, USA
[9]Center for RNA Systems Biology, University of California, San Francisco, San Francisco, CA 94158, USA
[10]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
[11]Department of Medicine and Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94158, USA
[12]Present address: Regenerative Medicine Project, Tokyo Metropolitan Institute of Medical Science, Tokyo, 156-8506, Japan
*Correspondence: mo.mandegar@gladstone.ucsf.edu (M.A.M.), bconklin@gladstone.ucsf.edu (B.R.C.)
http://dx.doi.org/10.1016/j.stem.2016.01.022

SUMMARY

Developing technologies for efficient and scalable disruption of gene expression will provide powerful tools for studying gene function, developmental pathways, and disease mechanisms. Here, we develop clustered regularly interspaced short palindromic repeat interference (CRISPRi) to repress gene expression in human induced pluripotent stem cells (iPSCs). CRISPRi, in which a doxycycline-inducible deactivated Cas9 is fused to a KRAB repression domain, can specifically and reversibly inhibit gene expression in iPSCs and iPSC-derived cardiac progenitors, cardiomyocytes, and T lymphocytes. This gene repression system is tunable and has the potential to silence single alleles. Compared with CRISPR nuclease (CRISPRn), CRISPRi gene repression is more efficient and homogenous across cell populations. The CRISPRi system in iPSCs provides a powerful platform to perform genome-scale screens in a wide range of iPSC-derived cell types, dissect developmental pathways, and model disease.

INTRODUCTION

To understand the biological roles of genes in development and disease, we must decipher the relationships between genotype and phenotype. Until recently, RNAi has been the most commonly used loss-of-function tool to study human biology (Boettcher and McManus, 2015). However, RNAi suffers from off-target effects and incomplete silencing of the desired gene (Jackson et al., 2003; Kim et al., 2013b; Krueger et al., 2007). Alternatively, programmable nucleases, such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), allow more precise gene editing in model organisms, particularly in mammalian and human systems (Gaj et al., 2013; Kim and Kim, 2014). While ZFNs and TALENs are efficient tools for targeting single alleles, they cannot be easily used for library-scale loss-of-function studies.

In 2012, clustered regularly interspaced short palindromic repeat (CRISPR) technology emerged as a new tool for gene editing. This technology is a microbial adaptive-immune system that uses RNA-guided nucleases to recognize and cleave foreign genetic elements (Doudna and Charpentier, 2014; Wiedenheft et al., 2012). The recently engineered CRISPR/Cas9 system consists of two components: a single-chimeric guide RNA (gRNA) that provides target specificity and a CRISPR-associated protein (Cas9) that acts as a helicase and a nuclease to unwind and cut the target DNA (Cong et al., 2013; Mali et al., 2013). In this system, the only restriction for targeting a specific locus is the protospacer adjacent motif (PAM) sequence ("NGG" in the case of *Sp*Cas9) (Doudna and Charpentier, 2014).

CRISPR nuclease (CRISPRn) has been used for genome-scale screens to identify essential genes for cell viability in cancer and embryonic stem cells (Shalem et al., 2014) and human leukemic cell lines (Wang et al., 2014, 2015). However, CRISPRn may not be the most robust system for loss-of-function studies, because it is limited by the number of cells within a population that do not produce knockout phenotypes (González et al., 2014). In addition, partial loss- or gain-of-function phenotypes can be generated by Cas9-induced in-frame insertion/deletions (INDELs) and hypomorphic alleles (Shi et al., 2015), which can obscure the readout.

The nuclease deactivated version of Cas9 (dCas9) blocks transcription in prokaryotic and eukaryotic cells (known as CRISPR interference; CRISPRi) (Qi et al., 2013). More recently, dCas9 was fused to the Krüppel-associated box (KRAB) repression domain to generate dCas9-KRAB, producing a

more efficient transcriptional interference (Gilbert et al., 2013, 2014; Kearns et al., 2014). To further this effort, we aimed to use CRISPRi technology to efficiently repress genes to study early differentiation and model disease with human induced pluripotent stem cells (iPSCs) (Takahashi et al., 2007).

iPSCs are well suited to study early embryonic development and disease since they can produce different functional cell types in vitro (Sterneckert et al., 2014). Early embryonic development consists of a series of accurately timed events that affect gene activation and repression (Bolouri and Davidson, 2003). Therefore, precisely regulating the timing and dosage of transcription factors critically affects embryonic development (McFadden et al., 2005; Takeuchi et al., 2011), and dysregulation in the timing and dosage of transcripts can lead to disease development (Theodoris et al., 2015). In this study, we compared inducible CRISPR systems for gene knockout (using Cas9) or knockdown (using dCas9-KRAB) to enable temporal control of loss-of-function phenotypes in iPSCs and differentiated cell types.

RESULTS

Generation of CRISPRi and CRISPRn iPSC Lines

For loss-of-function studies, we independently derived multiple stable CRISPRi and CRISPRn human iPSC clones in two genetic backgrounds: wild-type B (WTB) and wild-type C (WTC) (Miyaoka et al., 2014). In separate targeting events, the CRISPRi and CRISPRn constructs (see Supplemental Experimental Procedures) were integrated into the AAVS1 locus of WTB and WTC iPSCs using a TALEN-assisted gene-trap approach (Figures 1A, 1B, and S1). Transgenes integrated at the AAVS1 locus remain transcriptionally active in both iPSCs and differentiated cell types (Hockemeyer et al., 2011; Lombardo et al., 2011). We generated several different versions of the CRISPRi system that are either inducible or constitutive; the inducible CRISPRi (Gen1 and Gen2) clones express dCas9-KRAB (KRAB domain fused at the N terminus) from the inducible TetO promoter, while the constitutive CRISPRi clones (Gen3) express dCas9-KRAB under the constitutively active CAG promoter. The CRISPRn (Gen1) clones express Cas9 under the inducible TetO promoter (Figure S1).

The average efficiency of forming stable clones was ~350 colonies per million iPSCs transfected with AAVS1 TALENs and donor plasmid (data not shown). From each condition, multiple independent colonies were isolated and expanded. A subset of the stable colonies from each targeting vector was screened using junction PCR. Two putative colonies from each targeting event were further characterized by stably introducing an *OCT4*-specific gRNA and performing knockdown or knockout assays with immunofluorescence and western blot analysis. All putative CRISPRi clones containing an *OCT4*-specific gRNA showed efficient knockdown (>95%) of OCT4 in bulk populations, while a significant fraction of the CRISPRn cells remained OCT4 positive (~30%–40%) in bulk populations containing *OCT4*-specific gRNA (Figure S1). One clone each from CRISPRi and CRISPRn (Gen1 lines in the WTC genetic background) were subsequently used as lead clones for further studies.

To enable non-invasive and high-throughput phenotypic analysis in iPSC-derived cardiomyocytes (iPS-CMs), we performed a second targeting event that introduced the green fluorescent calcium-modulated protein 6 fast type (GCaMP) calcium sensor (Chen et al., 2013) into the other AAVS1 locus of the CRISPRi cell line. The GCaMP transgene is driven off the strong, constitutive CAG promoter (Figure S1). We found that CRISPRi iPSCs could differentiate into iPS-CMs, so that we could measure calcium transients based on the GCaMP-fluorescent intensity (Movie S1) (Huebsch et al., 2015). Lead CRISPRi and CRISPRn iPSCs were karyotypically normal (Figures S2A and S2B) and expressed pluripotency markers, as expected (Figures S2C and S2D).

RNA-sequencing (RNA-seq) analysis indicated that expression of dCas9-KRAB or Cas9 was undetectable in the absence of doxycycline, and addition of doxycycline without any gRNA resulted in robust selective induction of dCas9-KRAB or Cas9, while the rest of the transcriptome remained virtually unchanged (Figures S2E and S2F). Furthermore, the RNA-seq data suggest that the addition of the KRAB domain has no detectable off-target effects when compared to expression of Cas9 alone. Remarkably the one gene that appeared to be upregulated upon doxycycline induction (without gRNA) was the same gene (Vimentin; *VIM*) for both CRISPRi and CRISPRn cells (Figures S2E and S2F). Since the same gene is upregulated for CRISPRi and CRISPRn cells, we suspect it may represent an off-target activity of the doxycycline-induced transactivator. Importantly, our experiments suggest that the expression of dCas9-KRAB alone has no additional effects on gene expression.

We also expressed dCas9-KRAB and Cas9 by continuously culturing CRISPRi and CRISPRn lines with doxycycline for 3 weeks (four passages). With this long-term treatment, we observed no cytotoxicity, decrease in proliferation, or change in morphology in these cells (Figures S2G and S2H). Using a droplet digital PCR (ddPCR)-based copy-number assay, we measured the number of integration events (Figure S2I). We further validated on-target integration sites on the lead CRISPRi and CRISPRn clones with junction PCR (Figure S2J) and verified their sequences (data not shown).

To further ensure there was no leaky expression of the single doxycycline-inducible vector, we measured the protein levels of dCas9-KRAB and Cas9 in iPSCs. With immunostaining, flow cytometry and western blots did not detect dCas9-KRAB or Cas9 protein without doxycycline in either CRISPRi or CRISPRn iPSCs, indicating that the TetO promoter has high fidelity in the AAVS1 locus. After doxycycline treatment, all cells in the CRISPRi and CRISPRn lines expressed dCas9-KRAB or Cas9 within 48 hr, respectively (Figures 1C–1H). dCas9-KRAB and Cas9 were expressed at similar levels after induction, and both proteins rapidly degraded after removing doxycycline (Figures 1F, 1H, and S2K). These data showed that dCas9-KRAB and Cas9 expression could be tightly regulated with the TetO promoter, which would support studies that rely on precisely timing gene knockdown or knockout.

Comparison of Loss of Function between CRISPRi and CRISPRn

To compare CRISPRi and CRISPRn for loss-of-function studies, we designed a gRNA that targets the first exon of *NANOG*, a transcription factor necessary for maintaining the pluripotency network. We selected *NANOG* as our first target gene because its deficiency is sufficient to give an immediate readout, as

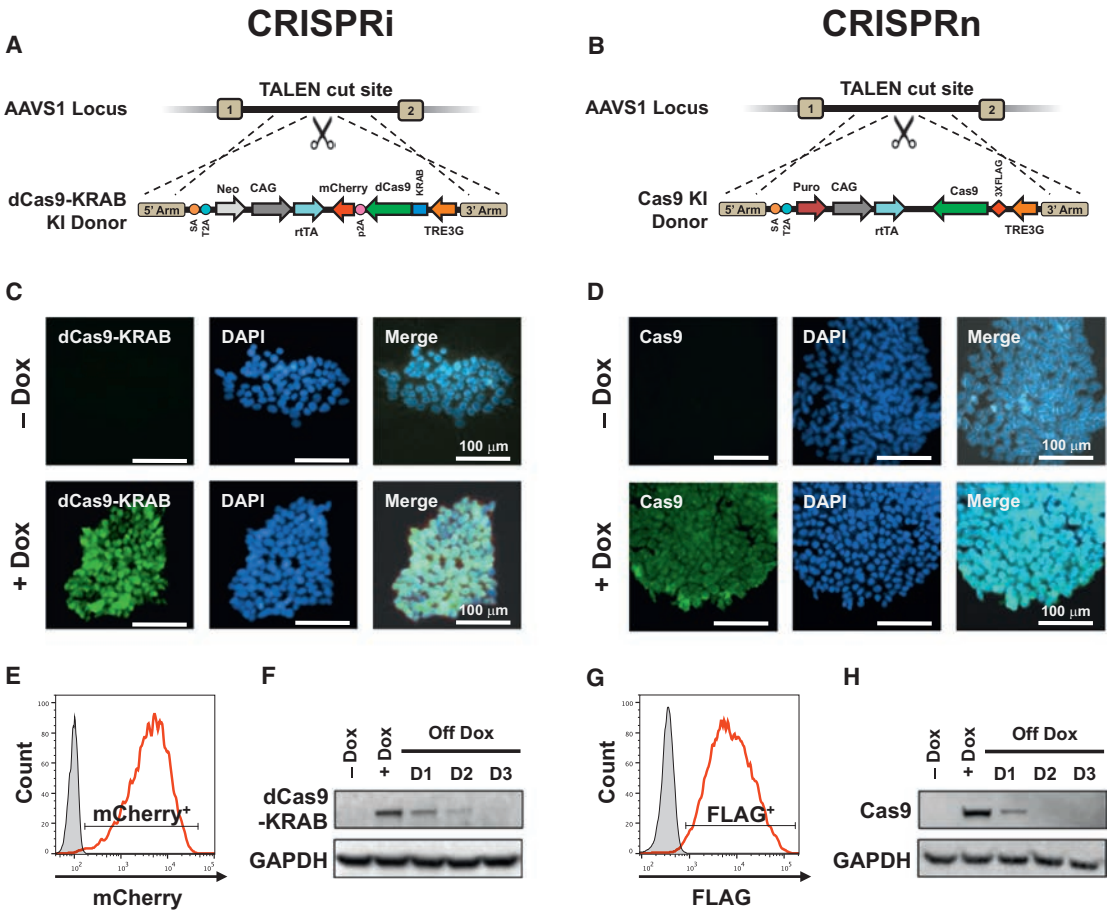


Figure 1. Generation and Characterization of Inducible CRISPRi and CRISPRn iPSCs

(A and B) Schematic overview of the strategy for TALEN-mediated targeting to the AAVS1 locus to generate the CRISPRi and CRISPRn iPSC lines. The doxycycline-controlled reverse transcriptional activator (rtTA) is driven by a strong constitutive promoter (CAG). The third-generation doxycycline-response element (TRE3G) drives transcription of either Cas9 (CRISPRn) or dCas9-KRAB-P2A-mCherry (CRISPRi) and is oriented in the opposite direction of the transactivator to ensure no leaky expression without doxycycline treatment.

(C and D) Immunostaining of CRISPRi and CRISPRn colonies before and after 48 hr of doxycycline treatment with an antibody against Cas9 (green). Nuclei are stained with DAPI (blue). All nuclei showed expression of dCas9-KRAB or Cas9 after adding doxycycline.

(E and G) Flow cytometry analysis of CRISPRi and CRISPRn iPSC lines before and after 48 hr of doxycycline treatment. Doxycycline treatment of CRISPRi and CRISPRn produced expression of mCherry and FLAG in all cells, respectively. The doxycycline-untreated sample is plotted in gray.

(F and H) CRISPRi and CRISPRn iPSC lines were treated with doxycycline (2 μM) for 24 hr, which was then removed to measure the protein half-life of dCas9-KRAB and Cas9. Total protein was extracted from samples and analyzed by western blot with antibodies against Cas9 and GAPDH as a loading control. Both the CRISPRi and CRISPRn clones express dCas9-KRAB and Cas9 at similar levels after doxycycline treatment, and the half-life of both proteins was ~12 hr in iPSCs. Scale bars, 100 μm.

indicated by a clear loss of pluripotent cell morphology (Hayashi et al., 2015). In general, Cas9 can disrupt gene function at any given exon (Doench et al., 2014), while dCas9-KRAB knocks down gene expression only when gRNAs are targeted to the transcription start site (TSS) (Gilbert et al., 2014). Hence, for this comparative study, we used the same gRNA sequence for both CRISPRi and CRISPRn. Here, we introduced a gRNA targeting 358 bp downstream of the *NANOG* TSS (142 bp into exon 1 of *NANOG*) into the CRISPRi and CRISPRn clones and selected subclones (as described in Experimental Procedures). We then treated multiple independent subclones of CRISPRi and CRISPRn iPSCs containing the *NANOG* gRNA-expression vector (as indicated by mKate2 expression) with doxycycline (Figure 2).

With CRISPRi, we found that NANOG expression was completely lost (>99%) in multiple independent iPSC subclones after doxycycline treatment (Figures 2A, 2C, 2E, S3A, and S3C). However, with CRISPRn, only 60%–70% of the cells lost NANOG expression in multiple independent subclones post-doxycycline induction (Figures 2B, 2D, 2G, S3B, and S3D). Next, we extracted genomic DNA from *NANOG* gRNA-containing CRISPRi and CRISPRn iPSCs and performed sequence analysis. As expected, we found that CRISPRi iPSCs did not harbor any mutations in the *NANOG* locus pre- or post-doxycycline treatment (Figure 2F). However, with CRISPRn, after 12–17 days of continuous doxycycline treatment, among the mutated alleles, 30%–50% of the sequences contained in-frame INDELs at the cut site (a total of 77 sequenced clones) (Figure 2H).

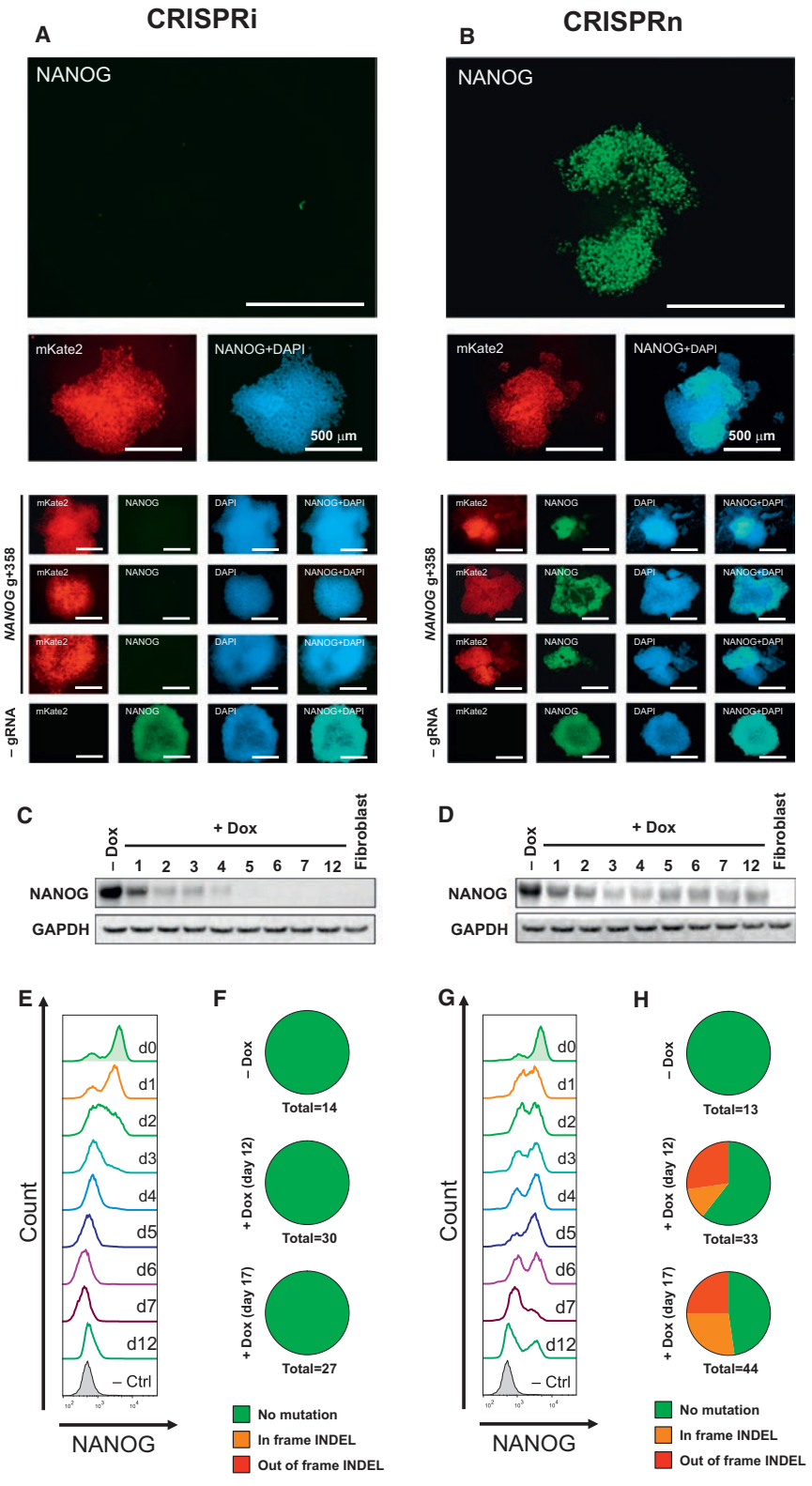


Figure 2. Comparison of the Efficiency of CRISPRi Knockdown and CRISPRn Knockout

(A and B) Immunostaining of representative (A) CRISPRi and (B) CRISPRn stable clones, each containing the same gRNA targeting the first exon of *NANOG* (*NANOG* g+358). After 7 days of doxycycline treatment, *NANOG* expression (green) was completely lost in all CRISPRi clones but showed a variegated pattern of knockout in multiple independent CRISPRn clones. The mKate2 signal indicates the presence of the gRNA-expression vector in all cells within the clone. Nuclei are counterstained with DAPI.

(C, D, E, and G) Western blot and flow cytometry analyses of (C and E) CRISPRi and (D and G) CRISPRn stable clones that contain the same gRNA against the first exon of *NANOG*. With CRISPRi, *NANOG* expression was uniformly decreased during doxycycline treatment and did not increase thereafter; however, with CRISPRn, the percentage of NANOG-positive cells fluctuated during doxycycline treatment. Even after 12 days of continuous doxycycline treatment, ~30% of the population stained positive for NANOG.

(F and H) Genomic DNA was extracted from (F) CRISPRi and (H) CRISPRn stable lines containing a gRNA against *NANOG* before and after continuous doxycycline treatment for up to 17 days and subjected to sequencing. Red, out-of-frame INDELs; orange, in-frame INDELs; green, non-mutated alleles. Even after 12–17 days of continuous doxycycline treatment, 50%–70% of sequenced alleles from CRISPRn contained no mutation, and 30%–50% of mutated alleles were in-frame INDELs. No mutations were observed in either CRISPRi or CRISPRn without doxycycline, and the CRISPRi clones did not contain any mutations after doxycycline treatment. The total number of sequenced colonies is listed below each pie graph.

Scale bars, 500 μm.

To further compare CRISPRi with CRISPRn, we targeted another pluripotency transcription factor, *OCT4*, with two independent gRNAs. Similar to our findings with *NANOG*, *OCT4* was completely knocked down in independent CRISPRi clones expressing the gRNA vector after doxycycline treatment (Figure S3E). In contrast, the attempted knockout of OCT4 with CRISPRn again yielded incomplete effects (Figure S3F). These findings were also replicated in a completely different iPSC line (WTB genetic background; CRISPRi Gen1B and CRISPRn Gen1B) (Figures S1D and S1F). We analyzed the genomic DNA of CRISPRn cells after 14 days of continuous doxycycline treatment and found 30%–40% of the mutated alleles had in-frame INDELs (a total of 91 sequenced clones) (Figure S3G). These results suggested that, in the context of targeting pluripotency factors, CRISPRi more rapidly generates loss-of-function phenotypes in bulk populations than CRISPRn. CRISPRi caused a complete

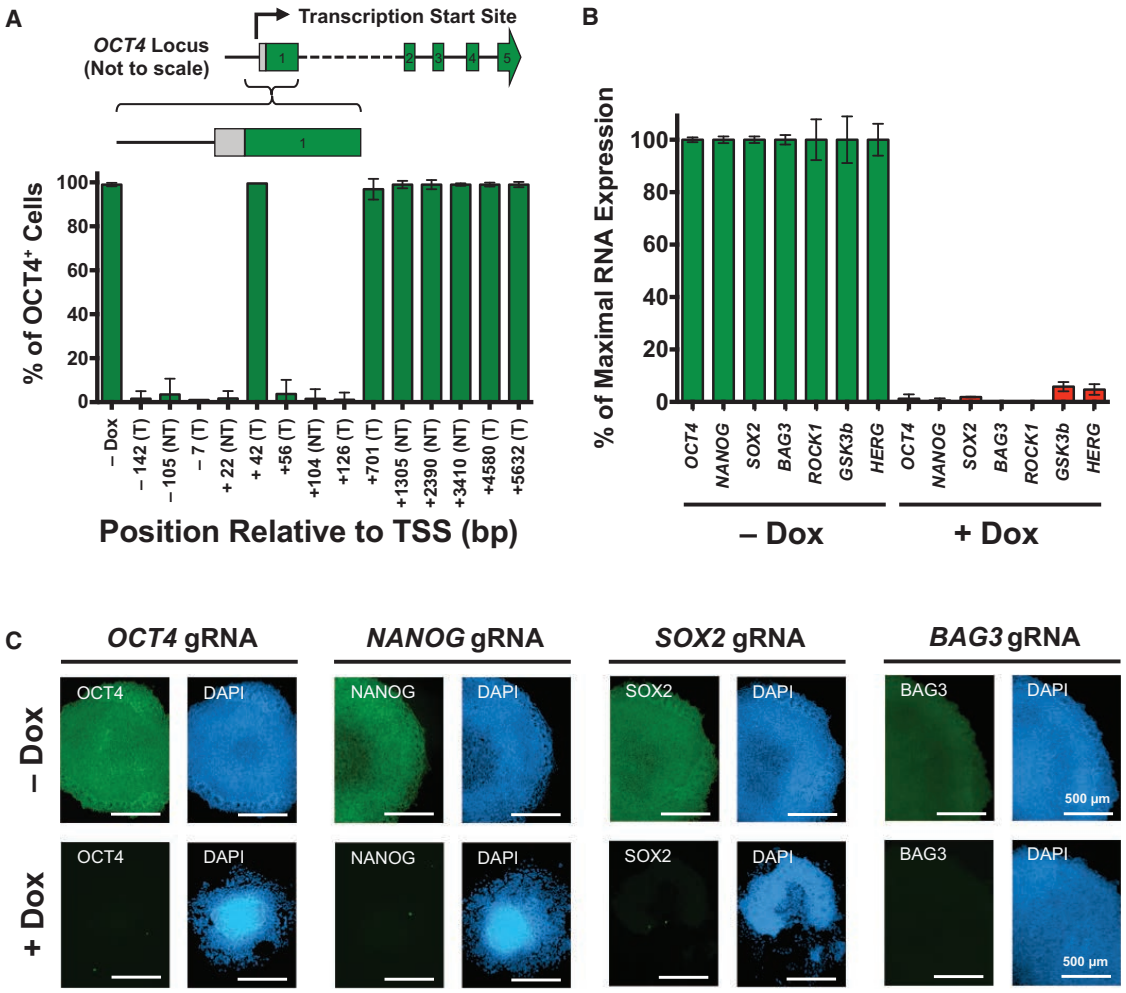


Figure 3. CRISPRi Knockdown Is Efficient in iPSCs

(A) Efficiency of gRNA knockdown based on proximity to the transcription start site (TSS). The binding location of each gRNA is indicated relative to the TSS of the *OCT4* locus and whether it targets the template (T) or non-template (NT) strand. Only gRNAs targeting near the TSS (approximately ±150 bp) effectively knocked down *OCT4*.

(B) TaqMan qPCR analysis of stable iPSCs containing gRNA against the gene of interest showed greater than 90% knockdown efficiency after 7 days of doxycycline induction in different endogenous genetic loci.

(C) Immunostaining of stable clones containing a single gRNA against the gene of interest (*OCT4, SOX2*, *NANOG,* and *BAG3*). After 7 days of doxycycline treatment, there was a complete knockdown of the protein of interest (green). As expected, DAPI staining revealed that knocking down *OCT4*, *NANOG* and *SOX2* resulted in loss of pluripotency and clear morphological changes. Also, knocking down *BAG3* did not cause a loss of pluripotent morphology, as indicated by the distinct and round colony edges.

Error bar represents SD.

loss of transcript expression and rapid cell differentiation when targeting *NANOG* and *OCT4* within 5–7 days of knockdown initiation. With CRISPRn, even after ~2 weeks of doxycycline treatment, a significant fraction (30%–40%) of the cells remained NANOG and OCT4 positive and maintained their pluripotency. Therefore, we focused on using CRISPRi as a loss-of-function tool in subsequent experiments.

CRISPRi Is Most Effective near the TSS

To further test the efficacy of gRNAs in CRISPRi, we designed multiple gRNAs that target near the TSS of OCT4. With flow cytometry assays for OCT4 staining (Figure 3A), we found that most gRNAs targeting near the TSS (approximately −150 bp to +150 bp around the TSS in this study) were highly effective at gene knockdown, but gRNAs targeting significantly (>700 bp) downstream of the TSS were not. This result agrees with previous data (Gilbert et al., 2014) and suggests that CRISPRi primarily blocks transcription at initiation, which reduces the likelihood of off-target effects from transcript interference elsewhere in the genome. Following these design criteria, for subsequent gene targets, we designed gRNAs to target near the TSS.

CRISPRi Efficiently Knocks Down a Broad Range of Genetic Loci

To test the efficiency of CRISPRi across a broad range of genetic loci in both iPSCs and differentiating/differentiated cell types, we

designed gRNAs against a total of nine genomic loci. The loci included core pluripotency transcription factors (*OCT4*, *NANOG*, and *SOX2*), kinases (*ROCK1* and *GSK3-β*), a cardiac mesoderm-transcription factor (*MESP1*), and cardiac disease-associated genes (*BAG3*, *MYBPC3*, and *HERG*). Except for *MESP1* (expressed only transiently in cardiac mesoderm cells) and *MYBPC3* (expressed only in cardiomyocytes), all other genes are expressed in iPSCs at different levels. We generated populations of CRISPRi iPSCs containing stably integrated gRNA-expression constructs. We then cultured these stable polyclones or clonal populations either with or without doxycycline for at least 7 days.

Three to five gRNAs were designed to target near the TSS of each gene and initially were tested individually in polyclonal populations. Approximately half of the tested gRNAs were active in polyclonal populations with a silencing activity of over 70% (Figure S4A). We did not observe a difference in the knockdown efficiency between gRNAs targeting either the template or non-template strands (Figures 3A, S4A, and S4B). The most active gRNA-containing polyclonal line was further passaged and subcloned for more detailed knockdown analysis. Using the most active gRNA, we achieved 90%–99% knockdown of the gene of interest in a selected population of iPSCs after doxycycline treatment (Figure 3B). As expected, when we subcloned polyclonal populations via single-cell cloning, we observed a higher percentage of transcriptional knockdown. With immunofluorescence analysis we found that iPSC clones expressing gRNAs against *OCT4*, *NANOG*, *SOX2*, and *BAG3* showed complete loss of target protein expression 7 days after doxycycline induction. In cells expressing gRNAs against the core pluripotency transcription factors *OCT4*, *NANOG*, and *SOX2*, we observed clear morphological changes and a loss of pluripotency after doxycycline induction; however, loss of a non-pluripotency gene (*BAG3*) did not affect pluripotent morphology (Figure 3C).

Using the Gen1 CRISPRi knockin vector, we targeted non-iPSCs with a different genetic background to determine how broadly this technology can be applied to other cell types. A T-lymphocyte (CEM) CRISPRi line was generated, as described in Experimental Procedures. Similar to the iPSC experiments, gRNAs were introduced to the stable CEM CRISPRi cell line, and cells cultured in either the presence or absence of doxycycline for 10 days. Three gRNAs were tested to knock down *CD4* in CEM-CRISPRi cells, and all showed greater than 70% knockdown efficiency in polyclonal populations (Figure S4B). The most active gRNA-containing polyclone was subcloned, and three independent clonal lines were isolated and assayed for knockdown, where greater than 95% knockdown efficiency was observed (Figure S4C). These results clearly demonstrate the doxycycline-inducible CRISPRi vector system is highly versatile and transportable to other cell lines and shows high efficiency of knockdown across a range of cell types and genetic loci.

CRISPRi Knockdown Is Reversible and Tunable and Can Be Allele Specific

GCaMP is a calcium-sensitive modified GFP and, thus, can be used as a fluorescent reporter under steady-state levels of cytoplasmic Ca^{2+} (Apáti et al., 2013). Using GCaMP (driven off the strong constitutive promoter, CAG), we monitored the green-fluorescence signal in iPSCs to determine if we could knock down GCaMP and then reverse its expression by removing doxycycline from the culture. We found that adding doxycycline for 7 days knocked down GCaMP expression by 98%, which was completely restored after removing doxycycline for 14 days (Figure 4A). Similarly, we targeted the *BAG3* endogenous locus and achieved efficient transcript knockdown post-doxycycline treatment. *BAG3* expression was fully restored after doxycycline withdrawal (Figure 4B). These findings indicate that CRISPRi knockdown is fully reversible in iPSCs.

To determine if we could achieve variable levels of knockdown with different gRNA sequences, we tested two additional gRNAs targeting GCaMP (g+24 and g+91) (Figure 4C). These gRNAs knocked down GCaMP expression by only ~30% and ~50%, as measured by flow cytometry (Figures 4D and 4E). Therefore, by changing the location of the gRNA-binding site, we can tune the level of knockdown when trying to mimic haploinsufficiency or reduced protein levels (rather than complete loss of function). In addition, we tested whether the knockdown level is tunable by titrating the doxycycline concentration. Careful titration of the doxycycline concentration enabled homogenous modulation of GCaMP expression (Figure S5).

We next sought to further test the tunability of knockdown with CRISPRi. We determined if we could use single-nucleotide polymorphisms (SNPs) to specifically target one allele for knockdown to achieve a heterozygous-like state. In our CRISPRi iPSCs, there is a SNP near the TSS of *OCT4*. Thus, we designed a gRNA in which the heterozygous SNP is located in the PAM sequence (AGG versus AGA). Because an "NGG" sequence is required for dCas9 to target DNA, we could selectively target only one *OCT4* allele (Figure 4F). After doxycycline induction, we found that the iPSC population carrying the SNP-specific *OCT4* gRNA (*OCT4* g−4) remained OCT4 positive (~97%) by flow cytometry analysis. However, the median intensity of OCT4 staining was reduced by ~40% after 7 days of doxycycline treatment, implying that OCT4 expression was homogeneously reduced in all cells and not just a fraction of them (Figures 4G and 4H). We confirmed this finding with immunocytochemistry and TaqMan qPCR (data not shown).

CRISPRi Knockdown Is Highly Specific

To assess the specificity of CRISPRi targeting, we designed a gRNA that targets the GCaMP transgene, since its silencing should have few downstream transcriptional and cellular consequences. Indeed, expression of the GCaMP transcript was over 30-fold lower in the doxycycline-treated condition, while few other endogenous transcripts changed expression level with the exception of *VIM* as previously discussed (Figure 5A).

CRISPRi to Promote iPSC Differentiation

To show that our system can release iPSCs from their pluripotent state to promote differentiation, we tested the efficiency of CRISPRi in knocking down core pluripotency transcription factors (*OCT4*, *SOX2*, and *NANOG*) without adding small molecules or cytokines to the mTeSR media. We targeted gRNA against these genes and performed a time-course analysis of a selected number of transcripts by TaqMan qPCR (Figure 5B). We found that knocking down these target transcripts caused cell differentiation, as indicated by morphological changes and transient expression of the lineage-specific transcript *T* (mesoderm marker), and expression of *PAX6* (neuronal progenitor marker). After 3 days

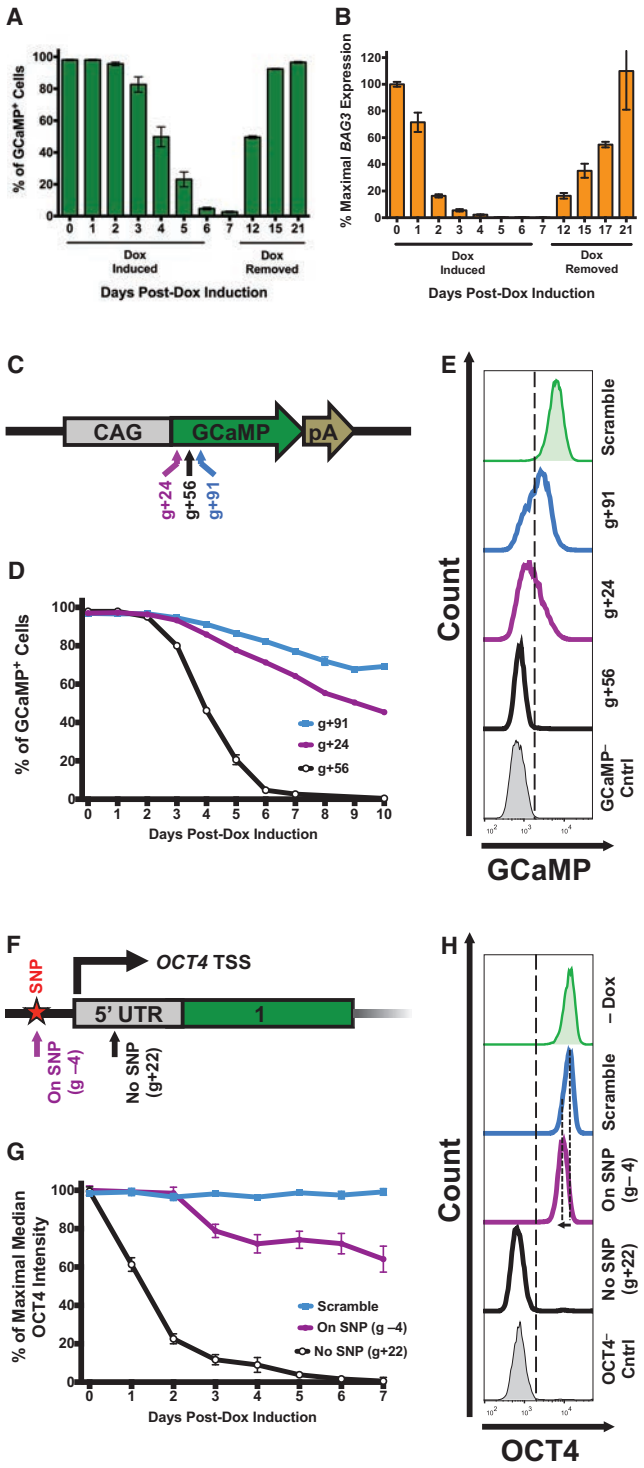


Figure 4. CRISPRi Knockdown Is Reversible and Tunable

A CRISPRi clone containing gRNA against the GCaMP transgene (GCaMP g+56) and endogenous *BAG3* locus were used to test the knockdown efficiency and reversibility of the CRISPRi system in iPSCs.

(A) Flow cytometry analysis of GCaMP expression showed that after 7 days of doxycycline induction, GCaMP was knocked down by ~99% and was completely restored after doxycycline withdrawal for 14 days.

(B) Using TaqMan qPCR, *BAG3* transcript levels were knocked down to nearly undetectable levels, and expression was restored after doxycycline withdrawal.

(C) Schematic diagram of the GCaMP-expression vector in which the GCaMP open reading frame (ORF) is driven off the CAG promoter. The locations of three gRNAs (g+24, g+56, and g+91) are schematically highlighted on the GCaMP ORF. The coordinates of GCaMP gRNA are based on the translation start site. pA, poly A signal.

(D) Three stable CRISPRi colonies, each containing a different gRNA against GCaMP, were selected using blasticidin and cultured with doxycycline for 10 days. The percentage of GCaMP-positive cells for each gRNA-containing clone was plotted as a function of time based on flow cytometry analysis. Variable levels of GCaMP knockdown (~30%, ~50%, and ~99%) were achieved with different gRNA sequences. n = 1–3 technical replicates for each time point.

(E) Flow cytometry plots of GCaMP fluorescence of stable CRISPRi clones on day 10 of doxycycline treatment. Using different gRNAs that target near the same region, variable levels of knockdown can be achieved. A scramble gRNA-containing CRISPRi and a GCaMP-negative iPSC population are displayed as controls.

(F) Partial schematic diagram of the *OCT4* locus marked with the location of the TSS and two gRNA-binding locations. Asterisk, an SNP; green box, exon 1; gray box, 5′ UTR.

(G) Three stable CRISPRi colonies, two with different gRNAs against *OCT4* and one with a scrambled control, were selected with blasticidin. Stable clones that contain either a scramble gRNA, a gRNA that targets a PAM sequence containing a SNP (*OCT4* g–4), or a gRNA that does not target a SNP (*OCT4* g+22) were treated with doxycycline. The percentage of the maximal median intensity of OCT4 staining for each gRNA-containing clone is plotted as a function of time by flow cytometry analysis. Complete loss of *OCT4* expression (>98% knockdown) was observed after 7 days of doxycycline induction only when both alleles were targeted using *OCT4* g+22. While using *OCT4* g–4, which targets only one *OCT4* allele (due to SNP in the PAM sequence), a gradual loss of OCT4 staining intensity is observed over time (down by ~40% by day 7). Error bars represent SD; n = 1–3 technical replicates for each time point.

(H) Flow cytometry plots of OCT4 staining on day 7 of doxycycline treatment. Dashed lines highlight the loss of OCT4-staining intensity (~40%) when using *OCT4* g–4 compared to the scramble control. By targeting only one allele of *OCT4*, the OCT4-staining intensity homogeneously shifts (while remaining OCT4-positive), indicating that each cell experiences approximately the same level of knockdown. Note that the x axis is a log-scale of OCT4 intensity. Differentiated iPSC-derived fibroblasts (OCT4$^-$ Cntrl) and a non-doxycycline-treated (–Dox) sample are displayed as controls.

Error bars represent SD.

of doxycycline treatment, over 80% of the target transcript was depleted, indicating that CRISPRi can precisely and temporally control efficient knockdown of the transcript of interest.

CRISPRi Knockdown in Cardiac Mesoderm and iPS-CMs

To determine if loss-of-function approaches using CRISPRi can be applied in differentiated cell types, we targeted the cardiac mesoderm-specific transcription factor (*MESP1*) and two known cardiac-related disease-causing genes (*MYBPC3* and *HERG*). We established stable polyclonal lines of iPSCs containing gRNA against these three genes and differentiated them into cardiac mesoderm or iPS-CMs as described in Experimental Procedures (Figures S6A and S6B). Using a gRNA against these genes, *MESP1* was knocked down by ~90% in cardiac progenitor cells, and *MYBPC3* and *HERG* by ~90% and 60%, respectively, in lactate-purified iPS-CMs (Figure 6A). With western blots and immunocytochemistry, we observed ~90% MYBPC3 protein knockdown on day-35 lactate-purified iPS-CMs (Figures 6B and 6C).

Using flow cytometry, we analyzed the doxycycline response of CRISPRi cells based on mCherry expression (as a surrogate

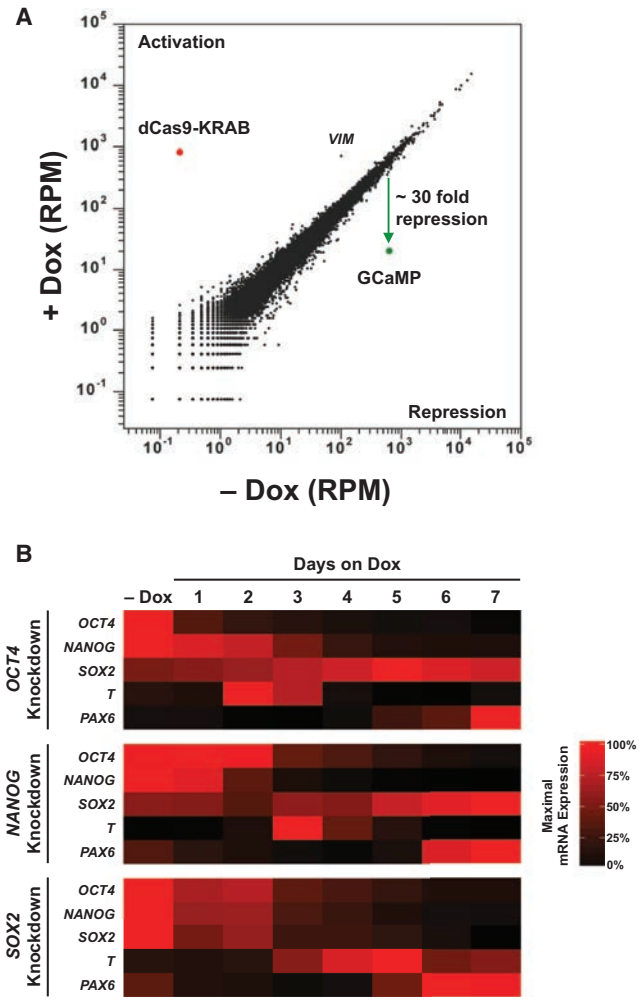


Figure 5. RNA-Seq and TaqMan qPCR Analysis

(A) RNA-sequencing RPMs (reads per million) are plotted for CRISPRi cells stably expressing a gRNA targeting the GCaMP transgene (GCaMP g+56) cultured in the absence or presence of doxycycline. CRISPRi knockdown is specific to the GCaMP transcript, and few off-target transcriptional changes were observed. Data represent two independent biological replicates.

(B) Heatmap of TaqMan qPCR of stable clones containing a single gRNA against the gene of interest (*OCT4*, *NANOG*, and *SOX2*) as a function of days after doxycycline treatment. Analysis shows that by day 3, over 80% of the target transcript is depleted. Three housekeeping genes (*18S*, *GAPDH*, and *UBC*) were used to measure relative transcript levels. Each data point is an average of two to four technical replicates. TaqMan probes are listed in Supplemental Experimental Procedures.

for dCas9-KRAB expression; Figure S5A). There was no silencing of the TetO promoter in low-passage and high-passage iPSCs, suggesting that long-term culturing (>3 months) does not cause silencing. However, cardiac progenitors (day 5) and iPS-CMs (day 15) lose ~20% and 50%–80% of the doxycycline response, respectively. Prolonging the duration of doxycycline treatment (from 2 to 7 days) and splitting the cells improved doxycycline response (as measured by mCherry expression) in iPS-CMs (Figure S6C). For this reason, we initiated all of our knockdowns on day 5 post-differentiation to obtain the maximum amount of target gene silencing. It is worth noting that with CRISPRi, only minute amounts of the dCas9-KRAB protein are necessary to induce a knockdown. Hence, knockdown might occur even in cells that do not show detectable mCherry expression (Figure S5).

The knockdown of the *HERG* potassium channel in iPSCs was highly efficient (>95%), while in iPS-CMs it was only 60% effective. We hypothesize that the reduction in the efficiency of *HERG* knockdown is partially due to activation of other *HERG* isoforms in iPS-CMs. We further investigated whether knocking down the *HERG* potassium channel in iPS-CMs would recapitulate a physiologically relevant cellular phenotype. We found that knocking down *HERG* in iPS-CMs lead to a prolonged beat duration and the appearance of a shoulder during the downstroke, as measured using the GCaMP signal (which can be used as a surrogate for the action potential) (Huebsch et al., 2015) (Figures 6D and 6E). We confirmed the prolongation of action potential duration by patch-clamp electrophysiology in the *HERG* knockdown samples (Figures 6F). We expected this result, because the *HERG* potassium channel pumps potassium ions out of cells to lower the inner membrane potential during diastole. This cellular phenotype recapitulates aspects of the phenotype observed in LQT patients and their iPS-CMs (Schwartz et al., 2012; Spencer et al., 2014).

DISCUSSION

In this study, we combined the power of human iPSC technology, which generates functional human cells, with inducible CRISPR-based genome editing and modulation technologies. Using the TetO inducible system, we deploy the newly developed CRISPRi system in the AAVS1 safe–harbor locus of human iPSCs to enable precise control of transcript silencing upon addition of doxycycline. With this approach, we rapidly and efficiently generated loss-of-function phenotypes in iPSCs and their cell-type derivatives to study mechanisms in development and disease. We introduced a single doxycycline-inducible vector system into the AAVS1 safe-harbor locus to gain tight transcriptional control of dCas9-KRAB (for CRISPRi) and Cas9 (for CRISPRn) for gene knockdown and knockout studies, respectively. This inducible vector system helped us precisely control the timing of knocking down the expression of target genes in a clonal iPSC line carrying the gRNA of interest. We were also able to efficiently target the CRISPRi vector into non-iPSC human cells (T-lymphocytes) and show efficient levels of transgene knockdown, which demonstrates the versatility of using the CRISPRi system in a wide range of cell types. This system can be readily targeted to other human cellular models in vitro and also to mouse models (Soriano, 1999) by exchanging the AAVS1-homology arms with the ROSA26-specific knockin arms.

We found that in iPSC populations, CRISPRi produced a homogeneous and rapid loss-of-function phenotype compared to CRISPRn. CRISPRi avoids potential complications associated with incomplete loss-of-function and gain-of-function phenotypes in cell populations produced by Cas9-induced hypomorphic alleles. Therefore, CRISPRi represents a powerful technology for repressing gene expression in bulk populations and especially when performing genome-scale phenotypic screens. Every CRISPRi iPSC that contained a target-specific gRNA

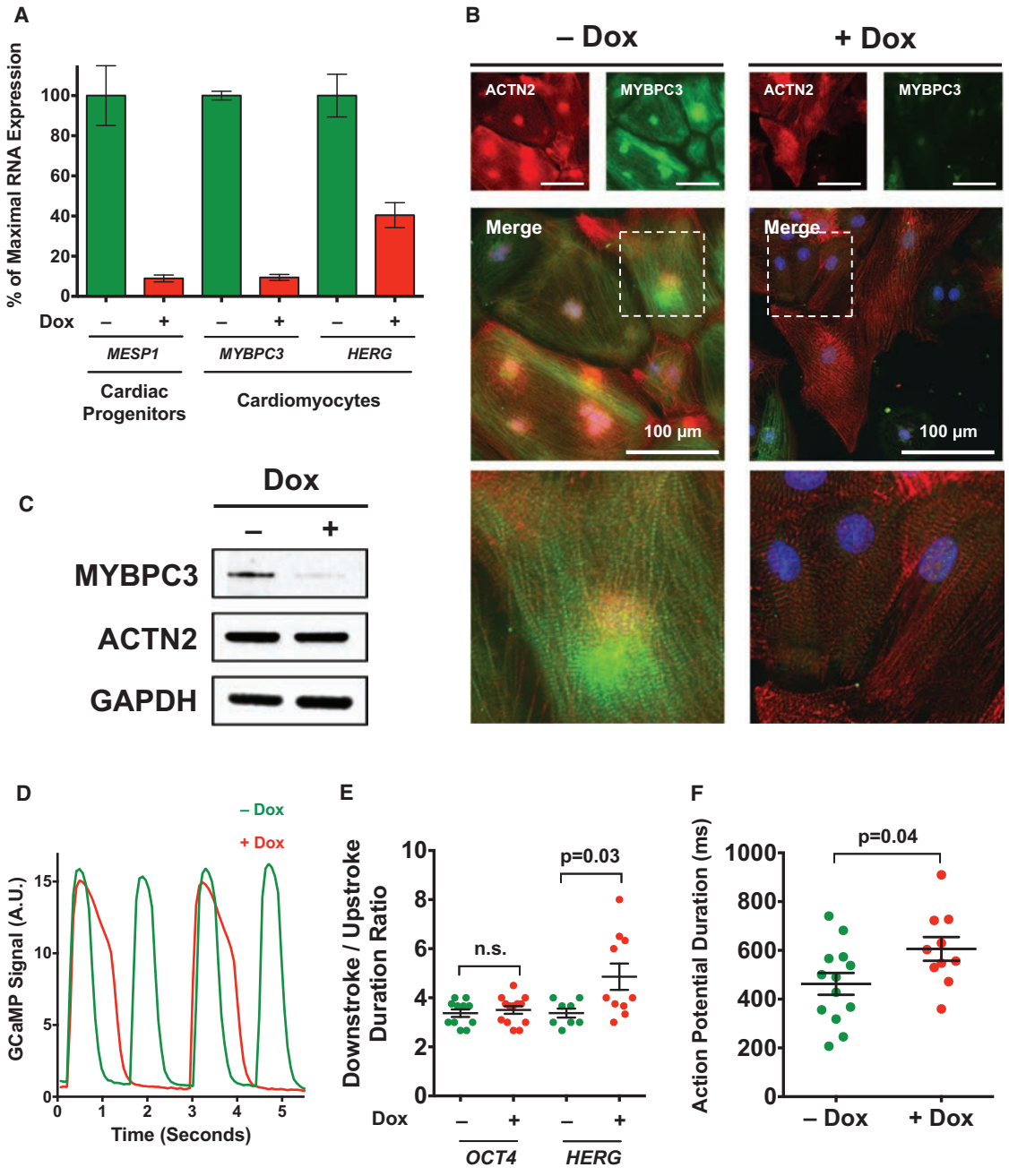


Figure 6. CRISPRi Knockdown in Differentiated Cell Types and Cardiac Disease Modeling

(A) Using CRISPRi, *MESP1* was knocked down by ~90% in polyclonal cardiac progenitors, and *MYBPC3* and *HERG* were knocked down by ~90% and 60% in polyclonal iPS-CMs, respectively.

(B) Immunostaining of day-35 lactate-purified iPS-CMs stained with antibodies against MYBPC3 (green) and ACTN2 (red). Using CRISPRi knockdown, loss of MYBPC3 was observed in over 85% of analyzed cells in a polyclonal population. Nuclei were counterstained with DAPI. Scale bar, 100 μm.

(C) Western blot of day-35 lactate-purified iPS-CMs with antibodies against MYBPC3, ACTN2, and GAPDH. Using CRISPRi, MYBPC3 protein was knocked down by ~90%.

(D) GCaMP fluorescence in iPS-CMs containing gRNA against *HERG* and cultured in doxycycline (red). Recordings show a prolonged beat duration compared to untreated controls (green).

(E) Quantified ratio of the downstroke-to-upstroke duration of doxycycline-treated iPS-CMs shows a significant difference in untreated iPS-CMs containing a gRNA against *HERG*, but not in iPS-CMs containing gRNA against *OCT4* (negative control).

(F) Patch-clamp recordings from single iPS-CMs show prolonged action potential durations in doxycycline-treated samples containing *HERG* gRNA.

Error bars represent SD.

displayed a rapid, uniform, and efficient transcriptional knockdown. This result was also validated across multiple endogenous loci in iPSCs, cardiac progenitors, and iPS-CMs. By contrast, using CRISPRn, we found that while all cells harbored the gRNA-expression vector and had continuous expression of Cas9, they did not all display complete loss-of-function phenotypes. Indeed, up to one-third of the cells maintained expression of the target gene. When we sequenced the target alleles, we found that of the mutated alleles, over one-third had in-frame INDELs, potentially resulting in a hypomorphic protein encoded by a gene that is now resistant to further Cas9 cutting using the target gRNA. Statistically, we expect that one-third of the INDELs generated by double-strand breaks induced by Cas9 through the non-homologous end-joining pathway would produce in-frame mutations. This effect could cause partial loss-of-function or gain-of-function phenotypes. Additionally, the location and size of the in-frame INDEL might not change the function of the mutated protein compared with the wild-type protein (Boettcher and McManus, 2015; Shi et al., 2015; Sung et al., 2013).

CRISPRi gRNAs were only effective at promoter regions close to the TSS, which may reduce the likelihood of off-target effects by transcriptional interference elsewhere in the genome. Indeed, RNA-seq analysis showed that the knockdown of GCaMP was highly specific. Furthermore, expression of dCas9-KRAB did not cause significant off-target transcriptional changes as compared to Cas9 expression alone. Although CRISPRi is highly effective, there are cases when other genetic tools such as CRISPRn, TALENs, and RNAi may have advantages. For instance, we and others (Gilbert et al., 2014) have shown that CRISPRi gRNAs are only effective near the TSS, which restricts the efficiency of transcript for genes that have poorly defined or multiple TSSs. CRISPRn and TALENs can be effective at any exon as long as the genomic region is accessible (Doench et al., 2014; Kim et al., 2013b). Additionally, RNAi can target any constitutive portion of the mRNA and has already been approved for human therapy (Davidson and McCray, 2011; Haussecker, 2012); however, RNAi has been shown to have many off-target effects (Jackson et al., 2003; Kim et al., 2013b; Krueger et al., 2007).

We also demonstrated the feasibility of allele-specific interference and the tunable nature of CRISPRi-based knockdown, which can be used to study the dose-dependent effects of a gene involved in development and disease. The dosage of transcription factors plays a significant role during development and organogenesis (McFadden et al., 2005; Takeuchi et al., 2011). In addition, many human diseases result from haploinsufficiency in which a mutation in a single copy of a gene produces the disease phenotype (Armanios et al., 2005; Marston et al., 2012; Minami et al., 2014; Theodoris et al., 2015). Therefore, to study the dose-dependent effects of transcription factors in development and disease, CRISPRi can be used to homogeneously tune the level of repression in cells by either choosing the relevant gRNA sequences or empirically titrating the levels of doxycycline to achieve the desired knockdown level. Alternatively, introducing a single point mutation at different positions in the gRNA sequence (which leads to mismatches between the RNA-DNA homology sequence) can be used to tune CRISPRi knockdown activity (Gilbert et al., 2014). Finally, CRISPRi knockdown was reversible in iPSCs upon doxycycline withdrawal, which would support studies involving transient knockdown of transcripts within a specific window during cell differentiation.

Our studies with CRISPRi in iPSCs showed that knocking down transcripts involved in maintaining pluripotency is highly efficient and rapidly causes a complete loss of pluripotent morphology, followed by cell differentiation in all cells expressing the appropriate gRNA. We also used this approach to knock down the *HERG* potassium channel to mimic an LQT2-type phenotype in iPS-CMs. We found that the inducible TetO promoter is partially silenced during the cardiac differentiation process, which has been reported to be due to methylation at CpG dinucleotides (Oyer et al., 2009). This silencing is independent of integration at the AAVS1 locus, as CAG-driven transgenes integrated at the AAVS1 locus remain active after differentiation. To avoid the effects of promoter silencing, we initiated transcript knockdown in the iPSC state or progenitor cells (day 5 of differentiation), where the vast majority of the cells respond to doxycycline. This strategy has proved highly effective at transgene knockdown in cardiac progenitors and iPS-CMs. To circumvent issues with silencing in future studies, we generated a non-inducible CRISPRi iPSC line (Gen3; in which dCas9-KRAB is driven off the CAG promoter), and the knockdown can be initiated upon introduction of gRNA. With this cell line, we expect to achieve highly efficient knockdown in differentiated cell types, such as iPS-CMs.

Several groups have used the CRISPR/Cas9 system for loss-of-function genetic screens in human cells (Shalem et al., 2014; Wang et al., 2014). Furthermore, some groups have used genome-scale screens with CRISPRi and CRISPR activation (CRISPRa) to identify known and novel genes that control cell growth and sensitivity to cholera-diphtheria toxin (Gilbert et al., 2014). In this study, we present our CRISPRi iPSC lines as suitable model systems for performing screens to identify novel transcripts of pluripotency, drug resistance, and cell survival at the pluripotent stem cell stage. With genome-scale screens, we can identify factors that improve cell-specific differentiation into functional cell types that have been traditionally hard to obtain, and we can more rapidly generate mature functional cell types that better mimic in vivo cell counterparts. In addition, with CRISPRi, we can repress putative disease-associated genes in a medium- to high-throughput manner to unravel the molecular mechanisms underlying human disease in vitro. Finally, we can build on the current power of CRISPRi for developmental screens by using an orthogonal dCas9-effector system for gene activation via CRISPRa, which can synergistically modulate gene knockdown and activation and direct cell fate toward a particular lineage.

EXPERIMENTAL PROCEDURES

iPSC Culture

WTB and WTC iPSCs and derivative lines were maintained under feeder-free conditions on growth factor-reduced Matrigel (BD Biosciences) and fed daily with mTeSR medium (STEMCELL Technologies) (Ludwig et al., 2006). Accutase (STEMCELL Technologies) was used to enzymatically dissociate iPSCs into single cells. To promote cell survival during enzymatic passaging, cells were passaged with the p160-Rho-associated coiled-coil kinase (ROCK) inhibitor Y-27632 (10 μM; Selleckchem) (Watanabe et al., 2007). iPSCs were frozen in 90% fetal bovine serum (HyClone) and 10%

DMSO (Sigma). The committee on Human Research at the University of California, San Francisco approved the iPSC research protocol (#10-02521).

Generation of Stable CRISPRi and CRISPRn iPSC Lines

iPSCs were singularized with accutase, resuspended in PBS, and counted with a Countess automated cell counter (Life Technologies). For plasmid transfections, the human stem cell nucleofector kit 1 solution was used on the Amaxa nucleofector 2b device (program A-23; Lonza). To generate the CRISPRi and CRISPRn iPSC lines, two million WTC or WTB iPSCs were nucleofected with the appropriate knockin vector (5 μg) and each AAVS1 TALEN pair (2 μg). Cells were then seeded in six-well plates in serial dilutions in mTeSR supplemented with Y-27632 (10 μM). Selection was applied 3 days post-nucleofection with the appropriate antibiotic in mTeSR plus Y-27632 (10 μM). To knock in the CRISPRi construct (carrying the Neomycin resistance gene cassette), Geneticin (Life Technologies) was applied at 100 μg/ml. To knock in the CRISPRn and GCaMP constructs (carrying the Puromycin resistance gene cassette), 0.5 μg/ml Puromycin (Life Technologies) was added. Selection was maintained for ~10 days until stable colonies appeared. Colonies with a diameter of greater than ~500 μm were manually picked using a P200 pipette tip under an EVOS FL picking microscope (Life Technologies) and transferred to individual wells of a 24-well plate containing mTeSR medium supplemented with Y-27632 (10 μM). Clones were then expanded into larger vessel formats.

Generation of CEM CRISPRi Cell Line

CEM CRISPRi cells were generated by electroporation of 0.5 μg of each AAVS1 TALEN pair and 1 μg of the Gen1 CRISPRi vector with an Amaxa nucleofector 2b device and Amaxa cell line nucleofector kit C (Lonza). Cells were selected in 1 μg/μl G418, and clonal lines were generated by dilution in 96-well plates. Clonal populations were selected based on doxycycline induction of mCherry expression. Oligos encoding the CD4 protospacer were annealed and cloned into the pSLQ1371 vector using restriction sites BstXI and BlpI, and lentivirus was produced in HEK293T cells (Gilbert et al., 2014). To compare performance of CD4 gRNAs, each was transduced into CEM-CRISPRi cells. Transduced populations were incubated for 96 hr with doxycycline (2 μM). Knockdown efficiency was calculated by gating all mCherry-expressing cells, and comparing cell-surface CD4 expression in the presence or absence of gRNA-expressing cells (BFP^+). Three independent stable CEM CRISPRi clones were selected with 0.6 μg/ml Puromycin and incubated in the presence or absence of doxycycline (2 μM) for 14 days to assess maximal CD4 knockdown. Cells were stained using anti-CD4 APC-conjugated antibody and cell surface CD4 staining was quantified using a BD LSRII flow cytometer. CD4 knockdown was quantified as percent reduction relative to no doxycycline treatment condition.

gRNA Design and Cloning into the gRNA-Expression Vector

For CRISPRi, three to five gRNAs were designed to target near the TSS of the gene of interest (250 bp upstream and downstream, respectively). The location of the TSS was determined using NCBI (http://www.ncbi.nlm.nih.gov/). gRNA oligos were designed, phosphorylated, annealed, and cloned into the pgRNA-CKB vector using BsmBI ligation strategy. Additional details and a list of gRNA sequences are listed in supplemental experimental procedures.

gRNA Nucleofection and Selection of Stable CRISPRi and CRISPRn Clones

The gRNA-expression vector (pgRNA-CKB) was transfected into either the CRISPRi or CRISPRn cells with the human stem cell nucleofector kit 1 solution on the Amaxa nucleofector 2b device (program A-23; Lonza). Two million CRISPRi or CRISPRn iPSCs and 5 μg of the circular gRNA-expression plasmid were used per nucleofection. Nucleofected cells were then seeded in a single well of a six-well plate in mTeSR supplemented with Y-27632 (10 μM). Blasticidin selection (10 μg/ml) was applied 24 hr post-nucleofection in mTeSR supplemented with Y-27632 (10 μM) for 7–10 days, until stable colonies appeared. Stable colonies were then pooled and passaged at least three times in mTeSR plus Blasticidin and Y-27632 to enrich for cells with integration at transcriptionally active sites (Figure S3).

RNA Sequencing

For each sample, 1 μg of total RNA was prepared using TRIzol as previously described. Strand-specific mRNaseq libraries were prepared using TruSeq Stranded mRNA Library Prep Kit (Illumina). Upon completion, libraries were quantified and pooled using Qubit dsDNA HS assay and Agilent's Bioanalyzer high-sensitivity DNA assay. The indexed libraries were pooled and sequenced on Illumina HiSeq 4000 as 50-bp single-end reads. Reads were aligned to the hg19 genome assembly using the Ensembl 75 reference transcriptome customized to include the GCaMP6f constructs using TopHat2 (Kim et al., 2013a). Unaligned reads were subsequently aligned to the CRISPRi or CRISPRn knockin constructs where appropriate. Transcript alignments were then counted using SubRead v1.4.6 and analyzed with custom scripts written in Python (Liao et al., 2013). All data are displayed as reads per million (RPM) with a pseudocount of 0.075.

iPS-CM Differentiation and Lactate Purification

iPSCs were differentiated into iPS-CMs using the WNT modulation-differentiation method (Lian et al., 2012) (Figure S5A). iPS-CMs were purified via a modified version of the lactate metabolic-selection method (Tohyama et al., 2013). Additional details are outlined in Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession number for the RNA-seq data reported in this paper is GEO: PRJNA307261.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and one movie and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2016.01.022.

AUTHOR CONTRIBUTIONS

M.A.M. and B.R.C. were primarily responsible for conception, design, and interpretation of the experiments. M.A.M. conducted most experiments with help from N.H., E.F., E.S., A.T., M.P.O., T.V.E., K.H., and L.M.J. Y.M. and A.H.C. generated the CRISPRn Gen1C iPSC line. C.I.S. performed electrophysiology experiments. D.E.G. generated the CEM CRISPRi cell line and provided knockdown analysis. L.A.G., J.S.W., and L.S.Q. provided technical expertise, the CRISPRi fusion cassette, and gRNA expression constructs. J.E.V. and M.A.H. conducted and analyzed the RNA-seq experiments. M.A.M., P.L.S., and B.R.C. wrote the manuscript with support from all authors.

ACKNOWLEDGMENTS

We thank members of the Conklin laboratory, Gladstone Institute of Cardiovascular Disease, Roddenberry Stem Cell Core, BioFulcrum, a Gladstone Institutes Enterprise, and Innovative Genomics Institute for technical assistance and helpful comments on the manuscript. We thank Tim Rand and Knut Woltjen for valuable discussions and helpful comments on the manuscript. We thank S. John Liu for RNA-seq analysis advice. We thank Jen Berman and Samantha Cooper at Bio-Rad for assistance with designing ddPCR probe-primer sets. Summer students Matthew Keller and Monique Morrison assisted with preliminary experiments. CEM $CD4^+$ cells were obtained from Dr. J.P. Jacobs through the AIDS Reagent Program, Division of AIDS, NIAID, NIH. M.A.M. is supported by the Canadian Institutes of Health Research postdoctoral fellowship 129844. N.H. was supported by the CIRM training program TG2-01160 and T32 HL007544. E.F. was supported by a Bridges to Stem Cell Training grant TB1-01188 from CIRM. L.M.J. is supported by the CIRM Training Grant TG2-01160 and NICHD Career Development Award 1K12HD072222. Y.M. received fellowships from the Uehara Memorial Foundation Research and Gladstone-CIRM. D.G. is supported by UCSF-Gladstone Center for AIDS Research (CFAR), an NIH-funded program (P30 AI027763). T.V.E. was supported by Carlsberg Travel Grant (2013-01-0423), The Lundbeck Foundation (R140-2013-13348), and OUH Internationalisation Foundation. J.E.V., M.A.H., L.A.G., and J.S.W. were supported by the Howard Hughes

Medical Institute and the National Institutes of Health (R01 DA036858). B.R.C. received support from the US National Heart, Lung, and Blood Institute, the National Institutes of Health (U01-HL100406, U01-GM09614, R01-HL108677, U01-HL098179, U01-HL099997, P01-HL089707, R01HL130533, and R01-HL060664), and an Agilent University Relations Grant.

Received: August 20, 2015
Revised: December 21, 2015
Accepted: January 24, 2016
Published: March 10, 2016

REFERENCES

Apáti, Á., Pászty, K., Hegedűs, L., Kolacsek, O., Orbán, T.I., Erdei, Z., Szebényi, K., Péntek, A., Enyedi, Á., and Sarkadi, B. (2013). Characterization of calcium signals in human embryonic stem cells and in their differentiated offspring by a stably integrated calcium indicator protein. Cell. Signal. *25*, 752–759.

Armanios, M., Chen, J.-L., Chang, Y.-P.C., Brodsky, R.A., Hawkins, A., Griffin, C.A., Eshleman, J.R., Cohen, A.R., Chakravarti, A., Hamosh, A., and Greider, C.W. (2005). Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. Proc. Natl. Acad. Sci. USA *102*, 15960–15964.

Boettcher, M., and McManus, M.T. (2015). Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. Mol. Cell *58*, 575–585.

Bolouri, H., and Davidson, E.H. (2003). Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. Proc. Natl. Acad. Sci. USA *100*, 9371–9376.

Chen, T.-W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature *499*, 295–300.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Davidson, B.L., and McCray, P.B., Jr. (2011). Current prospects for RNA interference-based therapies. Nat. Rev. Genet. *12*, 329–340.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat. Biotechnol. *32*, 1262–1267.

Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096.

Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. *31*, 397–405.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

González, F., Zhu, Z., Shi, Z.-D., Lelli, K., Verma, N., Li, Q.V., and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell Stem Cell *15*, 215–226.

Haussecker, D. (2012). The business of RNAi therapeutics in 2012. Mol. Ther. Nucleic Acids *1*, e8.

Hayashi, Y., Caboni, L., Das, D., Yumoto, F., Clayton, T., Deller, M.C., Nguyen, P., Farr, C.L., Chiu, H.-J., Miller, M.D., et al. (2015). Structure-based discovery of NANOG variant with enhanced properties to promote self-renewal and reprogramming of pluripotent stem cells. Proc. Natl. Acad. Sci. USA *112*, 4666–4671.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C., et al. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat. Biotechnol. *29*, 731–734.

Huebsch, N., Loskill, P., Mandegar, M.A., Marks, N.C., Sheehan, A.S., Ma, Z., Mathur, A., Nguyen, T.N., Yoo, J.C., Judge, L.M., et al. (2015). Automated video-based analysis of contractility and calcium flux in human-induced pluripotent stem cell-derived cardiomyocytes cultured over different spatial scales. Tissue Eng. Part C Methods *21*, 467–479.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. (2003). Expression profiling reveals off-target gene regulation by RNAi. Nat. Biotechnol. *21*, 635–637.

Kearns, N.A., Genga, R.M.J., Enuameh, M.S., Garber, M., Wolfe, S.A., and Maehr, R. (2014). Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development *141*, 219–223.

Kim, H., and Kim, J.-S. (2014). A guide to genome engineering with programmable nucleases. Nat. Rev. Genet. *15*, 321–334.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013a). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kim, Y., Kweon, J., Kim, A., Chon, J.K., Yoo, J.Y., Kim, H.J., Kim, S., Lee, C., Jeong, E., Chung, E., et al. (2013b). A library of TAL effector nucleases spanning the human genome. Nat. Biotechnol. *31*, 251–258.

Krueger, U., Bergauer, T., Kaufmann, B., Wolter, I., Pilk, S., Heider-Fabian, M., Kirch, S., Artz-Oppitz, C., Isselhorst, M., and Konrad, J. (2007). Insights into effective RNAi gained from large-scale siRNA validation screening. Oligonucleotides *17*, 237–250.

Lian, X., Hsiao, C., Wilson, G., Zhu, K., Hazeltine, L.B., Azarin, S.M., Raval, K.K., Zhang, J., Kamp, T.J., and Palecek, S.P. (2012). Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. Proc. Natl. Acad. Sci. USA *109*, E1848–E1857.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. *41*, e108.

Lombardo, A., Cesana, D., Genovese, P., Di Stefano, B., Provasi, E., Colombo, D.F., Neri, M., Magnani, Z., Cantore, A., Lo Riso, P., et al. (2011). Site-specific integration and tailoring of cassette design for sustainable gene transfer. Nat. Methods *8*, 861–869.

Ludwig, T.E., Bergendahl, V., Levenstein, M.E., Yu, J., Probasco, M.D., and Thomson, J.A. (2006). Feeder-independent culture of human embryonic stem cells. Nat. Methods *3*, 637–646.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Marston, S., Copeland, O., Gehmlich, K., Schlossarek, S., and Carrier, L. (2012). How do MYBPC3 mutations cause hypertrophic cardiomyopathy? J. Muscle Res. Cell Motil. *33*, 75–80.

McFadden, D.G., Barbosa, A.C., Richardson, J.A., Schneider, M.D., Srivastava, D., and Olson, E.N. (2005). The Hand1 and Hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner. Development *132*, 189–201.

Minami, S.S., Min, S.-W., Krabbe, G., Wang, C., Zhou, Y., Asgarov, R., Li, Y., Martens, L.H., Elia, L.P., Ward, M.E., et al. (2014). Progranulin protects against amyloid β deposition and toxicity in Alzheimer's disease mouse models. Nat. Med. *20*, 1157–1164.

Miyaoka, Y., Chan, A.H., Judge, L.M., Yoo, J., Huang, M., Nguyen, T.D., Lizarraga, P.P., So, P.-L., and Conklin, B.R. (2014). Isolation of single-base genome-edited human iPS cells without antibiotic selection. Nat. Methods *11*, 291–293.

Oyer, J.A., Chu, A., Brar, S., and Turker, M.S. (2009). Aberrant epigenetic silencing is triggered by a transient reduction in gene expression. PLoS ONE *4*, e4832.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Schwartz, P.J., Crotti, L., and Insolia, R. (2012). Long-QT syndrome: from genetics to management. Circ Arrhythm Electrophysiol *5*, 868–877.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87.

Shi, J., Wang, E., Milazzo, J.P., Wang, Z., Kinney, J.B., and Vakoc, C.R. (2015). Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. Nat. Biotechnol. *33*, 661–667.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat. Genet. *21*, 70–71.

Spencer, C.I., Baba, S., Nakamura, K., Hua, E.A., Sears, M.A.F., Fu, C.C., Zhang, J., Balijepalli, S., Tomoda, K., Hayashi, Y., et al. (2014). Calcium transients closely reflect prolonged action potentials in iPSC models of inherited cardiac arrhythmia. Stem Cell Reports *3*, 269–281.

Sterneckert, J.L., Reinhardt, P., and Schöler, H.R. (2014). Investigating human disease using stem cell models. Nat. Rev. Genet. *15*, 625–639.

Sung, Y.H., Baek, I.-J., Kim, D.H., Jeon, J., Lee, J., Lee, K., Jeong, D., Kim, J.-S., and Lee, H.-W. (2013). Knockout mice created by TALEN-mediated gene targeting. Nat. Biotechnol. *31*, 23–24.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Takeuchi, J.K., Lou, X., Alexander, J.M., Sugizaki, H., Delgado-Olguín, P., Holloway, A.K., Mori, A.D., Wylie, J.N., Munson, C., Zhu, Y., et al. (2011). Chromatin remodelling complex dosage modulates transcription factor function in heart development. Nat. Commun. *2*, 187.

Theodoris, C.V., Li, M., White, M.P., Liu, L., He, D., Pollard, K.S., Bruneau, B.G., and Srivastava, D. (2015). Human disease modeling reveals integrated transcriptional and epigenetic mechanisms of NOTCH1 haploinsufficiency. Cell *160*, 1072–1086.

Tohyama, S., Hattori, F., Sano, M., Hishiki, T., Nagahata, Y., Matsuura, T., Hashimoto, H., Suzuki, T., Yamashita, H., Satoh, Y., et al. (2013). Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. Cell Stem Cell *12*, 127–137.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR/Cas9 system. Science *343*, 80–84.

Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. Science *350*, 1096–1101.

Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J.B., Nishikawa, S., Nishikawa, S., Muguruma, K., and Sasai, Y. (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. Nat. Biotechnol. *25*, 681–686.

Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature *482*, 331–338.

CellPress

Cell Stem Cell
Short Article

Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells

Joshua B. Black,[1] Andrew F. Adler,[1,9] Hong-Gang Wang,[6,8,10] Anthony M. D'Ippolito,[2,3] Hunter A. Hutchinson,[1] Timothy E. Reddy,[2,4] Geoffrey S. Pitt,[6,7,8] Kam W. Leong,[1,11] and Charles A. Gersbach[1,2,5,*]
[1]Department of Biomedical Engineering, Duke University, Durham, NC 27708, USA
[2]Center for Genomic and Computational Biology, Duke University, Durham, NC 27708, USA
[3]University Program in Genetics and Genomics, Duke University Medical Center, Durham, NC 27710, USA
[4]Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC 27710, USA
[5]Department of Orthopaedic Surgery, Duke University Medical Center, Durham, NC 27710, USA
[6]Ion Channel Research Unit, Duke University Medical Center, Durham, NC 27710, USA
[7]Department of Neurobiology, Duke University Medical Center, Durham, NC 27710, USA
[8]Division of Cardiology, Department of Medicine, Duke University Medical Center, Durham, NC 27710, USA
[9]Present address: Department of Neurosciences, University of California, San Diego, La Jolla, CA 92093, USA
[10]Present address: Cardiovascular Research Institute, Weill Cornell Medicine, New York, NY 10021, USA
[11]Present address: Department of Biomedical Engineering, Columbia University, New York, NY 10027, USA
*Correspondence: charles.gersbach@duke.edu
http://dx.doi.org/10.1016/j.stem.2016.07.001

SUMMARY

Overexpression of exogenous fate-specifying transcription factors can directly reprogram differentiated somatic cells to target cell types. Here, we show that similar reprogramming can also be achieved through the direct activation of endogenous genes using engineered CRISPR/Cas9-based transcriptional activators. We use this approach to induce activation of the endogenous *Brn2*, *Ascl1*, and *Myt1l* genes (BAM factors) to convert mouse embryonic fibroblasts to induced neuronal cells. This direct activation of endogenous genes rapidly remodeled the epigenetic state of the target loci and induced sustained endogenous gene expression during reprogramming. Thus, transcriptional activation and epigenetic remodeling of endogenous master transcription factors are sufficient for conversion between cell types. The rapid and sustained activation of endogenous genes in their native chromatin context by this approach may facilitate reprogramming with transient methods that avoid genomic integration and provides a new strategy for overcoming epigenetic barriers to cell fate specification.

INTRODUCTION

Direct reprogramming of somatic cells has tremendous potential to advance applications in disease modeling, drug discovery, and gene and cell therapies. Common approaches to achieve cellular reprogramming rely on the ectopic expression of transgenes encoding lineage-specific transcription factors (Davis et al., 1987; Takahashi and Yamanaka, 2006; Vierbuchen et al., 2010). To demonstrate stable cellular reprogramming to an autonomous cell phenotype, the expression of exogenous transcription factors should be transient. Thus the establishment of positive feedback networks regulating endogenous genes is necessary to sustain a transgene-independent cellular identity (Vierbuchen and Wernig, 2011). In many cases, the endogenous genes are occluded by *cis*-acting repressive chromatin marks that are slow to remodel (Vierbuchen and Wernig, 2012). This slow remodeling process typically necessitates prolonged expression of the exogenous factors, limiting the efficacy of transient delivery methods, and poses a major bottleneck to improving the efficiency, speed, and robustness of reprogramming (Hanna et al., 2009).

The type II clustered regularly interspaced short palindromic repeat (CRISPR) system and the CRISPR-associated Cas9 nuclease have recently been repurposed from an adaptive immune system in bacteria and archaea to a gene editing tool (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013b) and transcriptional regulator (Cheng et al., 2013; Gilbert et al., 2013; Konermann et al., 2013; Maeder et al., 2013b; Mali et al., 2013a; Perez-Pinera et al., 2013; Qi et al., 2013) of endogenous genes in mammalian cells. The ability to program these transcription factors to target any genomic locus of interest through the simple exchange of the 20-nt targeting sequence of the guide RNA (gRNA) enables a simple, robust, and highly scalable method for control of complex transcriptional networks (Thakore et al., 2016). Furthermore, dCas9-based transcription factors can target stably silenced genes within compacted chromatin to initiate chromatin remodeling and transcriptional activation (Perez-Pinera et al., 2013; Polstein et al., 2015). Thus, this technology may provide a method to deterministically initiate expression of endogenous gene networks of alternate cell lineages.

The CRISPR/Cas9 system and other platforms for programmable transcriptional regulation have been incorporated into methods for cellular reprogramming in a few recent studies. Gao et al. used transcription activator-like effector (TALE)-based

transactivators targeting an enhancer of *Oct4* to generate mouse induced pluripotent stem cells. Notably, that study required co-delivery of vectors directly encoding ectopic *C-MYC*, *KLF4*, and *SOX2* to achieve pluripotency (Gao et al., 2013). More recently, we have demonstrated the direct conversion of primary mouse embryonic fibroblasts (PMEFs) to skeletal myocytes using a dCas9-based transactivator targeting the endogenous *Myod1* gene (Chakraborty et al., 2014). Several groups have also applied CRISPR/Cas9-based transcriptional regulation to direct the differentiation of human induced pluripotent and embryonic stem cells (Balboa et al., 2015; Chavez et al., 2015; Wei et al., 2016).

The above examples involve the targeted activation of a single transcription factor to guide reprogramming or differentiation, but many approaches require concurrent expression of multiple factors to efficiently establish a mature phenotype (Takahashi and Yamanaka, 2006; Vierbuchen et al., 2010). There have been no examples demonstrating multiplex endogenous gene activation to induce cellular reprogramming, and the versatility of that approach for direct conversion to other cell phenotypes is not known. Moreover, only the report of TALE transcription factors targeting *Oct4* evaluated changes to epigenetic marks at the target loci (Gao et al., 2013), and this group later reported that dCas9-based transcriptional activators were inefficient at endogenous gene activation and reprogramming (Gao et al., 2014). In this study, we tested the hypothesis that targeted epigenetic reprogramming of the regulatory elements controlling expression of lineage-specific transcription factors is sufficient for direct conversion between cell types by applying dCas9-based transactivators to the activation of endogenous genes that directly convert PMEFs to induced neuronal cells (iNs).

RESULTS

Multiplex Endogenous Gene Activation of Neurogenic Factors in PMEFs

Overexpression of transgenes encoding the transcription factors *Brn2, Ascl1*, and *Myt1l* (BAM factors) has been shown to directly convert cultured PMEFs to functional induced neuronal cells (Vierbuchen et al., 2010). We hypothesized that the targeted activation of the endogenous genes encoding these same factors in their native chromatin context via a dCas9-based transactivator could more rapidly and deterministically remodel the chromatin at the target loci and provide an alternate method to achieve the reprogramming of PMEFs to iNs (Figure 1A). To achieve targeted gene activation, we used a transactivator with both N-terminal and C-terminal VP64 transactivation domains (^{VP64}dCas9^{VP64}) (Chakraborty et al., 2014) that generated a ~10-fold improvement in activation of *ASCL1* in HEK293T cells at 3 days post-transfection compared to the first-generation dCas9 transcription factor with a single C-terminal VP64 domain (Maeder et al., 2013b; Perez-Pinera et al., 2013) (Figure 1B). We used ^{VP64}dCas9^{VP64} for the remainder of this study.

We used lentiviral delivery to constitutively express ^{VP64}dCas9^{VP64} in PMEFs. Initially, we delivered the gRNAs through transient transfection of plasmid DNA in order to assess stable reprogramming of cell phenotype following transient activity of transactivators. The induction of *Brn2* and *Ascl1* gene expression by ^{VP64}dCas9^{VP64} was attained by delivering four gRNAs targeted to the putative promoter region directly upstream of the transcription start site (TSS). The decision to deliver four gRNAs for each gene was based on the reported synergistic effects of multiple gRNAs on gene activation (Maeder et al., 2013b; Mali et al., 2013a; Perez-Pinera et al., 2013). The optimal gRNAs were selected from a pool of eight gRNAs through elimination screening (Figure S1A). The gRNAs targeting regions proximal to the TSS of the *Myt1l* locus did not induce detectable levels of activation, but targeting an intronic region directly upstream of the first coding exon of *Myt1l* was sufficient to activate expression (Figure S1B).

Co-transfection of 12 gRNA expression plasmids (CR-BAM), targeting each of the three endogenous BAM factors with 4 gRNAs, into PMEFs stably expressing ^{VP64}dCas9^{VP64} was sufficient to induce transcriptional upregulation of all three endogenous genes when compared to the transfection of a plasmid encoding firefly luciferase (pLuc; Figure 1C). We also detected Brn2 and Ascl1 protein expression by western blot (Figure S1C), although we could not detect Myt1l protein using commercially available antibodies. In addition to gRNA transfections, we transfected three plasmids encoding the BAM factor transgenes under the control of the EF1α/HTLV promoter (pBAM) into the same cells and observed a modest increase in the mRNA levels of the corresponding endogenous genes (Figure 1C).

To attain successful reprogramming, it is generally considered necessary to express the exogenous factors at high levels (Vierbuchen and Wernig, 2011). Therefore, we compared the total mRNA and protein levels of *Brn2*, *Ascl1*, and *Myt1l* produced 3 days after CR-BAM and pBAM plasmid transfections (Figures 1D–1F). Despite the higher levels of transcriptional activation from the endogenous loci by CR-BAM (Figure 1C), pBAM transfection generated significantly more total mRNA encoding each BAM factor than induction by CR-BAM, as determined by qRT-PCR (Figure 1D). Quantitation of single-cell protein levels from immunofluorescence staining also revealed significantly higher single-cell levels of Brn2 and Ascl1 in cells transfected with pBAM compared to those transfected with CR-BAM (Figures 1E and 1F).

Induction of Neuronal Cells from PMEFs via ^{VP64}dCas9^{VP64}-Mediated Gene Activation

Treated PMEFs were assayed for neuronal phenotypes as detailed schematically in Figure 2A. We observed an increase in mRNA of the early pan-neuronal marker βIII tubulin (*Tuj1*) 3 days after transfection with either pBAM or CR-BAM when compared to a pLuc control (Figure 2B). We cultured the cells for 2 weeks in neurogenic medium and analyzed expression of pan-neuronal markers by immunofluorescence staining. We identified cells with neuronal morphologies that expressed Tuj1 in populations transfected with CR-BAM (Figure 2C). A subset of Tuj1$^+$ cells also expressed the more mature pan-neuronal marker Map2 (Figure 2C). The generation of Tuj1$^+$ Map2$^+$ cells with neuronal morphologies following treatment with ^{VP64}dCas9^{VP64} and gRNAs was contingent on the addition of a small-molecule cocktail to the medium that has been used previously for neural differentiation of embryonic stem cells and has been shown to improve the efficiency of the direct conversion of human fibroblasts to neurons when used in parallel

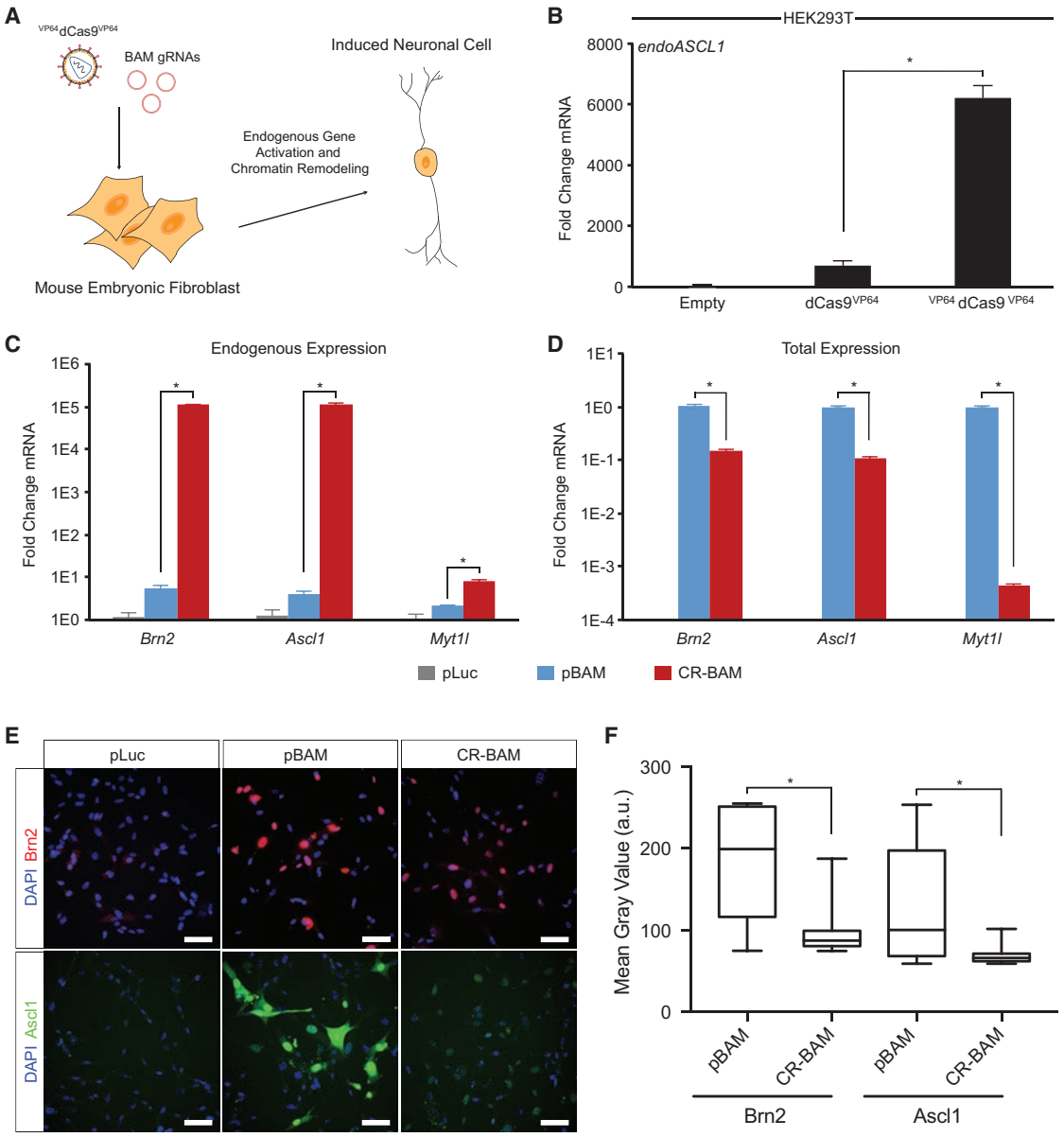


Figure 1. Endogenous Gene Activation of Neuronal Transcription Factors in PMEFs

(A) Reprogramming of PMEFs to neuronal cells via transduction of ^{VP64}dCas9^{VP64} and transfection of gRNA expression plasmids targeting the endogenous BAM factors.

(B) Transcriptional activation of *ASCL1* in HEK293T cells with dCas9^{VP64} or ^{VP64}dCas9^{VP64} (*$p < 0.05$).

(C and D) Endogenous expression (C) and total expression (D) of the BAM factors in PMEFs with targeted activation (CR-BAM) or ectopic overexpression (pBAM; *$p < 0.05$).

(E) Immunofluorescence staining of Brn2 and Ascl1 in PMEFs demonstrated protein expression through targeted activation of the endogenous loci or expression from ectopic plasmids (scale bar, 50 μm).

(F) Automated image analysis of fluorescence intensity revealed significantly more single-cell Brn2 and Ascl1 protein with pBAM transfection compared to CR-BAM (*$p < 0.05$ between distributions of single-cell mean fluorescence; Z-test).

All gRNAs used are listed in Table S1. All assays were performed on day 3 post-transfection. qRT-PCR data are presented as mean ± SEM for n = 3 biological replicates. p values for qRT-PCR data were determined by global one-way ANOVA with Holm-Bonferroni post hoc tests ($\alpha = 0.05$). See also Figure S1.

with ectopic expression of neural transcription factors (Ladewig et al., 2012).

We used a lentiviral fluorescent reporter encoding dsRed-Express under the control of the synapsin I promoter (Syn-RFP) as a proxy to define the most functionally mature iNs in the heterogeneous population of reprogrammed cells (Adler et al., 2012). We readily identified RFP$^+$ cells with elaborate arborizations in CR-BAM-transfected PMEFs (Figure 2C). We also identified rare cells with fibroblastic morphologies reactive to the Tuj1 antibody in PMEFs following pLuc transfection

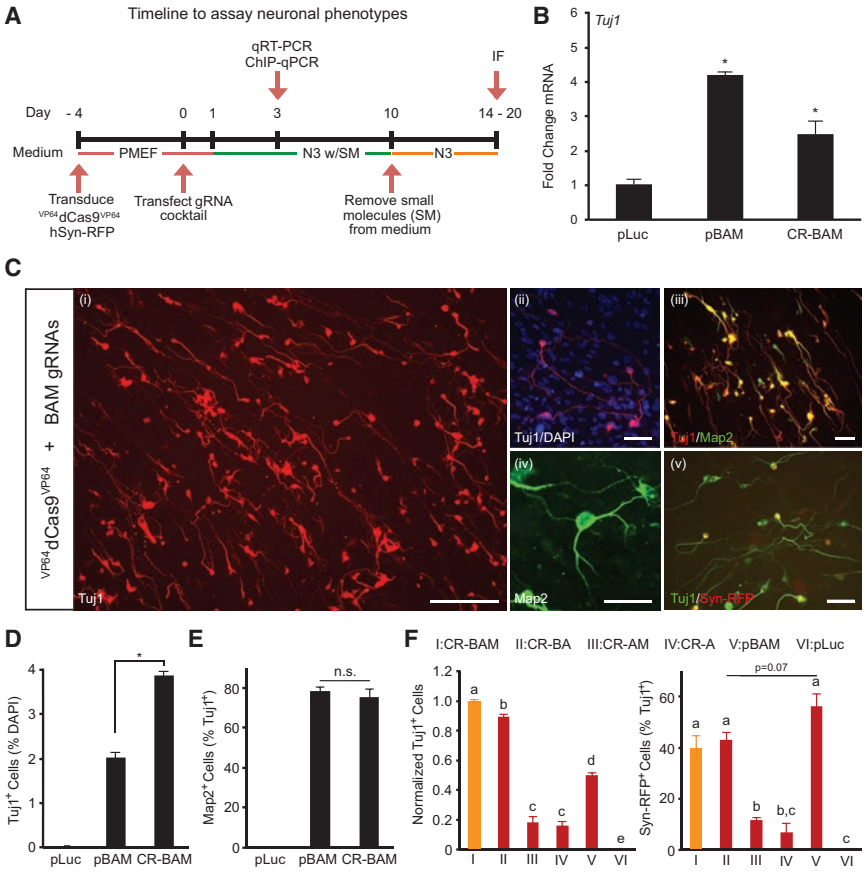


Figure 2. Induction of Neuronal Cells from PMEFs via ^{VP64}dCas9^{VP64}-Mediated Gene Activation

(A) PMEFs were transduced with a lentivirus encoding the ^{VP64}dCas9^{VP64} transactivator and subsequently transfected with gRNAs targeting *Brn2*, *Ascl1*, and *Myt1l*. Neuronal phenotypes were assayed as indicated.

(B) Transcriptional activation of *Tuj1* was detected in PMEFs at day 3 post-transfection of pBAM or CR-BAM (*$p < 0.05$ relative to transfection of a plasmid encoding firefly luciferase [pLuc]).

(C) Immunofluorescence staining revealed numerous Tuj1$^+$ cells with neuronal morphologies co-expressing Map2 at day 14 post-transfection of CR-BAM. The cells with the most elaborate neuronal morphologies activated the synapsin promoter in a Syn-RFP lentiviral reporter (scale bars, 100 μm [i], 50 μm [ii–v]).

(D) Quantitation of Tuj1$^+$ cells as percent nuclei at day 14 post-transfection of pLuc, pBAM, or CR-BAM (*$p < 0.05$).

(E) Quantitation of Map2$^+$ cells as percent Tuj1$^+$ cells at day 14 post-transfection of pLuc, pBAM, or CR-BAM (n.s., not significant).

(F) Quantitation of Tuj1$^+$ and RFP$^+$ cells with transfection of different combinations of gRNAs. Tuj1$^+$ cells are normalized to CR-BAM transfection. Conditions that share the same letter (a–e) are not significantly different.

p values were determined by global one-way ANOVA with Holm-Bonferroni post hoc tests (α = 0.05). See also Figure S2.

(Figure S2A), but these cells were never reactive to the Map2 antibody. Consistent with previous studies, direct overexpression of the ectopic BAM factors via transfection of constitutive expression plasmids generated Tuj1$^+$Map2$^+$ cells with neuronal morphologies (Figure S2B) (Adler et al., 2012; Vierbuchen et al., 2010).

Image analysis revealed that CR-BAM transfection generated a modest, but statistically significant and reproducible, increase in the number of Tuj1$^+$ cells compared to pBAM transfection after 14 days in culture post-transfection (Figure 2D), despite much lower overall expression of the BAM factors (Figures 1D–1F). There was no difference in the percentage of Tuj1$^+$ cells that also expressed Map2 (Figure 2E). To evaluate the contribution of each neurogenic factor to the generation of Tuj1$^+$ cells and to the level of neuronal maturation, we transfected gRNAs targeting different combinations of the endogenous factors. Removal of gRNAs targeting the *Brn2* locus attenuated iN production ~5-fold when compared to that generated with targeted activation of all three endogenous factors (Figure 2F). We detected a slight reduction in Tuj1$^+$ cell production with the removal of *Myt1l* gRNAs (Figure 2F). Neuronal maturity was assessed as the percentage of Tuj1$^+$ cells co-positive for the Syn-RFP reporter. Removal of *Brn2* gRNAs reduced the percentage of RFP$^+$ cells >2-fold, but no change was detected with removal of *Myt1l* gRNAs (Figure 2F). pBAM transfection generated a higher percentage of RFP$^+$ cells than CR-BAM transfection, though it was not statistically significant (Figure 2F).

Induction of Endogenous Gene Expression Is Rapid and Sustained

For any reprogramming strategy, activation of the endogenous genes encoding the master fate-specifying transcription factors is an important step to the successful reprogramming and stability of the new cellular phenotype (Vierbuchen and Wernig, 2011). Consequently, we compared the kinetics of endogenous gene expression through late stages of reprogramming with pBAM or CR-BAM transfection. We observed activation of all three endogenous genes as early as 1 day post-transfection with CR-BAM that remained at high levels through day 18 in culture (Figure S3A). Expression of the BAM factors from the endogenous loci was significantly higher with targeted activation via CR-BAM compared to ectopic overexpression via pBAM transfection throughout the time course of the experiment. Activation of the endogenous genes by pBAM transfection was delayed, and a significant and sustained increase over baseline levels was only detected for endogenous *Ascl1* and *Myt1l* (Figure S3A).

We next assessed the kinetics of expression of the downstream pan-neuronal marker *Tuj1*. Both pBAM and CR-BAM treatment generated a significant increase in *Tuj1* expression throughout the time course of the experiment (Figure S3B). At early time points, *Tuj1* levels were higher with pBAM treatment than CR-BAM. However, *Tuj1* levels with pBAM treatment peaked 7 days post-transfection and declined thereafter, whereas expression following CR-BAM treatment remained

stable through day 18 in culture (Figure S3B). Importantly, the exogenous BAM factors and gRNAs were significantly depleted by day 18 in culture after transient transfection (Figure S3C), though levels of activation from the endogenous genes remained high in cells treated with CR-BAM (Figure S3A).

Direct Activation via ^{VP64}dCas9^{VP64} Rapidly Remodels Chromatin at Target Loci

The kinetics of gene activation led us to speculate whether the rapid and sustained elevated levels of endogenous gene expression achieved with CR-BAM corresponded to an altered epigenetic program at the target loci. We used chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) data generated as part of the Encyclopedia of DNA Elements (ENCODE) Project (Mouse ENCODE Consortium, 2012) to identify histone modifications enriched at the transcriptionally active BAM factor loci in mouse embryonic brain tissue, including H3K27ac and H3K4me3 (Figures 3A, 3C, and S4A). We hypothesized that targeting the endogenous BAM factors for activation with ^{VP64}dCas9^{VP64} in PMEFs could recapitulate the chromatin signatures found at these loci in developing brain tissue.

To investigate the effects of BAM-factor induction on the epigenetic programming at the target loci, we performed chromatin immunoprecipitation (ChIP) qPCR in PMEFs transduced with ^{VP64}dCas9^{VP64} and transfected with pLuc, pBAM, or CR-BAM plasmids (Figures 3 and S4). We used qPCR primers tiled along intragenic and regulatory regions of the *Brn2, Ascl1,* and *Myt1l* loci. We detected a significant enrichment in H3K27ac and H3K4me3 at the *Brn2* and *Ascl1* loci on day 3 post-transfection of CR-BAM (Figures 3B and 3D). H3K4me3 was enriched along the gene bodies of *Brn2* and *Ascl1*. H3K27ac was enriched along the gene bodies and regions surrounding the putative promoter sequences of both genes. In contrast, targeted activation of *Myt1l* only induced modest detectable enrichment in H3K27ac at the gRNA target sites directly upstream of the first coding exon (Figure S4B). No significant change in H3K27ac or H3K4me3 was measured within the putative *Myt1l* promoter. Though overexpression of the BAM factors induced modest levels of expression of the endogenous genes by day 3 post-transfection (Figures 1C and S3A), we did not detect corresponding enrichment in H3K27ac and H3K4me3 at the endogenous loci (Figures 3B, 3D, and S4B).

Generation of Induced Neuronal Cells with Multiplex gRNA Lentiviral Vectors

To explore a strategy for stable expression of the CRISPR/Cas9 transcription factors, and to see if the same outcomes observed with transient expression held true with constitutive expression, we used a single lentiviral vector capable of expressing four gRNAs from four independent RNA polymerase III promoters (Kabadi et al., 2014) (Figure 4A). Co-transduction of lentiviruses encoding ^{VP64}dCas9^{VP64} and a set of four gRNAs targeting each of the three BAM factors (lentiCR-BAM) permitted concurrent activation of the endogenous BAM factors in PMEFs by day 6 post-transduction (Figure 4B). For comparison, we used lentiviral vectors directly encoding the BAM factors (lentiBAM), and demonstrated activation of the corresponding endogenous genes by day 6 post-transduction (Figure 4B). Similar to the results we obtained with transient transfection of expression plasmids, targeted activation of the endogenous genes via lentiviral delivery generated significantly more endogenous transcript from the *Brn2* and *Ascl1* loci than that induced through ectopic expression of the BAM factors. However, unlike the transfection experiments, endogenous *Myt1l* expression was significantly higher with transduction of lentiBAM compared to lentiCR-BAM (Figure 4B).

Following extended culture for 2 weeks in neurogenic medium, we readily identified Tuj1$^+$Map2$^+$ cells with complex neuronal morphologies (Figure 4C). All Tuj1$^+$ cells identified also co-expressed Map2. To promote further neuronal maturation and for electrophysiological assessments, PMEFs were replated onto a previously established monolayer of primary rat astrocytes following transduction of ^{VP64}dCas9^{VP64} and gRNAs (Vierbuchen et al., 2010). Synapsin-RFP expression and cell morphology were used to select the most mature neuronal cells for patch-clamp analysis after 21 days in culture. In current-clamp mode, single or multiple action potentials were readily elicited in response to depolarizing current injections (six out of seven cells analyzed; Figure 4D). The same cells displayed voltage-dependent inward and outward currents. The transient inward currents were abolished in the presence of the voltage-gated Na^+ channel blocker tetrodotoxin (TTX; Figure 4E). The average resting membrane potential, action potential (AP) threshold and AP amplitude were -41 ± 3.8 mV, -33 ± 2.6 mV, and 49 ± 9.7 mV, respectively (mean ± SEMs, n = 7 cells).

In contrast to what we observed by transient transfection of the reprogramming factors, constitutive expression of the BAM factor transgenes via lentiviral vectors generated significantly more Tuj1$^+$Map2$^+$ cells than that detected with ^{VP64}dCas9^{VP64} (Figure 4F). We hypothesized that the prolonged and high levels of expression of the BAM factor transgenes enabled by lentiviral delivery permitted further epigenetic and transcriptional reprogramming that improved the efficiency of iN generation when compared to transient transfection methods. Consequently, we revisited the analysis of chromatin remodeling at the endogenous BAM factor loci in the context of lentiviral delivery of the reprograming factors. We found that, as shown with transient transfection, targeted activation of the endogenous genes via lentiCR-BAM transduction led to the rapid deposition of H3K27ac at the *Brn2* and *Ascl1* loci as early as day 3 post-transduction that persisted at day 6 (Figure 4G). Also, as seen with transient transfection, we did not detect enrichment of H3K27ac at the *Myt1l* locus with lentiCR-BAM transduction, although we did measure an increase in *Myt1l* mRNA (Figures 4B and 4G). In contrast to what we observed with transient transfection of the BAM factors, we detected enrichment of H3K27ac along regions of all three endogenous genes with lentiBAM transduction (Figure 4G). Furthermore, we only detected minor enrichment in H3K27ac at all three genes at day 3 post-transduction of lentiBAM; however, both *Ascl1* and *Myt1l* showed a substantial increase in H3K27ac deposition by day 6 post-transduction (Figure 4G).

DISCUSSION

In this study, we demonstrate direct cellular reprogramming to induced neuronal cells through targeted activation of endogenous genes. We utilized the CRISPR/Cas9 system as a programmable, locus-specific transcriptional regulator for the

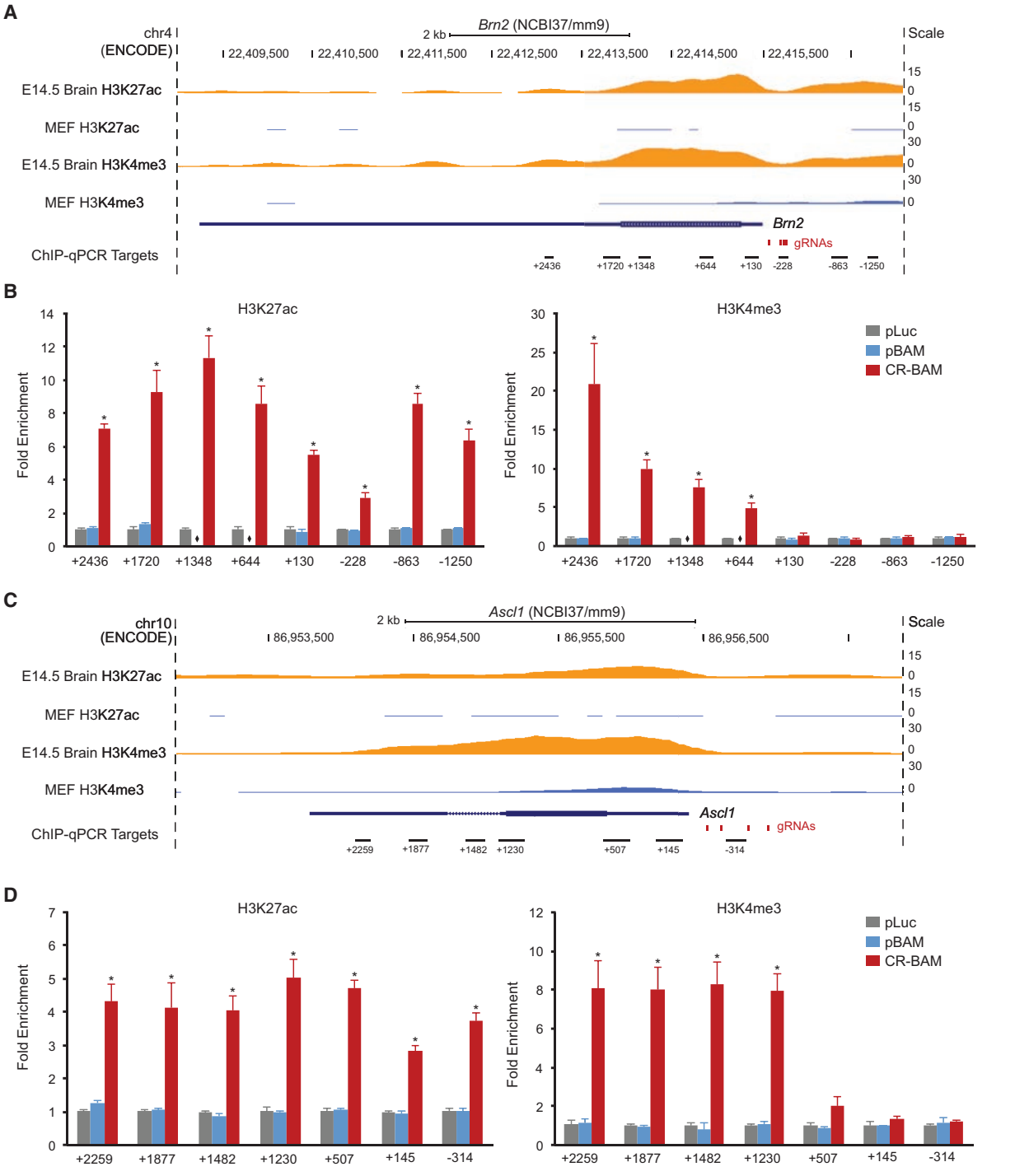


Figure 3. ^{VP64}dCas9^{VP64} Rapidly Remodels Epigenetic Marks at Target Loci

(A and C) Mouse genomic tracks depicting histone H3 modifications H3K27ac and H3K4me3 at the *Brn2* and *Ascl1* loci in embryonic brain tissue and fibroblasts (data from Mouse ENCODE; GEO: GSE31039). Red bars indicate gRNA target sites near the transcription start site, and black bars indicate the location of ChIP-qPCR amplicons along the gene locus.

(B and D) Targeted activation of endogenous *Brn2* and *Ascl1* in PMEFs induced significant enrichment of H3K27ac and H3K4me3 at multiple sites along the genomic loci at day 3 post-transfection (*$p < 0.05$, one-way ANOVA with Holm-Bonferroni post hoc tests, $n = 3$ biological replicates). Overexpression of the BAM factors via transfection of expression plasmids encoding BAM factor transgenes did not induce a significant change in these chromatin marks. qPCR primers targeting coding regions of the genes are not included for the pBAM transfection condition, as contaminating plasmid DNA biased enrichment values, and are marked with diamonds in (B). All fold enrichments are relative to transfection of a plasmid encoding firefly luciferase and normalized to a region of the *Gapdh* locus. See also Figure S3 and S4.

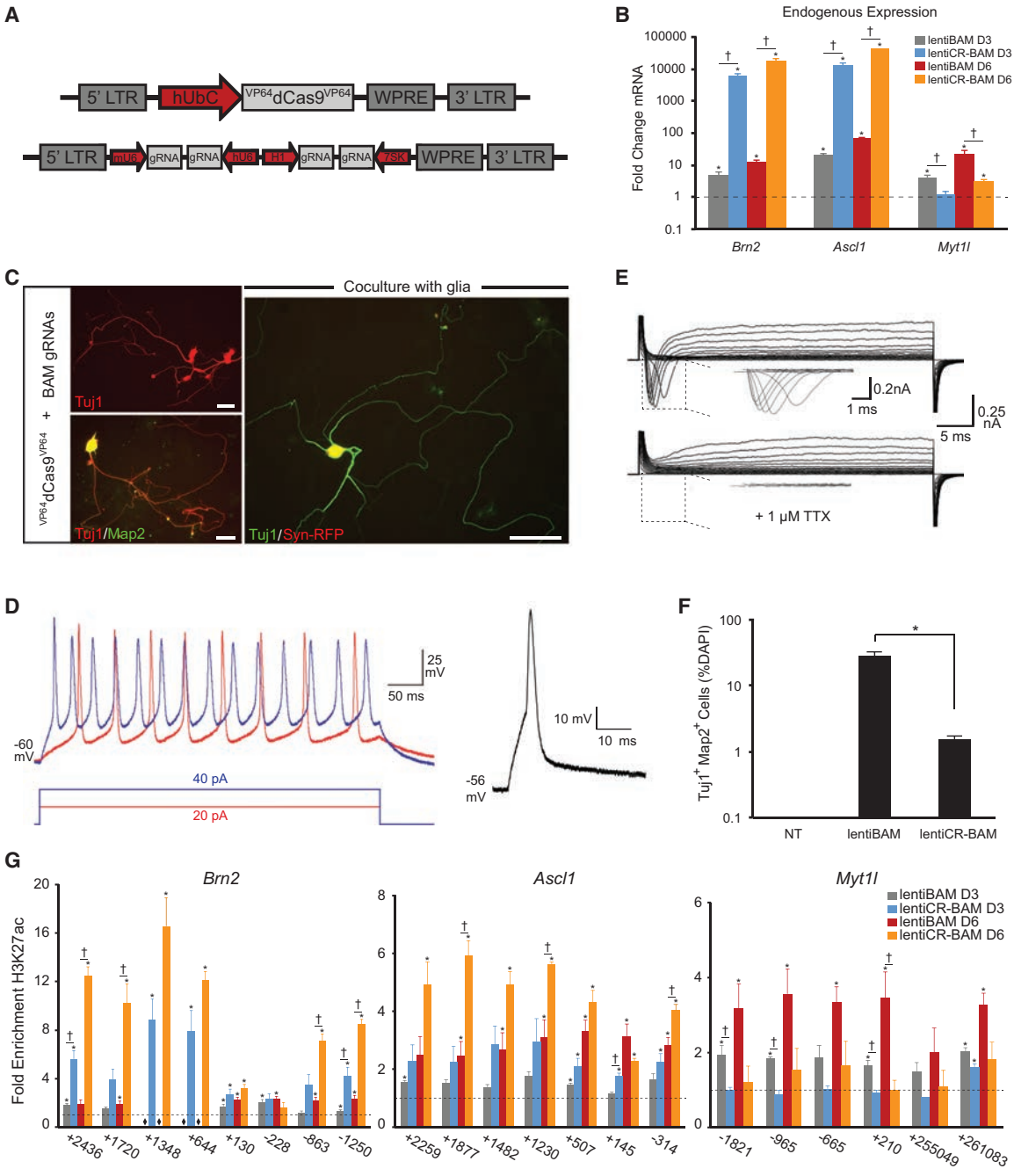


Figure 4. Generation of Functionally Mature iNs with Multiplex gRNA Vectors

(A) Schematic of ^{VP64}dCas9^{VP64} and multiplex gRNA lentiviral constructs used to enable stable integration and constitutive expression.

(B) Relative mRNA expression of the endogenous BAM factors following transduction of transgenes encoding the BAM factors (lentiBAM) or ^{VP64}dCas9^{VP64} and gRNAs targeting the endogenous BAM factors (lentiCR-BAM; *p < 0.05 relative to non-treated PMEFs; †p < 0.05 between lentiBAM versus lentiCR-BAM transduction).

(C) Immunofluorescence staining of PMEFs following transduction of lentiCR-BAM. Cells were co-positive for Tuj1 and Map2 and exhibited complex neuronal morphologies (scale bar, 50 μm).

(D) Action potentials were evoked from ^{VP64}dCas9^{VP64}-induced neuronal cells in response to 5-ms (right) or 500-ms (left) step depolarizing current injection (six out of seven cells analyzed) after empiric hyperpolarizing current injection to hold membrane potential at ~−60 mV.

(E) Representative whole-cell currents recorded with or without perfusion of 1 μM tetrodotoxin (TTX).

(F) Quantitation of Tuj1$^+$Map2$^+$ cells as percent nuclei (*p < 0.05 between lentiBAM versus lentiCR-BAM transduction; NT, non-treated PMEFs).

(G) Time course of H3K27ac enrichment along the *Brn2*, *Ascl1*, and *Mytl1* loci (*p < 0.05 relative to non-treated PMEFs; †p < 0.05 between lentiBAM versus lentiCR-BAM transduction).

All p values calculated by global ANOVA with Holm-Bonferroni post hoc tests ($\alpha = 0.05$).

multiplex activation of the neurogenic factors *Brn2*, *Ascl1*, and *Myt1l* (BAM factors). We hypothesized that targeted activation of the endogenous genes in PMEFs, as opposed to the forced overexpression of the corresponding transgenes, could more directly access the endogenous loci and rapidly remodel their epigenetic signatures, thus potentially reflecting a more natural mechanism of action and serving as an alternate method to achieve cell lineage conversion.

In PMEFs, the *cis*-repressive chromatin landscape at neuronal loci may preclude binding of regulatory factors, in turn impeding transcriptional activation. As a result, expression of the BAM factors in PMEFs from exogenous vectors likely relies on stochastic processes for reactivation of the corresponding endogenous genes. Furthermore, transient delivery of the BAM factors, as done in our initial experiments (Figures 1, 2, and 3), limits the time window within which the endogenous networks and positive feedback loops can be established. We demonstrated that targeting the endogenous genes directly induced the enrichment of histone H3 modifications H3K27ac and H3K4me3 at the *Brn2* and *Ascl1* loci at 3 days post-transfection, whereas transgene overexpression via transfection of plasmids encoding the reprogramming factors did not alter these chromatin marks (Figures 3 and S4). Additionally, we observed sustained high levels of expression from the endogenous genes at later stages of reprogramming despite the transient delivery of the gRNA plasmids (Figure S3).

In contrast, we found that stable integration and constitutive expression of the exogenous reprogramming factors via lentiviral delivery led to the eventual deposition of H3K27ac at their endogenous loci with a concomitant improvement in reprogramming capacity (Figures 4F and 4G). We did not observe a similar improvement with constitutive expression of ^{VP64}dCas9^{VP64} and gRNAs, which is possibly attributable to the lower levels of overall expression of the neuronal transcription factors achieved by transactivation of the endogenous genes compared to ectopic overexpression. Consequently, the direct activation of the endogenous genes via ^{VP64}dCas9^{VP64} may be more amenable to transient delivery approaches that avoid undesired consequences of vector integration into the genome. Such transient methods, including the direct delivery of ribonucleoprotein Cas9-gRNA complexes, may be a more clinically translatable method of generating reprogrammed cells that are genetically unmodified.

Achieving robust and well-defined reprogrammed cell populations is still a central challenge. Reprogrammed cells often fail to acquire completely mature phenotypes and can retain epigenetic remnants of the native cell type (Kim et al., 2010). Moreover, a recent study demonstrated that reprogramming efficiency can be limited by divergence to a competing cell identity (Treutlein et al., 2016). The molecular mechanisms and practical consequences of these limitations are largely unknown. As the toolkit of designer transcription factors expands to precisely modify the epigenome (Hilton et al., 2015; Kearns et al., 2015; Maeder et al., 2013a; Mendenhall et al., 2013; Thakore et al., 2016), these tools may be used to prime specific genomic loci in diverse cell types, promote endogenous transcription factor binding, and directly correct regions of epigenetic remnants that prove to be problematic for a given application. This may lead to improved reprogramming fidelity and extension of the breadth of donor cells amenable to reprogramming.

EXPERIMENTAL PROCEDURES

Cell Culture, Transfections, and Viral Transductions

PMEFs were maintained in high serum media during transduction and transfection of expression plasmids and subsequently cultured in neurogenic serum-free medium for the duration of the experiments to promote neuronal survival and maturation. Lentivirus was produced in HEK293T cells using the calcium phosphate precipitation method. All transfections were performed using Lipofectamine 3000 (Invitrogen) in accordance with the manufacturer's protocol. All expression plasmids used in this study can be found in Table S2.

Immunofluorescence Staining and qRT-PCR

All sequences for qRT-PCR primers can be found in Table S3. Total RNA was isolated using the QIAGEN RNeasy and QIAshredder kits, reverse transcribed using the SuperScript VILO Reverse Transcription Kit (Invitrogen), and analyzed using Perfecta SYBR Green Fastmix (Quanta BioSciences). All qRT-PCR data are presented as fold change in RNA normalized to *Gapdh* expression. For immunofluorescence staining, cells were fixed with 4% paraformaldehyde, permeabilized with 0.2% Triton X-100, and incubated with primary and secondary antibodies.

Electrophysiology

A synapsin-RFP lentiviral reporter was used to identify cells in co-culture with primary rat astroglia for patch-clamp analysis at indicated time points. Action potentials and inward and outward currents were recorded in whole-cell configuration. Data were analyzed and prepared for publication using pCLAMP and MATLAB.

Chromatin Immunoprecipitation qPCR

Chromatin was immunoprecipitated using antibodies against H3K27ac and H3K4me3, and gDNA was purified for qPCR analysis. All sequences for ChIP-qPCR primers can be found in Table S3. qPCR was performed using SYBR green Fastmix (Quanta BioSciences), and the data are presented as fold change gDNA relative to negative control and normalized to a region of the *Gapdh* locus.

Mouse ENCODE ChIP-Sequencing Datasets

H3K4me3 and H3K27ac ChIP-sequencing data from C57BL/6 E14.5 whole brain and mouse embryonic fibroblasts (GSE31039) were acquired from the Mouse ENCODE Consortium generated in Bing Ren's laboratory at the Ludwig Institute for Cancer Research.

Statistical Methods

Statistical analysis was done using GraphPad Prism 7. All data were analyzed with at least three biological replicates and presented as mean ± SEM. See figure legends for details on specific statistical tests run and p values calculated for each experiment.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2016.07.001.

AUTHOR CONTRIBUTIONS

J.B.B., A.F.A., K.W.L., and C.A.G. designed experiments. J.B.B., A.F.A., H.-G.W., A.M.D., and H.A.H. performed the experiments. All authors analyzed the data. J.B.B. and C.A.G. wrote the manuscript with contributions by all authors.

CONFLICTS OF INTEREST

C.A.G. and J.B.B. are inventors on filed patent applications related to this work.

ACKNOWLEDGMENTS

The authors thank Ami Kabadi, Pablo Perez-Pinera, and Syandan Chakraborty for providing plasmid constructs; Christopher Grigsby for protocols for PMEF

transfections and providing the pUNO-Brn2 plasmid; Rui Dai for technical assistance; and Jeff Stogsdill and Cagla Eroglu for supplying astrocytes and providing protocols for co-culture experiments. This work was supported by US NIH grants to C.A.G. and T.E.R. (R01DA036865 and U01HG007900), an NIH Director's New Innovator Award (DP2OD008586) and National Science Foundation (NSF) Faculty Early Career Development (CAREER) Award (CBET-1151035) to C.A.G., and NIH grant P30AR066527. J.B.B. was supported by an NIH Biotechnology Training Grant (T32GM008555).

Received: May 13, 2015
Revised: May 11, 2016
Accepted: June 30, 2016
Published: August 11, 2016

REFERENCES

Adler, A.F., Grigsby, C.L., Kulangara, K., Wang, H., Yasuda, R., and Leong, K.W. (2012). Nonviral direct conversion of primary mouse embryonic fibroblasts to neuronal cells. Mol. Ther. Nucleic Acids *1*, e32.

Balboa, D., Weltner, J., Eurola, S., Trokovic, R., Wartiovaara, K., and Otonkoski, T. (2015). Conditionally stabilized dCas9 activator for controlling gene expression in human cell reprogramming and differentiation. Stem Cell Reports *5*, 448–459.

Chakraborty, S., Ji, H., Kabadi, A.M., Gersbach, C.A., Christoforou, N., and Leong, K.W. (2014). A CRISPR/Cas9-based system for reprogramming cell lineage specification. Stem Cell Reports *3*, 940–947.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. Nat. Methods *12*, 326–328.

Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. *23*, 1163–1171.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Mouse ENCODE Consortium (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biol. *13*, 418.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987–1000.

Gao, X., Yang, J., Tsang, J.C., Ooi, J., Wu, D., and Liu, P. (2013). Reprogramming to pluripotency using designer TALE transcription factors targeting enhancers. Stem Cell Reports *1*, 183–197.

Gao, X., Tsang, J.C., Gaba, F., Wu, D., Lu, L., and Liu, P. (2014). Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. Nucleic Acids Res. *42*, e155.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *462*, 595–601.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Kabadi, A.M., Ousterout, D.G., Hilton, I.B., and Gersbach, C.A. (2014). Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. Nucleic Acids Res. *42*, e147.

Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat. Methods. *12*, 401–413.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I., et al. (2010). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285–290.

Konermann, S., Brigham, M.D., Trevino, A.E., Hsu, P.D., Heidenreich, M., Cong, L., Platt, R.J., Scott, D.A., Church, G.M., and Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. Nature *500*, 472–476.

Ladewig, J., Mertens, J., Kesavan, J., Doerr, J., Poppe, D., Glaue, F., Herms, S., Wernet, P., Kögler, G., Müller, F.J., et al. (2012). Small molecules enable highly efficient neuronal conversion of human fibroblasts. Nat. Methods *9*, 575–578.

Maeder, M.L., Angstman, J.F., Richardson, M.E., Linder, S.J., Cascio, V.M., Tsai, S.Q., Ho, Q.H., Sander, J.D., Reyon, D., Bernstein, B.E., et al. (2013a). Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat. Biotechnol. *31*, 1137–1142.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013b). CRISPR RNA-guided activation of endogenous human genes. Nat. Methods *10*, 977–979.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. *31*, 833–838.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013b). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Mendenhall, E.M., Williamson, K.E., Reyon, D., Zou, J.Y., Ram, O., Joung, J.K., and Bernstein, B.E. (2013). Locus-specific editing of histone modifications at endogenous enhancers. Nat. Biotechnol. *31*, 1133–1136.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., et al. (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat. Methods *10*, 973–976.

Polstein, L.R., Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Bledsoe, P., Song, L., Safi, A., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators. Genome Res. *25*, 1158–1169.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

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

Thakore, P.I., Black, J.B., Hilton, I.B., and Gersbach, C.A. (2016). Editing the epigenome: technologies for programmable transcription and epigenetic modulation. Nat. Methods *13*, 127–137.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M., and Quake, S.R. (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature *534*, 391–395.

Vierbuchen, T., and Wernig, M. (2011). Direct lineage conversions: unnatural but useful? Nat. Biotechnol. *29*, 892–907.

Vierbuchen, T., and Wernig, M. (2012). Molecular roadblocks for cellular reprogramming. Mol. Cell *47*, 827–838.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. Nature *463*, 1035–1041.

Wei, S., Zou, Q., Lai, S., Zhang, Q., Li, L., Yan, Q., Zhou, X., Zhong, H., and Lai, L. (2016). Conversion of embryonic stem cells into extraembryonic lineages by CRISPR-mediated activators. Sci. Rep. *6*, 19648.

Cell

Interspecies Chimerism with Mammalian Pluripotent Stem Cells

Jun Wu,[1] Aida Platero-Luengo,[1] Masahiro Sakurai,[1] Atsushi Sugawara,[1] Maria Antonia Gil,[2] Takayoshi Yamauchi,[1] Keiichiro Suzuki,[1] Yanina Soledad Bogliotti,[3] Cristina Cuello,[2] Mariana Morales Valencia,[1] Daiji Okumura,[1,7] Jingping Luo,[1] Marcela Vilariño,[3] Inmaculada Parrilla,[2] Delia Alba Soto,[3] Cristina A. Martinez,[2] Tomoaki Hishida,[1] Sonia Sánchez-Bautista,[4] M. Llanos Martinez-Martinez,[4] Huili Wang,[3] Alicia Nohalez,[2] Emi Aizawa,[1] Paloma Martinez-Redondo,[1] Alejandro Ocampo,[1] Pradeep Reddy,[1] Jordi Roca,[2] Elizabeth A. Maga,[3] Concepcion Rodriguez Esteban,[1] W. Travis Berggren,[1] Estrella Nuñez Delicado,[4] Jeronimo Lajara,[4] Isabel Guillen,[5] Pedro Guillen,[4,5] Josep M. Campistol,[6] Emilio A. Martinez,[2] Pablo Juan Ross,[3] and Juan Carlos Izpisua Belmonte[1,8,*]

[1]Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
[2]Department of Animal Medicine and Surgery, University of Murcia Campus de Espinardo, 30100 Murcia, Spain
[3]Department of Animal Science, University of California Davis, One Shields Avenue, Davis, CA 95616, USA
[4]Universidad Católica San Antonio de Murcia (UCAM) Campus de los Jerónimos, N° 135 Guadalupe 30107 Murcia, Spain
[5]Clinica Centro Fundación Pedro Guillén, Clínica CEMTRO, Avenida Ventisquero de la Condesa 42, 28035 Madrid, Spain
[6]Hospital Clínico de Barcelona-IDIBAPS, Universitat de Barcelona, 08007 Barcelona, Spain
[7]Present address: Graduate School of Agriculture, Department of Advanced Bioscience, Kinki University, 3327-204 Nakamachi, Nara 631-8505, Japan
[8]Lead Contact
*Correspondence: belmonte@salk.edu
http://dx.doi.org/10.1016/j.cell.2016.12.036

SUMMARY

Interspecies blastocyst complementation enables organ-specific enrichment of xenogenic pluripotent stem cell (PSC) derivatives. Here, we establish a versatile blastocyst complementation platform based on CRISPR-Cas9-mediated zygote genome editing and show enrichment of rat PSC-derivatives in several tissues of gene-edited organogenesis-disabled mice. Besides gaining insights into species evolution, embryogenesis, and human disease, interspecies blastocyst complementation might allow human organ generation in animals whose organ size, anatomy, and physiology are closer to humans. To date, however, whether human PSCs (hPSCs) can contribute to chimera formation in non-rodent species remains unknown. We systematically evaluate the chimeric competency of several types of hPSCs using a more diversified clade of mammals, the ungulates. We find that naïve hPSCs robustly engraft in both pig and cattle pre-implantation blastocysts but show limited contribution to post-implantation pig embryos. Instead, an intermediate hPSC type exhibits higher degree of chimerism and is able to generate differentiated progenies in post-implantation pig embryos.

INTRODUCTION

Embryonic pluripotency has been captured in vitro at a spectrum of different states, ranging from the naive state, which reflects unbiased developmental potential, to the primed state, in which cells are poised for lineage differentiation (Weinberger et al., 2016; Wu and Izpisua Belmonte, 2016). When attempting to introduce cultured pluripotent stem cells (PSCs) into a developing embryo of the same species, recent studies demonstrated that matching developmental timing is critical for successful chimera formation. For example, naive mouse embryonic stem cells (mESCs) contribute to chimera formation when injected into a blastocyst, whereas primed mouse epiblast stem cells (mEpiSCs) efficiently engraft into mouse gastrula-stage embryos, but not vice versa (Huang et al., 2012; Wu et al., 2015). Live rodent interspecies chimeras have also been generated using naive PSCs (Isotani et al., 2011; Kobayashi et al., 2010; Xiang et al., 2008). However, it remains unclear whether naive PSCs can be used to generate chimeras between more distantly related species.

The successful derivation of human PSCs (hPSCs), including ESCs from pre-implantation human embryos (Reubinoff et al., 2000; Thomson et al., 1998), as well as the generation of induced pluripotent stem cells (iPSCs) from somatic cells through cellular reprograming (Takahashi et al., 2007; Park et al., 2008; Wernig et al., 2007; Yu et al., 2007; Aasen et al., 2008), has revolutionized the way we study human development and is heralding a new age of regenerative medicine. Several lines of evidence indicate that conventional hPSCs are in the primed pluripotent state, similar to mEpiSCs (Tesar et al., 2007; Wu et al., 2015). A number of recent studies have also reported the generation of putative naive hPSCs that molecularly resemble mESCs (Gafni et al., 2013; Takashima et al., 2014; Theunissen et al., 2014). These naive hPSCs have already provided practical and experimental advantages, including high single-cell cloning efficiency and facile genome editing (Gafni et al., 2013). Despite these advances, it remains unclear how the putative higher developmental potential of naive hPSCs can be used to better

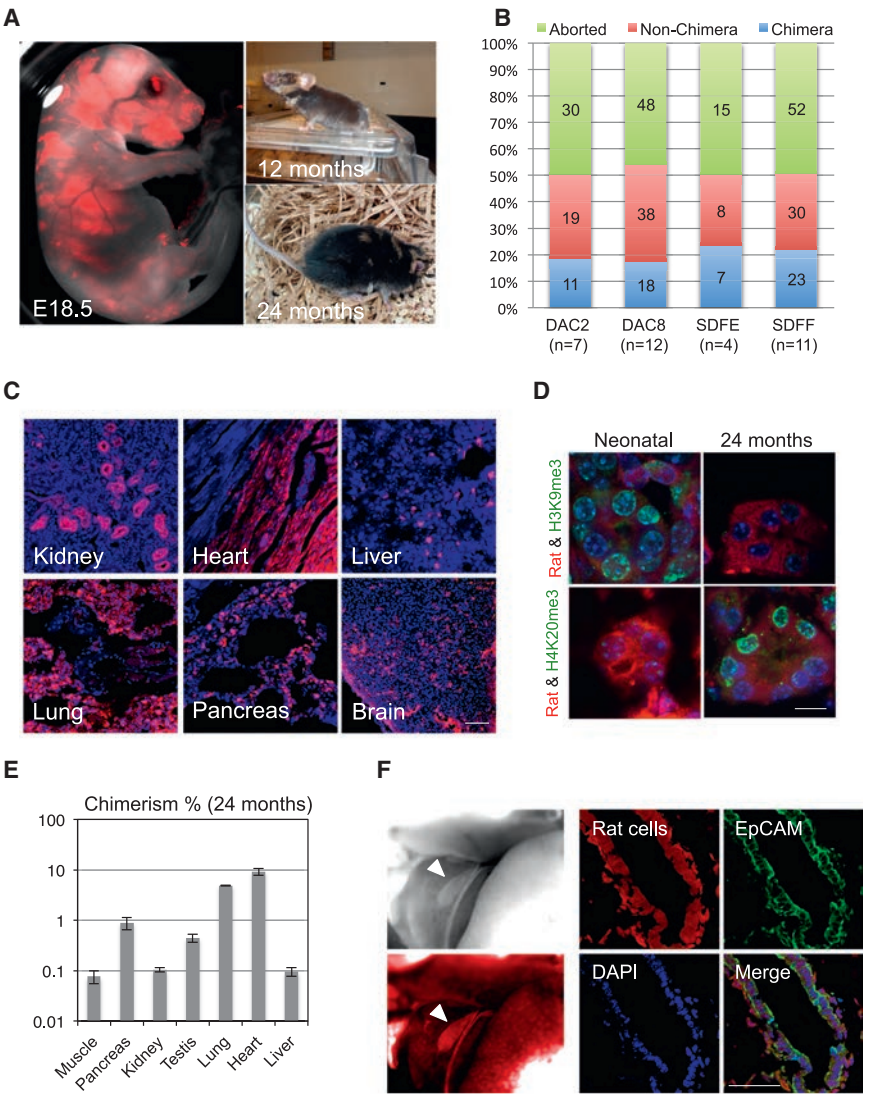


Figure 1. Interspecies Rat-Mouse Chimeras Derived from Rat PSCs

(A) Rat-mouse chimeras generated by rat ESCs (DAC2). Left, an E18.5 rat-mouse chimeric fetus. Red, hKO-labeled rat cells. Right, a 12-month-old (top) and 24-month-old (bottom) rat-mouse chimera.

(B) Chimera forming efficiencies with rat ESC lines (DAC2 and DAC8) and rat iPSC lines (SDFE and SDFF). n, number of embryo transfers.

(C) Representative fluorescence images showing hKO-labeled rat ESCs (DAC2) contributed to different tissues in the 24-month-old rat-mouse chimera. Red, hKO-labeled rat cells. Blue, DAPI. Scale bar, 100 μm.

(D) Representative immunofluorescence images showing the expression of aging-related histone marks, including H3K9me3 and H4K20me3, in the kidney tissue of neonatal and 24-month-old chimeras. Scale bar, 10 μm.

(E) Levels of chimerism of rat ESCs (DAC2) in different tissues of the 24-month-old rat-mouse chimera. Error bars indicate SD.

(F) Rat iPSCs (SDFE) contributed to the neonatal mouse gall bladder. Left, bright-field (top) and fluorescence (bottom) images showing a neonatal mouse gallbladder contained cells derived from rat iPSCs. White arrowheads indicate the gallbladder. Right, representative immunofluorescence images showing the expression of a gallbladder epithelium marker (EpCAM) by rat cells. Red, hKO-labeled rat cells; blue, DAPI. Scale bar, 50 μm.

See also Figure S1 and Table S2.

understand human embryogenesis and to develop regenerative therapies for treating patients.

Like naive rodent PSCs, naive hPSCs can potentially be used to generate interspecies chimeras for studying human development and disease, and producing functional human tissues via interspecies blastocyst complementation. To date, however, all reported attempts on generating hPSC-derived interspecies chimeras have used the mouse as the host animal, and the results obtained suggest that this process is rather inefficient (Gafni et al., 2013; Theunissen et al., 2014, 2016). Although the mouse is one of the most important experimental models for stem cell research, there are considerable differences between humans and mice (e.g., early post-implantation development, embryo size, gestational length, and developmental speed), which may hinder not only the efficiency but also the usefulness of human-mouse chimeric studies. Thus, expanding the repertoire of host species may complement this incipient but promising area of research in the field of regenerative medicine. In particular, interspecies chimera research of hPSCs using ungulates, e.g., pigs, cattle, and sheep, could lead to improved research models, as well as novel in vivo strategies for (1) generating human organs and tissues, (2) designing new drug screening methodologies, and (3) developing new human disease models (Wu and Izpisua Belmonte, 2015). Experiments to empirically test and evaluate the chimeric contribution of various types of hPSCs in the ungulates are thus imperative, but currently lacking. To start filling this void, we tested different types of hPSCs for their chimeric contribution potential in two ungulate species, pigs and cattle.

RESULTS

Naive Rat PSCs Robustly Contribute to Rat-Mouse Interspecies Chimera Formation

We first used rodent models to gain a better understanding of the factors and caveats underlying interspecies chimerism with PSCs. To this end, we used two chimeric-competent rat ESC lines, DAC2 and DAC8 (Li et al., 2008). We labeled both lines with a fluorescent marker, humanized kusabira orange (hKO), for cell tracking and injected them into mouse blastocysts. Following embryo transfer (ET) into surrogate mouse mothers, both DAC2 and DAC8 lines gave rise to live rat-mouse chimeras (Figures 1A and S1A). Many of the chimeras developed into

adulthood, and one chimera reached 2 years of age (Figure 1A), indicating that the xenogeneic rat cells sustained the physiological requirements of the mouse host without compromising its life span. We also generated two rat iPSC lines (SDFE and SDFF) from tail tip fibroblasts (TTFs) isolated from a neonatal Sprague-Dawley rat and used them to generate rat-mouse chimeras. Similar to rat ESCs, rat iPSCs could also robustly contribute to chimera formation in mice (Figure S1B). Overall, the chimera forming efficiencies of all rat PSC lines tested were ~20%, consistent with a previous report (Figure 1B) (Kobayashi et al., 2010).

We observed contribution of rat cells to a wide range of tissues and organs in both neonatal and aged rat-mouse chimeras (Figures 1C, S1A, and S1B). We examined aging-related histone marks in both neonatal and aged chimeras and found that the 2-year-old chimera exhibited histone signatures characteristic of aging (Figure 1D). We quantified the degree of chimerism in different organs of the aged chimera via quantitative qPCR analysis of genomic DNA using a rat-specific primer (Table S2). We found that different tissues contained different percentages of rat cells, with the highest contribution observed in the heart (~10%) (Figure 1E).

One anatomical difference between mice and rats is that rats lack a gallbladder. In agreement with a previous report (Kobayashi et al., 2010), we also observed the presence of gallbladders in rat-mouse chimeras (chimeras derived from injecting rat PSCs into a mouse blastocyst). Interestingly, rat cells contributed to the chimeric gallbladder and expressed the gallbladder epithelium marker EpCAM (Figures 1F and S1C), which suggests that the mouse embryonic microenvironment was able to unlock a gallbladder developmental program in rat PSCs that is normally suppressed during rat development.

A Versatile CRISPR-Cas9-Mediated Interspecies Blastocyst Complementation System

Chimeric contribution of PSCs is random and varies among different host blastocysts and donor cell lines used. To selectively enrich chimerism in a specific organ, a strategy called blastocyst complementation has been developed where the host blastocysts are obtained from a mutant mouse strain in which a gene critical for the development of a particular lineage is disabled (Chen et al., 1993; Kobayashi et al., 2010; Wu and Izpisua Belmonte, 2015). Mutant blastocysts used for complementation experiments were previously obtained from existing lines of knockout mice, which were generated by gene targeting in germ-line-competent mouse ESCs—a time-consuming process. To relieve the dependence on existing mutant strains, we developed a blastocyst complementation platform based on targeted genome editing in zygotes. We chose to use the CRISPR-Cas9 system, which has been harnessed for the efficient generation of knockout mouse models (Wang et al., 2013) (Figure 2A).

For proof-of-concept, we knocked out the *Pdx1* gene in mouse by co-injecting Cas9 mRNA and *Pdx1* single-guide RNA (sgRNA) into mouse zygotes. During mouse development, *Pdx1* expression is restricted to the developing pancreatic anlagen and is a key player in pancreatic development. Mice homozygous for a targeted mutation in *Pdx1* lack a pancreas and die within a few days after birth (Jonsson et al., 1994; Offield et al., 1996). Similarly, *Pdx1*$^{-/-}$ mice generated by the zygotic co-injection of Cas9 mRNA and *Pdx1* sgRNA were apancreatic, whereas other internal organs appeared normal (Figure S2A). These mice survived only a few days after birth. We observed the efficiency for obtaining *Pdx1*$^{-/-}$ mouse via CRISPR-Cas9 zygote genome editing was ~60% (Figure S2F). Next, we combined zygotic co-injection of Cas9/sgRNA with blastocyst injection of rat PSCs, and found that rat PSC-derivatives were enriched in the neonatal pancreas of *Pdx1*$^{-/-}$ mice and expressed α-AMYLAYSE, a pancreatic enzyme that helps digest carbohydrates (Figures 2B and S2B). Of note is that in these animals the pancreatic endothelial cells were still mostly of mouse origin, as revealed by staining with an anti-CD31 antibody (Figure 2B). Importantly, pancreas enriched with rat cells supported the successful development of *Pdx1*$^{-/-}$ mouse host into adulthood (>7 months), and maintained normal serum glucose levels in response to glucose loading, as determined using the glucose tolerance test (GTT) (Figure S2C).

Taking advantage of the flexibility of the CRISPR-Cas9 zygotic genome editing, we next sought to enrich xenogenic rat cells toward other lineages. *Nkx2.5* plays a critical role in early stages of cardiogenesis, and its deficiency leads to severe growth retardation with abnormal cardiac looping morphogenesis, an important process that leads to chamber and valve formation (Lyons et al., 1995; Tanaka et al., 1999). Mice lacking *Nkx2.5* typically die around E10.5 (Lyons et al., 1995; Tanaka et al., 1999). Consistent with previous observations, CRISPR-Cas9 mediated inactivation of *Nkx2.5* resulted in marked growth-retardation and severe malformation of the heart at E10.5 (Figure S2D). In contrast, when complemented with rat PSCs, the resultant *Nkx2.5*$^{-/-}$ mouse hearts were enriched with rat cells and displayed a normal morphology, and the embryo size was restored to normal (Figures 2C and S2D). Of note is that although rat PSCs rescued embryo growth and cardiac formation in E10.5 *Nkx2.5*$^{-/-}$ mouse embryos, to date we still have not obtained a live rescued chimera (n = 12, where n is the number of ETs). *Pax6* is a transcription factor that plays key roles in development of the eye, nose and brain. Mice homozygous for a *Pax6* loss-of-function mutation lack eyes, nasal cavities, and olfactory bulbs, and exhibit abnormal cortical plate formation, among other phenotypes (Gehring and Ikeo, 1999). *Pax6* is best known for its conserved function in eye development across all species examined (Gehring and Ikeo, 1999). In agreement with the published work, CRISPR-Cas9 mediated *Pax6* inactivation disrupted eye formation in the E15.5 mouse embryo (Figure S2E). When complemented with rat PSCs, we observed the formation of chimeric eyes enriched with rat cells in *Pax6*$^{-/-}$ mouse neonate (Figures 2D and S2E). Similar to *Pdx1*$^{-/-}$, we observed efficient generation of homozygous *Nkx2.5*$^{-/-}$ and *Pax6*$^{-/-}$ mouse embryos via zygotic co-injection of Cas9 mRNA and sgRNAs (Figure S2F). All DNA sequencing results of CRISPR-Cas9 mediated gene knockouts and gRNA sequences are summarized in Tables S1 and S2, respectively.

In sum, for the pancreas, heart, and eye, as well as several other organs (data not shown), we successfully generated chimerized organs that were enriched with rat cells, demonstrating

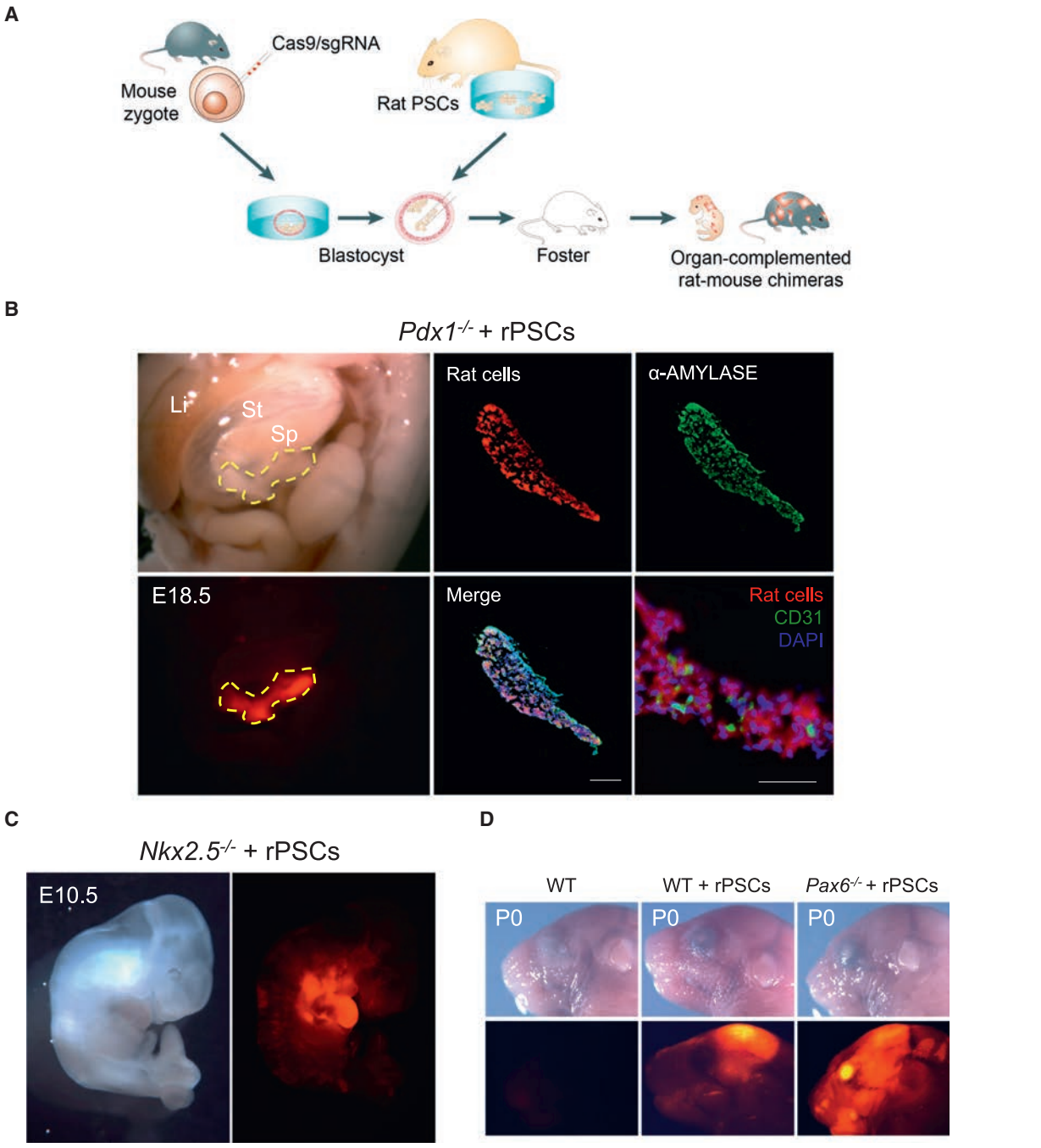


Figure 2. Interspecies Blastocyst Complementation via CRISPR-Cas9-Mediated Zygote Genome Editing

(A) Schematic of the CRISPR-Cas9 mediated rat-mouse blastocyst complementation strategy.

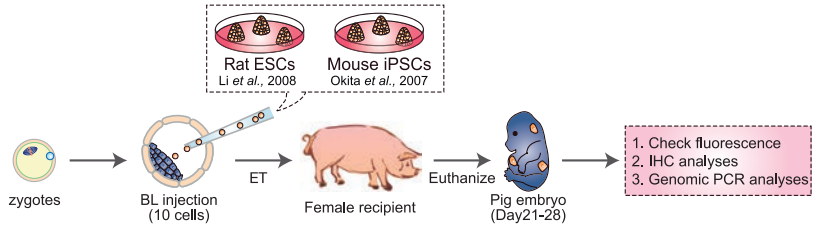
(B) Left, bright-field (top) and fluorescence (bottom) images showing the enrichment of rat cells in the pancreas of an E18.5 *Pdx1*$^{-/-}$ mouse. Li, liver; St, stomach; Sp, spleen. Yellow-dotted line encircles the pancreas. Red, hKO-labeled rat cells. Middle and right (top), representative immunofluorescence images showing rat cells expressed α-amylase in the *Pdx1*$^{-/-}$ mouse pancreas. Blue, DAPI. Right (bottom), a representative immunofluorescence image showing that some pancreatic endothelial cells, as marked by a CD31 antibody, were not derived from rat PSCs. Scale bar, 100 μm.

(C) Bright field (left) and fluorescence (right) images showing the enrichment of rat cells in the heart of an E10.5 Nkx2.5$^{-/-}$ mouse. Red, hKO-labeled rat cells.

(D) Bright field (top) and fluorescence (bottom) images showing the enrichment of rat cells in the eye of a neonatal Pax6$^{-/-}$ mouse. Red, hKO-labeled rat cells. WT, mouse control; WT+rPSCs, control rat-mouse chimera without Cas9/sgRNA injection.

See also Figure S2 and Tables S1 and S2.

A



B

Cell line	BL origin	Normal size	Growth retarded	Fluorescence positive (FO+)	
				Normal size	Growth retarded
Mouse iPSCs	2C	7	12	0	0
	ZG	4	2	0	0
	2C	0	1	0	0
	Total	11	15	0	0
Rat ESCs	2C	10	6	0	0
	2C	0	3	0	0
	Total	10	9	0	0

2C, 2-cell embryo; ZG, zygote; BL, blastocyst

C

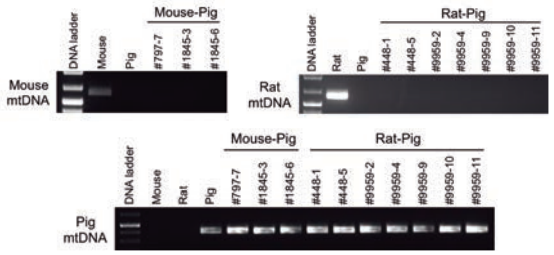


Figure 3. Naive Rodent PSCs Fail to Contribute to Chimera Formation in Pigs

(A) Schematic of the generation and analyses of post-implantation pig embryos derived from blastocyst injection of naive rodent PSCs.

(B) Summary of the pig embryos recovered between day 21–28 of pregnancy.

(C) Genomic PCR analyses of pig embryos derived from blastocyst injection of mouse iPSCs or rat ESCs. Mouse- and rat- specific mtDNA primers were used for the detection of chimeric contribution from mouse iPSCs and rat ESCs, respectively. Pig-specific mtDNA primers were used for the control.

See also Tables S2 and S3.

the efficacy and versatility of the CRISPR-Cas9 mediated interspecies blastocyst complementation platform.

Naive Rodent PSCs Do Not Contribute to Chimera Formation in Pigs

It is commonly accepted that the key functional feature of naive PSCs is their ability to generate intraspecies germline chimeras (Nichols and Smith, 2009). Studies in rodents also support the notion that attaining the naive pluripotent state is the key step in enabling chimera formation across species boundaries (Xiang et al., 2008; Isotani et al., 2011; Kobayashi et al., 2010). However, it has not yet been tested whether naive rodent PSCs can contribute to chimera formation when using a non-rodent host. To further examine the relationship between naive PSCs and interspecies chimerism, we injected rat ESCs into pig blastocysts followed by ET to recipient sows. In addition to rat ESCs, we also used a germline competent mouse iPSC line (Okita et al., 2007). Several criteria were used to determine the chimeric contribution of rodent cells in pig embryos, namely, (1) detection of fluorescence (hKO) signal, (2) immunohistochemical (IHC) labeling of embryo sections with an anti-hKO antibody, and (3) genomic PCR with mouse- or rat-specific primers targeting mitochondrial DNA (mtDNA) (Figure 3A). We terminated the pregnancy between day 21–28 of pig development and collected embryos derived from the injection of mouse iPSCs or rat ESCs into pig blastocyst (26 and 19 embryos, respectively) (Figure 3B; Table S3). We failed to detect any hKO signal in both normal size and growth retarded embryos (Figure 3B). We next sectioned the pig embryos and stained them with an antibody against hKO. Similarly, we did not detect any hKO-positive cells in the embryonic sections examined (data not shown). Finally, we employed a more sensitive test, using genomic PCR to amplify rat- or mouse-specific mtDNA sequences (pig-specific mtDNA primers served as the loading control) (Table S2). Consistently, genomic PCR analyses did not detect any rodent contribution to the pig embryos (Figure 3C). Taken together, although naive rodent PSCs can robustly contribute to rodent-specific interspecies chimeras, our results show that these cells are incapable of contributing to normal embryonic development in pigs.

Generation of Naive, Intermediate, and Primed hiPSCs

Next, we sought to systematically evaluate the chimeric competency of hPSCs in ungulate embryos. We generated hiPSCs using several reported naive PSC culture methods, a culture protocol supporting a putative intermediate pluripotent state between naive mESCs and primed mEpiSCs (Tsukiyama and Ohinata, 2014), and a primed culture condition (Figure 4A). Mouse ground state culture condition (2iL) induces the differentiation of primed hPSCs. However, when combined with the forced expression of NANOG and KLF2 (NK2), transcription factors that help to maintain murine naive pluripotency, 2iL culture can stabilize hPSCs in an immature state (Takashima et al., 2014; Theunissen et al., 2014). We generated doxycycline (DOX)-inducible NK2-expressing naive hiPSCs cultured in 2iL medium from primed hiPSCs (2iLD-hiPSCs). Transgene-free primed hiPSCs were reprogramed from human foreskin fibroblasts (HFFs) using episomal vectors (Okita et al., 2011). For comparison, we also generated naive hiPSCs from HFFs using the NHSM culture condition (Gafni et al., 2013) (NHSM-hiPSCs). It has been shown that cells grown in 4i medium, a

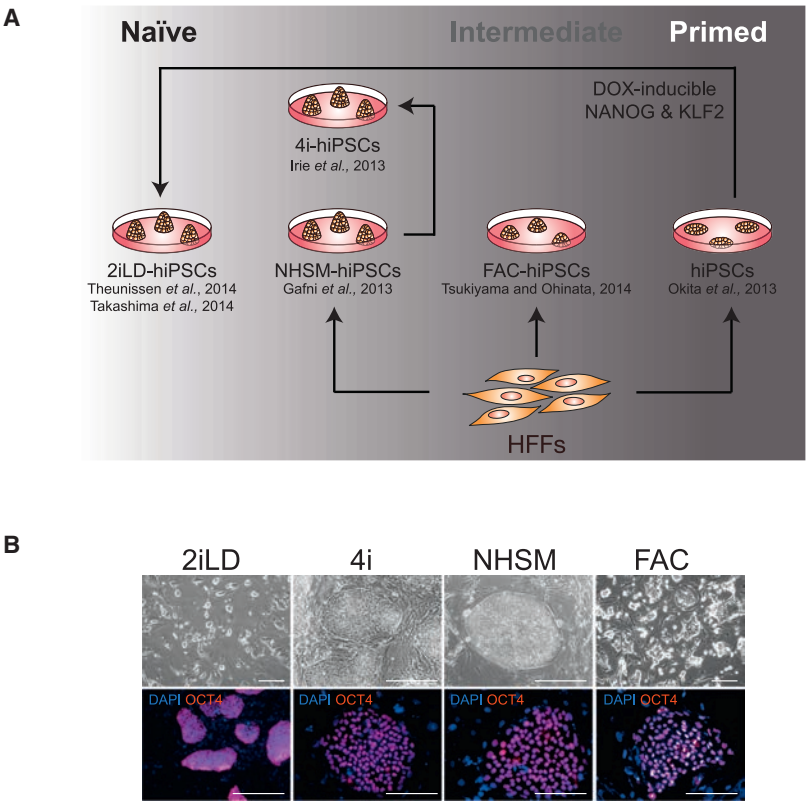


Figure 4. Generation and Interspecies ICM Incorporation of Different Types of hiPSCs

(A) Schematic of the strategy for generating naive, intermediate, and primed hiPSCs.

(B) (Top) Representative bright-field images showing the colony morphologies of naive (2iLD-, 4i-, and NHSM-hiPSCs) and intermediate (FAC-hiPSCs) hiPSCs. Bottom, representative immunofluorescence images of naive and intermediate hiPSCs stained with an anti-OCT4 antibody. Red, OCT4; blue, DAPI. Scale bar, 100 μm.

(C) Schematic of the experimental procedures for producing cattle and pig blastocysts obtained from in vitro fertilization (IVF) and parthenoactivation, respectively. Blastocysts were subsequently used for laser-assisted blastocyst injection of hiPSCs. After hiPSC injection, blastocysts were cultured in vitro for 2 days before fixation and analyzed by immunostaining with an anti-HuNu and an anti-SOX2 antibodies. Criteria to evaluate the survival of human cells, as well as the degree and efficiency of ICM incorporation are shown in the blue box.

(D) Number of hiPSCs that integrated into the cattle (left) and pig (right) ICMs after ten hiPSCs were injected into the blastocyst followed by 2 days of in vitro culture. Red line, the average number of ICM-incorporated hiPSCs. Blue dot, the number of ICM-incorporated hiPSCs in each blastocyst.

See also Figure S3 and Table S4.

simplified version of NHSM, have a significant potential for germ cell induction, a distinguishing feature between naive mESCs and primed mEpiSCs (Irie et al., 2015). Thus, we also culture-adapted NHSM-hiPSCs in 4i medium (4i-hiPSCs), resulting in stable 4i-hiPSCs with similar morphological and molecular characteristics to parental NHSM-hiPSCs (Figure 4B). In addition, we generated another type of hiPSC by direct reprogramming of HFFs in a modified mEpiSC medium containing bFGF, Activin-A, and CHIR99021 (FAC; Figure 4A). mEpiSCs cultured in FAC medium exhibited features characteristic of both naive mESCs and primed mEpiSCs, supporting an intermediate pluripotent state (Tsukiyama and Ohinata, 2014). hiPSCs generated and cultured in FAC medium (FAC-hiPSCs) displayed a colony morphology intermediate between that of 2iLD- and primed hiPSCs, with less defined borders (Figure 4B). 2iLD-hiPSCs, NHSM-hiPSCs, 4i-hiPSCs, and FAC-hiPSCs could all be stably maintained long term in culture, preserving normal karyotypes and the homogeneous, nuclear localization of OCT4 protein (Figure 4B; data not shown). Notably, similar to hiPSCs grown in naive cultures (2iLD-hiPSCs, NHSM-hiPSCs, 4i-hiPSCs), FAC-hiPSCs could also be efficiently propagated by single-cell dissociation without using a ROCK kinase inhibitor. After injecting cells into the kidney capsule of immunodeficient NSG mice, all of these hiPSCs formed teratomas that consisted of tissues from all three germ layers: endoderm, mesoderm, and ectoderm (Figure S3A). To facilitate the identification of human cells in subsequent chimera experiments, we labeled hiPSCs with either green fluorescence protein (GFP) or hKO fluorescence markers.

Chimeric Contribution of hiPSCs to Pig and Cattle Blastocysts

The ability to integrate into the inner cell mass (ICM) of a blastocyst is informative for evaluating whether hiPSCs are compatible with pre-implantation epiblasts of the ungulate species. This is also one of the earliest indicators of chimeric capability. We therefore evaluated interspecies chimeric ICM formation by injecting hiPSCs into blastocysts from two ungulate species, pig and cattle.

Cattle-assisted reproductive technologies, such as in vitro embryo production, are well established given the commercial benefits of improving the genetics of these animals. Cattle also serve as a research model because of several similarities to human pre-implantation development (Hansen, 2014; Hasler, 2014). Using techniques for producing cattle embryos in vitro, we developed a system for testing the ability and efficiency of hiPSCs to survive in the blastocyst environment and to integrate into the cattle ICM (Figure 4C). Cattle embryos were obtained by in vitro fertilization (IVF) using in vitro matured oocytes collected from ovaries obtained from a local slaughterhouse. The tightly connected cells of the blastocyst trophectoderm from large livestock species, such as pig and cattle, form a barrier that complicates cell microinjection into the blastocoel. Thus, microinjection often results in embryo collapse and the inability to deposit the cells into the embryo. To facilitate cell injection we employed a laser-assisted approach, using the laser to perforate the zona pellucida and to induce damage to a limited number of trophectoderm cells. This allowed for easy access into the blastocyst cavity for transferring the human cells (Figure S3B). Furthermore, the zona ablation and trophectoderm access allowed use a blunt-end pipette for cell transfer, thus minimizing further embryo damage. This method resulted in a nearly 100% injection effectiveness and >90% embryo survival.

To determine whether hiPSCs could engraft into the cattle ICM, we injected ten cells from each condition into cattle blastocysts collected 7 days after fertilization. After injection, we cultured these blastocysts for additional 2 days before analysis. We used several criteria to evaluate the chimeric contribution of hiPSCs to cattle blastocysts: (1) average number of human cells in each blastocyst, (2) average number of human cells in each ICM, (3) percentage of blastocysts with the presence of human cells in the ICM, (4) percentage of SOX2+ human cells in the ICM, and (5) percentage of human cells in the ICM that are SOX2+ (Figure 4C). Our results indicated that both naive and intermediate (but not primed) hiPSCs could survive and integrate into cattle ICMs, albeit with variable efficiencies (Figures 4D and S3C–S3E; Table S4). Compared with other cell types, 4i-hiPSCs exhibited the best survival (22/23 blastocysts contained human cells), but the majority of these cells lost SOX2 expression (only 13.6% of human cells remained SOX2+). On average, 3.64 4i-hiPSCs were incorporated into the ICM. NHSM-hiPSCs were detected in 46 of 59 injected blastocysts, with 14.41 cells per ICM. Of these, 89.7% remained SOX2+. For 2iLD-hiPSCs, 40 of 52 injected blastocysts contained human cells, with 5.11 cells per ICM, and 69.9% of the ICM-incorporated human cells remained SOX2+. FAC-hiPSCs exhibited moderate survival rate (65/101) and ICM incorporation efficiency (39/101), with an average of 2.31 cells incorporated into the ICM, and 89.3% remaining SOX2+.

We also performed ICM incorporation assays by injecting hiPSCs into pig blastocysts. Because certain complications are frequently associated with pig IVF (Abeydeera, 2002; Grupen, 2014) (e.g., high levels of polyspermic fertilization), we used a parthenogenetic activation model, which enabled us to efficiently produce embryos that developed into blastocysts (King et al., 2002). Pig oocytes were obtained from ovaries collected at a local slaughterhouse. Once the oocytes were matured in vitro, we removed the cumulus cells and artificially activated the oocytes using electrical stimulation. They were then cultured to blastocyst stage (Figure 4C). We injected ten hiPSCs into each pig parthenogenetic blastocyst and evaluated their chimeric contribution after 2 days of in vitro culture (Figures 4C and S3C–S3E; Table S4). Similar to the results in cattle, we found that hiPSCs cultured in 4i and NHSM media survived better and yielded a higher percentage of blastocysts harboring human cells (28/35 and 37/44, respectively). Also, among all blastocysts containing human cells, we observed an average of 9.5 cells per blastocyst for 4i-hiPSCs and 9.97 cells for NHSM-hiPSCs. For NHSM-hiPSCs, 19/44 blastocysts had human cells incorporated into the ICM. In contrast, only 6/35 blastocysts had 4i-hiPSCs localized to the ICM. For 2iLD-hiPSCs, we observed an average of 5.7 cells per blastocyst, with 2.25 human cells localized to the ICM. For FAC-hiPSCs, an average of 3.96 and 1.62 human cells were found in the blastocyst and ICM, respectively. Once incorporated into the ICM, 82.2%, 72%, 60.9%, and 40% of 2iLD-, 4i-, NHSM-, and FAC-hiPSCs, respectively, stained positive for the pluripotency

marker SOX2. These results indicate that both naive and intermediate hiPSCs seem to perform better when injected into cattle than pig blastocysts. This suggests a different in vivo blastocyst environment in pig and cattle, with the cattle blastocysts providing an environment that is more permissive for hiPSC integration and survival.

Chimeric Contribution of hiPSCs to Post-implantation Pig Embryos

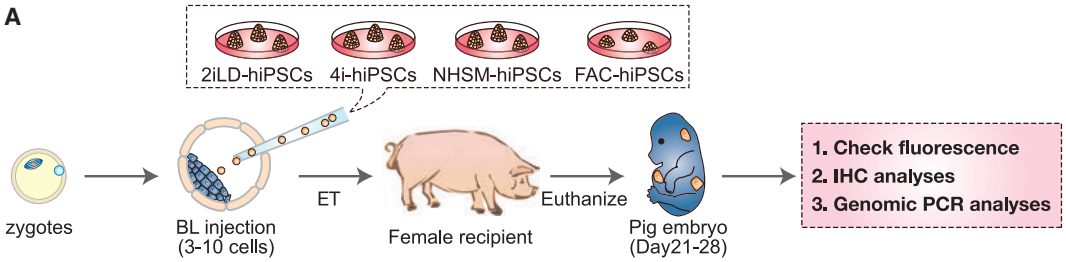
Although ICM incorporation of hiPSCs is the necessary first step to contribute to the embryo proper of host animals, it has limited predictive value for post-implantation chimera formation, as other factors are involved. Next, we investigated if any of the naive and intermediate hiPSCs that we generated, which showed robust ICM incorporation in pre-implantation blastocysts, could contribute to post-implantation development following ET. The pig has certain advantages over cattle for experiments involving post-implantation embryos, as they are a polytocus species, and are commonly used as a translational model given their similarities to humans concerning organ physiology, size, and anatomy. We thus chose the pig for these experiments. Since there was little to no contribution of primed hiPSCs, even at the pre-implantation blastocyst stage, we excluded these cells from the ET experiments. Pig embryos were derived in vivo or through parthenogenesis. A total of 167 embryo donors were used in this study, from which we collected 1,298 zygotes, 1,004 two-cell embryos and 91 morulae (Table S5). Embryos were cultured in vitro until they reached the blastocyst stage (Figures S4AA and S4B). Overall, 2,181 good quality blastocysts with a well-defined ICM were selected for subsequent blastocyst injections, of which 1,052 were derived from zygotes, 897 from two-cell embryos, 91 from morulae, and 141 from parthenogenetic activation (Table S5). We injected 3-10 hiPSCs into the blastocoel of each of these blastocysts (Figures 5A, S4A, and S4C; Table S6). After in vitro embryo culture, a total of 2,075 embryos (1,466 for hiPSCs; Table S6; 477 for rodent PSCs; Table S3) that retained good quality were transferred to surrogate sows. A total of 41 surrogate sows received 30–50 embryos each, resulting in 18 pregnancies (Table S6). Collection of embryos between day 21-28 of development resulted in the harvesting of 186 embryos: 43 from 2iLD-hiPSCs, 64 from FAC-hiPSCs, 39 from 4i-hiPSCs, and 40 from NHSM-hiPSCs (Figures 5B, S4A, S4D, and S4F). In addition, 17 control embryos were collected from an artificially inseminated sow (Figure 5B).

Following evaluating the developmental status of the obtained embryos, more than half showed retarded growth and were smaller than control embryos (Figures 5B and S4B), as was seen when pig blastocysts were injected with rodent PSCs (Figure 3B). Among different hiPSCs, embryos injected with FAC-hiPSCs were more frequently found to be normal size (Figure 5C). From the recovered embryos, and based on fluorescence imaging (GFP for 2iLD-hiPSCs and FAC-hiPSCs; hKO for 4i-hiPSCs and NHSM-hiPSCs), we observed positive fluorescence signal (FO+) in 67 embryos among which 17 showed a normal size and morphology, whereas the rest were morphologically underdeveloped (Figures 5B). In contrast, among fluorescence negative embryos we found the majority (82/119) appeared normal size (Figure 5E), suggesting contribution of hiPSCs might have interfered with normal pig development. Closer examination of the underdeveloped embryos revealed that 50 out of 87 were FO+ (Figures 5B). Among all the FO+ embryos the distribution of normal size versus growth retarded embryos for each cell lines was: 3:19 for 2iLD-hiPSCs, 7:14 for FAC-hiPSCs, 2:12 for 4i-hiPSCs, and 5:5 for NHSM-hiPSCs (Figure 5D). Among normal size embryos we found 3/13 from 2iLD-hiPSCs, 7/47 from FAC-hiPSCs, 2/14 from 4i-hiPSCs, and 5/25 from NHSM-hiPSCs that were FO+ (Figure 5B). All normal size FO+ embryos derived from 2iLD-hiPSCs, 4i-hiPSCs, or NHSM-hiPSCs showed a very limited fluorescence signal (Figure S5A). In contrast, normal size FO+ FAC-hiPSC-derived embryos typically exhibited a more robust fluorescence signal (Figures 6A and S5A).

Detecting fluorescence signal alone is insufficient to claim chimeric contribution of donor hiPSCs to these embryos, as auto-fluorescence from certain tissues and apoptotic cells can yield false positives, especially when chimerism is low. We thus sectioned all normal size embryos deemed positive based on the presence of fluorescence signal and subjected them to IHC analyses with antibodies detecting GFP or hKO. For 2iLD-hiPSC-, 4i-hiPSC-, and NHSM-hiPSC-derived embryos, in agreement with fluorescence signals observed in whole-embryo analysis, we detected only a few hKO- or GFP-positive cells in limited number of sections (Figure S5A). This precluded us from conducting further IHC analysis using lineage markers. For FAC-hiPSC-derived embryos, we confirmed via IHC analysis (using an anti-GFP antibody) that they contained more human cells (Figures 6A, S5A, and S5B). We then stained additional sections using antibodies against TUJ1, EPCAM, SMA, CK8, and HNF3β (Figures 6B and S5C) and observed differentiation of FAC-hiPSCs into different cell lineages. In addition, these cells were found negative for OCT4, a pluripotency marker (data not shown). Moreover, the presence of human cells was further verified with a human-specific HuNu antibody staining (Figure 6B) and a sensitive genomic PCR assay using a human specific *Alu* sequence primer (Figure 6C; Table S2). Together, these results indicate that naive hiPSCs injected into pig blastocysts inefficiently contribute to chimera formation, and are only rarely detected in post-implantation pig embryos. An intermediate hPSC type (FAC-hiPSCs) showed better chimeric contribution and differentiated to several cell types in post-implantation human-pig chimeric embryos. It should be noted that the levels of chimerism from all hiPSCs, including the FAC-hiPSCs, in pig embryos were much lower when compare to rat-mouse chimeras (Figures 1C, 1E, S1A, and 1B), which may reflect the larger evolutionary distance between human-pig than between rat-mouse.

DISCUSSION

Our study confirms that live rat-mouse chimeras with extensive contribution from naive rat PSCs can be generated. This is in contrast to earlier work in which rat ICMs were injected into mouse blastocysts (Gardner and Johnson, 1973). One possible explanation for this discrepancy is that cultured PSCs acquire artificial features that make them more proliferative and/or better able to survive than embryonic ICM cells, which in turn leads to their more robust xeno-engraftment capability in a mouse host.



B

Cell line	Blastocyst origin	Total		Fluorescence positive (FO+)	
		Normal size	Growth retarded	Normal size	Growth retarded
2iLD	Morula/EB	6	13	1	8
	2C	5	0	2	0
	ZG	2	17	0	11
	subtotal	13	30	3	19
FAC	BI	20	6	3	5
	2C	9	4	3	4
	2C	7	2	0	2
	2C	7	1	0	1
	ZG	0	1	0	0
	ZG	2	0	0	0
	PT	2	3	1	2
	subtotal	47	17	7	14
4i	ZG	6	6	0	1
	ZG	6	13	1	5
	ZG	2	6	1	6
	subtotal	14	25	2	12
NHSM	ZG	1	6	0	4
	2C	4	2	1	0
	2C	10	3	2	0
	2C	8	3	1	1
	PT	2	1	1	0
	subtotal	25	15	5	5
Total		**99**	**87**	**17**	**50**
Non-injected		17	0	0	0

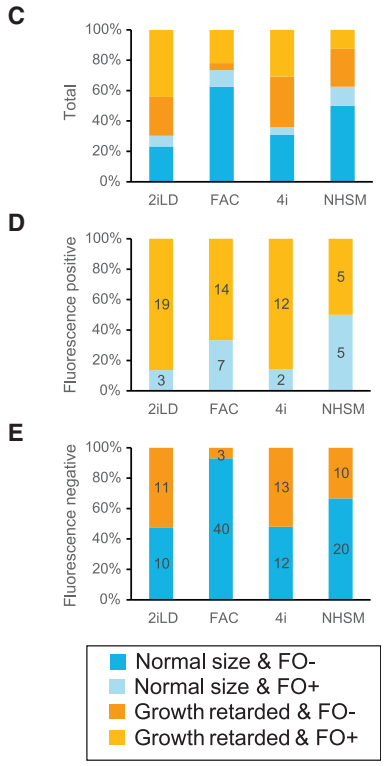


Figure 5. Generation of Post-implantation Human-Pig Chimeric Embryos

(A) Schematic of the experimental procedures for the generation and analyses of post-implantation pig embryos derived from blastocyst injection of naive and intermediate hiPSCs.

(B) Summary of the pig embryos recovered between day 21–28 of pregnancy.

(C) Bar graph showing proportions of normal size and growth retarded embryos, as well as the proportion of fluorescence-positive and -negative embryos, generated from different types of hiPSCs.

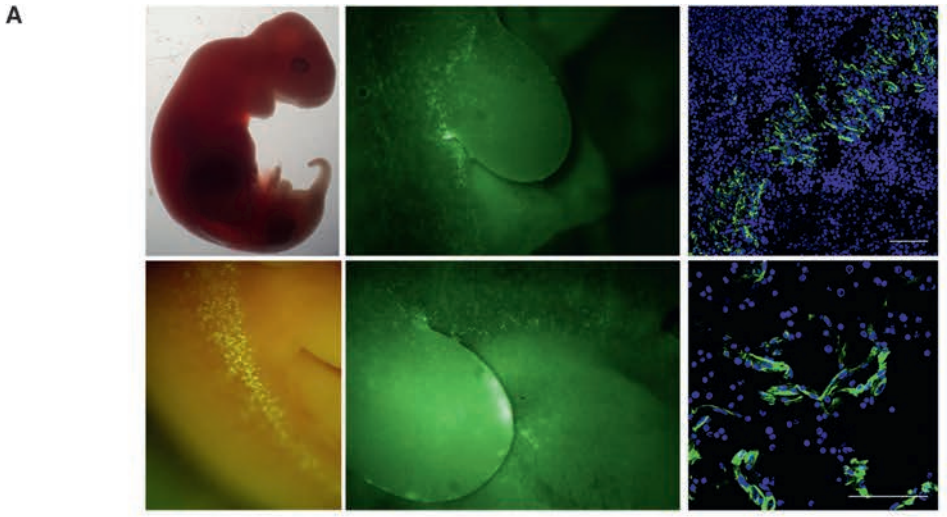
(D) Bar graph showing the proportion of normal size and growth-retarded embryos (among those exhibiting a fluorescence signal) generated from different types of hiPSCs.

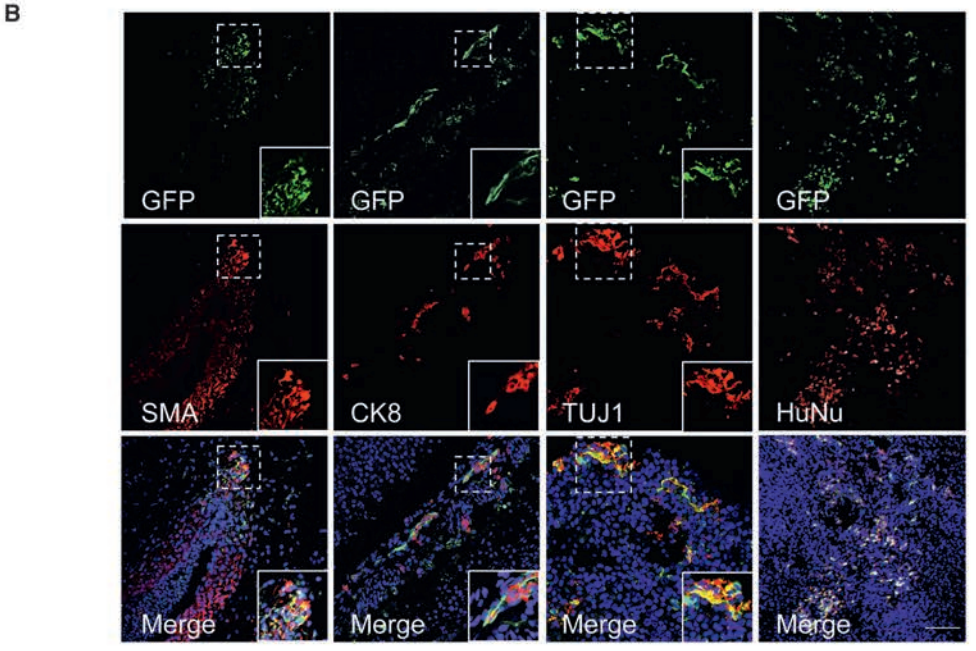
(E) Bar graph showing the proportion of normal-sized and growth-retarded embryos (among those without exhibiting a fluorescence signal) generated from different types of hiPSCs.

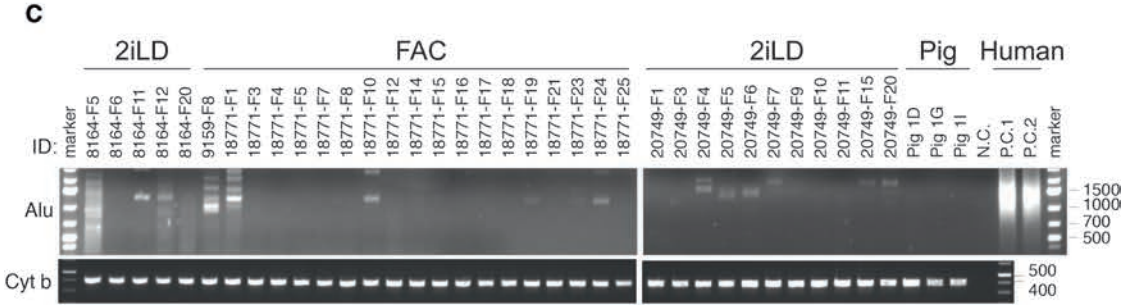
See also Figure S4 and Tables S5 and S6.

Rat-mouse chimeras generated by injecting donor rat PSCs into a mouse host were mouse-sized and developed into adulthood with apparently normal appearance and physiology. We further show in this study that a rat-mouse chimera could live a full mouse lifespan (about 2 years) and exhibit molecular signatures characteristic of aged cells. This demonstrates that cells from two different species, which diverged ~18 million years ago, can live in a symbiotic environment and are able to support normal organismal aging. The fact that rat PSCs were able to contribute to the mouse gallbladder, an organ that is absent in the rat, highlights the importance of embryonic niches in orchestrating the specification, proliferation, and morphogenesis of tissues and organs during organismal development and evolutionary speciation (Izpisúa-Belmonte et al., 1992).

Previous interspecies blastocyst complementation experiments generated host embryos by crossing heterozygous mutant mouse strains, which were themselves generated through targeted gene disruption in germline competent ESCs.







(legend on next page)

These experiments are labor intensive and time consuming. Moreover, only ~25% of blastocysts derived from genetic crosses are homozygous mutants, posing a limitation for efficient complementation. CRISPR-Cas9 mediated zygote genome editing offers a faster and more efficient one-step process for generating mice carrying homozygous mutations, thereby providing a robust interspecies blastocyst complementation platform. Additionally, the multiplexing capability of CRISPR-Cas9 (Cong et al., 2013; Yang et al., 2015) could potentially be harnessed for multilineage complementation. For example, in the case of the pancreas, one might hope to eliminate both the pancreatic parenchyma and vasculature of the host to generate a more complete xenogeneic pancreas. Despite the advantages, there are several technical limitations of the CRISPR-Cas9 blastocyst complementation system that need to be overcome before unlocking its full potential. First, gene inactivation relies on the error-prone, non-homologous end joining (NHEJ) pathway, which is often unpredictable. In-frame mutations and mosaicism are among the factors that may affect outcomes. A more predictable targeted gene inactivation strategy that utilizes homologous recombination (HR) is still inefficient in the zygote. Second, each embryo must be injected twice when using this system and embryos must be cultured in vitro for several days before ET, thereby compromising embryo quality. Technical advancements that include a more robust gene-disruption strategy (e.g., targeted generation of frameshift mutations via homology independent targeted integration [Suzuki et al., 2016]), alternative CRISPR/Cas9 delivery methods, and improved culture conditions for manipulated embryos will likely help improve and optimize the generation of organogenesis-disabled hosts.

We observed a slower clearance of an intraperitoneally injected glucose load for *Pdx1*$^{-/-}$ than *Pdx1*$^{+/-}$ rat-mouse chimeras, while both were slower than wild-type mouse controls (Figure S2C). While this result may seem to contradict a previous report (Kobayashi et al., 2010), the discrepancy is likely due to the development of autoimmune type inflammation that is often observed in adult rat-mouse (chimeras made by injection of rat PSCs into mouse blastocyst, data not shown) (>7 months, this study) and mouse-rat chimeras (chimeras made by injection of mouse PSCs into rat blastocyst; H. Nakauchi, personal communication), which is less evident in young chimeras (~8 weeks; Kobayashi et al. 2010). Interestingly though, we did observe a similarly slower clearance of glucose load in wild-type rats, although the initial spike was much lower in rats compared to mice or chimeras (Figure S2C). Thus, the rat cellular origin might also have played a role in the different GTT responses observed.

Rodent ESCs/iPSCs, considered as the gold standard cells for defining naive pluripotency, can robustly contribute to intra- and inter-species chimeras within rodent species. These and other results have led to the assumption that naive PSCs are the cells of choice when attempting to generate interspecies chimeras involving more disparate species. Here, we show that rodent PSCs fail to contribute to chimera formation when injected into pig blastocysts. This highlights the importance of other contributing factors underlying interspecies chimerism that may include, but not limited to, species-specific differences in epiblast and trophectoderm development, developmental kinetics, and maternal microenvironment.

To date, and taking into consideration all published studies that have used the mouse as the host species, it is probably appropriate to conclude that interspecies chimera formation involving hPSCs is inefficient (De Los Angeles et al., 2015). It has been argued that this apparent inefficiency results from species-specific differences between human and mouse embryogenesis. Therefore, studies utilizing other animal hosts would help address this important question. Here we focused on two species, pig and cattle, from a more diverse clade of mammals and found that naive and intermediate, but not primed, hiPSCs could robustly incorporate into pre-implantation host ICMs. Following ET, we observed, in general and similar to the mouse studies, low chimera forming efficiencies for all hiPSCs tested. Interestingly, injected hiPSCs seemed to negatively affect normal pig development as evidenced by the high proportion of growth retarded embryos. Nonetheless, we observed that FAC-hiPSCs, a putative intermediate PSC type between naive and primed pluripotent states, displayed a higher level of chimerism in post-implantation pig embryos. IHC analyses revealed that FAC-hiPSCs integrated and subsequently differentiated in host pig embryos (as shown by the expression of different lineage markers, and the lack of expression of the pluripotency marker OCT4). Whether the degree of chimerism conferred by FAC-hiPSCs could be sufficient for eliciting a successful interspecies human-pig blastocyst complementation, as demonstrated herein between rats and mice, remains to be demonstrated. Studies and approaches to improve the efficiency and level of hPSC interspecies chimerism (Wu et al., 2016), such as matching developmental timing, providing a selective advantage for donor hPSCs, generating diverse hPSCs with a higher chimeric potential and selecting a species evolutionarily closer to humans, among others parameters, will be needed.

The procedures and observations reported here on the capability of human pluripotent stem cells to integrate and differentiate in a ungulate embryo, albeit at a low level and efficiency, when

Figure 6. Chimeric Contribution of hiPSCs to Post-implantation Pig Embryos

(A) Representative bright field (left top) fluorescence (left bottom and middle) and immunofluorescence (right) images of GFP-labeled FAC-hiPSCs derivatives in a normal size day 28 pig embryo (FAC #1). Scale bar, 100 μm.

(B) Representative immunofluorescence images showing chimeric contribution and differentiation of FAC-hiPSCs in a normal size, day 28 pig embryo (FAC #1). FAC-hiPSC derivatives are visualized by antibodies against GFP (top), TUJ1, SMA, CK8 and HuNu (middle). (Bottom) Merged images with DAPI. Insets are higher magnification images of boxed regions. Scale bar, 100 μm.

(C) Representative gel images showing genomic PCR analyses of pig embryos derived from blastocyst injection of 2iLD-iPSCs (surrogates #8164 and #20749) and FAC-hiPSCs (surrogates #9159 and #18771) using a human specific Alu primer. A pig specific primer Cyt b was used for loading control. nc, negative control with no genomic DNA loaded. pc, positive controls with human cells. Pig 1D, 1G, and 1I, pig controls. ID, surrogate and pig embryos.

See also Figure S5 and Table S2.

optimized, may constitute a first step towards realizing the potential of interspecies blastocyst complementation with hPSCs. In particular, they may provide a better understanding of human embryogenesis, facilitate the development and implementation of humanized animal drug test platforms, as well as offer new insights on the onset and progression of human diseases in an in vivo setting. Ultimately, these observations also raise the possibility of xeno-generating transplantable human tissues and organs towards addressing the worldwide shortage of organ donors.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Rodents
  - Pigs
  - Human iPSC Culture Media
  - Culture and maintenance of rat ESCs/iPSCs and mouse iPSCs
- METHOD DETAILS
  - Chemicals Unless Otherwise Indicated, Chemicals Were Obtained from Sigma-Aldrich
  - Rat iPSC Generation
  - Human iPSC Generation
  - Generation of Fluorescently Labeled Rat PSCs and hiPSCs
  - Mouse Embryo Collection
  - sgRNA Design and In Vitro Transcription
  - Microinjection of Cas9 mRNA and sgRNAs to Mouse Zygotes
  - Microinjection of Rat PSCs to Mouse Blastocysts
  - Mouse Embryo Transfer
  - Genomic PCR
  - Quantitative Genomic PCR
  - Genotyping and DNA Sequencing
  - Glucose Tolerance Test
  - Cattle In Vitro Embryo Production
  - Pig Parthenogenetic Embryo Production
  - Microinjection of PSCs to Cattle and Pig Blastocysts and Embryo Culture
  - Pig and Cattle Blastocyst Immunostaining
  - Pig In Vivo Embryo Recovery and Transfer
  - Pig Parthenogenetic Embryo Transfer
  - Immunocytochemistry
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.12.036.

AUTHOR CONTRIBUTIONS

J.W. and J.C.I.B. conceived the study. J.W. generated and characterized all naive and intermediate hiPSC lines. K.S. generated and characterized primed hiPSCs. J.W. and T.H. generated rat iPSCs. J.W., A.P.-L., T.Y., M.M.V., D.O., A.O., P.R., C.R.E., J.W., and P.M.R. performed immunohistochemistry analyses of mouse and pig embryos. K.S., T.Y., E.S., A.P.-L., and M.M.V. performed genotyping, genomic PCR, and genomic qPCR analyses. A.S., M.S., and J.P.L. performed mouse Cas9/sgRNA injection, blastocyst injection, and embryo transfer. Y.S.B., M.S., and M.V. prepared hiPSCs, performed morulae and blastocyst injections, and analyzed hiPSC contribution to cattle and ppig ICMs. H.W. produced parthenogenetic pig embryos. D.A.S., Y.S.B., and M.V. produced cattle embryos. Work at UC Davis and University of Murcia was performed under the supervision of P.J.R. and E.A.M., respectively. E.A.M., M.A.G., C.C., I.P., C.A.M., S.S.B., A.N., and J.R. designed, coordinated, performed, and analyzed data related to pig embryo collection, embryo culture, blastocyst injection, embryo transfer, and embryo recover. E.N.D., J.L., I.G., P.G., T.B., M.L.M.-M., and J.M.C. coordinated work between Salk, and University of Murcia. J.W., P.J.R., and J.C.I.B. wrote the manuscript.

ACKNOWLEDGMENTS

J.C.I.B. dedicates this paper to Dr. Rafael Matesanz, Director of the Spain's National Organ Transplant Organization. Rafael's work has helped save thousands of patients in need of an organ. He constitutes a relentless inspiration for those of us trying to understand and alleviate human disease. The authors are grateful to Xiomara Lucas, Maria Dolores Ortega, Moises Gonzalvez, Jose Antonio Godinez, and Jesus Gomis for their assistance throughout this work. We thank the staff of the Agropor S.A. and Porcisan S.A. piggeries (Murcia, Spain) for the help and excellent management of animals. We thank Joan Rowe, Bret McNabb, Aaron Prinz, and Kent Parker and their crews for excellent assistance with embryo transfers and pig care at UC Davis. We thank Mako Yamamoto for help with mouse embryo dissection. We would like to thank Uri Manor of the Salk Waitt Advanced Biophotonics Core for technical advice on imaging analysis. We would like to thank the Salk Stem Cell Core for providing cell culture reagents. We would like to thank May Schwarz and Peter Schwarz for administrative help. We thank David O'Keefe for critical reading and editing of the manuscript. This experimental study was supported by The Fundación Séneca (GERM 19892/GERM/15), Murcia, Spain. The MINECO is acknowledged for their grant-based support (BES-2013-064087 and BES-2013-064069) (to C.A.M and A.N.). P.J.R was supported by a UC Davis Academic Senate New Research grant. Work in the laboratory of J.C.I.B. was supported by the UCAM, Fundacion Dr. Pedro Guillen, G. Harold and Leila Y. Mathers Charitable Foundation, and The Moxie Foundation.

Received: February 2, 2016
Revised: October 30, 2016
Accepted: December 22, 2016
Published: January 26, 2017

REFERENCES

Aasen, T., Raya, A., Barrero, M.J., Garreta, E., Consiglio, A., González, F., Vassena, R., Bilić, J., Pekarik, V., Tiscornia, G., et al. (2008). Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. Nat. Biotechnol. *26*, 1276–1284.

Abeydeera, L.R. (2002). In vitro production of embryos in swine. Theriogenology *57*, 256–273.

Chen, J., Lansford, R., Stewart, V., Young, F., and Alt, F.W. (1993). RAG-2-deficient blastocyst complementation: an assay of gene function in lymphocyte development. Proc. Natl. Acad. Sci. USA *90*, 4528–4532.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

De Los Angeles, A., Ferrari, F., Xi, R., Fujiwara, Y., Benvenisty, N., Deng, H., Hochedlinger, K., Jaenisch, R., Lee, S., Leitch, H.G., et al. (2015). Hallmarks of pluripotency. Nature *525*, 469–478.

Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., and Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. J. Virol. *72*, 8463–8471.

Funahashi, H., Ekwall, H., and Rodriguez-Martinez, H. (2000). Zona reaction in porcine oocytes fertilized in vivo and in vitro as seen with scanning electron microscopy. Biol. Reprod. *63*, 1437–1442.

Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., et al. (2013). Derivation of novel human ground state naive pluripotent stem cells. Nature *504*, 282–286.

Gardner, R.L., and Johnson, M.H. (1973). Investigation of early mammalian development using interspecific chimaeras between rat and mouse. Nat. New Biol. *246*, 86–89.

Gehring, W.J., and Ikeo, K. (1999). Pax 6: mastering eye morphogenesis and eye evolution. Trends Genet. *15*, 371–377.

Grupen, C.G. (2014). The evolution of porcine embryo in vitro production. Theriogenology *81*, 24–37.

Hansen, P.J. (2014). Current and future assisted reproductive technologies for mammalian farm animals. In Current and Future Reproductive Technologies and World Food Production, G. Cliff Lamb and N. DiLorenzo, eds. (Springer New York), pp. 1–22.

Hasler, J.F. (2014). Forty years of embryo transfer in cattle: a review focusing on the journal Theriogenology, the growth of the industry in North America, and personal reminisces. Theriogenology *81*, 152–169.

Hishida, T., Nozaki, Y., Nakachi, Y., Mizuno, Y., Okazaki, Y., Ema, M., Takahashi, S., Nishimoto, M., and Okuda, A. (2011). Indefinite self-renewal of ESCs through Myc/Max transcriptional complex-independent mechanisms. Cell Stem Cell *9*, 37–49.

Hockemeyer, D., Soldner, F., Cook, E.G., Gao, Q., Mitalipova, M., and Jaenisch, R. (2008). A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. Cell Stem Cell *3*, 346–353.

Holm, P., Booth, P.J., Schmidt, M.H., Greve, T., and Callesen, H. (1999). High bovine blastocyst development in a static in vitro production system using SOFaa medium supplemented with sodium citrate and myo-inositol with or without serum-proteins. Theriogenology *52*, 683–700.

Huang, Y., Osorno, R., Tsakiridis, A., and Wilson, V. (2012). In vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation. Cell Rep. *2*, 1571–1578.

Irie, N., Weinberger, L., Tang, W.W.C., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J.H., and Surani, M.A. (2015). SOX17 is a critical specifier of human primordial germ cell fate. Cell *160*, 253–268.

Isotani, A., Hatayama, H., Kaseda, K., Ikawa, M., and Okabe, M. (2011). Formation of a thymus from rat ES cells in xenogeneic nude mouse↔rat ES chimeras. Genes Cells *16*, 397–405.

Izpisúa-Belmonte, J.C., Brown, J.M., Crawley, A., Duboule, D., and Tickle, C. (1992). Hox-4 gene expression in mouse/chicken heterospecific grafts of signalling regions to limb buds reveals similarities in patterning mechanisms. Development *115*, 553–560.

Jonsson, J., Carlsson, L., Edlund, T., and Edlund, H. (1994). Insulin-promoter-factor 1 is required for pancreas development in mice. Nature *371*, 606–609.

King, T.J., Dobrinsky, J.R., Zhu, J., Finlayson, H.A., Bosma, W., Harkness, L., Ritchie, W.A., Travers, A., McCorquodale, C., Day, B.N., et al. (2002). Embryo development and establishment of pregnancy after embryo transfer in pigs: coping with limitations in the availability of viable embryos. Reproduction *123*, 507–515.

Kobayashi, T., Yamaguchi, T., Hamanaka, S., Kato-Itoh, M., Yamazaki, Y., Ibata, M., Sato, H., Lee, Y.-S., Usui, J., Knisely, A.S., et al. (2010). Generation of rat pancreas in mouse by interspecific blastocyst injection of pluripotent stem cells. Cell *142*, 787–799.

Kutner, R.H., Zhang, X.-Y., and Reiser, J. (2009). Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. Nat. Protoc. *4*, 495–505.

Li, P., Tong, C., Mehrian-Shai, R., Jia, L., Wu, N., Yan, Y., Maxson, R.E., Schulze, E.N., Song, H., Hsieh, C.-L., et al. (2008). Germline competent embryonic stem cells derived from rat blastocysts. Cell *135*, 1299–1310.

Lyons, I., Parsons, L.M., Hartley, L., Li, R., Andrews, J.E., Robb, L., and Harvey, R.P. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. Genes Dev. *9*, 1654–1666.

Martinez, E.A., Angel, M.A., Cuello, C., Sanchez-Osorio, J., Gomis, J., Parrilla, I., Vila, J., Colina, I., Diaz, M., Reixach, J., et al. (2014). Successful non-surgical deep uterine transfer of porcine morulae after 24 hour culture in a chemically defined medium. PLoS ONE *9*, e104696.

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell Stem Cell *4*, 487–492.

Offield, M.F., Jetton, T.L., Labosky, P.A., Ray, M., Stein, R.W., Magnuson, M.A., Hogan, B.L., and Wright, C.V. (1996). PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. Development *122*, 983–995.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313–317.

Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K., et al. (2011). A more efficient method to generate integration-free human iPS cells. Nat. Methods *8*, 409–412.

Park, I.H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141–146.

Parrish, J.J., Susko-Parrish, J.L., Leibfried-Rutledge, M.L., Critser, E.S., Eyestone, W.H., and First, N.L. (1986). Bovine in vitro fertilization with frozen-thawed semen. Theriogenology *25*, 591–600.

Parrish, J.J., Susko-Parrish, J., Winer, M.A., and First, N.L. (1988). Capacitation of bovine sperm by heparin. Biol. Reprod. *38*, 1171–1180.

Petters, R.M., and Wells, K.D. (1993). Culture of pig embryos. J. Reprod. Fertil. Suppl. *48*, 61–73.

Pursel, V.G., and Johnson, L.A. (1975). Freezing of boar spermatozoa: fertilizing capacity with concentrated semen and a new thawing procedure. J. Anim. Sci. *40*, 99–102.

Reubinoff, B.E., Pera, M.F., Fong, C.Y., Trounson, A., and Bongso, A. (2000). Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. Nat. Biotechnol. *18*, 399–404.

Ross, P.J., Ragina, N.P., Rodriguez, R.M., Iager, A.E., Siripattarapravat, K., Lopez-Corrales, N., and Cibelli, J.B. (2008). Polycomb gene expression and histone H3 lysine 27 trimethylation changes during bovine preimplantation development. Reproduction *136*, 777–785.

Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E.J., Hatanaka, F., Yamamoto, M., Araoka, T., Li, Z., et al. (2016). In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature *540*, 144–149.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell *158*, 1254–1269.

Tanaka, M., Chen, Z., Bartunkova, S., Yamasaki, N., and Izumo, S. (1999). The cardiac homeobox gene Csx/Nkx2.5 lies genetically upstream of multiple genes essential for heart development. Development *126*, 1269–1280.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196–199.

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic

identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell *15*, 471–487.

Theunissen, T.W., Friedli, M., He, Y., Planet, E., O'Neil, R.C., Markoulaki, S., Pontis, J., Wang, H., Iouranova, A., Imbeault, M., et al. (2016). Molecular criteria for defining the naive human pluripotent state. Cell Stem Cell *19*, 502–515.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–1147.

Tsukiyama, T., and Ohinata, Y. (2014). A modified EpiSC culture condition containing a GSK3 inhibitor can support germline-competent pluripotency in mice. PLoS ONE *9*, e95329.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910–918.

Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J.H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nat. Rev. Mol. Cell Biol. *17*, 155–169.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318–324.

Wu, J., and Izpisua Belmonte, J.C. (2015). Dynamic pluripotent stem cell states and their applications. Cell Stem Cell *17*, 509–525.

Wu, J., and Izpisua Belmonte, J.C. (2016). Stem cells: a renaissance in human biology research. Cell *165*, 1572–1585.

Wu, J., Okamura, D., Li, M., Suzuki, K., Luo, C., Ma, L., He, Y., Li, Z., Benner, C., Tamura, I., et al. (2015). An alternative pluripotent state confers interspecies chimaeric competency. Nature *521*, 316–321.

Wu, J., Greely, H.T., Jaenisch, R., Nakauchi, H., Rossant, J., and Belmonte, J.C. (2016). Stem cells and interspecies chimaeras. Nature *540*, 51–59.

Xiang, A.P., Mao, F.F., Li, W.-Q., Park, D., Ma, B.-F., Wang, T., Vallender, T.W., Vallender, E.J., Zhang, L., Lee, J., et al. (2008). Extensive contribution of embryonic stem cells to the development of an evolutionarily divergent host. Hum. Mol. Genet. *17*, 27–37.

Yang, L., Güell, M., Niu, D., George, H., Lesha, E., Grishin, D., Aach, J., Shrock, E., Xu, W., Poci, J., et al. (2015). Genome-wide inactivation of porcine endogenous retroviruses (PERVs). Science *350*, 1101–1104.

Yoshioka, K., Noguchi, M., and Suzuki, C. (2012). Production of piglets from in vitro-produced embryos following non-surgical transfer. Anim. Reprod. Sci. *131*, 23–29.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917–1920.

Zou, J., Maeder, M.L., Mali, P., Pruett-Miller, S.M., Thibodeau-Beganny, S., Chou, B.-K., Chen, G., Ye, Z., Park, I.H., Daley, G.Q., et al. (2009). Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. Cell Stem Cell *5*, 97–110.

Molecular Cell
Article

A G-Rich Motif in the lncRNA *Braveheart* Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage

Zhihong Xue,[1] Scott Hennelly,[2,3] Boryana Doyle,[4,5] Arune A. Gulati,[1] Irina V. Novikova,[2,6] Karissa Y. Sanbonmatsu,[2,3] and Laurie A. Boyer[1,7,8,*]
[1]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Los Alamos National Laboratory, Theoretical Biology and Biophysics Group, Los Alamos, NM 87545, USA
[3]New Mexico Consortium, Los Alamos, NM 87544, USA
[4]Undergraduate Research Opportunities Program
[5]Department of Physics
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[6]Pacific Northwest National Laboratory, Environmental Molecular Sciences Laboratory, Richmond, WA 99354, USA
[7]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[8]Lead Contact
*Correspondence: lboyer@mit.edu
http://dx.doi.org/10.1016/j.molcel.2016.08.010

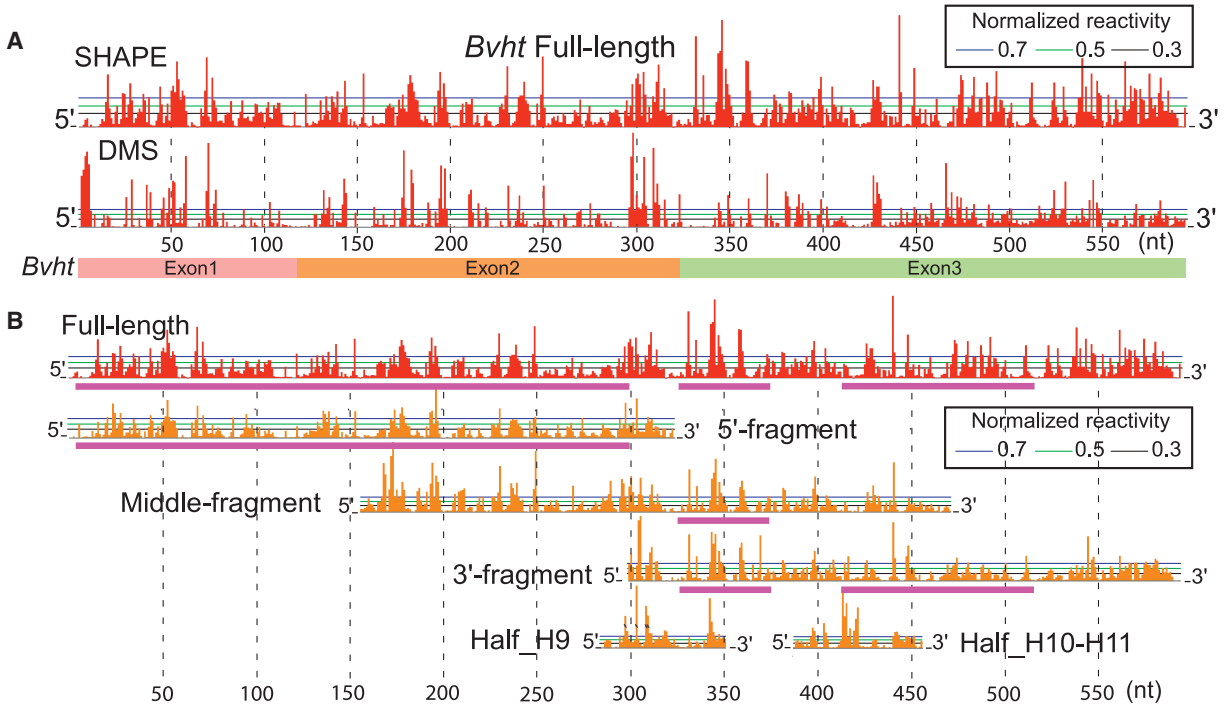
SUMMARY

Long non-coding RNAs (lncRNAs) are an emerging class of transcripts that can modulate gene expression; however, their mechanisms of action remain poorly understood. Here, we experimentally determine the secondary structure of *Braveheart* (*Bvht*) using chemical probing methods and show that this ~590 nt transcript has a modular fold. Using CRISPR/Cas9-mediated editing of mouse embryonic stem cells, we find that deletion of 11 nt in a 5′ asymmetric G-rich internal loop (AGIL) of *Bvht* (*bvht*^{dAGIL}) dramatically impairs cardiomyocyte differentiation. We demonstrate a specific interaction between AGIL and cellular nucleic acid binding protein (CNBP/ZNF9), a zinc-finger protein known to bind single-stranded G-rich sequences. We further show that CNBP deletion partially rescues the *bvht*^{dAGIL} mutant phenotype by restoring differentiation capacity. Together, our work shows that *Bvht* functions with CNBP through a well-defined RNA motif to regulate cardiovascular lineage commitment, opening the door for exploring broader roles of RNA structure in development and disease.

INTRODUCTION

Long non-coding RNAs (lncRNAs) have emerged as important regulators of development and disease. These transcripts are typically >200 nt in length and are often polyadenylated, capped, and alternatively spliced but lack coding potential (Ulitsky and Bartel, 2013). Although biochemical and biophysical studies of lncRNAs are in their early stages, proposed mechanisms of action include chromatin scaffolding, Polycomb complex (PRC2) recruitment to chromatin, mRNA decay, and decoys for proteins and micro RNAs (miRNAs) (Geisler and Coller, 2013; Quinn and Chang, 2016). Studies have highlighted diverse cellular roles for lncRNAs across eukaryotes such as X chromosome inactivation, genomic imprinting, cell-cycle regulation, embryonic stem cell (ESC) pluripotency, and lineage commitment (Flynn and Chang, 2014; Lee and Bartolomei, 2013). In metazoans, there is a growing number of lncRNAs that function in lineage commitment and differentiation with key examples in the cardiovascular system (Grote et al., 2013; Han et al., 2014; Klattenhoff et al., 2013), including many that show differential expression in cardiac disease (Fatica and Bozzoni, 2014; Rizki and Boyer, 2015). Thus, it remains a critical goal to understand how long non-coding transcripts contribute to regulation of cell fate and disease.

Comparative sequence analysis has facilitated RNA secondary structure predictions and has helped to reveal the functions of ribonuclease P and riboswitches (Gutell et al., 2002; Mian, 1997; Parsch et al., 2000). These structural predictions are also experimentally supported by chemical probing methods (e.g., in-line, SHAPE, DMS), NMR, and X-ray crystallography (Mondragón, 2013; Noller, 1984; Serganov and Patel, 2007). In contrast, predicting lncRNA secondary structure has been more complicated because these transcripts appear to be rapidly evolving and generally display low sequence conservation (Ponting et al., 2009). Recently, chemical probing methods have been exploited for studying lncRNA secondary structure. For example, selective 2′ hydroxyl acylation analyzed by primer extension (SHAPE) probing of in vitro transcripts showed that the lncRNAs *SRA* and *HOTAIR* display a complex structural organization that comprises a variety of elements comparable to well-folded RNAs like group II introns and ribosomal RNAs (Novikova et al., 2012; Somarowthu et al., 2015). Genome-wide probing of RNA secondary structure using dimethylsulfate sequencing (DMS-seq) or in vivo click SHAPE sequencing (icSHAPE-seq) has also been performed in living cells, revealing active unfolding of mRNA structures, suggesting that RNA structures contribute to



(legend on next page)

global RNA processing and translation (Ding et al., 2014; Rouskin et al., 2014; Spitale et al., 2015). Most lncRNAs, however, are not sufficiently abundant for detection in vivo, and in vivo secondary structure studies can be obfuscated in the cell by the binding of proteins to RNA. Overall, detailed analysis of the native structure of individual lncRNAs is still largely lacking and is necessary to gain deeper insights into their precise roles.

Our prior work identified the mouse lncRNA *Braveheart* (*Bvht*), which appears to act in *trans* to regulate cardiovascular lineage commitment (Klattenhoff et al., 2013). Given that lncRNAs are generally lowly conserved by sequence and that many of these transcripts are species specific (Johnsson et al., 2014; Ponting et al., 2009), RNA secondary structure is key for understanding their broader roles. To investigate the molecular mechanism of *Bvht* action, here we determined the secondary structure of in vitro transcribed full-length *Bvht* (~590 nt) using SHAPE and DMS probing and find that the transcript is organized into a highly modular structure including a 5′ asymmetric G-rich internal loop (AGIL). Using CRISPR/Cas9-mediated homology-directed repair (HDR), we deleted this loop (denoted $bvht^{dAGIL}$) in mouse ESCs (mESCs) and show that the AGIL motif is necessary for cardiomyocyte (CM) differentiation. Similar to short hairpin RNA (shRNA)-mediated *bvht* depletion, key cardiac transcription factors (TFs) fail to activate during the transition from nascent mesoderm to the cardiac progenitor (CP) state. Using a protein microarray platform, we demonstrate that the AGIL motif interacts with a small subset of factors including the heart-expressed zinc-finger TF cellular nucleic acid binding protein (CNBP/ZNF9), known to bind G-rich single-stranded nucleic acids (Calcaterra et al., 2010; Chen et al., 2007). Finally, we find that CNBP represses CM differentiation and that loss of CNBP partially rescues the $bvht^{dAGIL}$ phenotype, suggesting that these factors function together to specify the cardiovascular lineage. Our results show how a small RNA motif in *Bvht* can direct cell fate and demonstrate that structural studies combined with genetic perturbation can provide critical insights into lncRNA function.

RESULTS

Braveheart Is Organized into a Highly Modular Structure

RNA can form complex structures that have catalytic activity or that act as scaffolds for the binding of metal ions, small molecules, nucleic acids, and proteins (Mondragón, 2013; Noller, 1984; Serganov and Patel, 2007). To obtain the secondary structure of *Bvht*, we used the shotgun secondary structure determination strategy (3S) (Novikova et al., 2013), with the goal of obtaining more detailed mechanistic insight into *Bvht* function. First, we performed SHAPE probing (Deigan et al., 2009) (Figure 1A, top) and DMS probing (Tijerina et al., 2007) (Figure 1A, bottom) on in vitro transcribed full-length *Bvht* (Figure S1; Table S1, available online). We next repeated the SHAPE and DMS probing on shorter fragments (Table S1) to identify sub-domains of *Bvht*. When a region's reactivity in shorter fragments shows similarity to the profile in the full-length RNA, it suggests that this region adopts a modular fold in the context of full-length RNA structure. As shown in Figure 1B, we generated overlapping fragments and performed SHAPE probing as above. Detailed comparisons between each fragment and the full-length transcript revealed several regions of similar reactivity (Figure 1B). For example, the ~55 nt stretch at the 3′ end of *Bvht* exhibited high reactivity using both SHAPE and DMS probing, indicating a low probability of being structured, and was left out of the analysis. We obtained the fold for *Bvht* by coordinating the modular sub-folds.

The overall secondary structure shown in Figure 1C is most consistent with both our SHAPE and DMS analysis of full-length *Bvht* and of the shorter fragments. *Bvht* consists of 12 helices, 8 terminal loops, 5 sizeable (>5 nt) internal loops, and a five-way junction (5WJ). *Bvht* appears to be organized into three domains, roughly corresponding to its three exons: the 5′ domain (H1–H2), central domain (H3–H8), and 3′ domain (H9–H12) (Figure 1C). The 5′ domain contains an AGIL between H1 and H2, consisting of a large single-stranded region (14 nt) on the 5′ side and very short single-stranded region (3 nt) on the 3′ side. The central domain consists of a 5WJ (H4, H5, H6, H7, and H8) connected to the 5′ domain by H3. The 3′ domain contains four helices (H9, H10, H11, and H12).

Braveheart AGIL Motif Is Necessary for Proper ESC Differentiation

To date, lncRNA function has largely been determined by transcript knockdown or by genetic deletion of large regions that may encompass regulatory elements confounding phenotypic interpretation. We focused on dissecting the function of the AGIL region because it appeared to be less commonly represented in known RNA secondary structure databases and because G-rich regions often play regulatory roles in the genome (Aguilera and García-Muse, 2012; Rhodes and Lipps, 2015). For example, after searching the Gutell database of secondary structures of ribosomal and RNase P RNAs (Cannone et al., 2002), we found that only 13 of >400,000 asymmetric 5′ internal loops had similar size and asymmetry. The crystal structure of one such loop was recently solved, forming an intricate tightly packed configuration of purines (Ren et al., 2016). Thus, using

Figure 1. *Bvht* Secondary Structure Determination by Chemical Probing

(A) Normalized SHAPE (top) and DMS (bottom) probing reactivity profiles of full-length *Bvht*. Horizontal lines indicate normalized dimensionless reactivity. Both traces were normalized by the reactivities for highly reactive nucleotides. Nucleotides that have a normalized reactivity >0.5 are considered as highly flexible and likely represent single-stranded regions. Positions of *Bvht* exons are labeled below the reactivity profile.

(B) Shotgun secondary structure (3S) analysis of *Bvht*. Normalized SHAPE probing reactivity of indicated *Bvht* fragments is compared to full-length transcript. Full length, 1–590; 5′ fragment, 1–325; middle fragment, 155–475; 3′ fragment, 300–590; Half_H9, 282–349; and Half_H10-H11, 380–457. The sub-regions with highly similar reactivity patterns to full-length transcript are highlighted in purple under the reactivity profile.

(C) Secondary structure of *Bvht* was derived with 3S via SHAPE and DMS chemical probing experiments. The normalized SHAPE or DMS reactivity is represented by indicated colors. Circle, SHAPE; diamond, DMS. The AGIL motif is highlighted by red dashed lines. H1 to H12 indicates the helices.

See also Figure S1.

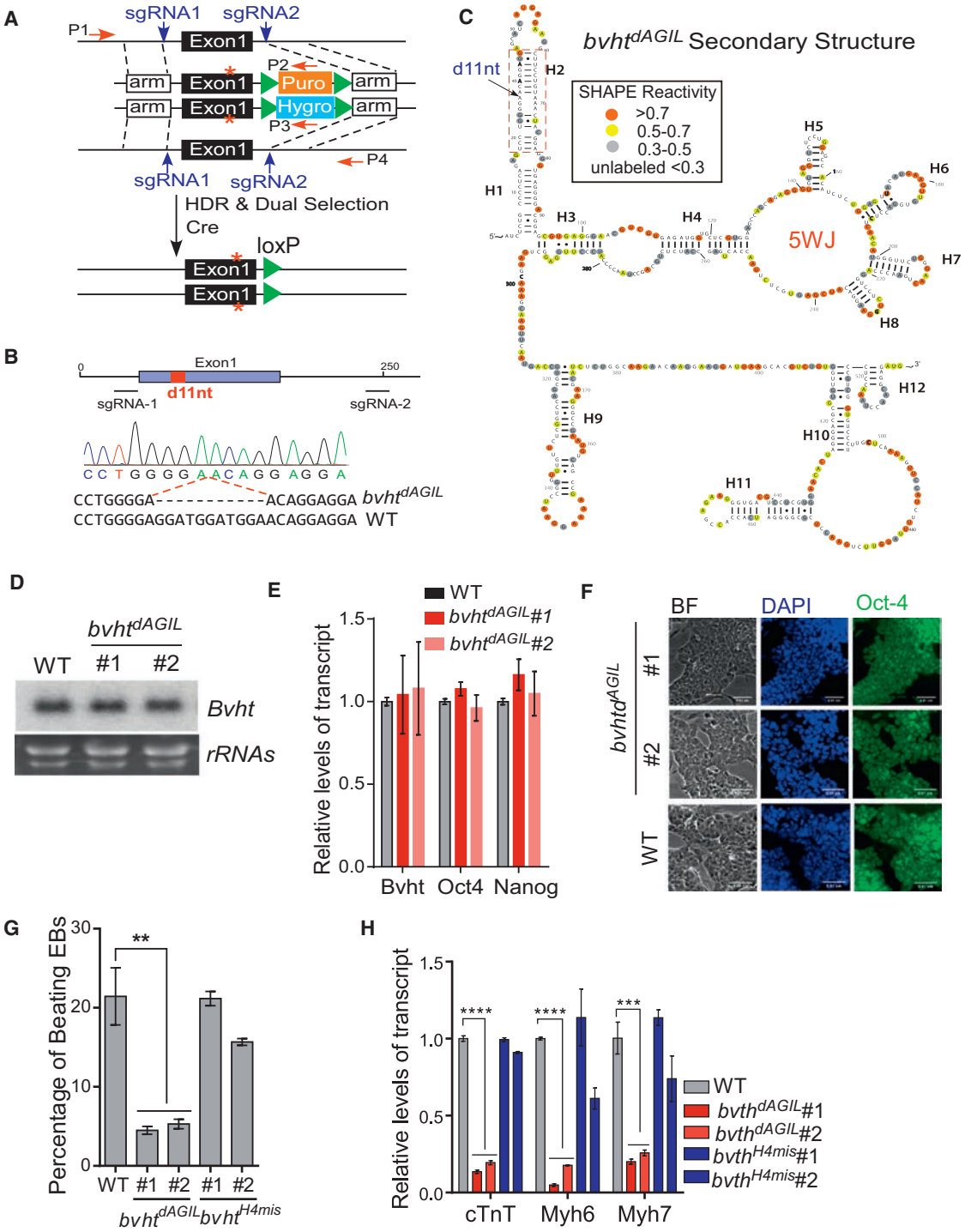


Figure 2. *Bvht* AGIL Motif Is Necessary for Formation of Contracting EBs

(A) Schematic showing the strategy of introducing mutations in *Bvht* endogenous locus. Two small guide RNAs (sgRNAs) and two repair templates including different selection cassettes (Puro or Hygro) as indicated are applied for CRISPR/Cas9-mediated HDR. After dual selection, both alleles will be mutated at designated loci. The selection cassettes are then removed by Cre recombinase-mediated recombination. Asterisk, mutations; triangle, loxP site; P1, P2, P3, and P4 are primers for PCR-based screening.

(B) Diagram showing the positions of sgRNAs and d11nt in *Bvht* endogenous locus. Partial DNA sequencing trace of the PCR product of *bvht*^{dAGIL} ESC genomic DNA.

(C) Secondary structure of *bvht*^{dAGIL} was derived from SHAPE probing experiment. d11nt indicates the deleted 11 nt sequences from AGIL motif.

(legend continued on next page)

CRISPR/Cas9 HDR, we generated an 11 nt deletion in AGIL (denoted *bvht*^{dAGIL}) at the endogenous *Bvht* locus in mESCs to disrupt this loop (Figures 2A and 2B). We used a dual selection strategy to facilitate recovery of homozygous clones (~20%–50% frequency) and expanded several clones for experimental evaluation.

SHAPE probing of *bvht*^{dAGIL} RNA shows deletion of the AGIL motif does not destabilize overall *Bvht* structure (Figures 2C and S2A). Thus, we next examined *Bvht* levels in ESCs and found that the mutant transcript was expressed at comparable levels to wild-type (WT) by northern blot and qRT-PCR (Figures 2D and 2E). Similar to shRNA-mediated depletion of *Bvht* (Klattenhoff et al., 2013), *bvht*^{dAGIL} did not affect expression of pluripotency markers such as Oct4 and Nanog, and mutant ESCs showed normal morphology and self-renewal properties as well as typical cell-cycle kinetics (Figures 2E, 2F, and S2B). We then tested whether *bvht*^{dAGIL} could form embryoid bodies (EBs), which give rise to derivatives of all three germ layers. Notably, CMs can form in EBs and can be visualized as beating cell clusters. We allowed WT and mutant ESCs to aggregate in the absence of pluripotency growth factors and then measured the percentage of spontaneously beating EBs at different time points. We found that *bvht*^{dAGIL} EBs show significantly reduced beating (~5%) compared to WT cells (~25%) at day 10, similar to our observations in *Bvht*-depleted EBs (Klattenhoff et al., 2013).

Helical junctions are often important for the structural and catalytic properties of RNAs (Bindewald et al., 2008). For example, a four-way junction promotes the functional folded state of the hairpin ribozyme (Tan et al., 2003). Thus, we also introduced mismatch mutations into the H4 region (*bvht*^{H4mis}) to destabilize the 5WJ by CRISPR/Cas9-mediated HDR and selected clones (Figures S2C–S2E). In contrast to *bvht*^{dAGIL}, alteration of the H4 region did not significantly affect the percentage of beating EBs (Figure 2G). *Bvht*^{dAGIL} EBs also displayed a failure to activate genes associated with the cardiac contractile apparatus such as cardiac troponin T (cTnT) and myosin heavy chain genes, whereas *bvht*^{H4mis} EBs showed normal expression comparable to WT controls (Figure 2H). In contrast, neuronal and endodermal genes were expressed normally in *bvht*^{dAGIL} EBs in response to retinoic acid treatment similar to WT and *bvht*^{H4mis} EBs (Figures S2F and S2G). Although these data do not preclude a secondary role for the 5WJ, our analysis suggests that the *Bvht* AGIL motif is required for formation of spontaneously contracting EBs.

Braveheart AGIL Motif Is Necessary for Cardiovascular Lineage Commitment

To further dissect AGIL function in the cardiovascular lineage, we employed a directed in vitro CM differentiation assay that permits isolation of cell populations at well-defined stages (ESCs, precardiac mesoderm [MES], CPs, and CMs) (Kattman et al., 2011; Wamstad et al., 2012) (Figure 3A). At each stage, cells are subject to fluorescence-activated cell sorting (FACS) using antibodies against specific markers to quantify differentiation efficiency. Using this approach, we routinely isolate a high percentage of Pdgfra+, Flk1+ (MES), Nkx2.5-GFP+ (CP), and cTnT+ (CM) cell populations (Figure 3B). In contrast, FACS of *bvht*^{dAGIL} cells showed a striking reduction in the percentage of CP and CM marked cells during differentiation. We also demonstrate that although *bvht*^{dAGIL} and WT cells showed similar morphology at day 4 (MES), immunofluorescence of the cultures at day 5.3 (CP) and day 10 (CM) using antibodies against Nkx2.5-GFP or cTnT, respectively, showed no staining in the mutant cells (Figure 3C). These results are highly reproducible among multiple independent *bvht*^{dAGIL} ESC clones and similar to shRNA depletion of *Bvht* (Figures S3A, S3B, and S3E), suggesting that the differentiation defects are not due to off-target effects.

We next analyzed the expression of a set of cardiac TFs that failed to activate upon shRNA-mediated depletion of *Bvht* (Figures S3C and S3D) (Klattenhoff et al., 2013). The mesodermal marker Brachyury showed higher expression at day 4 in *bvht*^{dAGIL} cells and sustained expression at day 5.3 compared to WT controls (Figure 3D). MesP1 is one of the earliest known markers of a common multi-potent cardiovascular progenitor (Bondue et al., 2008; Lindsley et al., 2008) and showed decreased expression at day 4 (MES) in *bvht*-shRNA-depleted cells (Klattenhoff et al., 2013). Although MesP1 expression showed no change in the *bvht*^{dAGIL} mutant, we observed a failure to activate the cardiac TFs downstream of this factor, including Nkx2.5, Gata4, Gata6, Hand1, Hand2, Tbx5, and Mef2c, compared to WT cells, suggesting that distinct regions of *Bvht* contribute to its total activity (Figure 3E). These data are highly reproducible using multiple independent ESC clones (Figure S3F). Moreover, expression of WT *Bvht* from the ROSA26 locus in the *bvht*^{dAGIL} background (Figure S3G) rescued the CM differentiation defect, indicating that the phenotype is due to loss of AGIL function (Figures S3H–S3K). Together, our data point to a central role for the *Bvht* AGIL motif in specifying the cardiovascular lineage.

Braveheart AGIL Interacts with Factors that Bind G-Rich Nucleic Acids

A prevailing model suggests that lncRNAs act as molecular scaffolds, mediating interactions with proteins (Geisler and Coller, 2013; Quinn and Chang, 2016; Rinn and Chang, 2012). Although genome-wide studies support binding between lncRNAs and proteins, few studies have identified RNA structural motifs

(D) Northern blot analysis showing the levels of *Bvht* transcripts in indicated ESC lines. rRNAs are used for loading control.
(E) qRT-PCR analysis showing the levels of *Bvht* and ESC pluripotency markers Oct4 and Nanog in indicated ESC lines. Experiments were performed in triplicate and data are represented as mean values ± SD.
(F) Immunofluorescence staining of indicated ESCs using Oct4 antibody. Nuclei were stained with DAPI. BF, bright field. Scale bar, 100 μm.
(G) Percentage of spontaneously contracting embryoid bodies (EBs) at day 12 of differentiation (n > 200) from indicated ESCs.
(H) qRT-PCR analysis of EBs at day 12 showing the relative levels of CM markers from indicated ESC lines.
All experiments were performed in triplicate and data are represented as mean values ± SD. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001 (two-tailed Student's t test).
See also Figure S2.

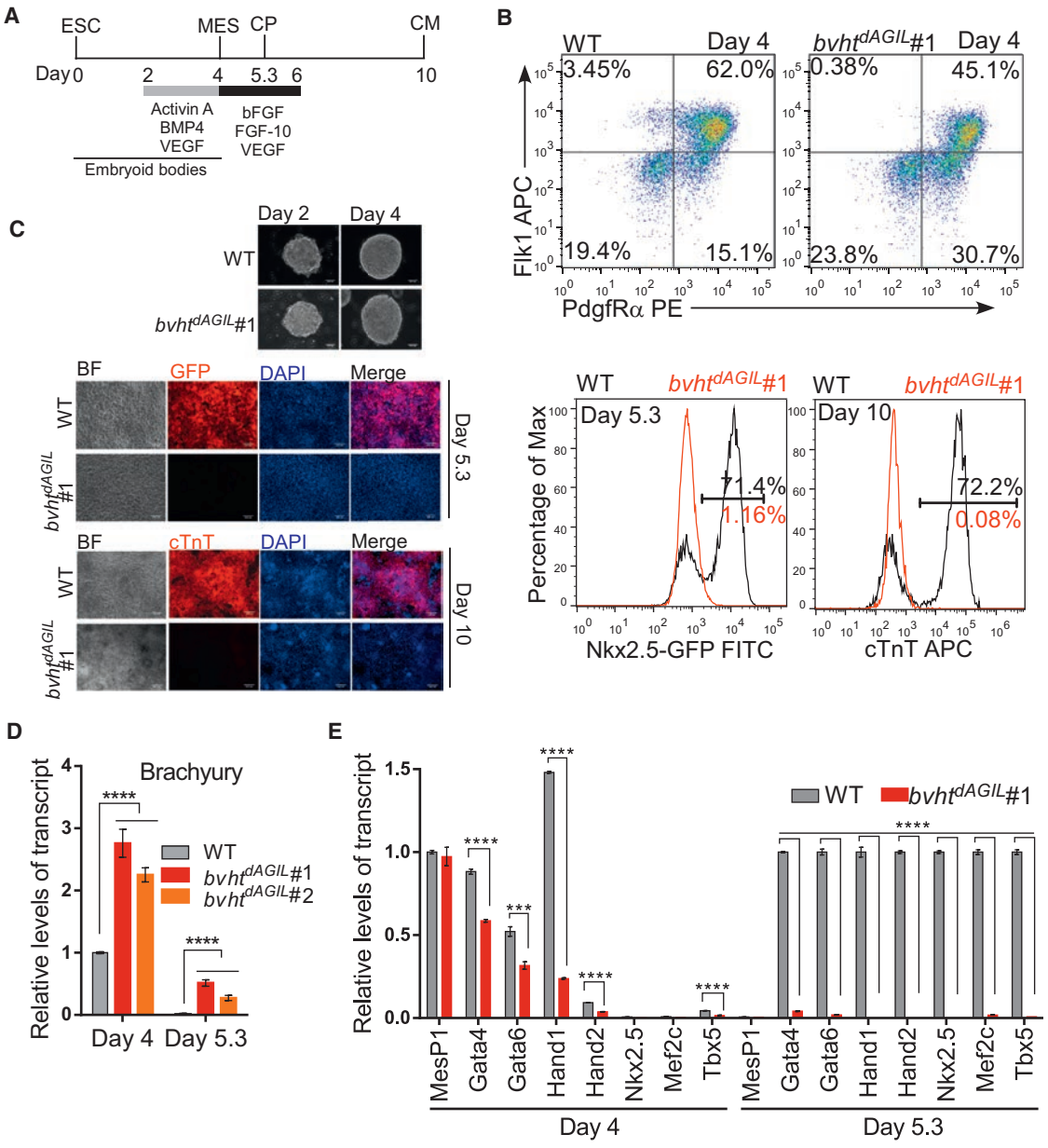


Figure 3. *Bvht* AGIL Motif Is Necessary for Cardiovascular Lineage Commitment

(A) Timeline of CM differentiation protocol. Black and gray bars represent the time period where differentiating cultures were treated with the growth factors listed below each respective bar.

(B) Cells at indicated time points were analyzed for marker expression by flow cytometry. Numbers in plots indicate percentage of gated populations.

(C) Immunofluorescence staining of indicated cells using anti-GFP (day 5.3) and anti-cTnT (day 10) antibodies. Nuclei were stained with DAPI. BF, bright field. Scale bar, 100 μm.

(D and E) qRT-PCR analysis showing the relative levels of Brachyury (D) and cardiac (E) TFs at day 4 and day 5.3 from indicated ESC lines. WT value at day 4 (D) and day 5.3 (E) is set to 1 for each gene.

All experiments were performed in triplicate and data are represented as mean values ± SD. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001 (two-tailed Student's t test).

See also Figure S3.

responsible for these interactions (Chu et al., 2015). To identify proteins that potentially interact with the *Bvht* AGIL motif, we used a human protein microarray platform that has successfully identified lncRNA-binding proteins (Kretz et al., 2013; Siprashvili et al., 2012). Full-length *Bvht* and *bvht*^{dAGIL} transcripts were generated by in vitro transcription and labeled with Cy5 (Figure S4A; Table S2). Equal concentrations of labeled transcript were then individually incubated with the protein microarray

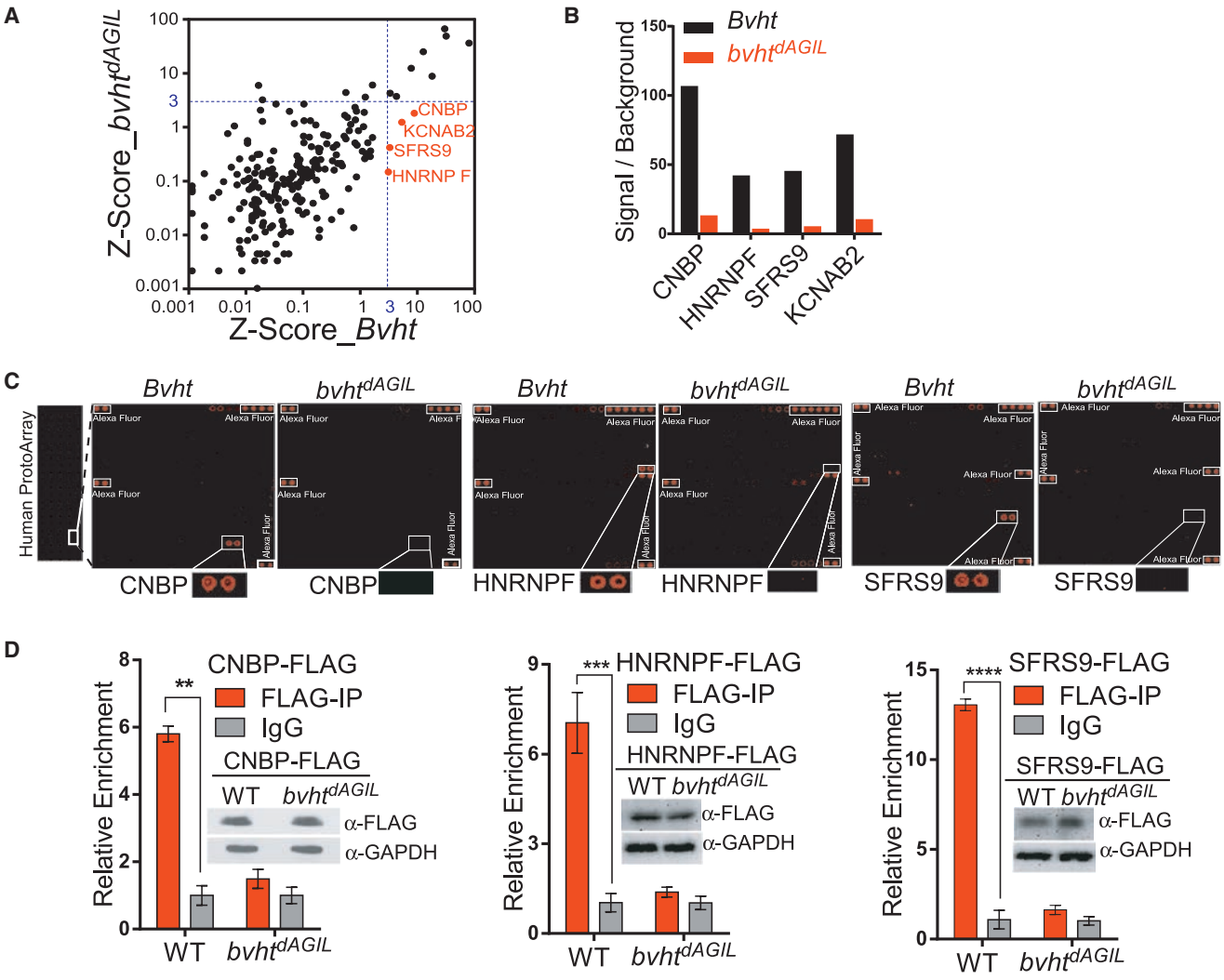


Figure 4. *Bvht* Interacts with CNBP, a Zinc-Finger TF

(A) Protein microarray analysis detecting *Bvht*-interacting proteins. Cy5-labeled *Bvht* and *bvht*^{dAGIL} transcripts were incubated with a human recombinant protein microarray. *Z* scores of fold-change signal intensity over background are depicted in the scatterplot. The dashed blue line represents the *Z* score cutoff used to select significant RNA-protein binding events. The significant *Bvht*-binding proteins are colored in red. The horizontal axis is *Bvht*, and the vertical axis for *bvht*^{dAGIL}. A logarithmic scale was used to display both axes.

(B) Quantification of human protein microarray showing fold changes of signal intensity over background for indicated proteins. Values are the average of duplicate protein spots.

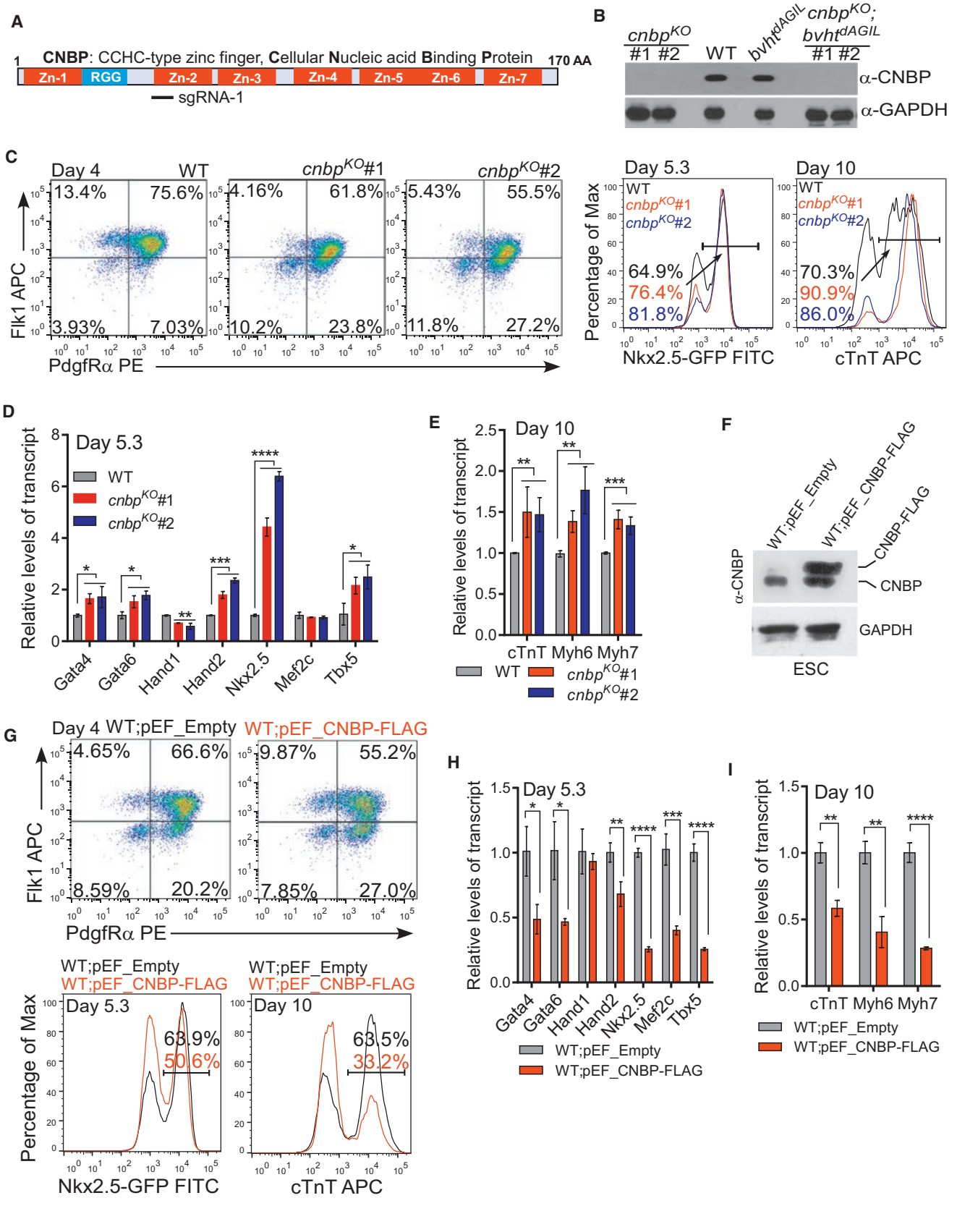
(C) Image of human protein microarray (left) and enlarged subarray (right) showing that mutation of AGIL motif dramatically reduces the interaction between *Bvht* and CNBP, HNRNPF, and SFRS9. Alexa Fluor 647-labeled rabbit anti-mouse IgG or anti-human IgA2 in corners and middle edge of each subarray are used for reference.

(D) RNA immunoprecipitation showing the interaction between *Bvht* and CNBP, HNRNPF, and SFRS9 in ESCs. Flag-tagged CNBP, HNRNPF, or SFRS9 was constitutively expressed in both WT and *bvht*^{dAGIL} ESCs. Immunoblot analysis using anti-Flag antibody shows equal expression levels of Flag-tagged CNBP, HNRNPF, and SFRS9 in indicated ESCs. Mouse IgG was used for negative control.

All experiments were performed in triplicate and data are represented as mean values ± SD. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001 (two-tailed Student's t test).

See also Figure S4 and Table S2.

containing ~9,400 recombinant human proteins (Human ProtoArray). Using a stringent cut-off (*Z* score > 3), we identified 12 candidates that strongly interacted with the WT transcript (Figures 4A and S4B; Table S2). Notably, four of these candidates (CNBP, HNRNPF, SFRS9, and KCNAB2) showed dramatically decreased binding when the array was probed with the *bvht*^{dAGIL} transcript (Figures 4B and 4C). These proteins are conserved between mouse and human and are highly expressed across the differentiation time course except for KCNAB2 (Figure S4C). We previously showed that *Bvht* interacts with the Polycomb



(legend on next page)

repressive complex (PRC2) (Klattenhoff et al., 2013); however, we did not detect any change in the interaction with PRC2 in the mutant ESCs by RNA immunoprecipitation (RIP) (data not shown), suggesting that AGIL is not required for this interaction and that cooperation between *Bvht* and PRC2 may be a later event in regulating CM differentiation.

We next validated the interaction between *Bvht* and mouse CNBP, HNRNPF, and SFRS9 by expressing mouse Flag-tagged versions of these factors in both WT and $bvht^{dAGIL}$ ESCs followed by immunoprecipitation using an anti-Flag antibody (Figure 4D). We found that all three candidates co-purified with WT *Bvht*, but not $bvht^{dAGIL}$, as shown by qRT-PCR. Upon analysis of ProtoArray results available for ~20 distinct non-coding RNAs (Kretz et al., 2013; Marques Howarth et al., 2014; Siprashvili et al., 2012), CNBP and HNRNPF binding appeared to be highly specific to *Bvht*, whereas SFRS9 interacted broadly with other non-coding RNAs. HNRNPF, a member of ubiquitously expressed heterogeneous nuclear ribonucleoproteins family, is an RNA-binding protein with roles in mRNA splicing and mRNA metabolism and transport, and can bind G-rich sequences (Matunis et al., 1994; Reznik et al., 2014; Wang et al., 2012). CNBP (ZNF9) is a zinc-finger TF containing seven CCHC-type zinc fingers and one RNA recognition motif (RGG) (Figure 5A) that also binds G-rich single-stranded DNA and RNA (Armas et al., 2008; Calcaterra et al., 2010). CNBP has roles in neural crest cell expansion, and null mice die around embryonic day 10.5 (E10.5) (Chen et al., 2003; Weiner et al., 2007, 2011); however, its overall function is poorly characterized. Notably, CNBP is highly expressed in heart and skeletal muscle, and heterozygous $cnbp^{+/-}$ mice exhibit severe dilated cardiomyopathy (Chen et al., 2007). Moreover, CNBP is currently the only known gene linked to myotonic dystrophy type 2 in human, and patients often display severe heart defects (Jones et al., 2011; Lee et al., 2012; Liquori et al., 2001). Thus, given its binding preference for single-stranded G-rich nucleic acids and its understudied roles in the heart, we focused on further characterization of CNBP.

CNBP Represses CM Differentiation

To test the function of CNBP in our system, we introduced small indels using CRISPR/Cas9 genome editing in both WT and $bvht^{dAGIL}$ ESCs, generating $cnbp^{KO}$ and $cnbp^{KO};bvht^{dAGIL}$ ESCs (Figures 5A and S5A). Clones were sequenced for the presence of the mutations and immunoblot confirmed loss of CNBP in both $cnbp^{KO}$ and $cnbp^{KO};bvht^{dAGIL}$ ESCs (Figure 5B). Importantly, neither disruption of the *Bvht* AGIL motif nor $cnbp^{KO}$ affected the expression of either CNBP or *Bvht*, respectively (Figures 5B and S5B). Moreover, loss of CNBP did not affect the expression of ESC pluripotency markers Oct4 and Nanog, similar to $bvht^{dAGIL}$ (Figure S5B).

We next tested two independent $cnbp^{KO}$ ESC clones for their ability to differentiate into CMs. As shown in Figure S5C, $cnbp^{KO}$ cells show similar morphologies to WT cells at both day 2 and day 4 of differentiation and are fully capable of differentiating into CPs at day 5.3 and CMs at day 10, as shown by immunofluorescence analysis of Nkx2.5-GFP and cTnT, respectively. In fact, $cnbp^{KO}$ cells generate significantly higher percentages of Nkx2.5-GPF+ cells (CP) at day 5.3 and cTnT+ cells (CM) at day 10 by FACS when compared to WT cells (Figure 5C). Moreover, qRT-PCR analysis showed that cardiac TFs (e.g., Nkx2.5, Gata4, Gata6, Hand2, and Tbx5) at day 5.3 and CM marker genes (e.g., cTnT, Myh6, and Myh7) at day 10 exhibit higher expression levels in $cnbp^{KO}$ cells compared to WT cells (Figures 5D and 5E).

To further test CNBP function, we constitutively overexpressed Flag-tagged CNBP in WT ESCs, which did not affect the expression levels of *Bvht* and ESC pluripotency markers Oct4 and Nanog (Figure S5D). In contrast to $cnbp^{KO}$ ESCs, cells expressing higher levels of CNBP produced significantly lower percentages of Nkx2.5-GPF+ cells (CP) at day 5.3 and cTnT+ cells (CM) at day 10 compared to control cells by FACS (Figures 5G and S5E). Consistent with these data, cardiac TFs and CM marker genes showed decreased expression levels by qRT-PCR upon CNBP overexpression (Figures 5H and 5I). Together, our data suggest that CNBP functions, in part, as a negative regulator of cardiovascular lineage commitment.

Loss of CNBP Partially Rescues the $bvht^{dAGIL}$ Phenotype

Based on the above results, we hypothesized that *Bvht* may functionally antagonize CNBP to promote cardiovascular lineage commitment, predicting that loss of CNBP would rescue the $bvht^{dAGIL}$ mutant phenotype. To test this idea, we first performed EB differentiation of $cnbp^{KO};bvht^{dAGIL}$ ESCs compared to WT ESCs. At day 12 of EB differentiation, the expression levels of CM marker genes including cTnT, Myh6, and Myh7 were significantly restored in the $cnbp^{KO};bvht^{dAGIL}$ double mutant cells compared to $bvht^{dAGIL}$ single mutant (Figure S6A). We then performed the CM differentiation assay and found that the $cnbp^{KO};bvht^{dAGIL}$ double mutants produced significantly increased percentages of CP and CM cells compared to the $bvht^{dAGIL}$ mutant alone (Figure 6A). Nkx2.5 is expressed throughout the CP-to-CM stages (Ma et al., 2008; Wamstad

Figure 5. CNBP Represses CM Differentiation

(A) Diagram of CNBP (Uniprot: P53996-2) functional domains, including seven CCHC zinc fingers (aa 4–21, 45–62, 65–82, 89–106, 110–127, 128–145, and 149–166) and RGG box of RNA binding (aa 22–35). The target sequence of CNBP_sgRNA-1 is labeled on the bottom.

(B) Immunoblot analysis with anti-CNBP antibody showing the protein levels of CNBP in indicated ESC lines. GAPDH was used as loading control.

(C) Cells at indicated time points were analyzed for marker expression by flow cytometry. Numbers in plots indicate percentage of gated populations.

(D and E) qRT-PCR analysis showing the relative levels of cardiac marker genes at day 5.3 (D) and day 10 (E) of CM differentiation.

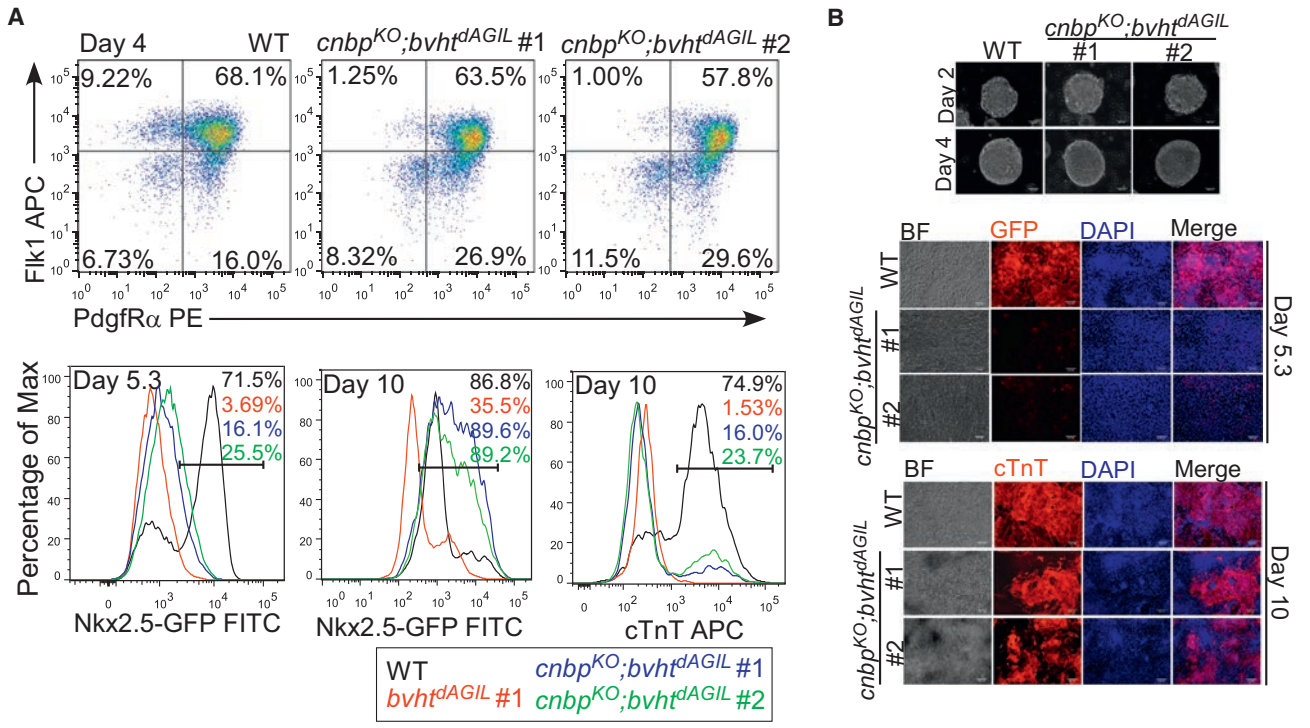
(F) Immunoblot analysis with anti-CNBP antibody showing the protein levels of endogenous CNBP and recombinant CNBP-FLAG in ESCs. GAPDH was used as loading control.

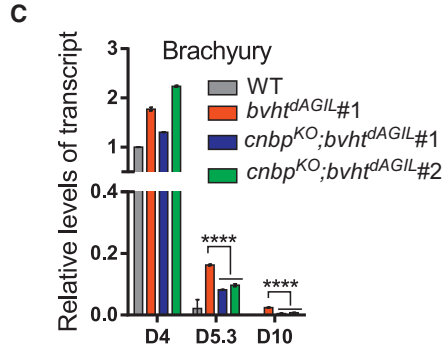
(G) Cells at indicated time points were analyzed for marker expression by flow cytometry. Numbers in plots indicate percentage of gated populations.

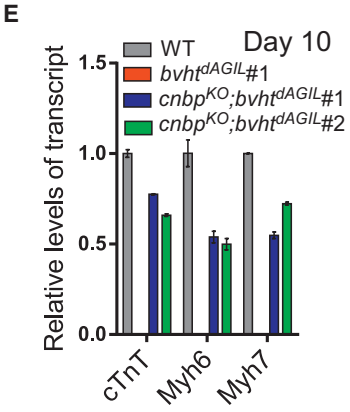
(H and I) qRT-PCR analysis showing the relative levels of cardiac marker genes at day 5.3 (H) and day 10 (I) of CM differentiation.

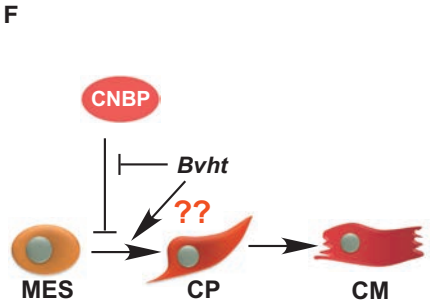
All experiments were performed in triplicate and data are represented as mean values ± SD. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001 (two-tailed Student's t test).

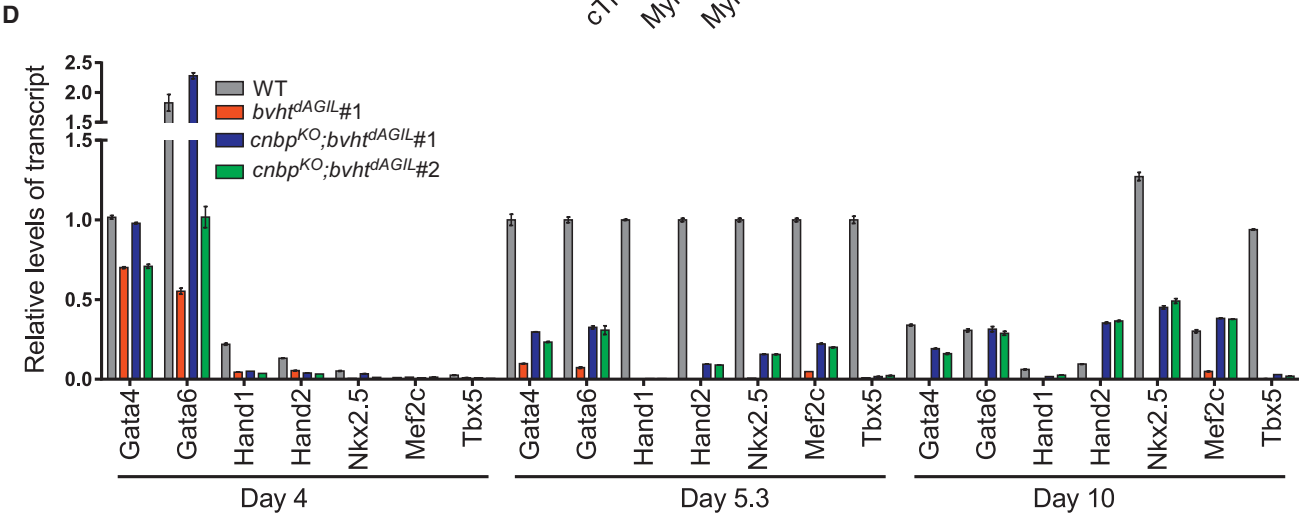
See also Figure S5.











(legend on next page)

et al., 2012). Markedly, *cnbp*^{KO};*bvht*^{dRHT} mutant cells generate percentages of Nkx2.5-GFP+ cells at day 10 (CM) comparable to WT cells. Coinciding with our FACS analysis, the *cnbp*^{KO};*bvht*^{dAGIL} cells also showed significant levels of Nkx2.5 and cTnT by immunofluorescence staining (Figures 6B and 6E), whereas levels were undetectable in *bvht*^{dAGIL} mutant alone (Figure 3C).

We then analyzed the expression of the mesodermal marker Brachyury and key cardiac TFs by qRT-PCR (Figures 6C and 6D). Brachyury levels in *cnbp*^{KO};*bvht*^{dAGIL} double mutants showed comparable expression levels to WT cells. Expression of Nkx2.5, Gata4, Gata6, Hand2, and Mef2c was also partially restored at both day 5.3 and day 10. Moreover, we observed that the CM-specific genes cTnT, Myh6, and Myh7 showed a significant increase in expression (50%∼70% relative to WT cells) in *cnbp*^{KO};*bvht*^{dAGIL} double mutants compared to the AGIL mutant alone (Figure 6E). Together, our data suggest that CNBP and *Bvht* function together to regulate cardiovascular lineage commitment.

DISCUSSION

Our work establishes that RNA secondary structure determination coupled with genetic studies can reveal important functional motifs required for lncRNA mechanisms of action. Our study revealed several important findings regarding the role of *Bvht* in cardiovascular lineage commitment. First, we show that *Bvht* adopts a modular secondary structure in vitro that harbors a 5′ AGIL. Remarkably, a small 11 nt deletion in the AGIL motif (*bvht*^{dAGIL}) within the ∼590 nt non-coding transcript prevents the transition from nascent mesoderm to the CP state in our in vitro differentiation assay. Second, we found that the zinc-finger TF CNBP specifically interacts with *Bvht*. We also show that CNBP acts as a negative regulator of the cardiac developmental program and that genetic ablation of CNBP partially rescues the differentiation defect of *bvht*^{dAGIL} mutant cells. Collectively, these data suggest that *Bvht* functionally antagonizes CNBP to promote cardiovascular lineage commitment (Figure 6F).

In some cases, lncRNAs such as GAS5, PANDA, NF-YA, and NORAD have been reported to function as molecular decoys to titrate interacting proteins away from their regulatory targets through competitive binding (Hung et al., 2011; Kino et al., 2010; Lee et al., 2016). However, the low abundance of *Bvht* transcript makes the molecular decoy model unlikely to explain its mode of action. Expression of *Bvht* from the ROSA26 locus using its endogenous promoter largely rescues the AGIL mutant phenotype, suggesting that low copy number is sufficient to mediate its function in *trans* in a locus-specific manner (Figures S3G–S3K). Recently, lncRNAs including *Fendrr*, *PRNA*, and *PARTICL* were found to target specific genomic loci through directly hybridizing to nascent DNA via sequence complementarity or DNA:DNA:RNA (Grote et al., 2013; O'Leary et al., 2015; Schmitz et al., 2010). In addition, it has been proposed that low-abundance RNAs such as the RNA component of telomerase (TERC), which can perform multiple turnover reactions, could accomplish super-stoichometric functionalities (Goff and Rinn, 2015; Mozdy and Cech, 2006; Zappulla and Cech, 2004), providing another potential model for studying the molecular mechanisms of low-abundance lncRNAs such as *Bvht* in future studies.

Our results suggest CNBP is a critical component of *Bvht*'s mode of action in cardiovascular lineage commitment. CNBP is highly conserved among vertebrates and can bind single-stranded G-rich DNA or RNA (Calcaterra et al., 2010). It has been proposed that CNBP acts as a nucleic acid chaperone and can promote the formation of G-quadruplex (G4) structures in which four guanines are assembled in a planar arrangement by Hoogsteen hydrogen bonding followed by intra- or inter-molecular folding of the tetramers (Armas et al., 2008; Borgognone et al., 2010; Rhodes and Lipps, 2015). For example, CNBP represses the expression of heterogeneous ribonucleoprotein K (hnRNPK) in fibrosarcoma cells and c-Myc in human HeLa cells through its conversion of promoter G-rich sequences into G4 DNA (Chen et al., 2013; Qiu et al., 2014). We found that different algorithms including QGRS Mapper, QGRS-H Predictor, and TetraplexFinder all predict that the *Bvht* AGIL motif can form a G4 structure (Figure S6B) (Kikin et al., 2006; Menendez et al., 2012; Yadav et al., 2008). Notably, G4 motifs have been identified in the promoters or UTRs of cardiac genes such Nkx2.5, Gata4, and Mef2d (Nie et al., 2015; Zhang et al., 2008). Moreover, the specific inactivation of the G4-resolving RNA helicase RHAU in either cardiac mesoderm or progenitors leads to abnormal heart development (Nie et al., 2015). Thus, it is possible that *Bvht* and CNBP function together to regulate cardiac gene expression through control of G4 structures. Our probing studies indicate that the stems that flank the AGIL motif may be important for maintaining the G-rich loop in a single-stranded conformation, which could be important for facilitating CNBP binding to this region. Thus, detailed mechanistic follow-up of this and other models, as well as dissecting the function of additional *Bvht* AGIL-interacting proteins, will be a focus of future investigation.

Figure 6. Loss of CNBP Partially Rescues the *bvht*^{dAGIL} Phenotype

(A) Cells at indicated time points during CM differentiation were analyzed for marker expression by flow cytometry. Numbers in plots indicate percentage of gated populations.

(B) Immunofluorescence staining of indicated cells using anti-GFP (day 5.3) and anti-cTnT (day 10) antibodies. Nuclei were stained with DAPI. BF, bright field. Scale bar, 100 μm.

(C and D) qRT-PCR analysis showing the relative levels of Brachyury (C) and core cardiac (D) TFs. WT value at day 4 (C) or at day 5.3 (D) was set to 1 for each gene.

(E) qRT-PCR analysis showing the relative levels of CM marker genes at day 10.

(F) Model of *Bvht* and CNBP regulating cardiovascular lineage commitment. *Bvht* functionally antagonizes the repression of CNBP on the transition from cardiac mesoderm to progenitors. Potential additional factors working together with *Bvht* remained to be elucidated.

All experiments were performed in triplicate and data are represented as mean values ± SD. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001 (two-tailed Student's t test).

See also Figure S6.

Our work demonstrates that the identification of specific motifs by carefully dissecting individual lncRNAs is critical for understanding overall lncRNA function and can explain why these transcripts are overall lowly conserved at the sequence level. Recently, in vivo DMS and SHAPE methods were developed to directly probe RNA structure in living cells (Ding et al., 2014; Rouskin et al., 2014; Spitale et al., 2013); however, the low abundance of many lncRNAs makes it difficult probe their structures in vivo, and the complex binding of proteins to RNAs can make interpretation of probing signals more complicated. Thus, secondary structure maps of free RNA molecules are necessary to facilitate a complete understanding of how these structures contribute to lncRNA modes of action under complex conditions. Together, determination of lncRNA motifs using both in vitro and in vivo probing results could be used to identify motif sequence fingerprints and homologs across species through phylogenetic sequence alignments and covariance analysis (Sanbonmatsu, 2016). Recent studies show that lncRNAs may undergo multiple secondary structure conformations in vivo (Lu et al., 2016). Our combined 3S and functional analyses confidently identify the AGIL motif; however, we have not eliminated the possibility of alternative folds. In some cases, RNA also forms higher-order structures composed of tightly packed secondary structure elements (Leontis et al., 2006; Weeks, 2010). Thus, dissecting tertiary structures of lncRNAs under physiological conditions also represents an important area for future investigation. In vitro secondary structures of lncRNAs are also an important first step toward crystallographic and cryoelectron microscopy (cryo-EM) 3D structures. Ultimately, studies aimed at mechanistic dissection of lncRNA structures are expected to facilitate a detailed understanding of how these transcripts contribute to fundamental biological processes and open the door to exploiting RNA motifs as biological and therapeutic tools.

EXPERIMENTAL PROCEDURES

Detailed experimental and analysis methods can be found in the Supplemental Experimental Procedures.

Chemical Probing

SHAPE probing was performed using fast-acting 1M7 reagent (Deigan et al., 2009), and DMS probing was performed as described (Tijerina et al., 2007).

ESC Lines and Growth Conditions

mESCs were cultured on irradiated mouse embryonic fibroblasts (MEFs) using standard conditions as previously described (Wamstad et al., 2012). NKX2.5-GFP mESCs (Hsiao et al., 2008) were used as WT ESCs in this study.

Generation of ESC Lines with CRISPR/Cas9

CRISPR/Cas9-mediated homology-directed repair or non-homologous end joining was performed as described (Ran et al., 2013) using a bicistronic expression vector expressing Cas9 and sgRNA (px330, Addgene #42230).

ESC Differentiation

EB formation and directed differentiation were performed as described (Klattenhoff et al., 2013; Wamstad et al., 2012).

Immunostaining ESCs and Differentiated Cell Types

Cells were fixed and stained according to our previous studies (Klattenhoff et al., 2013; Wamstad et al., 2012).

Flag-Tagged CNBP, HNRNPF, and SFSR9

Flag-tagged CNBP/HNRNPF/SFRS9 cassette was cloned into pEGIP (Addgene #26777). Lentiviral production and ESC infection were performed using protocols from the RNAi Consortium (Broad Institute).

RNA Immunoprecipitation

Cells were UV cross-linked and RNA immunoprecipitation was performed as described (Jeon and Lee, 2011; Lai et al., 2013).

ProtoArray Processing and Analysis

In vitro RNA production and labeling followed by probing the ProtoArray Human Protein Microarray v5.0 (Life Technologies cat# PAH0525101) were performed as described (Siprashvili et al., 2012).

ACCESSION NUMBERS

The accession number for the ProtoArray raw data reported in this paper is ArrayExpress (http://www.ebi.ac.uk/arrayexpress): E-MTAB-4995.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2016.08.010.

AUTHOR CONTRIBUTIONS

S.H. designed and performed chemical probing experiments; K.Y.S., S.H., and I.V.N. analyzed the chemical probing data and determined the structure. Z.X. and L.A.B. designed all other experiments and interpreted the results. Z.X., B.D., and A.A.G. performed these experiments. L.A.B. and Z.X. wrote the manuscript.

ACKNOWLEDGMENTS

We thank members of the L.A.B. lab, David Bartel, Igor Ulitski, and Matt Taliaferro for insightful discussions and for critical evaluation of the manuscript. We also thank Chikdu Shivalila for technical support in CRIPSR/Cas9 experiments, Lionel Lam for help scanning the protein microarray, and the flow cytometry core for technical support. K.Y.S., S.H., and I.V.N. were supported by LANL-LDRD. A.A.G. was supported by the Sarnoff Cardiovascular Research Foundation. This work was also supported in part by the Koch Institute Core grant (P30CA014051) and the NHLBI Bench to Bassinet Program (U01HL098179) to L.A.B.

Received: March 17, 2016
Revised: June 21, 2016
Accepted: August 5, 2016
Published: September 8, 2016

REFERENCES

Aguilera, A., and García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Mol. Cell *46*, 115–124.

Armas, P., Nasif, S., and Calcaterra, N.B. (2008). Cellular nucleic acid binding protein binds G-rich single-stranded nucleic acids and may function as a nucleic acid chaperone. J. Cell. Biochem. *103*, 1013–1036.

Bindewald, E., Hayes, R., Yingling, Y.G., Kasprzak, W., and Shapiro, B.A. (2008). RNAJunction: a database of RNA junctions and kissing loops for three-dimensional structural analysis and nanodesign. Nucleic Acids Res. *36*, D392–D397.

Bondue, A., Lapouge, G., Paulissen, C., Semeraro, C., Iacovino, M., Kyba, M., and Blanpain, C. (2008). Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. Cell Stem Cell *3*, 69–84.

Borgognone, M., Armas, P., and Calcaterra, N.B. (2010). Cellular nucleic-acid-binding protein, a transcriptional enhancer of c-Myc, promotes the formation of parallel G-quadruplexes. Biochem. J. *428*, 491–498.

Calcaterra, N.B., Armas, P., Weiner, A.M., and Borgognone, M. (2010). CNBP: a multifunctional nucleic acid chaperone involved in cell death and proliferation control. IUBMB Life *62*, 707–714.

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Müller, K.M., et al. (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics *3*, 2.

Chen, W., Liang, Y., Deng, W., Shimizu, K., Ashique, A.M., Li, E., and Li, Y.P. (2003). The zinc-finger protein CNBP is required for forebrain formation in the mouse. Development *130*, 1367–1379.

Chen, W., Wang, Y., Abe, Y., Cheney, L., Udd, B., and Li, Y.P. (2007). Haploinsuffciency for Znf9 in Znf9+/- mice is associated with multiorgan abnormalities resembling myotonic dystrophy. J. Mol. Biol. *368*, 8–17.

Chen, S., Su, L., Qiu, J., Xiao, N., Lin, J., Tan, J.H., Ou, T.M., Gu, L.Q., Huang, Z.S., and Li, D. (2013). Mechanistic studies for the role of cellular nucleic-acid-binding protein (CNBP) in regulation of c-myc transcription. Biochim. Biophys. Acta *1830*, 4769–4777.

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell *161*, 404–416.

Deigan, K.E., Li, T.W., Mathews, D.H., and Weeks, K.M. (2009). Accurate SHAPE-directed RNA structure determination. Proc. Natl. Acad. Sci. USA *106*, 97–102.

Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature *505*, 696–700.

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. Nat. Rev. Genet. *15*, 7–21.

Flynn, R.A., and Chang, H.Y. (2014). Long noncoding RNAs in cell-fate programming and reprogramming. Cell Stem Cell *14*, 752–761.

Geisler, S., and Coller, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat. Rev. Mol. Cell Biol. *14*, 699–712.

Goff, L.A., and Rinn, J.L. (2015). Linking RNA biology to lncRNAs. Genome Res. *25*, 1456–1465.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M., and Herrmann, B.G. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev. Cell *24*, 206–214.

Gutell, R.R., Lee, J.C., and Cannone, J.J. (2002). The accuracy of ribosomal RNA comparative structure models. Curr. Opin. Struct. Biol. *12*, 301–310.

Han, P., Li, W., Lin, C.-H., Yang, J., Shang, C., Nurnberg, S.T., Jin, K.K., Xu, W., Lin, C.-Y., Lin, C.-J., et al. (2014). A long noncoding RNA protects the heart from pathological hypertrophy. Nature *514*, 102–106.

Hsiao, E.C., Yoshinaga, Y., Nguyen, T.D., Musone, S.L., Kim, J.E., Swinton, P., Espineda, I., Manalac, C., deJong, P.J., and Conklin, B.R. (2008). Marking embryonic stem cells with a 2A self-cleaving peptide: a NKX2-5 emerald GFP BAC reporter. PLoS ONE *3*, e2532.

Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C., Wang, P., et al. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat. Genet. *43*, 621–629.

Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119–133.

Johnsson, P., Lipovich, L., Grandér, D., and Morris, K.V. (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function. Biochim. Biophys. Acta *1840*, 1063–1071.

Jones, K., Jin, B., Iakova, P., Huichalaf, C., Sarkar, P., Schneider-Gold, C., Schoser, B., Meola, G., Shyu, A.B., Timchenko, N., and Timchenko, L. (2011). RNA Foci, CUGBP1, and ZNF9 are the primary targets of the mutant CUG and CCUG repeats expanded in myotonic dystrophies type 1 and type 2. Am. J. Pathol. *179*, 2475–2489.

Kattman, S.J., Witty, A.D., Gagliardi, M., Dubois, N.C., Niapour, M., Hotta, A., Ellis, J., and Keller, G. (2011). Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. Cell Stem Cell *8*, 228–240.

Kikin, O., D'Antonio, L., and Bagga, P.S. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. *34*, W676–W682.

Kino, T., Hurt, D.E., Ichijo, T., Nader, N., and Chrousos, G.P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci. Signal. *3*, ra8.

Klattenhoff, C.A., Scheuermann, J.C., Surface, L.E., Bradley, R.K., Fields, P.A., Steinhauser, M.L., Ding, H., Butty, V.L., Torrey, L., Haas, S., et al. (2013). Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell *152*, 570–583.

Kretz, M., Siprashvili, Z., Chu, C., Webster, D.E., Zehnder, A., Qu, K., Lee, C.S., Flockhart, R.J., Groff, A.F., Chow, J., et al. (2013). Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature *493*, 231–235.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497–501.

Lee, J.T., and Bartolomei, M.S. (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell *152*, 1308–1323.

Lee, T.M., Maurer, M.S., Karbassi, I., Braastad, C., Batish, S.D., and Chung, W.K. (2012). Severe dilated cardiomyopathy in a patient with myotonic dystrophy type 2 and homozygous repeat expansion in ZNF9. Congest. Heart Fail. *18*, 183–186.

Lee, S., Kopp, F., Chang, T.C., Sataluri, A., Chen, B., Sivakumar, S., Yu, H., Xie, Y., and Mendell, J.T. (2016). Noncoding RNA NORAD regulates genomic stability by sequestering PUMILIO proteins. Cell *164*, 69–80.

Leontis, N.B., Lescoute, A., and Westhof, E. (2006). The building blocks and motifs of RNA architecture. Curr. Opin. Struct. Biol. *16*, 279–287.

Lindsley, R.C., Gill, J.G., Murphy, T.L., Langer, E.M., Cai, M., Mashayekhi, M., Wang, W., Niwa, N., Nerbonne, J.M., Kyba, M., and Murphy, K.M. (2008). Mesp1 coordinately regulates cardiovascular fate restriction and epithelial-mesenchymal transition in differentiating ESCs. Cell Stem Cell *3*, 55–68.

Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., and Ranum, L.P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science *293*, 864–867.

Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., et al. (2016). RNA duplex map in living cells reveals higher-order transcriptome structure. Cell *165*, 1267–1279.

Ma, Q., Zhou, B., and Pu, W.T. (2008). Reassessment of Isl1 and Nkx2-5 cardiac fate maps using a Gata4-based reporter of Cre activity. Dev. Biol. *323*, 98–104.

Marques Howarth, M., Simpson, D., Ngok, S.P., Nieves, B., Chen, R., Siprashvili, Z., Vaka, D., Breese, M.R., Crompton, B.D., Alexe, G., et al. (2014). Long noncoding RNA EWSAT1-mediated gene repression facilitates Ewing sarcoma oncogenesis. J. Clin. Invest. *124*, 5275–5290.

Matunis, M.J., Xing, J., and Dreyfuss, G. (1994). The hnRNP F protein: unique primary structure, nucleic acid-binding properties, and subcellular localization. Nucleic Acids Res. *22*, 1059–1067.

Menendez, C., Frees, S., and Bagga, P.S. (2012). QGRS-H Predictor: a web server for predicting homologous quadruplex forming G-rich sequence motifs in nucleotide sequences. Nucleic Acids Res. *40*, W96–W103.

Mian, I.S. (1997). Comparative sequence analysis of ribonucleases HII, III, II PH and D. Nucleic Acids Res. *25*, 3187–3195.

Mondragón, A. (2013). Structural studies of RNase P. Annu. Rev. Biophys. *42*, 537–557.

Mozdy, A.D., and Cech, T.R. (2006). Low abundance of telomerase in yeast: implications for telomerase haploinsufficiency. RNA *12*, 1721–1737.

Nie, J., Jiang, M., Zhang, X., Tang, H., Jin, H., Huang, X., Yuan, B., Zhang, C., Lai, J.C., Nagamine, Y., et al. (2015). Post-transcriptional regulation of Nkx2-5 by RHAU in heart development. Cell Rep. *13*, 723–732.

Noller, H.F. (1984). Structure of ribosomal RNA. Annu. Rev. Biochem. *53*, 119–162.

Novikova, I.V., Hennelly, S.P., and Sanbonmatsu, K.Y. (2012). Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. Nucleic Acids Res. *40*, 5034–5051.

Novikova, I.V., Dharap, A., Hennelly, S.P., and Sanbonmatsu, K.Y. (2013). 3S: shotgun secondary structure determination of long non-coding RNAs. Methods *63*, 170–177.

O'Leary, V.B., Ovsepian, S.V., Carrascosa, L.G., Buske, F.A., Radulovic, V., Niyazi, M., Moertl, S., Trau, M., Atkinson, M.J., and Anastasov, N. (2015). PARTICLE, a triplex-forming long ncRNA, regulates locus-specific methylation in response to low-dose irradiation. Cell Rep. *11*, 474–485.

Parsch, J., Braverman, J.M., and Stephan, W. (2000). Comparative sequence analysis and patterns of covariation in RNA secondary structures. Genetics *154*, 909–921.

Ponting, C.P., Oliver, P.L., and Reik, W. (2009). Evolution and functions of long noncoding RNAs. Cell *136*, 629–641.

Qiu, J., Chen, S., Su, L., Liu, J., Xiao, N., Ou, T.M., Tan, J.H., Gu, L.Q., Huang, Z.S., and Li, D. (2014). Cellular nucleic acid binding protein suppresses tumor cell metastasis and induces tumor cell death by downregulating heterogeneous ribonucleoprotein K in fibrosarcoma cells. Biochim. Biophys. Acta *1840*, 2244–2252.

Quinn, J.J., and Chang, H.Y. (2016). Unique features of long non-coding RNA biogenesis and function. Nat. Rev. Genet. *17*, 47–62.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Ren, A., Vušurović, N., Gebetsberger, J., Gao, P., Juen, M., Kreutz, C., Micura, R., and Patel, D.J. (2016). Pistol ribozyme adopts a pseudoknot fold facilitating site-specific in-line cleavage. Nat. Chem. Biol. http://dx.doi.org/10.1038/nchembio.2125.

Reznik, B., Clement, S.L., and Lykke-Andersen, J. (2014). hnRNP F complexes with tristetraprolin and stimulates ARE-mRNA decay. PLoS ONE *9*, e100992.

Rhodes, D., and Lipps, H.J. (2015). G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. *43*, 8627–8637.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Rizki, G., and Boyer, L.A. (2015). Lncing epigenetic control of transcription to cardiovascular development and disease. Circ. Res. *117*, 192–206.

Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature *505*, 701–705.

Sanbonmatsu, K.Y. (2016). Towards structural classification of long non-coding RNAs. Biochim. Biophys. Acta *1859*, 41–45.

Schmitz, K.M., Mayer, C., Postepska, A., and Grummt, I. (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev. *24*, 2264–2269.

Serganov, A., and Patel, D.J. (2007). Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. Nat. Rev. Genet. *8*, 776–790.

Siprashvili, Z., Webster, D.E., Kretz, M., Johnston, D., Rinn, J.L., Chang, H.Y., and Khavari, P.A. (2012). Identification of proteins binding coding and non-coding human RNAs using protein microarrays. BMC Genomics *13*, 633.

Somarowthu, S., Legiewicz, M., Chillón, I., Marcia, M., Liu, F., and Pyle, A.M. (2015). HOTAIR forms an intricate and modular secondary structure. Mol. Cell *58*, 353–361.

Spitale, R.C., Crisalli, P., Flynn, R.A., Torre, E.A., Kool, E.T., and Chang, H.Y. (2013). RNA SHAPE analysis in living cells. Nat. Chem. Biol. *9*, 18–20.

Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T., and Chang, H.Y. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. Nature *519*, 486–490.

Tan, E., Wilson, T.J., Nahas, M.K., Clegg, R.M., Lilley, D.M., and Ha, T. (2003). A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. Proc. Natl. Acad. Sci. USA *100*, 9308–9313.

Tijerina, P., Mohr, S., and Russell, R. (2007). DMS footprinting of structured RNAs and RNA-protein complexes. Nat. Protoc. *2*, 2608–2623.

Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. Cell *154*, 26–46.

Wamstad, J.A., Alexander, J.M., Truty, R.M., Shrikumar, A., Li, F., Eilertson, K.E., Ding, H., Wylie, J.N., Pico, A.R., Capra, J.A., et al. (2012). Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell *151*, 206–220.

Wang, Y., Ma, M., Xiao, X., and Wang, Z. (2012). Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. Nat. Struct. Mol. Biol. *19*, 1044–1052.

Weeks, K.M. (2010). Advances in RNA structure analysis by chemical probing. Curr. Opin. Struct. Biol. *20*, 295–304.

Weiner, A.M., Allende, M.L., Becker, T.S., and Calcaterra, N.B. (2007). CNBP mediates neural crest cell expansion by controlling cell proliferation and cell survival during rostral head development. J. Cell. Biochem. *102*, 1553–1570.

Weiner, A.M., Sdrigotti, M.A., Kelsh, R.N., and Calcaterra, N.B. (2011). Deciphering the cellular and molecular roles of cellular nucleic acid binding protein during cranial neural crest development. Dev. Growth Differ. *53*, 934–947.

Yadav, V.K., Abraham, J.K., Mani, P., Kulshrestha, R., and Chowdhury, S. (2008). QuadBase: genome-wide database of G4 DNA–occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. Nucleic Acids Res. *36*, D381–D385.

Zappulla, D.C., and Cech, T.R. (2004). Yeast telomerase RNA: a flexible scaffold for protein subunits. Proc. Natl. Acad. Sci. USA *101*, 10024–10029.

Zhang, R., Lin, Y., and Zhang, C.T. (2008). Greglist: a database listing potential G-quadruplex regulated genes. Nucleic Acids Res. *36*, D372–D376.

Find your way with Cell Press Reviews

Whether you're looking for answers to fundamental questions, brushing up on a new subject or looking for a synthesis on a hot topic, turn to Cell Press Reviews. We publish reviews in a vast array of disciplines spanning the life sciences.

Propel your research forward, faster with Cell Press Reviews. Our insightful, authoritative reviews—published in our primary research and Trends journals—go beyond synthesis and offer a point of view.

www.cell.com/reviews

CellPress
Your work is our life

CellPress

Cell Stem Cell
Article

An Isogenic Human ESC Platform for Functional Evaluation of Genome-wide-Association-Study-Identified Diabetes Genes and Drug Discovery

Hui Zeng,[1,3,9] Min Guo,[2,3,9] Ting Zhou,[3] Lei Tan,[3] Chi Nok Chong,[3] Tuo Zhang,[5] Xue Dong,[3] Jenny Zhaoying Xiang,[5] Albert S. Yu,[6] Lixia Yue,[6] Qibin Qi,[7] Todd Evans,[3] Johannes Graumann,[4,8] and Shuibing Chen[3,4,*]

[1]Department of Hematology in Xiangya Hospital
[2]Department of Endocrinology in Xiangya Hospital
Central South University, 87 Xiangya Road, Changsha, Hunan 410008, China
[3]Department of Surgery
[4]Department of Biochemistry
[5]Genomic Core
Weill Cornell Medical College, 1300 York Avenue, New York, NY 10065, USA
[6]Calhoun Cardiology Center and Department of Cell Biology, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06030, USA
[7]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA
[8]Research Division, Weill Cornell Medical College in Qatar, Doha, State of Qatar
[9]Co-first author
*Correspondence: shc2034@med.cornell.edu
http://dx.doi.org/10.1016/j.stem.2016.07.002

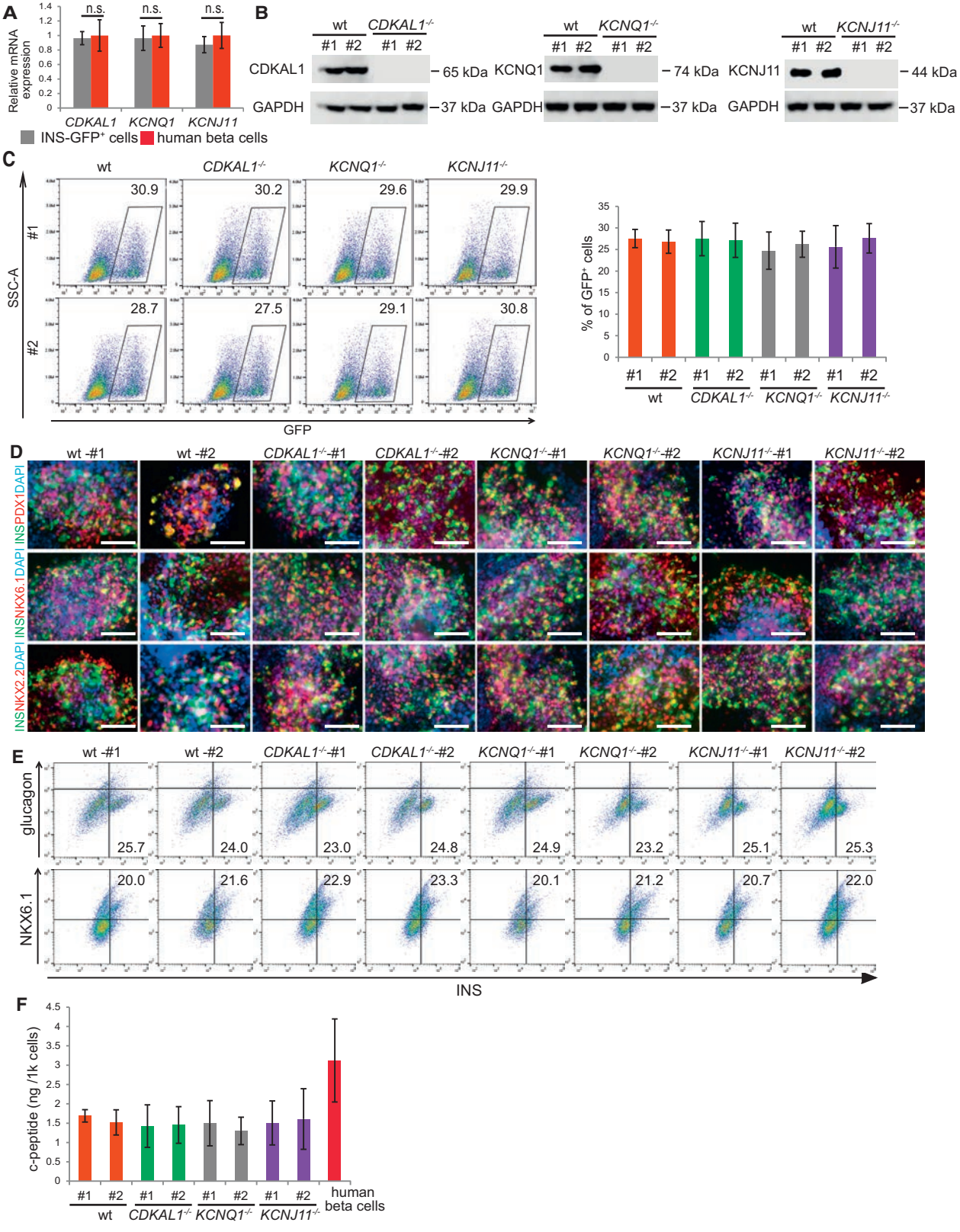
SUMMARY

Genome-wide association studies (GWASs) have increased our knowledge of loci associated with a range of human diseases. However, applying such findings to elucidate pathophysiology and promote drug discovery remains challenging. Here, we created isogenic human ESCs (hESCs) with mutations in GWAS-identified susceptibility genes for type 2 diabetes. In pancreatic beta-like cells differentiated from these lines, we found that mutations in CDKAL1, KCNQ1, and KCNJ11 led to impaired glucose secretion in vitro and in vivo, coinciding with defective glucose homeostasis. CDKAL1 mutant insulin+ cells were also hypersensitive to glucolipotoxicity. A high-content chemical screen identified a candidate drug that rescued CDKAL1-specific defects in vitro and in vivo by inhibiting the FOS/JUN pathway. Our approach of a proof-of-principle platform, which uses isogenic hESCs for functional evaluation of GWAS-identified loci and identification of a drug candidate that rescues gene-specific defects, paves the way for precision therapy of metabolic diseases.

INTRODUCTION

Multiple genome-wide association studies (GWASs) have correlated type 2 diabetes mellitus (T2DM) with genetic variants, yielding a large number of loci and associated gene products that are linked to the disease phenotype—often with little or no insight into the mechanism underlying that link (Hivert et al., 2014). The current challenge is to establish robust systems to systematically evaluate the role of these loci using disease-relevant cells. Previous studies have used patient samples, cell lines, or animal models to seek mechanistic insight but with significant limitations. Large variation is observed in primary patient samples, perhaps due to genetic heterogeneity, whereas animal models present major physiological and metabolic differences that hamper understanding of the precise function of human genes in T2DM. Therefore, a robust system to systematically evaluate the role of T2DM-associated genes using disease-relevant human cells will provide an important tool for diabetes research and spur the development of precision (allele-specific) therapies, exemplified by the use of sulfonylurea drugs to treat patients carrying certain *KCNJ11* mutations (Gloyn et al., 2004).

Human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs) provide platforms to recapitulate cellular pathology of human diseases. Whereas two iPSC models have been used to mimic pancreatic beta cell defects in neonatal and inherited forms of diabetes, maturity onset diabetes of young 2 (Hua et al., 2013) and Wolfram syndrome patients (Shang et al., 2014), there is no robust model reported for T2DM-associated loci in the literature. Here, we focused on *CDKAL1*, *KCNQ1*, and *KCNJ11* loci that were identified and confirmed through the first wave of T2DM GWASs. Risk alleles of the genetic variants at these loci are associated with aspects of beta cell function (HOMA-B) rather than insulin resistance (HOMA-IR) (Saxena et al., 2007; Scott et al., 2007; Steinthorsdottir et al., 2007; Unoki et al., 2008; Yasuda et al., 2008). Some studies suggested potential roles of these genes in pancreatic beta cell function or survival. For example, knockdown of *Cdkal1* enhanced endoplasmic reticulum (ER) stress in insulinoma cells (Brambillasca et al., 2012), whereas *Cdkal1*$^{-/-}$ mice show reduced first-phase insulin exocytosis (Ohara-Imaizumi et al.,



(legend on next page)

2010) and are hypersensitive to high-fat-diet-induced ER stress (Wei et al., 2011) and defects in glucose-stimulated insulin secretion (Okamura et al., 2012).

For other T2DM-linked genes, population and rodent studies reported mixed or even conflicting results. The risk allele of lead SNP at *KCNQ1* is associated with impaired insulin secretion (Unoki et al., 2008; Yasuda et al., 2008) and reduced insulin exocytosis in patients (Rosengren et al., 2012), and *Kcnq1*$^{-/-}$ mice have impaired glucose-stimulated insulin secretion (GSIS) (Boini et al., 2009). However, forced expression of *Kcnq1* in an insulinoma cell line resulted in impairment of insulin secretion (Yamagata et al., 2011), and islets isolated from *Kcnq1*$^{-/-}$ mice revealed no difference in insulin secretion compared to wild-type islets (Asahara et al., 2015). Several activating mutations in *KCNJ11* result in permanent neonatal DM (Massa et al., 2005; Proks et al., 2004; Shimomura et al., 2006), and a polymorphism E23K is consistently linked with T2DM (Gloyn et al., 2003; Nielsen et al., 2003). A number of heterozygous mutations result in congenital hyperinsulinism (Bitner-Glindzicz et al., 2000). Heterozygous loss of murine *Kcnj11* causes a hyperinsulinemic phenotype, whereas complete loss underlies eventual secretory failure (Remedi et al., 2006). These mixed results suggest that GWAS-identified genes may play a context-dependent role in human pancreatic beta cells. Furthermore, using mouse models, it can be challenging to differentiate whether the GWAS-associated alteration causes cell-autonomous defects or acts indirectly through extra-pancreatic tissues.

We built on recent work deriving glucose-responsive pancreatic beta-like cells from hESCs/iPSCs (Pagliuca et al., 2014; Rezania et al., 2014) and used isogenic hESC-derived glucose-responding cells to systematically examine the role of several GWAS-identified genes in pancreatic beta cell function and survival. Whereas the mutations do not affect the generation of insulin^{+} cells, they impaired insulin secretion both in vitro and in vivo, coinciding with defective glucose homeostasis. *CDKAL1*$^{-/-}$ insulin^{+} cells also displayed hypersensitivity to glucolipotoxicity. A high-content chemical screen identified a candidate drug that rescued *CDKAL1*$^{-/-}$-specific defects by inhibiting the *FOS*/*JUN* pathway. These studies represent a proof of principle for the use of isogenic hESC-derived cells to define the precise role of genes associated with disease though GWASs in human pancreatic beta cells, as well as the lead-compound identification for pharmacological intervention of T2DM.

RESULTS

Generation of Biallelic Mutant hESC Lines by CRISPR-Cas9 Gene Targeting

We targeted indel mutations to *CDKAL1*, *KCNQ1*, or *KCNJ11* in INS$^{GFP/W}$ HES3 cells, because this reporter line allows for the purification of insulin-producing (insulin^{+}) cells (Micallef et al., 2012). First, qRT-PCR was used to monitor the expression of the targeted genes in insulin-GFP^{+} cells derived from INS$^{GFP/W}$ HES3 cells. The transcript levels of *CDKAL1*, *KCNQ1*, and *KCNJ11* were detected at levels comparable to those observed in primary human adult beta cells (Figure 1A), suggesting that these genes are likely to function in the insulin^{+} cells.

To mutate each gene, INS$^{GFP/W}$ HES3 cells were electroporated with a vector expressing Cas9 and a specific small guide RNA (sgRNA) targeted to the first or second exon of each gene (Table S1). After sub-cloning, an efficiency of 11%–15% (Table S2) was observed for the creation of biallelic mutant lines. Multiple independent clones for each mutation were expanded. All established clones have typical hESC colony morphology and express pluripotency markers, including OCT3/4, NANOG, TRA-1-60, and TRA-1-81 (Figure S1A). To account for possible variation between different clones, two clones (no. 1 and no. 2) were chosen of each mutant line for further analysis. Biallelic indel mutations for each of the targeted genes were verified by genomic DNA sequencing (Figure S1B). Each indel mutation creates an early frame shift that is predicted to generate null alleles. Western blotting experiments further validated the knockout (KO) of the target genes in day 30 (D30) differentiated cells derived from each mutant hESC line (Figure 1B).

Biallelic Mutation of *CDKAL1*, *KCNQ1*, or *KCNJ11* Does Not Affect the Stepwise Differentiation toward Insulin^{+} Cells or the Expression of More-Mature Beta Cell Markers

The isogenic lines were differentiated using a strategy modified slightly from a previously reported protocol (Rezania et al., 2014), which is summarized in Table S3. Immunocytochemistry analysis with antibodies against stage-specific markers was used to quantify differentiation efficiency. No significant difference was detected between wild-type and any of the isogenic mutant lines with respect to their capacity to differentiate toward definitive endoderm (SOX17^{+}/FOXA2^{+}; DE; Figures S1C and S1D) or pancreatic progenitors (PDX1^{+}/NKX6.1^{+}/SOX9^{+}; PP; Figures S1E–S1H). Flow cytometry analysis showed the

Figure 1. Biallelic Mutation of *CDKAL1*, *KCNQ1*, or *KCNJ11* Does Not Affect Differentiation or Expression of Mature Pancreatic Beta Cell Markers

(A) qRT-PCR experiments confirmed the expression of *CDKAL1*, *KCNQ1*, and *KCNJ11* in insulin-GFP^{+} (INS-GFP^{+}) cells derived from INS$^{GFP/W}$ HES3 cells (n = 4 independent experiments; error bars indicate SD). The expression level of *CDKAL1*, *KCNQ1*, and *KCNJ11* transcripts in primary human beta cells was calculated by dividing the expression level in primary human islets by the percentage of insulin^{+} cells.

(B) Western blotting analysis of wild-type and isogenic mutant hESC-derived D30 cells.

(C) Representative flow cytometry analysis and quantification of WT and isogenic mutant hESC-derived cells at day 30; n = 3.

(D) Immunocytochemistry analysis of WT and isogenic mutant hESC-derived D30 cells. The insulin^{+} cells express mature beta cell markers, including PDX1, NKX6.1, and NKX2.2. The scale bar represents 100 μm.

(E) Intracellular FACS analysis of D30 cells.

(F) Total c-peptide content per 1 k insulin-GFP^{+} cells as measured by ELISA; n = 3. Total c-peptide content in primary human beta cells was calculated by dividing the total c-peptide in primary human islets by the percentage of insulin^{+} cells.

Clones no. 1 and no. 2 are two independent isogenic hESC clones carrying different frameshift mutations. hESCs were differentiated using protocol 2. The data are presented as mean ± SD. See also Figure S1.

percentage of insulin-GFP$^+$ cells in D30 populations to be indistinguishable between wild-type and the isogenic mutant lines (Figures 1C and S1I). Together, these data suggest that biallelic mutation of *CDKAL1*, *KCNQ1*, or *KCNJ11* does not affect the stepwise differentiation of insulin$^+$ cells.

The expression of pancreatic beta cell makers in D30 hESC-derived insulin$^+$ cells was analyzed by immunocytochemistry, and all cells, regardless of genotype, were found to express markers indicative of mature pancreatic beta cells, including PDX1, NKX6.1, and NKX2.2 (Figure 1D). Intracellular fluorescence-activated cell sorting (FACS) analysis showed that most hESC-derived insulin$^+$ cells express the mature beta cell marker NKX6.1, but not the alpha cell marker glucagon (Figures 1E and S1J). Wild-type and isogenic mutant cell lines did not differ with respect to the NKX6.1$^+$/insulin$^+$ cell or insulin$^+$/glucagon$^-$ cell fractions (Figures 1E and S1J). Next, insulin-GFP$^+$ cells were purified by cell sorting and analyzed for transcript expression levels with qRT-PCR (Figure S1K). Undifferentiated hESCs served as a negative control and primary human islets as a positive control. Transcripts encoding mature pancreatic beta cells markers, including *NKX6.1*, *NKX2.2*, *PDX1*, *ISLET1*, *PAX6*, *NEUROD1*, *GCK*, *G6PC2*, *UCN3*, and *MAFA* are highly expressed at levels comparable to human islets in hESC-derived insulin-GFP$^+$ cells. No significant difference was observed between wild-type and isogenic mutant insulin-GFP$^+$ cells (Figure S1K). The total c-peptide level of wild-type and mutant hESC-derived insulin-GFP$^+$ cells, as measured by ELISA, was comparable to levels in primary human islets (Figure 1F; Table S4). Thus, mutation of *CDKAL1*, *KCNQ1*, or *KCNJ11* does not significantly affect the generation of mature beta-like cells or insulin production.

Mutation of *CDKAL1*, *KCNQ1*, or *KCNJ11* Differentially Impairs Insulin Secretion in Response to Multiple Secretagogues

The major function of pancreatic beta cells is to secrete insulin/c-peptide upon induction by secretagogues. D30-differentiated wild-type or mutant cells were stimulated with 30 mM KCl, and secreted human c-peptide was measured by ELISA. Wild-type cells respond with a 4.5- ± 1.6-fold induction of c-peptide secretion (Figures 2A, 2B, and S2A). *CDKAL1*$^{-/-}$ cells showed a small but insignificant decreased response, whereas *KCNQ1*$^{-/-}$ and *KCNJ11*$^{-/-}$ cells were severely and significantly impaired in their response to KCl stimulation (Figures 2A and 2B). The cells were further queried for their response to 10 mM arginine. Again, both wild-type and *CDKAL1*$^{-/-}$ D30 cells responded well, whereas *KCNQ1*$^{-/-}$ and *KCNJ11*$^{-/-}$ cells failed to respond (Figures 2C, 2D, and S2B). D30 cells were also stimulated with 20 μM forskolin or 50 μM IBMX to measure cyclic AMP (cAMP)-induced insulin secretion. Wild-type cells responded well to both, yielding 7.2- ± 1.9- and 6.2-± 1.8-fold induction of c-peptide secretion, respectively (Figures 2E, 2F, and S2C). Cells carrying the three mutant alleles were able to respond to both forskolin and IBMX stimulation (Figure 2E), but compared to wild-type cells, the fold induction was significantly decreased (Figure 2F). Finally, wild-type cells stimulated with 2 mM (low) or 20 mM (high) D-glucose responded to high glucose with a 2.3- ± 0.8-fold induction of c-peptide secretion, whereas all three mutant genotypes failed to respond (Figures 2G, 2H, and S2D). Thus, loss of *KCNQ1* or *KCNJ11* affects insulin secretion. Because *CDKAL1*$^{-/-}$ cells respond to KCl and arginine (Figures 2A and 2C), but not cAMP or glucose stimulation (Figures 2E and 2G), CDKAL1 may be involved in cAMP and glucose sensing rather than exocytosis of insulin granules.

Patch-clamp experiments were used to determine K_{ATP} channel activity in *KCNJ11*$^{-/-}$ cells. To perform K_{ATP} current recordings, wild-type insulin-GFP$^+$ cells were held at 0 mV to inactivate any voltage-gated ion channels, and K_{ATP} currents were elicited by depolarization from holding potential (HP) = 0 mV to +80 mV. K_{ATP} channels were activated by the K_{ATP}-channel-specific activator diazoxide (Pasyk et al., 2004; Figure S2E) and inhibited by K_{ATP}-channel-specific blocker glybenclamide. The effect of diazoxide was reversible. After washout of diazoxide, glybenclamide further reduced current amplitude from 400 pA to ~200 pA (Figure S2F), suggesting that, in the absence of diazoxide, there were basal K_{ATP} channel activities, which was likely induced by the pipette solution. Whereas K_{ATP} currents were recorded in wild-type insulin-GFP$^+$ cells, diazoxide and glybenclamide did not produce any effects in the recordings from insulin-GFP$^+$ *KCNJ11*$^{-/-}$ mutant cells (Figure S2G), suggesting the absence of K_{ATP} channel activity.

CDKAL1$^{-/-}$ Insulin-GFP$^+$ Cells Are Hypersensitive to Glucolipotoxicity

Hyperglycemia and hyperlipidemia are two major risk factors associated with pancreatic beta cell death in diabetic patients. Wild-type and isogenic *CDKAL1*$^{-/-}$, *KCNQ1*$^{-/-}$, and *KCNJ11*$^{-/-}$ D30 insulin-GFP$^+$ cells were cultured in the presence of 35 mM D-glucose for 96 hr or 1 mM palmitate for 48 hr. Cells were stained with propidium iodide (PI) to determine the cell death rate (Figure 3A). No significant difference was detected between wild-type and mutant insulin$^+$ cells under control conditions. However, the percentage of PI$^+$/insulin$^+$ cells in *CDKAL1*$^{-/-}$ insulin$^+$ cells was significantly higher compared to wild-type insulin$^+$ cells exposed to 35 mM D-glucose or 1 mM palmitate, indicating that *CDKAL1*$^{-/-}$ insulin$^+$ cells are hypersensitive to glucotoxicity and lipotoxicity (Figure 3B). In contrast, neither *KCNQ1*$^{-/-}$ nor *KCNJ11*$^{-/-}$ insulin$^+$ cells showed increased sensitivity to glucotoxicity or lipotoxicity (Figure S3A). Treated cells were stained with the apoptosis marker annexin V, as well as the cell death marker 7AAD, and evaluated by flow cytometry to measure apoptosis in insulin-GFP$^+$ cells (Figures 3C and S3B). Consistent with the PI staining results, the percentage of annexin V$^+$/7AAD$^-$ cells in *CDKAL1*$^{-/-}$ insulin-GFP$^+$ cells was significantly higher than wild-type (Figure 3D), *KCNQ1*$^{-/-}$, or *KCNJ11*$^{-/-}$ insulin-GFP$^+$ cells when cultured in the presence of 35 mM D-glucose or 1 mM palmitate (Figures S3C and S3D). We also measured the proliferation rate of wild-type and *CDKAL1*$^{-/-}$ insulin-GFP$^+$ cells (Figure S3E), which showed no significant difference (Figure S3F). RNA sequencing (RNA-seq) was used to compare the gene expression profiles in wild-type and *CDKAL1*$^{-/-}$ cells cultured in the presence or absence of palmitate. ER-stress-related genes were found significantly upregulated in *CDKAL1*$^{-/-}$ cells cultured under palmitate conditions (Figures 3E and 3F). This suggests, consistent with the literature (Brambillasca et al., 2012; Wei et al., 2011), that loss of CDKAL1 induces elevated ER stress under exposure to high levels of fatty acids.

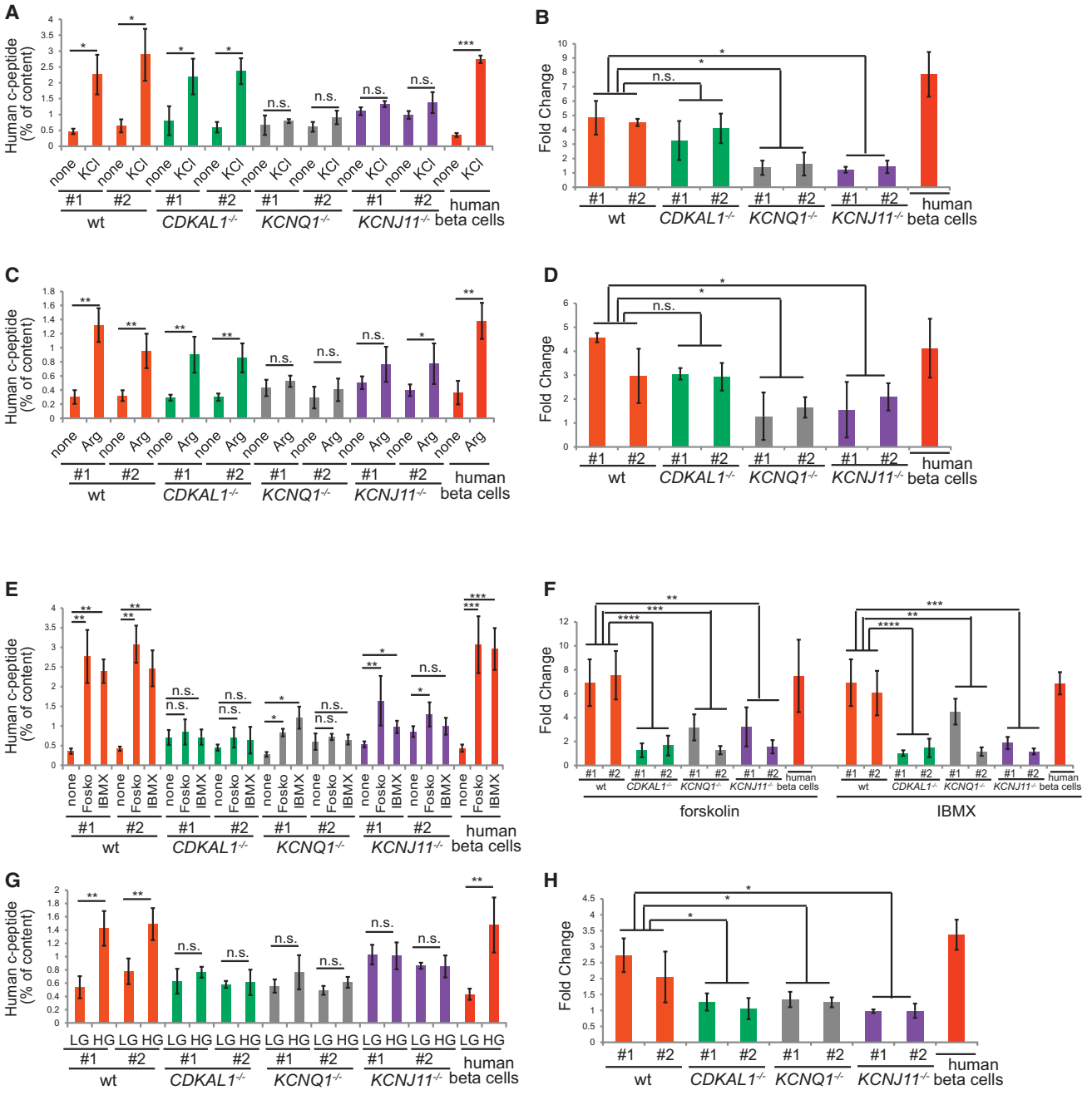


Figure 2. Biallelic Mutation of *CDKAL1*, *KCNQ1*, or *KCNJ11* Impairs Insulin Secretion upon Various Stimulations

(A and B) Human c-peptide (% of content; A) and fold change (B) of wt and isogenic mutant cells at day 30 with or without 30 mM KCl stimulation in the presence of 2 mM D-glucose; n = 3.

(C and D) Human c-peptide (% of content; C) and fold change (D) of wt and isogenic mutant cells at day 30 with and without 10 mM arginine stimulation in the presence of 2 mM D-glucose; n = 3.

(E and F) Human c-peptide (% of content; E) and fold change (F) of wt and isogenic mutant cells at day 30 with or without 20 μM forskolin and 50 μM IBMX stimulation in the presence of 2 mM D-glucose; n = 3.

(G and H) Human c-peptide (% of content; G) and fold change (H) of wt and isogenic mutant cells at day 30 with 2 mM or 20 mM D-glucose; n = 3.

Arg, arginine; forsk, forskolin; LG, 2 mM D-glucose; HG, 20 mM D-glucose. Human c-peptide secretion was calculated by dividing the secreted c-peptide by the total c-peptide of insulin-GFP^{+} cells or primary human beta cells. Clones no. 1 and no. 2 are two independent isogenic hESC clones carrying different frameshift mutations. hESCs were differentiated using protocol 2. The data are presented as mean ± SD. n.s. indicates a non-significant difference. p values calculated by unpaired two-tailed Student's t test were *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001. See also Figure S2.

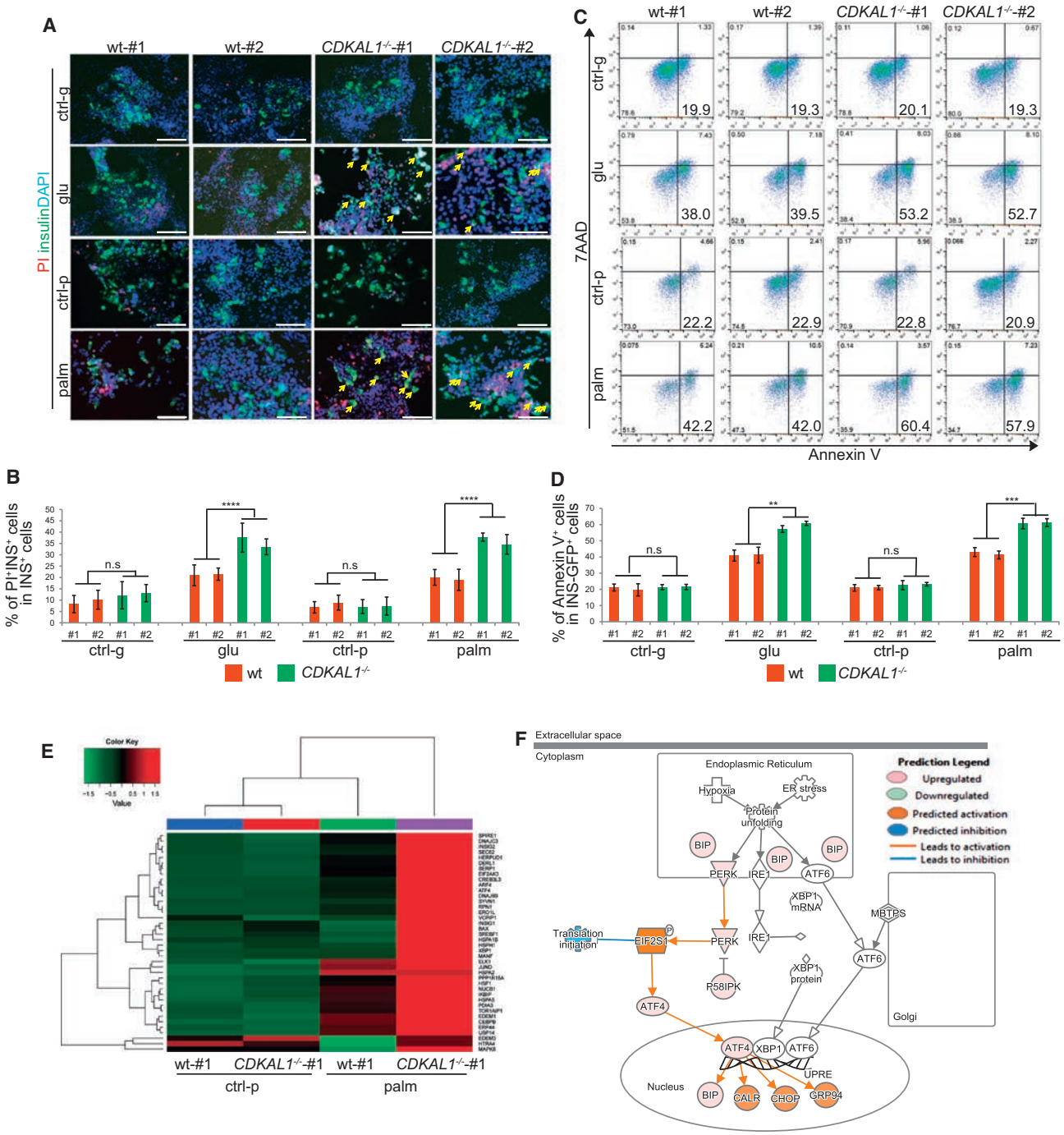


Figure 3. *CDKAL1*$^{-/-}$ Insulin-GFP^{+} Cells Are Hypersensitive to Glucotoxicity and Lipotoxicity

(A and B) Immunocytochemistry analysis (A) and quantification of the percentage (B) of PI^{+}/insulin^{+} cells in wt or *CDKAL1*$^{-/-}$ insulin^{+} cells cultured in the presence of 2 mM D-glucose (ctrl-g), 35 mM D-glucose (glu), no palmitate (ctrl-p), or 1 mM palmitate (palm). PI^{+}/insulin^{+} cells are highlighted by arrows.

(C and D) Flow cytometry analysis (C) and quantification of the percentage (D) of annexin V^{+} cells in wt and *CDKAL1*$^{-/-}$insulin-GFP^{+} cells cultured as in (A).

(E) Heatmap representing the expression profiles of ER-stress-related genes comparing wt and *CDKAL1*$^{-/-}$ insulin^{+} cells cultured in the absence or presence of 1 mM palmitate.

(F) Ingenuity pathway analysis of genes that are >2-fold upregulated in *CDKAL1*$^{-/-}$ insulin^{+} cells cultured in the presence of 1 mM palmitate.

INS, insulin; PI, propidium iodide. n = 3 independent biological replicates. n.s. indicates a non-significant difference. Clones no. 1 and no. 2 are two independent isogenic hESC clones carrying different frameshift mutations. hESCs were differentiated using protocol 2. p values calculated by unpaired two-tailed Student's t test were *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001. The scale bar represents 100 μm. See also Figure S3.

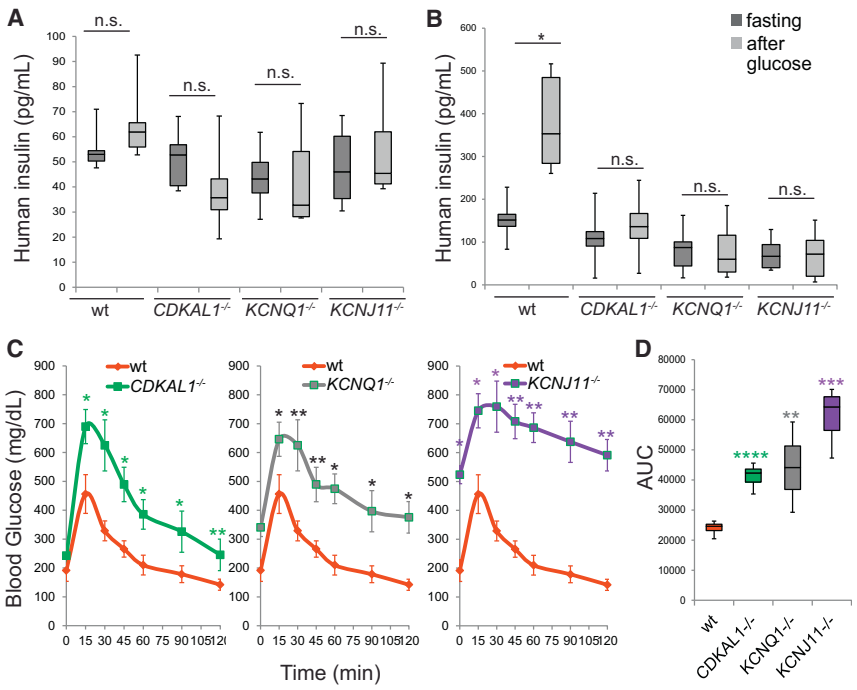


Figure 4. *CDKAL1*$^{-/-}$, *KCNQ1*$^{-/-}$, and *KCNJ11*$^{-/-}$ Cells Show Defective Glucose-Stimulated Insulin Secretion and Impaired Ability to Maintain Glucose Homeostasis after Transplantation into Streptozotocin-Treated Immuno-deficient Mice

(A) Human insulin GSIS at 2 weeks after transplantation of the mutant cells compared to wt cells. (B) GSIS secretion of SCID-beige mice carrying human cells at 6 weeks after transplantation. p values calculated by one-way repeated-measures ANOVA.

(C and D) Intraperitoneal glucose tolerance test (IPGTT) (C) and area under the curve (AUC) (D) of STZ-treated mice 6 weeks after transplantation. p values calculated by two-way repeated-measures ANOVA with a Bonferroni test for multiple comparisons between WT and mutant cells. n = 8 mice for each condition. hESCs were differentiated using protocol 2. n.s. indicates a non-significant difference. p values were *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001. See also Figure S4.

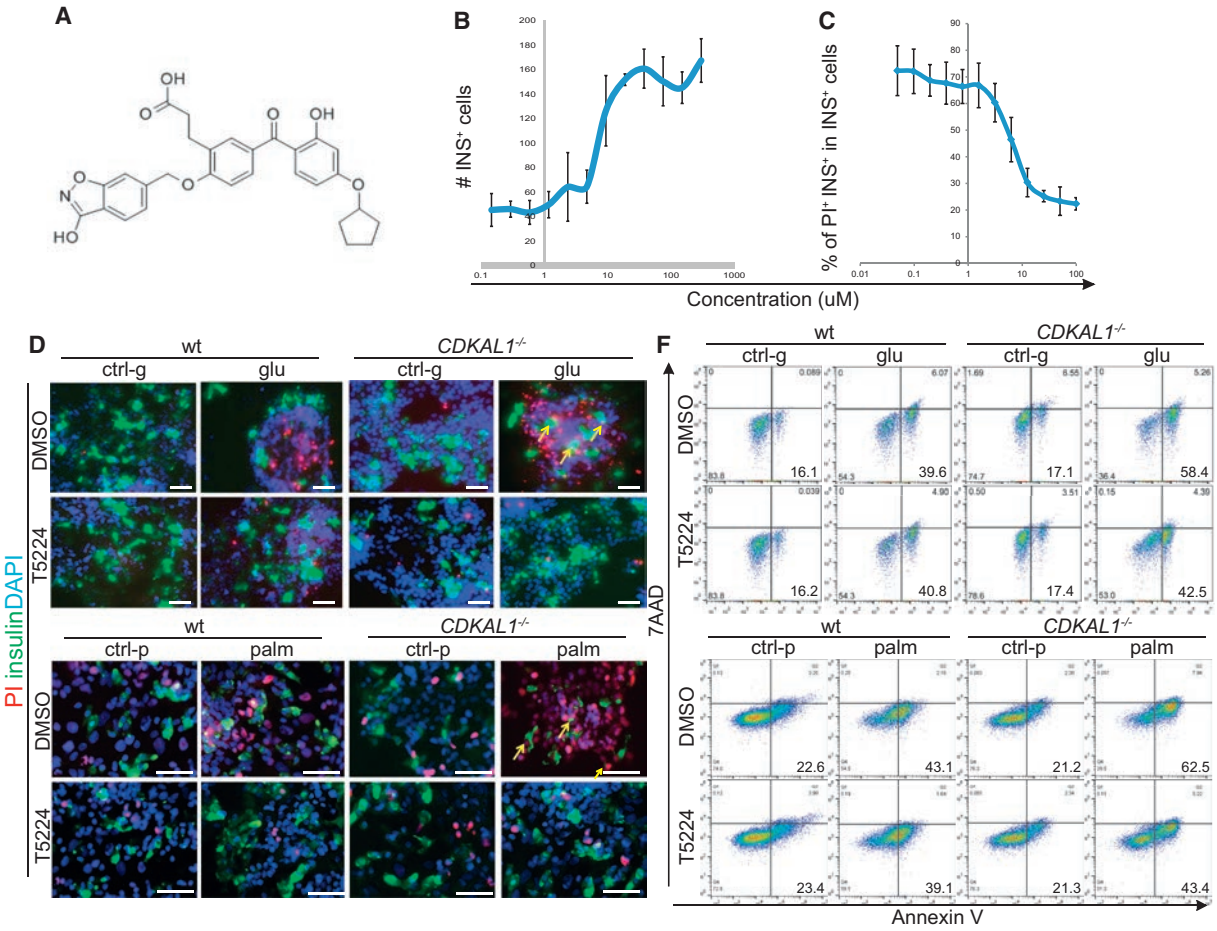
CDKAL1$^{-/-}$, *KCNQ1*$^{-/-}$, and *KCNJ11*$^{-/-}$ hESC-Derived Beta-like Cells Show Defective GSIS and Impaired Capacity to Maintain Glucose Homeostasis In Vivo

To determine the survival and functional capacities of *CDKAL1*$^{-/-}$, *KCNQ1*$^{-/-}$, and *KCNJ11*$^{-/-}$ hESC-derived beta-like cells in vivo, wild-type and isogenic mutant glucose-responding cells were transplanted under the kidney capsule of immuno-deficient severe combined immunodeficiency (SCID)-beige mice. Two days after transplantation, the mice were treated with 200 mg/kg streptozotocin (STZ) to chemically ablate endogenous murine pancreatic beta cells (Figure S4A). After STZ treatment, the levels of mouse insulin are below the detection limit of the ELISA kit (Figure S4B). Two weeks post-transplantation, SCID-beige mice carrying human cells were fasted overnight and monitored for GSIS, measuring by ELISA human insulin in serum at fasting and 30 min after stimulation with 3 g/kg glucose (Figures 4A and S4C). SCID-beige mice transplanted with wild-type or mutant cells displayed indistinguishable concentrations of human insulin (Figure 4A). By 6 weeks post-transplantation, SCID-beige mice carrying wild-type cells showed significantly increased insulin secretion after glucose stimulation (Figures 4B and S4D), whereas SCID-beige mice carrying *CDKAL1*$^{-/-}$, *KCNQ1*$^{-/-}$, or *KCNJ11*$^{-/-}$ cells continued to fail to respond to glucose stimulation (Figure 4B). Because SCID-beige mice transplanted with wild-type or mutant cells displayed indistinguishable concentrations of human insulin at 2 weeks after transplantation (Figure 4A), the failed GSIS of mice carrying mutant cells at 6 weeks after transplantation is due to the impaired function of the transplanted cells rather than unsuccessful transplantation. These results validate in vivo the impaired glucose response measured in mutant cells in vitro (Figure 2G).

To monitor the capacity of the transplanted cells to maintain glucose homeostasis in STZ-treated mice beyond an acute glucose response, an intraperitoneal glucose tolerance test (IPGTT) with 2 g/kg glucose was used. In contrast to SCID-beige mice carrying wild-type cells, those transplanted with *CDKAL1*$^{-/-}$, *KCNQ1*$^{-/-}$, or *KCNJ11*$^{-/-}$ cells show glucose intolerance (Figures 4C and S4E). The area under the curve (AUC) for the glucose tolerance test in SCID-beige mice carrying mutant cells was significantly higher compared to that of SCID-beige mice carrying wild-type cells (Figures 4D and S4F). Immunohistochemistry was used to document the persistence of human beta-like cells in transplanted human grafts. Mature pancreatic beta cells markers, including PDX1, NKX6.1, NKX2.2, and insulin, were detected in the grafts regardless of genotype (Figure S4G). Taken together, beta-like cells derived from *CDKAL1*$^{-/-}$, *KCNQ1*$^{-/-}$, or *KCNJ11*$^{-/-}$ hESCs present with impaired glucose-induced insulin secretion as well as glucose tolerance in SCID-beige mice carrying glucose-responding cells.

A High-Content Chemical Screen Identifies a Candidate Drug that Rescues *CDKAL1*$^{-/-}$-Specific Glucolipotoxicity and Impaired GSIS

A high-content chemical screen was performed to identify drug candidates capable of rescuing *CDKAL1*$^{-/-}$-specific glucolipotoxicity. D30-differentiated *CDKAL1*$^{-/-}$ cells were replated in 384-well plates and treated for 48 hr with chemicals from a collection of US Food and Drug Administration (FDA)-approved drugs and drug candidates in clinical trials at 10 μM in the presence of 1 mM palmitate. We screened 2,000 compounds for the capacity to decrease cell death by at least 80% in *CDKAL1*$^{-/-}$-derived beta-like cells exposed to glucolipotoxicity while also increasing the number of insulin^{+} cells at least 2-fold (Figure S5A). Of six initial lead hits, one compound, T5224 (Figure 5A), was validated to protect *CDKAL1*$^{-/-}$ insulin^{+} cells from glucolipotoxicity in follow-up experiments. Using the same platform as for the primary screening (1 mM palmitate), addition of T5224 caused increased numbers of insulin^{+} cells (Figure 5B) and a decreased percentage of PI^{+}/INS^{+} cells in *CDKAL1*$^{-/-}$ insulin^{+}



(legend on next page)

cells (Figures 5C) in a dose-dependent manner with a concentration for 50% of maximum effect (EC_{50}) of 16.2 μM. In addition, T5224 rescued the increased cell death rate in *CDKAL1*$^{-/-}$ insulin^{+} cells when cultured with high glucose or high palmitate (Figures 5D and 5E). As measured using the annexin V assay for apoptosis, T5224 also rescued the increased apoptotic rate in *CDKAL1*$^{-/-}$ insulin^{+} cells under conditions of high fatty acid concentration without affecting the rate in wild-type insulin^{+} cells (Figures 5F, 5G, and S5B), thus blunting hypersensitivity to glucolipotoxicity.

The *CDKAL1*$^{-/-}$ cells were treated with 30 μM T5224 for 48 hr and examined for impaired response to forskolin or glucose-stimulated insulin secretion (FSIS and GSIS). Remarkably, the mutant cells treated with T5224 showed increased insulin secretion in response to forskolin treatment (Figures 5H and S5C), significantly elevated compared to cells treated with DMSO and at a level of insulin secretion comparable to wild-type cells (Figure 5H). Similarly, T5224 treatment also rescued the impaired GSIS of *CDKAL1*$^{-/-}$ cells (Figures 5I and S5D). Notably, T5224 treatment did not significantly affect FSIS or GSIS in wild-type cells.

T5224 Rescues *CDKAL1*$^{-/-}$-Induced Beta Cell Defects through Inhibition of the *FOS/JUN* Pathway

T5224 was reported to be an inhibitor of FOS/JUN activator protein-1 (AP-1) (Aikawa et al., 2008). To explore this potential mechanism of action, RNA-seq was used to compare the global gene expression profiles in *CDKAL1*$^{-/-}$ and wild-type insulin-GFP^{+} cells. Pathway enrichment analysis highlighted the *FOS*/*JUN* and focal adhesion pathways as highly changed in *CDKAL1*$^{-/-}$ insulin-GFP^{+} cells (Figure 6A). Genes associated with the focal adhesion Gene Ontology (GO) term were consistently downregulated (Figures 6B and S6A) whereas the *FOS*/*JUN* pathway (Figure 6C) was consistently upregulated in *CDKAL1*$^{-/-}$ insulin-GFP^{+} cells. Among the top 20 genes showing relatively increased expression in *CDKAL1*$^{-/-}$ insulin-GFP^{+} cells are *FOSB* (6.3-fold), *FOS* (3.5-fold), and *JUNB* (2.4-fold; Figure 6D), which was confirmed by qRT-PCR (Figure 6E). Finally, western blotting experiments validated the relatively increased expression of FOS protein in mutant cells (Figure 6F).

To determine whether the mutation of *CDKAL1* induces pancreatic beta cell defects through activation of the *FOS*/*JUN* pathway, two sgRNAs and two scrambled sgRNAs were designed to knock out human *FOS* (Table S6). Wild-type and *CDKAL1*$^{-/-}$ hESC-derived day 10 PPs were infected with lentivirus expressing either sgFOS or a scrambled sgRNA, and following 4–6 days selection with puromycin, the cells were differentiated to beta-like cells for an additional 16–20 days. In cells expressing sgFOS, the expression of FOS was decreased by more than 99% based on western blotting experiments, validating the targeting efficiency (Figure S6B). The cells were cultured in the absence or presence of 35 mM D-glucose or 1 mM palmitate and analyzed with respect to the rates of cell death and apoptosis by PI staining and annexin V, respectively. Mutation of *FOS* using sgRNA rescues the increased cell death rate of *CDKAL1*$^{-/-}$ insulin-GFP^{+} cells (Figure 6G) and cell apoptotic rate of *CDKAL1*$^{-/-}$ insulin-GFP^{+} cells (Figures 6H, 6I, and S6C). In contrast, mutation of *FOS* does not affect cell death (Figure 6G) or apoptosis in wild-type insulin-GFP^{+} cells (Figures 6H and 6I). In addition to sgRNA, two short hairpin RNAs (shRNAs) against *FOS* were cloned into a lentiviral vector and used to knock down *FOS*. The knockdown efficiency in day 10 PPs is more than 50% based on western blotting experiments (Figure S6D). Consistent with the KO using sgFOS, knockdown of FOS using shRNAs rescued the increased cell apoptotic rate of *CDKAL1*$^{-/-}$ insulin-GFP^{+} cells when cultured in high fatty acid condition (Figures S6E and S6F).

Likewise, wild-type and *CDKAL1*$^{-/-}$ hESC-derived PPs infected with lentivirus expressing sgFOS or a scrambled sgRNA were differentiated for 20 days and measured for FSIS and GSIS. *CDKAL1*$^{-/-}$ cells infected with lentivirus containing scrambled sgRNA showed impaired FSIS and GSIS compared to wild-type cells. Transfection with lentivirus expressing sgFOS rescued those phenotypes (Figures 6J, 6K, S6G, and S6H). However, KO of FOS did not affect FSIS (Figure 6J) or GSIS (Figure 6K) in wild-type cells. Consistently, knockdown of *FOS* using shRNAs rescued the impaired FSIS (Figures S6I and S6K) and GSIS (Figures S6J and S6L) in *CDKAL1*$^{-/-}$ cells without affecting wild-type cells. Together, this suggests that loss of *CDKAL1* causes hypersensitivity to glucolipotoxicity and impairs FSIS and GSIS through the *FOS*/*JUN* pathway.

T5224 and Loss of *FOS* Rescues the Function of *CDKAL1*$^{-/-}$ Cells In Vivo

To examine the effect of T5224 on *CDKAL1*$^{-/-}$ cells in vivo, mice transplanted with wild-type and *CDKAL1*$^{-/-}$ cells were examined for GSIS at 10 weeks after transplantation. Consistent with the 6-week results reported above, mice transplanted with wild-type cells respond well to glucose stimulation. In contrast, mice transplanted with *CDKAL1*$^{-/-}$ cells showed impaired GSIS (Figures 7A and S7A). After glucose stimulation, the insulin level of mice transplanted with *CDKAL1*$^{-/-}$ cells was significantly lower than for mice transplanted with wild-type cells.

Subsequently, mice were treated with 300 mg/kg T5224 orally and measured for GSIS 48 hr after treatment. Mice treated with

Figure 5. A High-Content Chemical Screen Identifies a Drug Candidate that Rescues Glucolipotoxicity Caused Specifically by Mutations in *CDKAL1*

(A) Chemical structure of T5224.

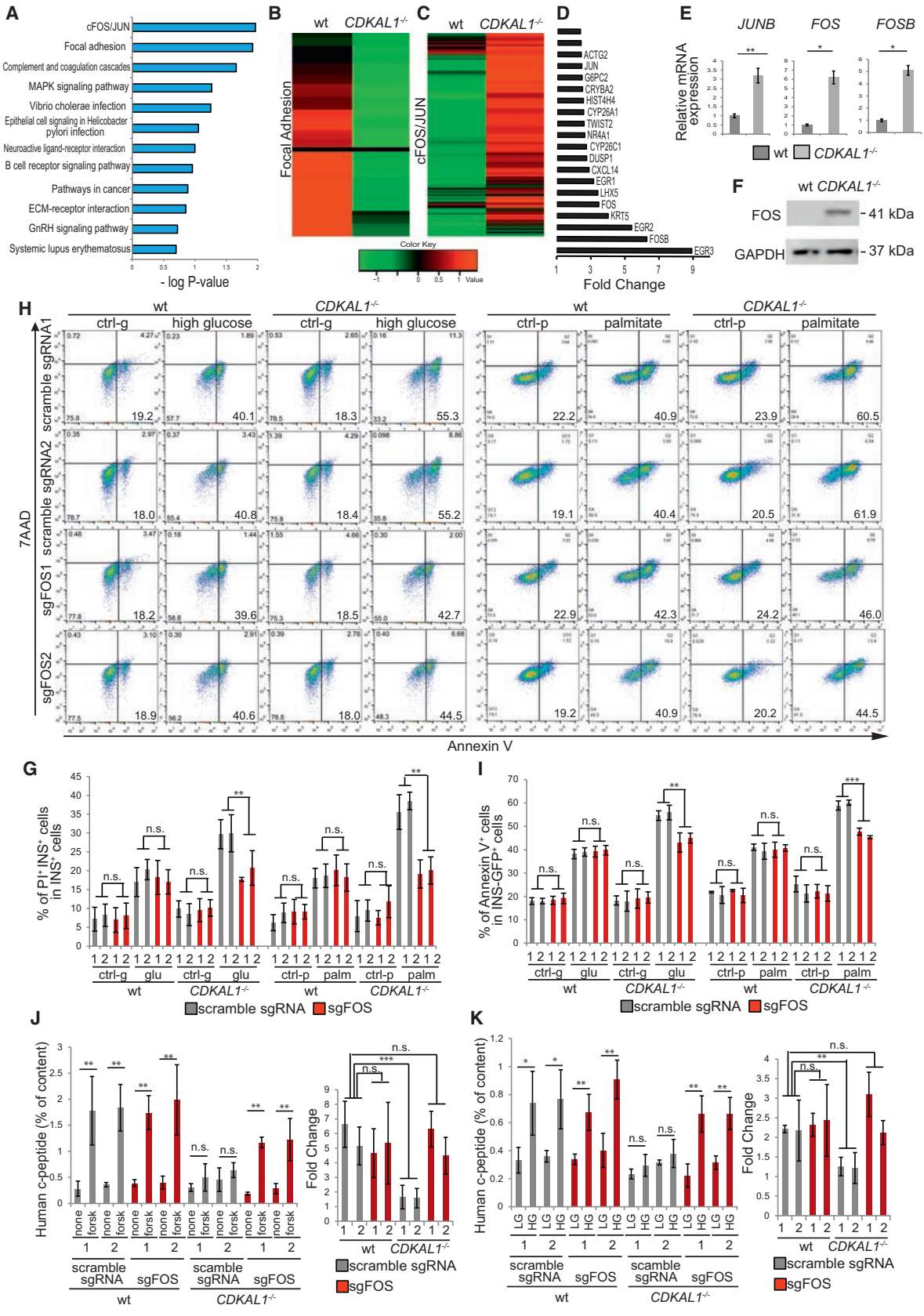
(B and C) Efficacy curve of T5224 on the number of insulin^{+} cells (B) and the percentage of PI^{+}INS^{+} cells (C).

(D and E) Immunocytochemistry analysis (D) and quantification of the percentage (E) of PI^{+}/insulin^{+} cells in wt and *CDKAL1*$^{-/-}$, insulin^{+} cells treated with 30 μM T5224 when cultured in the presence of 2 mM D-glucose (ctrl-g), 35 mM D-glucose (glu), no palmitate (ctrl-p), or 1 mM palmitate (palm). PI^{+}/insulin^{+} cells are highlighted by arrows.

(F and G) Flow cytometry analysis (F) and quantification (G) of apoptotic rate for WT or *CDKAL1*$^{-/-}$ insulin-GFP^{+} cells treated with DMSO or T5224.

(H and I) T5224 also rescues the impaired forskolin-induced (H) and glucose-induced insulin secretion (I).

Experiments in (A)–(C) were performed using cells derived from protocol 1. Experiments in (D)–(I) were performed using cells derived from protocol 2. n = 3 independent biological replicates for each condition. n.s. indicates a non-significant difference. p values calculated by unpaired two-tailed Student's t test were $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$. The scale bar represents 50 μm. See also Figure S5.



(legend on next page)

T5224 restored the capacity to respond to glucose stimulation (Figures 7B and S7B). T5224 treatment significantly increased the level of insulin secretion after glucose stimulation. In addition, the mice carrying *CDKAL1*$^{-/-}$ cells showed glucose intolerance. T5224 treatment restored the capacity of the SCID-beige mice carrying human cells to maintain glucose homeostasis (Figures 7C and S7C). The AUC for mice after T5224 treatment was significantly lower than for mice treated with control vehicle (Figures 7D and S7D). T5224 treatment was also examined using mice carrying wild-type cells. Consistent with the in vitro results (Figure 5), T5224 treatment affects neither GSIS (Figures S7E and S7F) nor glucose tolerance of mice carrying wild-type cells (Figures S7G–S7J). To determine the long-term effect of T5224, mice carrying *CDKAL1*$^{-/-}$ cells were treated with 300 mg/kg T5224 orally twice a week and measured for GSIS and glucose tolerance 4 weeks after treatment. The long-term treatment of T5224 restored both GSIS (Figures 7E and S7K) and glucose tolerance (Figures 7F, 7G, S7L, and S7M) for mice carrying *CDKAL1*$^{-/-}$ cells. Finally, D30 *CDKAL1*$^{-/-}$ cells carrying scrambled sgRNA and D30 *CDKAL1*$^{-/-}$ cells carrying sgFOS were transplanted into mice that were than measured for function in vivo 6 weeks after transplantation. Consistent with in vitro results (Figure 6), mice with *CDKAL1*$^{-/-}$ cells carrying sgFOS showed improved GSIS (Figures 7H and S7N) and a stronger ability to maintain glucose homeostasis (Figures 7I, 7J, S7O, and S7P) than mice transplanted with *CDKAL1*$^{-/-}$ cells carrying scrambled sgRNA. Together, these data suggest that T5224 or loss of FOS rescues the function of *CDKAL1*$^{-/-}$ cells in vivo.

DISCUSSION

With more than 80 loci associated with T2DM identified by GWASs, a robust platform to evaluate the role of these loci using disease-relevant cells is urgently needed. Here, we report proof of principle for using isogenic hESC-derived glucose-responding cells to evaluate the role of these loci in the function and survival of human pancreatic beta cells under conditions mimicking both health and disease. The derived glucose-responding cells share the same genetic background, providing a unique resource to determine the precise role of genes or loci in human pancreatic beta cells independent of complications from genetic heterogeneity implied by other approaches, such as patient-derived iPSCs.

We found that mutation of *KCNJ11* resulted in impaired insulin secretion upon KCl, arginine, forskolin, IBMX, and glucose stimulation, suggesting that *KCNJ11* plays an essential role in insulin secretion, which is consistent with results in homozygous *Kcnj11*$^{-/-}$ KO mice, as well as in homozygous *Kcnj11*$^{-/-}$-null mice (Remedi et al., 2006; Boini et al., 2009). In the context of reports that forced expression of *KCNQ1* in a mouse beta cell line results in impairment of insulin secretion (Yamagata et al., 2011) and islets isolated from *Kcnq1*$^{-/-}$ mice reveal no difference in the extent of basal or stimulated insulin secretion compared to islet from wild-type mice (Asahara et al., 2015), we were surprised to find impaired insulin secretion in *KCNQ1*$^{-/-}$ insulin-secreting cells. This apparent discrepancy may suggest dose- and/or species-specific roles in pancreatic beta cell function, highlighting the importance of using human-relevant cell types.

An ultimate goal of exploring loci or genetic variants associated with disease through GWASs is to identify locus-/variant-specific treatments. Risk alleles of SNPs at the *CDKAL1* locus associated with diabetes are thought to be loss-of-function alleles, which we modeled, generating null mutations. We found that *CDKAL1*$^{-/-}$ insulin^{+} cells showed impaired FSIS and GSIS, which is consistent with *Cdkal1*$^{-/-}$ mice showing reduced first-phase insulin exocytosis (Ohara-Imaizumi et al., 2010). *CDKAL1*$^{-/-}$ insulin^{+} cells also show increased ER stress, cell apoptosis, and death when cultured in high-glucose and high-fatty-acid conditions. Although there are papers describing the potential contribution of lipotoxicity in T2DM, direct evidence that lipotoxicity affects pancreatic beta cell death in vivo under normal physiological and pathological conditions needs to be further explored. Here, we found that *CDKAL1*$^{-/-}$ insulin^{+} cells are hypersensitive to both high-glucose- and high-fatty-acid-induced pancreatic beta-like cell death. Moreover, *CDKAL1*$^{-/-}$ insulin^{+} cells display defective GSIS and impaired ability to maintain glucose homeostasis following transplantation into STZ-treated mice. This is consistent with the in vitro functional defects of *CDKAL1*$^{-/-}$ insulin^{+} cells. Because the mice are hyperglycemic after STZ treatment, the observed glucotoxicity may further worsen the defects of *CDKAL1*$^{-/-}$ insulin^{+} cells. From a high-content chemical screen, T5224 was found to rescue the *CDKAL1* mutation-mediated pancreatic beta cell defects. T5224 has been investigated in clinical trials for patients with rheumatoid arthritis (Pharmaceutical Medicine, 2014) and may have the potential to be repurposed for *CDKAL1*-specific

Figure 6. T5224 Rescues Beta Cell Defects Caused by *CDKAL1* Mutation through Inhibiting the *FOS*/*JUN* Pathway

(A) Pathway enrichment analysis on up/downregulated genes in *CDKAL1*$^{-/-}$ insulin-GFP^{+} cells using the DAVID function annotation tool.

(B) Heatmap of focal-adhesion-pathway-associated genes comparing wt and *CDKAL1*$^{-/-}$ insulin-GFP^{+} cells.

(C) Heatmap of *FOS*/*JUN*-pathway-associated genes comparing WT and *CDKAL1*$^{-/-}$ insulin-GFP^{+} cells.

(D) Top 20 upregulated genes in *CDKAL1*$^{-/-}$ insulin-GFP^{+} cells as compared to wild-type cells.

(E) qRT-PCR analysis of *JUNB*, *FOS*, and *FOSB* expression in wt and *CDKAL1*$^{-/-}$ insulin-GFP^{+} cells.

(F) Western blotting analysis of FOS protein in wt and *CDKAL1*$^{-/-}$ cells at D30 of differentiation.

(G) Targeted mutation of *FOS* rescues the high death rate in *CDKAL1*$^{-/-}$ insulin-GFP^{+} cells in the presence of 35 mM D-glucose or 1 mM palmitate.

(H and I) Flow cytometry analysis (H) and quantification of apoptotic rate (I) of *CDKAL1*$^{-/-}$ insulin-GFP^{+} cells expressing Cas9 and either scrambled sgRNA or sgFOS.

(J and K) Mutation of *FOS* rescues the impaired forskolin-induced (J) and glucose-induced (K) insulin secretion that is caused by mutation of *CDKAL1*.

sgFOS no. 1 and no. 2 represent two independent sgRNAs targeting different locations of exon1 of *c-FOS*. Scramble sgRNA no. 1 and scramble no. 2 "target" controls were designed to have low homology to the human genome and are used as non-targeting controls. hESCs were differentiated using protocol 2. The data are presented as mean ± SD. n.s. indicates a non-significant difference. p values calculated by unpaired two-tailed Student's t test were *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. See also Figure S6.

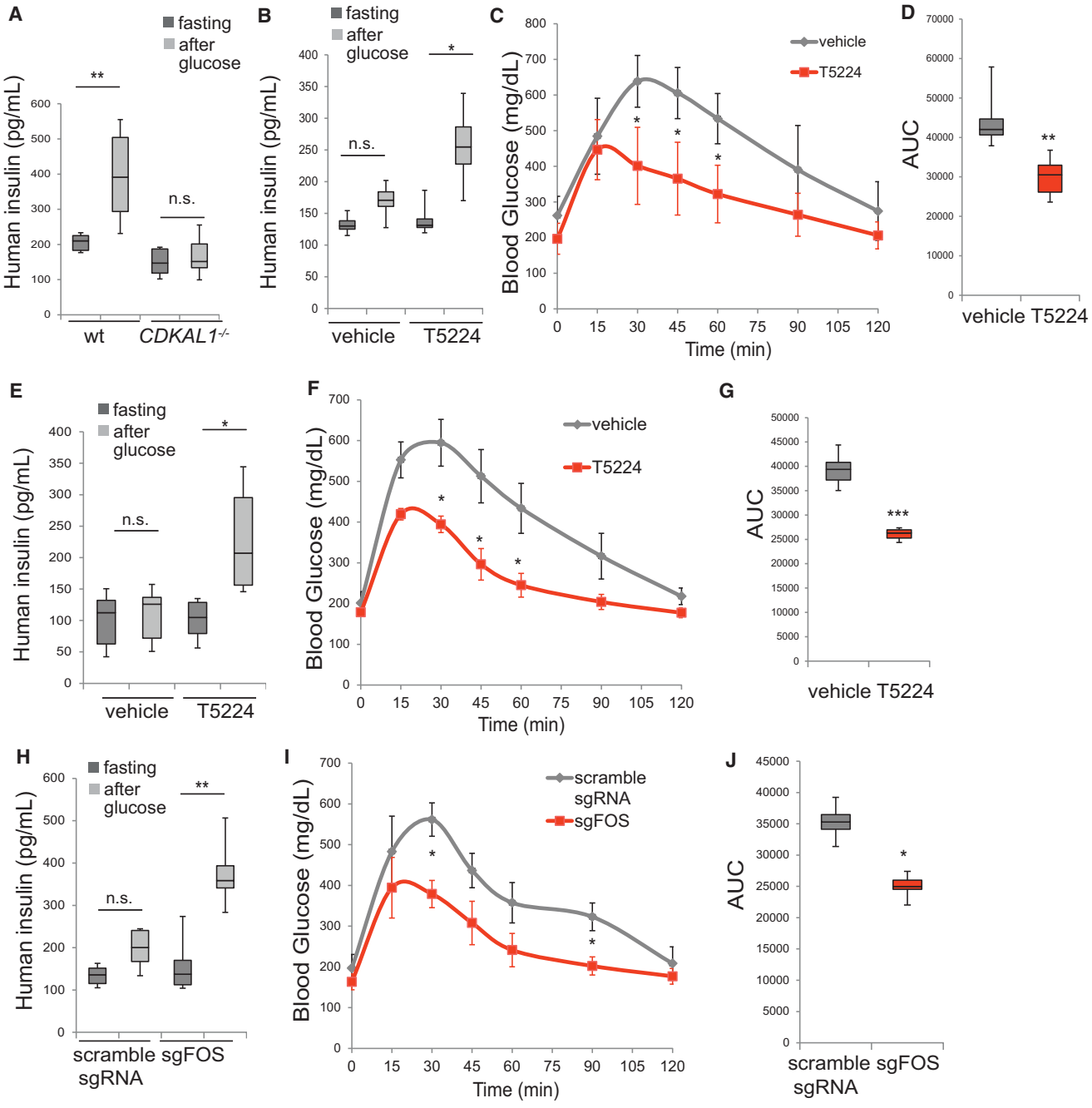


Figure 7. T5224 or Loss of *FOS* Rescues the Function of *CDKAL1*$^{-/-}$ Cells in SCID-Beige Mice Carrying Human Cells

(A) Human insulin GSIS at 10 weeks after transplantation of mutant cells compared to wt cells.

(B) GSIS secretion of SCID-beige mice carrying human cells after glucose stimulation 48 hr after treatment with 300 mg/kg T5224 or vehicle.

(C and D) IPGTT (C) and AUC (D) of mice transplanted with *CDKAL1*$^{-/-}$ cells treated with 300 mg/kg T5224 or vehicle.

(E) GSIS secretion of SCID-beige mice carrying human cells after glucose stimulation after treatment with T5224 or vehicle twice a week for 4 weeks.

(F and G) IPGTT (F) and AUC (G) of mice transplanted with *CDKAL1*$^{-/-}$ cells treated with 300 mg/kg T5224 or vehicle twice a week for 4 weeks.

(H) GSIS secretion of SCID-beige mice transplanted with *CDKAL1*$^{-/-}$ cells carrying scramble sgRNA or *CDKAL1*$^{-/-}$ cells carrying sgFOS.

(I and J) IPGTT (I) and AUC (J) of mice transplanted with *CDKAL1*$^{-/-}$ cells carrying scramble sgRNA or *CDKAL1*$^{-/-}$ cells carrying sgFOS at 6 weeks after transplantation.

n = 8 mice for each condition. hESCs were differentiated using protocol 2. In GSIS assay, p values were calculated by one-way repeated-measures ANOVA. In IPGTT assay, p values were calculated by two-way repeated-measures ANOVA with a Bonferroni test for multiple comparisons between DMSO and T5224 treated conditions. p values were *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. See also Figure S7.

treatment of T2DM. T5224 is able to strikingly rescue *CDKAL1*-mutation-mediated pancreatic beta cell dysfunction in vivo, which is a proof of concept for a T2DM drug candidate rescuing a gene-specific defect in vivo.

By combining high-content chemical screening and RNA-seq, we found the *FOS*/*JUN* pathway to be significantly upregulated in *CDKAL1*$^{-/-}$ insulin^{+} cells and that reducing *FOS*/*JUN* pathway activity either chemically or genetically rescued *CDKAL1* mutation-induced defects. Previous studies have shown that *FOS*/*JUN* activation is involved in cytokine and mechanical-stress-induced beta cell death (Abdelli et al., 2007; Hughes et al., 1990) and amylin-induced apoptosis (Zhang et al., 2002). Here, we found that *CDKAL1*$^{-/-}$-mediated activation of the *FOS*/*JUN* pathway through fatty acids may be a further effector of *FOS*/*JUN*-regulated beta cell survival, providing mechanistic insight into how *CDKAL1* locus may contribute to diabetes progression.

In summary, we established an isogenic hESC platform to systematically evaluate the role of disease-associated loci in the survival and function of human pancreatic beta-like cells in vitro and in vivo. The platform can be used to study other disease-associated loci/variants with respect to beta-like cell function. It is worth noting that the glucose-responding cells derived using the current reported protocols are not equivalent to primary human beta cells. Ca^{2+} flux assays suggested that approximately 30%–40% of the insulin-GFP^{+} cells show increased cytosolic Ca^{2+} concentrations in response to glucose stimulation (Figure S7Q), whereas robust glucose-induced signaling was observed in more than 70% of human beta cells based on the previous report (Rezania et al., 2014). The restricted functionality of pancreatic beta-like cells derived using current protocols might limit their application for evaluating subtle contributions of genes to glucose metabolism and Ca^{2+} signaling. Thus, additional work is needed to further improve the protocol to derive mature pancreatic beta-like cells. In addition, the platform established here can also be applied to study the role of disease-associated loci/variants in other diabetes-related cell types, such as hepatocytes, adipocytes, muscles, and/or intestinal neuroendocrine cells. Finally, the system may be used as a high-throughput/content chemical screening platform to identify candidate drugs correcting allele-specific defects for precision therapy of metabolic diseases.

EXPERIMENTAL PROCEDURES

Cell Culture and Chemicals

All experiments were performed using INS$^{GFP/W}$ HES3 cells. hESCs were grown on Matrigel-coated 6-well plates in mTeSR1 medium (STEMCELL Technologies). Cells were maintained at 37°C with 5% CO_2. T5224 was purchased from MedChem Express (HY-12270). Human islets were provided by IIDP (Integrated Islet Distribution Program).

Creation of Isogenic Mutant hESC Lines

To mutate the target genes, two sgRNAs targeting the first two exons of the target gene were designed, cloned into a vector carrying a CRISPR-Cas9 gene, and validated using the surveyor assay in 293T cells. After validation, *INS*$^{GFP/W}$ HES3 cells were dissociated using Accutase (STEMCELL Technologies) and transfected (8 × 10^5 cells per sample) in suspension using Human Stem Cell Nucleofector solution (Lonza) using electroporation and following the manufacturer's instructions. Cells were co-transfected with the vector expressing Cas9/sgRNA at 10 nM final concentration and a vector expressing puromycin. After replating, the transfected cells were treated with 500 ng/ml puromycin. After 2 days of puromycin selection, hESCs were dissociated into single cells by Accutase and replated at low density. Ten micromolar Y-27632 was added. After approximately 10 days, individual colonies were picked, mechanically disaggregated, and replated into two individual wells of 96-well plates. A portion of the cells was analyzed by genomic DNA sequencing. For biallelic frameshift mutants, we chose both homozygous mutants and compound heterozygous mutants. Wild-type clonal lines from the corresponding targeting experiments were included as wild-type controls to account for potential nonspecific effects associated with the gene-targeting process.

Stepwise Differentiation

Wild-type and isogenic mutant hESCs were differentiated using either of two slightly modified protocols from what was previously reported (Rezania et al., 2014). The details of protocol 1 and 2 are listed as Figure S1C and described in detail in the Supplemental Experimental Procedures.

In Vivo Transplantation, GSIS, and IPGTT

Wild-type and isogenic mutant hESCs at day 30 of differentiation were resuspended in 40 μl DMEM+B27 and transplanted under the kidney capsule of 6- to 8-week-old male SCID-beige mice. Two days after transplantation, the mice were treated with 200 mg/kg STZ. To perform GSIS, mice were starved for about 20 hr. Mouse blood was collected under fasting conditions and at 15 min after intraperitoneal injection with 3 g/kg glucose solution. The mouse sera were analyzed using the ultrasensitive human insulin ELISA kit (ALPCO; 80-INSHUU-E01.1). To perform IPGTT analysis, the mice were fasted overnight and treated with 2 g/kg glucose. Blood glucose level (mg/dl) in each animal was measured before and every 15 min in the first hour and every 30 min in the second hour after glucose injection. The mice transplanted with wild-type or *CDKAL1*$^{-/-}$ cells were orally treated with 300 mg/kg T5224 dissolved in polyvinylpyrrolidone K 60 solution (Sigma). After 48 hr treatment, the mice were examined for GSIS and IPGTT. The mice treated with polyvinylpyrrolidone K 60 solution (vehicle) were used as the controls. For long-term treatment, the mice were orally treated with 300 mg/kg T5224 twice a week for 4 weeks. GSIS and IPGTT were measured 48 hr after the last treatment.

High-Content Chemical Screening

To perform the high-content chemical screening, *CDKAL1*$^{-/-}$ D30 cells were plated on 804G-coated 384-well plates at 5,000 cells/40 μl medium/well. After overnight incubation, cells were treated at 10 μM with compounds from a chemical collection containing the Prestwick FDA-approved drug library and drugs in clinical trials. DMSO treatment was used as a negative control. After 48 hr incubation, cells were first stained with 100 μg/ml PI and then fixed and stained using an insulin antibody (Dako). Plates were analyzed using a Molecular Devices ImageXpress High-Content Analysis System. Two-dimensional analysis was used. Compounds decreasing the cell death rate in excess of 80% and increasing the number of insulin^{+} cells by 2-fold were picked as primary hits.

Statistical Analysis

n = 3 independent biological replicates if not otherwise specifically indicated. n.s. indicates non-significant difference. p values were calculated by unpaired two-tailed Student's t test if not otherwise specifically indicated. n = 8 mice for in vivo experiments if not otherwise specifically indicated. p values were calculated by one-way repeated-measures ANOVA or two-way repeated-measures ANOVA with a Bonferroni test for multiple comparisons between wild-type and KO cells. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2016.07.002.

AUTHOR CONTRIBUTIONS

S.C., H.Z., J.G., and T.E. designed the project; H.Z. and M.G. performed most experiments; T.Z., L.T., C.N.C., X.D., A.S.Y., and L.Y. performed other

necessary experiments; T.Z. and J.Z.X. performed the bioinformatics analysis; H.Z., M.G., and S.C. analyzed data; and H.Z., S.C., J.G., Q.Q., and T.E. wrote the manuscript.

CONFLICTS OF INTEREST

The authors have filed a patent entitled, "AP-1 inhibitors for precision therapy of diabetic patients."

ACKNOWLEDGMENTS

S.C. is funded by The New York Stem Cell Foundation (R-103), American Diabetes Association (1-12-JF-06), and Tri-institutional Starr Stem Cell Grant (2014-030). S.C. is a New York Stem Cell Foundation-Robertson Investigator. T.Z. is funded by a NYSTEM postdoctoral fellowship. A.S.Y. and L.Y. are supported by NIH HL078960 (to L.Y.) and AHA grant 16GRNT26430113 (to L.Y.). J.G. was supported by "Biomedical Research Program" funds at Weill Cornell Medical College in Qatar, a program funded by Qatar Foundation. Q.Q. is supported by a Scientist Development Award (K01HL129892) from the NHLBI. This study was also supported by a Shared Facility contract to T.E. and S.C. from the New York State Department of Health (NYSTEM C029156). NKX6.1 and NKX2.2 antibodies were provided by University of Iowa Hybridoma bank. Human islets were provided by The Integrated Islet Distribution Program. The plentiCRISPR v2 vector and Puro 2.0 were purchased from Addgene (plasmid no. 52961 and no. 24970). We are also very grateful for technical support and advice provided by Harold S. Ralph in the Cell Screening Core Facility and Jason McCormick in the Flow Cytometry Facility at Weill Cornell Medical College, NY.

Received: December 21, 2015
Revised: April 8, 2016
Accepted: July 1, 2016
Published: August 11, 2016

REFERENCES

Abdelli, S., Abderrahmani, A., Hering, B.J., Beckmann, J.S., and Bonny, C. (2007). The c-Jun N-terminal kinase JNK participates in cytokine- and isolation stress-induced rat pancreatic islet apoptosis. Diabetologia *50*, 1660–1669.

Aikawa, Y., Morimoto, K., Yamamoto, T., Chaki, H., Hashiramoto, A., Narita, H., Hirono, S., and Shiozawa, S. (2008). Treatment of arthritis with a selective inhibitor of c-Fos/activator protein-1. Nat. Biotechnol. *26*, 817–823.

Asahara, S., Etoh, H., Inoue, H., Teruyama, K., Shibutani, Y., Ihara, Y., Kawada, Y., Bartolome, A., Hashimoto, N., Matsuda, T., et al. (2015). Paternal allelic mutation at the Kcnq1 locus reduces pancreatic β-cell mass by epigenetic modification of Cdkn1c. Proc. Natl. Acad. Sci. USA *112*, 8332–8337.

Bitner-Glindzicz, M., Lindley, K.J., Rutland, P., Blaydon, D., Smith, V.V., Milla, P.J., Hussain, K., Furth-Lavi, J., Cosgrove, K.E., Shepherd, R.M., et al. (2000). A recessive contiguous gene deletion causing infantile hyperinsulinism, enteropathy and deafness identifies the Usher type 1C gene. Nat. Genet. *26*, 56–60.

Boini, K.M., Graf, D., Hennige, A.M., Koka, S., Kempe, D.S., Wang, K., Ackermann, T.F., Föller, M., Vallon, V., Pfeifer, K., et al. (2009). Enhanced insulin sensitivity of gene-targeted mice lacking functional KCNQ1. Am. J. Physiol. Regul. Integr. Comp. Physiol. *296*, R1695–R1701.

Brambillasca, S., Altkrueger, A., Colombo, S.F., Friederich, A., Eickelmann, P., Mark, M., Borgese, N., and Solimena, M. (2012). CDK5 regulatory subunit-associated protein 1-like 1 (CDKAL1) is a tail-anchored protein in the endoplasmic reticulum (ER) of insulinoma cells. J. Biol. Chem. *287*, 41808–41819.

Gloyn, A.L., Weedon, M.N., Owen, K.R., Turner, M.J., Knight, B.A., Hitman, G., Walker, M., Levy, J.C., Sampson, M., Halford, S., et al. (2003). Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. Diabetes *52*, 568–572.

Gloyn, A.L., Pearson, E.R., Antcliff, J.F., Proks, P., Bruining, G.J., Slingerland, A.S., Howard, N., Srinivasan, S., Silva, J.M., Molnes, J., et al. (2004). Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. N. Engl. J. Med. *350*, 1838–1849.

Hivert, M.F., Vassy, J.L., and Meigs, J.B. (2014). Susceptibility to type 2 diabetes mellitus–from genes to prevention. Nat. Rev. Endocrinol. *10*, 198–205.

Hua, H., Shang, L., Martinez, H., Freeby, M., Gallagher, M.P., Ludwig, T., Deng, L., Greenberg, E., Leduc, C., Chung, W.K., et al. (2013). iPSC-derived β cells model diabetes due to glucokinase deficiency. J. Clin. Invest. *123*, 3146–3153.

Hughes, J.H., Watson, M.A., Easom, R.A., Turk, J., and McDaniel, M.L. (1990). Interleukin-1 induces rapid and transient expression of the c-fos proto-oncogene in isolated pancreatic islets and in purified beta-cells. FEBS Lett. *266*, 33–36.

Massa, O., Iafusco, D., D'Amato, E., Gloyn, A.L., Hattersley, A.T., Pasquino, B., Tonini, G., Dammacco, F., Zanette, G., Meschi, F., et al.; Early Onset Diabetes Study Group of the Italian Society of Pediatric Endocrinology and Diabetology (2005). KCNJ11 activating mutations in Italian patients with permanent neonatal diabetes. Hum. Mutat. *25*, 22–27.

Micallef, S.J., Li, X., Schiesser, J.V., Hirst, C.E., Yu, Q.C., Lim, S.M., Nostro, M.C., Elliott, D.A., Sarangi, F., Harrison, L.C., et al. (2012). INS(GFP/w) human embryonic stem cells facilitate isolation of in vitro derived insulin-producing cells. Diabetologia *55*, 694–706.

Nielsen, E.M., Hansen, L., Carstensen, B., Echwald, S.M., Drivsholm, T., Glümer, C., Thorsteinsson, B., Borch-Johnsen, K., Hansen, T., and Pedersen, O. (2003). The E23K variant of Kir6.2 associates with impaired post-OGTT serum insulin response and increased risk of type 2 diabetes. Diabetes *52*, 573–577.

Ohara-Imaizumi, M., Yoshida, M., Aoyagi, K., Saito, T., Okamura, T., Takenaka, H., Akimoto, Y., Nakamichi, Y., Takanashi-Yanobu, R., Nishiwaki, C., et al. (2010). Deletion of CDKAL1 affects mitochondrial ATP generation and first-phase insulin exocytosis. PLoS ONE *5*, e15553.

Okamura, T., Yanobu-Takanashi, R., Takeuchi, F., Isono, M., Akiyama, K., Shimizu, Y., Goto, M., Liang, Y.Q., Yamamoto, K., Katsuya, T., et al. (2012). Deletion of CDKAL1 affects high-fat diet-induced fat accumulation and glucose-stimulated insulin secretion in mice, indicating relevance to diabetes. PLoS ONE *7*, e49055.

Pagliuca, F.W., Millman, J.R., Gürtler, M., Segel, M., Van Dervort, A., Ryu, J.H., Peterson, Q.P., Greiner, D., and Melton, D.A. (2014). Generation of functional human pancreatic β cells in vitro. Cell *159*, 428–439.

Pasyk, E.A., Kang, Y., Huang, X., Cui, N., Sheu, L., and Gaisano, H.Y. (2004). Syntaxin-1A binds the nucleotide-binding folds of sulphonylurea receptor 1 to regulate the KATP channel. J. Biol. Chem. *279*, 4234–4240.

Pharmaceutical Medicine (2014). Drugs in clinical development for rheumatoid arthritis summary and table. Pharmaceut. Med. *28*, 195–213.

Proks, P., Antcliff, J.F., Lippiat, J., Gloyn, A.L., Hattersley, A.T., and Ashcroft, F.M. (2004). Molecular basis of Kir6.2 mutations associated with neonatal diabetes or neonatal diabetes plus neurological features. Proc. Natl. Acad. Sci. USA *101*, 17539–17544.

Remedi, M.S., Rocheleau, J.V., Tong, A., Patton, B.L., McDaniel, M.L., Piston, D.W., Koster, J.C., and Nichols, C.G. (2006). Hyperinsulinism in mice with heterozygous loss of K(ATP) channels. Diabetologia *49*, 2368–2378.

Rezania, A., Bruin, J.E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T., et al. (2014). Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. Nat. Biotechnol. *32*, 1121–1133.

Rosengren, A.H., Braun, M., Mahdi, T., Andersson, S.A., Travers, M.E., Shigeto, M., Zhang, E., Almgren, P., Ladenvall, C., Axelsson, A.S., et al. (2012). Reduced insulin exocytosis in human pancreatic β-cells with gene variants linked to type 2 diabetes. Diabetes *61*, 1726–1733.

Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., Daly, M.J., et al.; Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science *316*, 1331–1336.

Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U., et al. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science *316*, 1341–1345.

Shang, L., Hua, H., Foo, K., Martinez, H., Watanabe, K., Zimmer, M., Kahler, D.J., Freeby, M., Chung, W., LeDuc, C., et al. (2014). β-cell dysfunction due to increased ER stress in a stem cell model of Wolfram syndrome. Diabetes *63*, 923–933.

Shimomura, K., Girard, C.A., Proks, P., Nazim, J., Lippiat, J.D., Cerutti, F., Lorini, R., Ellard, S., Hattersley, A.T., Barbetti, F., and Ashcroft, F.M. (2006). Mutations at the same residue (R50) of Kir6.2 (KCNJ11) that cause neonatal diabetes produce different functional effects. Diabetes *55*, 1705–1712.

Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Walters, G.B., Styrkarsdottir, U., Gretarsdottir, S., Emilsson, V., Ghosh, S., et al. (2007). A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat. Genet. *39*, 770–775.

Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G., Ng, D.P., Holmkvist, J., Borch-Johnsen, K., Jørgensen, T., et al. (2008). SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. Nat. Genet. *40*, 1098–1102.

Wei, F.Y., Suzuki, T., Watanabe, S., Kimura, S., Kaitsuka, T., Fujimura, A., Matsui, H., Atta, M., Michiue, H., Fontecave, M., et al. (2011). Deficit of tRNA(Lys) modification by Cdkal1 causes the development of type 2 diabetes in mice. J. Clin. Invest. *121*, 3598–3608.

Yamagata, K., Senokuchi, T., Lu, M., Takemoto, M., Fazlul Karim, M., Go, C., Sato, Y., Hatta, M., Yoshizawa, T., Araki, E., et al. (2011). Voltage-gated K+ channel KCNQ1 regulates insulin secretion in MIN6 β-cell line. Biochem. Biophys. Res. Commun. *407*, 620–625.

Yasuda, K., Miyake, K., Horikawa, Y., Hara, K., Osawa, H., Furuta, H., Hirota, Y., Mori, H., Jonsson, A., Sato, Y., et al. (2008). Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat. Genet. *40*, 1092–1097.

Zhang, S., Liu, J., MacGibbon, G., Dragunow, M., and Cooper, G.J. (2002). Increased expression and activation of c-Jun contributes to human amylin-induced apoptosis in pancreatic islet beta-cells. J. Mol. Biol. *324*, 271–285.

Cell Stem Cell
Article

Stage-Specific Human Induced Pluripotent Stem Cells Map the Progression of Myeloid Transformation to Transplantable Leukemia

Andriana G. Kotini,[1,2,3,4] Chan-Jung Chang,[1,2,3,4] Arthur Chow,[5,6,7,8] Han Yuan,[9] Tzu-Chieh Ho,[5,6,7,8] Tiansu Wang,[1,2,3,4] Shailee Vora,[1,2,3,4] Alexander Solovyov,[1,2,4,10] Chrystel Husser,[1,2,3,4] Malgorzata Olszewska,[1,2,3,4] Julie Teruya-Feldstein,[10] Deepak Perumal,[1,2,4] Virginia M. Klimek,[11] Alexandros Spyridonidis,[12] Raajit K. Rampal,[11] Lewis Silverman,[2,4] E. Premkumar Reddy,[2,4] Elli Papaemmanuil,[13] Samir Parekh,[1,2,4] Benjamin D. Greenbaum,[1,2,4,10] Christina S. Leslie,[9] Michael G. Kharas,[5,6,7,8,*] and Eirini P. Papapetrou[1,2,3,4,14,*]

[1]Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[2]Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[3]Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[4]Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[5]Molecular Pharmacology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[6]Center for Cell Engineering, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[7]Center for Stem Cell Biology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[8]Center for Experimental Therapeutics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[9]Computational Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[10]Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[11]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[12]School of Medicine, University of Patras, Patras 26504, Greece
[13]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[14]Lead Contact
*Correspondence: kharasm@mskcc.org (M.G.K.), eirini.papapetrou@mssm.edu (E.P.P.)
http://dx.doi.org/10.1016/j.stem.2017.01.009

SUMMARY

Myeloid malignancy is increasingly viewed as a disease spectrum, comprising hematopoietic disorders that extend across a phenotypic continuum ranging from clonal hematopoiesis to myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML). In this study, we derived a collection of induced pluripotent stem cell (iPSC) lines capturing a range of disease stages encompassing preleukemia, low-risk MDS, high-risk MDS, and secondary AML. Upon their differentiation, we found hematopoietic phenotypes of graded severity and/or stage specificity that together delineate a phenotypic roadmap of disease progression culminating in serially transplantable leukemia. We also show that disease stage transitions, both reversal and progression, can be modeled in this system using genetic correction or introduction of mutations via CRISPR/Cas9 and that this iPSC-based approach can be used to uncover disease-stage-specific responses to drugs. Our study therefore provides insight into the cellular events demarcating the initiation and progression of myeloid transformation and a new platform for testing genetic and pharmacological interventions.

INTRODUCTION

Human hematopoiesis is sustained by hematopoietic stem and progenitor cells (HSPCs) residing in the bone marrow (BM) through processes involving self-renewal, proliferation, and differentiation to distinct cell lineages ultimately giving rise to mature functional hematopoietic cells. Deregulation of these processes is believed to be central to the pathogenesis of hematopoietic disorders, which are typically grouped according to the two main blood lineages into myeloid and lymphoid, with the former generally classified as myeloproliferative disorders (MPDs), myelodysplastic syndromes (MDSs), syndromes with overlap of the two former categories (MDSs/MPDs), and the most dramatic, acute myeloid leukemia (AML). AML can develop de novo or from preexisting MPD or MDS. While the development of de novo AML from preleukemic hematopoietic stem cells (HSCs) and its progression from MPDs (mainly chronic myeloid leukemia [CML]) are better studied, the development of AML from MDS has not been well mapped due to the more limited biological models of MDS and the scarcity and poor growth of primary MDS cells, as opposed to cells from MPD and AML patients (Sperling et al., 2017).

Leukemogenesis has long been conceptualized as a multistep process. All current evidence points to a model whereby MDS and AML arise from HSPCs through the accumulation of multiple genetic (and potentially also epigenetic) changes (Elias et al., 2014). In recent years, deep characterization of the mutational landscape of myeloid disorders through large-scale DNA sequencing solidified a model of clonal evolution through the stepwise accumulation of mutations. Clonal tracking at high resolution

enabled by the identification of tens of recurrent gene mutations in MDS and AML has provided important insights into the nature and clonality status of myeloid disorders. First, it is now clear that clonal hematopoiesis is invariably established at the outset of MDS, and thus MDS is a preleukemic condition not fundamentally very different from AML (Papaemmanuil et al., 2013; Walter et al., 2012, 2013). Second, clonal hematopoiesis (termed clonal hematopoiesis of indeterminate potential [CHIP]) is found in healthy individuals with an age-dependent frequency and is associated with an increased risk of developing MDS, MPD, or AML (Genovese et al., 2014; Jaiswal et al., 2014; Steensma et al., 2015; Xie et al., 2014). This finding, in parallel with recent functional in vitro and in vivo studies, lends support to the existence of preleukemic HSCs that are functionally normal and have multilineage potential but harbor MDS- and AML-related mutations that may give them a clonal advantage (Jan et al., 2012; Shlush et al., 2014). These recent findings invite revisiting the boundaries among normal, premalignant, and malignant hematopoiesis and support an emerging view of myeloid malignancy as a disease spectrum comprising hematopoietic disorders that extend across a phenotypic continuum, ranging from normal hematopoiesis to clonal hematopoiesis or preleukemia to MDS and MDS/AML (Pandolfi et al., 2013; Steensma et al., 2015). However, the cellular events demarcating progression to overt leukemia through a premalignant myelodysplastic phase are not well defined.

Here, we generated patient-derived induced pluripotent stem cells (iPSCs) representative of a range of disease stages across the spectrum of myeloid malignancy, including familial predisposition, low-risk MDS, high-risk MDS, and MDS/AML. We characterized the hematopoiesis derived from this panel of iPSC lines and identified phenotypes of graded severity and/or stage specificity, which together delineate a phenotypic roadmap of disease progression, leading to the most dramatic phenotype of a serially transplantable leukemia. As proof of principle that transitions between stages (progression or reversal) can be modeled in our system, we show that a high-risk MDS-iPSC line can be phenotypically reverted to a premalignant state by correction of a chr7q deletion, whereas a preleukemic iPSC line can progress to either low-risk or high-risk MDS following CRISPR/Cas9-mediated inactivation of the second *GATA2* allele or deletion of chr7q, respectively. We also model the stepwise progression of normal cells to preleukemia and subsequent MDS through the sequential introduction of genetic lesions associated with earlier (ASXL1 truncation) and later (chr7q deletion) disease stages. We then use this model to uncover disease-stage-specific therapeutic effects of 5-AzaC, a drug used as first-line therapy in MDS and whose mechanism of action remains elusive, and rigosertib, a small-molecule inhibitor of RAS signaling. Our study provides insights into the pathophysiologic changes underlying the initiation and progression of myeloid transformation and a new platform to test genetic and pharmacologic interventions to reverse this process.

RESULTS

Integrating Cell Reprogramming with Mutational Analyses Enables the Generation of Disease-Stage-Specific iPSCs

We derived iPSC lines from four patients (patients 1–4) with low-risk MDS (refractory anemia [RA] by French-American-British classification [FAB]), high-risk MDS (refractory anemia with excess blasts [RAEB] by FAB) and secondary AML (sAML or MDS/AML, i.e., AML from preexisting MDS) (Figure 1; Table S1). For reprogramming, we used BM or peripheral blood (PB) mononuclear cells (BMMCs or PBMCs) (Table S1) and reasoned that it might be possible, taking advantage of the genetic and clonal heterogeneity of these cell populations, to derive iPSC lines from normal cells, cells of the major clone, as well as cells from minor subclones. We therefore performed a thorough genetic characterization (karyotype, fluorescence in situ hybridization [FISH], array comparative genomic hybridization [aCGH], and gene mutation analysis) to identify all known recurrent gene mutations and chromosomal abnormalities associated with myeloid neoplasms in the starting cells and the derivative iPSCs and used it to determine the provenance of each iPSC line (Figure 1). Thus, we were able to establish a variety of iPSC lines, which included: (1) iPSC lines derived from the dominant clone (i.e., harboring only genetic lesions present in the majority of the starting cells); (2) iPSC lines derived from sub-clones (i.e., harboring at least one genetic lesion present in a subset of the starting cells): AML-4.10, harboring a sub-clonal *KRAS* G12D mutation, and a second line harboring a sub-clonal *NRAS* Q61R mutation that could only be partially reprogrammed (Table S1); (3) iPSC lines derived from normal hematopoietic cells (i.e., harboring none of the somatic genetic lesions found in the starting cells); and (4) one iPSC line, N-3.10, derived from patient 3, harboring a germline *GATA2* T357N mutation predisposing to MDS/AML (Collin et al., 2015; Hahn et al., 2011) (Table S1). The MDS clone in this patient had acquired an additional somatic mutation in the other *GATA2* allele (*GATA2* 390delK) (Figure S1A), together with additional mutations and a t(1;7)(q10;p10) translocation, resulting in del(7q), a deletion commonly associated with germline *GATA2* mutations (Figure 1) (Wlodarski et al., 2016). All iPSC lines met all criteria of pluripotency for human cells (Figure S2). Reprogramming MDS and AML hematopoietic cells from BM or PB thus allows the derivation of iPSC lines capturing different disease stages, residual normal cells, and cells with predisposing mutations.

These reprogramming experiments, together with the genetic characterization of the original cells and derivative iPSCs, allowed us to make several additional observations.

First, detailed genetic analysis can pinpoint iPSC lines that originate from the same starting cell and are thus not truly different lines. iPSC lines MDS-3.4 and MDS-3.5 were found to both harbor the same *MYB* L51fs mutation, which was not detectable in the starting population, in addition to the somatic genetic lesions found in the starting MDS cells (t(1;7)(q10;p10), *GATA2* T357N, *GATA2* 390delK, *U2AF1* Q157R, *ETV6* S321fs) (Figure 1). This strongly suggested that these lines originated from the same cell, which we confirmed by integration site analysis of the lentiviral vector used for reprogramming (Kotini et al., 2015). Second, since our experiments entailed the parallel reprogramming of a mixed population of cells together with the ability to exclude lines that were not clonally independent (Figure S2D), we had a unique opportunity to directly compare the reprogramming efficiency of cells harboring malignancy-associated genetic lesions to that of normal cells of the same genetic background and determine how specific genetic lesions associated with myeloid malignancy may affect reprogramming

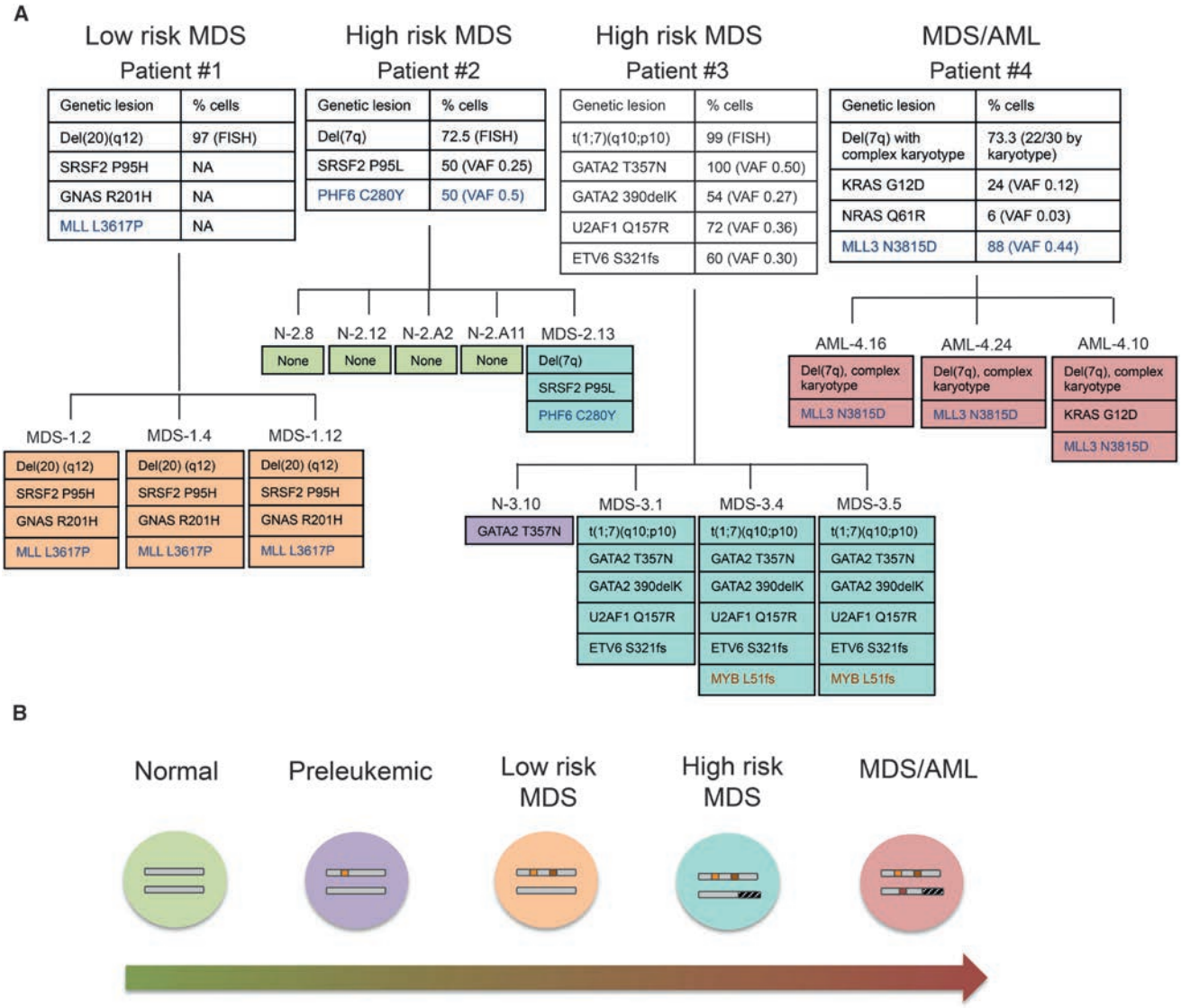


Figure 1. Generation of a Panel of Disease-Stage-Specific iPSCs

(A) iPSCs derived from four patients: one with low-risk MDS, two with high-risk MDS, and one with MDS/AML. The top panels show all recurrent gene mutations and chromosomal abnormalities detected in the starting cells used for reprogramming and their frequency. The bottom panels show the individual iPSC lines that were derived and their corresponding genetic profile. Blue font indicates gene mutations of uncertain significance. Brown font indicates mutations detected in the derivative iPSCs, but not in the starting cells. Patient 4 cells and the derivative iPSCs harbor a complex translocation among chromosomes 1, 7, and 14, resulting in a deletion of 7q (confirmed by aCGH; Figure S2B) and additional material of unknown origin on chromosome 15 (46,XX,der(1)t(1;7;14)(q32;p11p22;p11.1), der(7)del(7)(p11p22)inv(7)(p11q31),der(14)t(1;14)(q32;p11.1),add(15)(p11.1)). VAF, variant allele frequency.

(B) iPSC lines from (A) (note color code) capture distinct disease stages ranging from normal, preleukemic (i.e., cells with predisposing mutations), low-risk MDS, high-risk MDS, and MDS/AML.

See also Figures S1 and S2 and Tables S1 and S2.

efficiency. While we were able to readily reprogram del(7q)-MDS cells, as we have previously reported (Kotini et al., 2015), we were not able to reprogram cells from patients with del(5q)-MDS or monosomy 7 (patients 5–8; Table S1). After at least two attempts for each patient and using different aliquots of starting cells, we were only able to derive either no iPSC lines (patient 5) or only normal iPSC lines, even though normal cells comprised only a minority of the starting population. Since in other reprogramming experiments MDS or AML cells did not have a general reprogramming disadvantage over normal cells (patients 1, 3, and 4; Table S1), this bias is most likely determined by the specific genetic composition of the malignant clone in each patient. In agreement with this, reprogramming of patient 2 cells, in more than one independent reprogramming experiment, gave consistently more normal than MDS iPSCs (only 2 out of 17 iPSC lines were derived from the MDS cells,

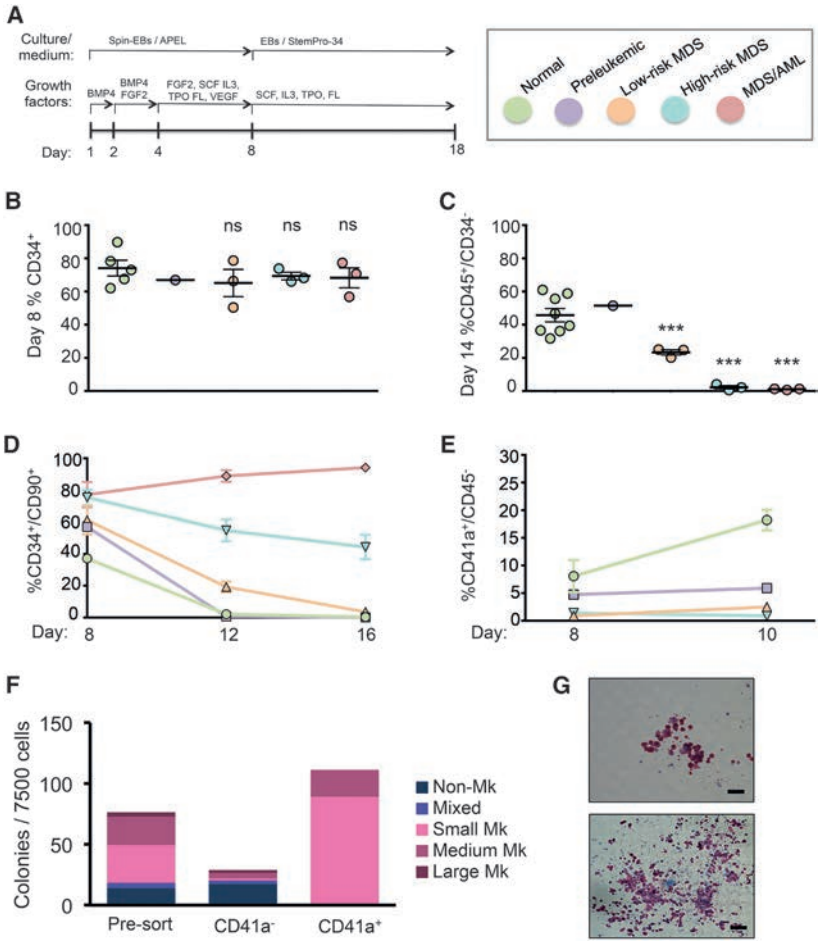


Figure 2. Disease-Stage-Specific iPSCs Capture Phenotypes of Graded Severity

(A) Scheme of hematopoietic differentiation protocol.

(B) Fraction of $CD34^+$ cells generated by all the different lines tested on day 8 of differentiation (extended data are shown in Figure S3C). Note color coding (key is shown in the top right panel of this figure). Mean and SEM of different lines are shown. For lines differentiated more than once (Figure S3C), the average value is shown.

(C) Fraction of more mature $CD45^+$ cells that have lost CD34 expression by day 14 of differentiation generated by the different iPSC lines (extended data are shown in Figure S3F). Mean and SEM of different lines are shown. For lines differentiated more than once, and the average value is shown.

(D) Fraction of $CD34^+$ cells maintaining CD90 expression at the indicated days of differentiation (extended data are shown in Figure S4B, top). For lines differentiated more than once, the average value is shown.

(E) Fraction of $CD41a^+/CD45^-$ cells, corresponding to megakaryocyte progenitors, at the indicated days of differentiation (extended data are shown in Figure S4B, bottom). For lines differentiated more than once, the average value is shown.

(F) Colony assays for megakaryocyte progenitors (CFU-Mk) in unsorted, $CD41a^+$, and $CD41a^-$ sorted cells from N-2.12 iPSCs at day 8 of differentiation. The average of two independent experiments is shown.

(G) Representative images of a medium (top) and large (bottom) CFU-Mk colony. Scale bars, 50 μm.

See also Figures S3 and S4 and Table S3.

with the remaining 15 derived from normal cells), whereas the del(7q) was present in over 70% of the starting cells (Table S1). To further investigate this and to determine whether the in vitro culture that is necessary to initiate reprogramming or the reprogramming process per se accounts for this skewing, we compared the clonal composition of cells from patient 2 before (day 0) and after in vitro culture (day 3) (Figures S1B and S1C). This showed that in vitro culture did not select for normal cells but rather resulted in preferential growth of the MDS clone over the normal cells, since the variant allele frequency (VAF) of both *SRSF2* P95L and *PHF6* C280Y clonal somatic mutations increased over time in culture. Similarly, cells from patients 7 and 8 analyzed for the del(5q) abnormality after culture and immediately before the initiation of reprogramming were found to consist mostly of clonal MDS cells, as the copy number of chr5q was almost 1 in both samples and 8 out of 10 metaphases of patient 7 cells harbored the del5q by karyotyping (Figures S1D and S1E). Finally, we compared the outcome of two different reprogramming protocols using in vitro expansion and reprogramming of either hematopoietic progenitors or erythroblasts, performed in parallel with the same aliquot of starting cells divided in two (Table S2). Erythroblast reprogramming, similarly to hematopoietic progenitor reprogramming, preferentially gave rise to normal iPSCs. These results show that the relative reprogramming disadvantage of MDS cells from some patients is not due to in vitro culture and suggest that some MDS-associated genetic lesions, but not others, exert a negative effect on or even completely abolish reprogramming potential.

Disease-Stage-Specific iPSCs Capture Cellular Phenotypes of Graded Severity or Disease Specificity

We selected a panel of iPSC lines representative of the different disease stages for phenotypic characterization following hematopoietic differentiation using a protocol that enables the derivation and study of hematopoiesis with definitive features (Figures 2A, S3A, and S3B). All lines gave rise to comparable percentages of $CD34^+$ cells early in differentiation (day 8), providing evidence against early developmental defects in mesoderm formation or hematopoietic lineage specification that could compound the identification of disease-relevant phenotypes (Figures 2B and S3C). In contrast, striking differences were observed in the timing and emergence of $CD45^+$ hematopoietic progenitor cells (HPCs) starting at the low-risk MDS stage (Figures 2C and S3D–S3F). Preleukemic cells, like normal cells, generated $CD45^+$ HPCs that comprised ~90% of the cells by day 14 of differentiation, with approximately half having lost CD34 expression, as a sign of further maturation beyond the progenitor stage (Figures 2C, S3E, and S3F). Low-risk MDS $CD45^+$ HPCs appeared later and matured later than normal HPCs, as evidenced by loss of CD34 (Figures 2C, S3E, and S3F). High-risk MDS iPSCs produced $CD45^+$ HPCs with a delay, as well as markedly

reduced overall efficiency (in contrast to low-risk MDS) (Figures 2C and S3D–S3F). On the other hand, MDS/AML-iPSCs gave rise to CD45$^+$ HPCs with efficiencies comparable to those of normal cells but failed to differentiate further and retained CD34 expression until day 18 and beyond (Figures 2C and S3D–S3F). Reciprocally, CD90 expression, normally lost by day 10–12 of differentiation, was retained by low-risk MDS, high-risk MDS, and MDS/AML cells in a stage-specific manner (Figures 2D, S3G, S4A, and S4B).

We found that megakaryocyte progenitors with the CD41a$^+$/CD45$^-$ surface phenotype that give rise to CFU-Mk colonies emerge in these cultures on or before day 8 in normal cells (Figures 2E–2G, S4B, and S4C). Strikingly, this population was severely decreased already in preleukemic cells and effectively abolished in MDS (Figures 2E, S4B, and S4C). While normal iPSCs gave rise to all types of hematopoietic colonies in methylcellulose cultures (Figures S5A and S5B), iPSCs from all disease stages exhibited reduced clonogenicity, with erythroid and multilineage colonies (burst-forming unit-erythrocyte [BFU-E] and colony-forming unit-granulocyte, erythrocyte, monocyte, megakaryocyte [CFU-GEMM]) primarily affected already in preleukemic and, more so, in low-risk MDS cells, while high-risk MDS generated very few or no colonies (Figures 3A and 3B). MDS/AML cells gave rise exclusively to myeloid colonies composed mostly of immature cells (Figures 3A and 3B). Morphologic assessment of HPCs on day 14 of differentiation and of more mature cells from methylcellulose cultures revealed dysplastic changes, which were milder and restricted to the erythroid lineage in preleukemic and low-risk MDS and more widespread and affecting all lineages in high-risk MDS cells (Figures 3B and S5C).

Finally, we measured the growth rate and viability of HPCs derived from the different disease stage iPSCs. Low-risk MDS showed a mild decrease in growth rate and viability, which was much more pronounced in high-risk MDS, whereas growth and viability of MDS/AML cells was completely restored to normal levels (Figures 3C, 3D, and S5D–S5F).

MDS/AML-Derived Hematopoietic Cells Give Rise to Serially Transplantable Leukemia

To assess in vivo engraftment potential, we then transplanted day 8–16 HPCs derived from iPSCs of the various disease stages into *NOD/SCID/IL-2Rγ*$^{-/-}$ (NSG) mice (Figure 4A). As expected from many previous studies, HPCs derived from normal iPSCs showed no detectable engraftment (Vo and Daley, 2015) (Figures 4B and 4C). Similarly, HPCs from MDS iPSCs (both low-risk and high-risk) did not exhibit engraftment potential (Figures 4B and 4C). In contrast, MDS/AML-HPCs showed high levels (up to 80%) of human engraftment in multiple animals (Figures 4B and 4C). The transplantable cells showed features of myeloid leukemia, including a predominantly myeloid immunophenotype and infiltration of the bone marrow and spleen by immature human CD45$^+$ cells with blast-like morphology, also found in the peripheral blood, which could be transplanted into secondary recipients (Figures 4D–4H). The latter readily succumbed to an AML-like disease within 3 weeks of transplantation.

In summary, our phenotypic analyses show that iPSCs derived from distinct disease stages across the myeloid malignancy spectrum capture hematopoietic phenotypes of graded severity and/or stage specificity that together delineate a phenotypic roadmap to myeloid transformation, ultimately leading to a fulminant serially transplantable myeloid leukemia (Figure 4I; Table S3).

Transcriptomes of Disease-Stage-Specific iPSC-Derived HPCs Recapitulate Features of Disease Progression

We performed RNA sequencing (RNA-seq) of sorted CD34$^+$ cells from three normal lines (two iPSC lines and the H1 hESC line) and two low-risk MDS, three high-risk MDS, and three MDS/AML iPSC lines (Figures 5 and S6A). By examining the gene expression profile among the different disease stages, we found a clear clustering of samples according to disease status by principal-component analysis, with the first principal component separating normal from MDS and the second separating the AML from the MDS samples (Figure 5A). Differential expression analyses identified 472 upregulated and 329 downregulated, 868 upregulated and 284 downregulated, and 760 upregulated and 439 downregulated genes among AML versus normal, high-risk MDS versus normal, and low-risk MDS versus normal, respectively ($\log2FC > 3$ or $\log2FC < -3$ and adjusted P value < 0.05; Figure S6B). Hierarchical clustering of all lines based on these differentially expressed genes recapitulated progression from normal to MDS/AML (Figure 5B). Based on analysis of Gene Ontology (GO) categories, genes involved in positive regulation of apoptosis and negative regulation of cell proliferation became upregulated at the transition from normal to low-risk MDS and subsequently downregulated upon transformation to MDS/AML, in agreement with our phenotypic analyses (Figure 5C). Similarly, negative regulation of differentiation was a category upregulated early on. By gene set enrichment analysis (GSEA), we identified significantly enriched disease-specific and shared functional pathways in the iPSC-derived HPCs representing the three different disease stages (Figure S6C; Table S4). Notably, both the high-risk MDS- and MDS/AML- iPSC-derived HPCs were significantly enriched for the high-risk MDS deletion 7q gene set, consistent with both their respective disease state and their specific genetic makeup (Figure 5D). Additionally, the MDS/AML-iPSC-HPCs showed specific enrichment for a gene set found in a subset of human AML patients that had a worse clinical prognosis and contained chromosome 7 abnormalities (Valk et al., 2004) (Figure 5E). Overall, these data suggest that gene expression programs found in HPCs derived from our iPSC panel recapitulate disease progression and capture gene expression signatures derived from primary samples from patients with myeloid malignancies.

Modeling Disease Stage Transitions

We next asked if this model and the phenotypes characterized therein could guide modeling transitions between disease stages, as readouts for disease progression or reversal. We first analyzed an iPSC line derived from the high-risk MDS line MDS-2.13 after spontaneous correction of the del(7q) (Kotini et al., 2015) (Figure 6A). Following correction of the del(7q), this line only harbors a *SRSF2* P95L mutation (and a *PHF6* mutation of uncertain significance). Since the *SRSF2* P95L mutation is an early event in MDS and alone not sufficient for the development of MDS, the corrected line (MDS-2.A3C) would be predicted to capture a preleukemic stage (Papaemmanuil et al., 2013).

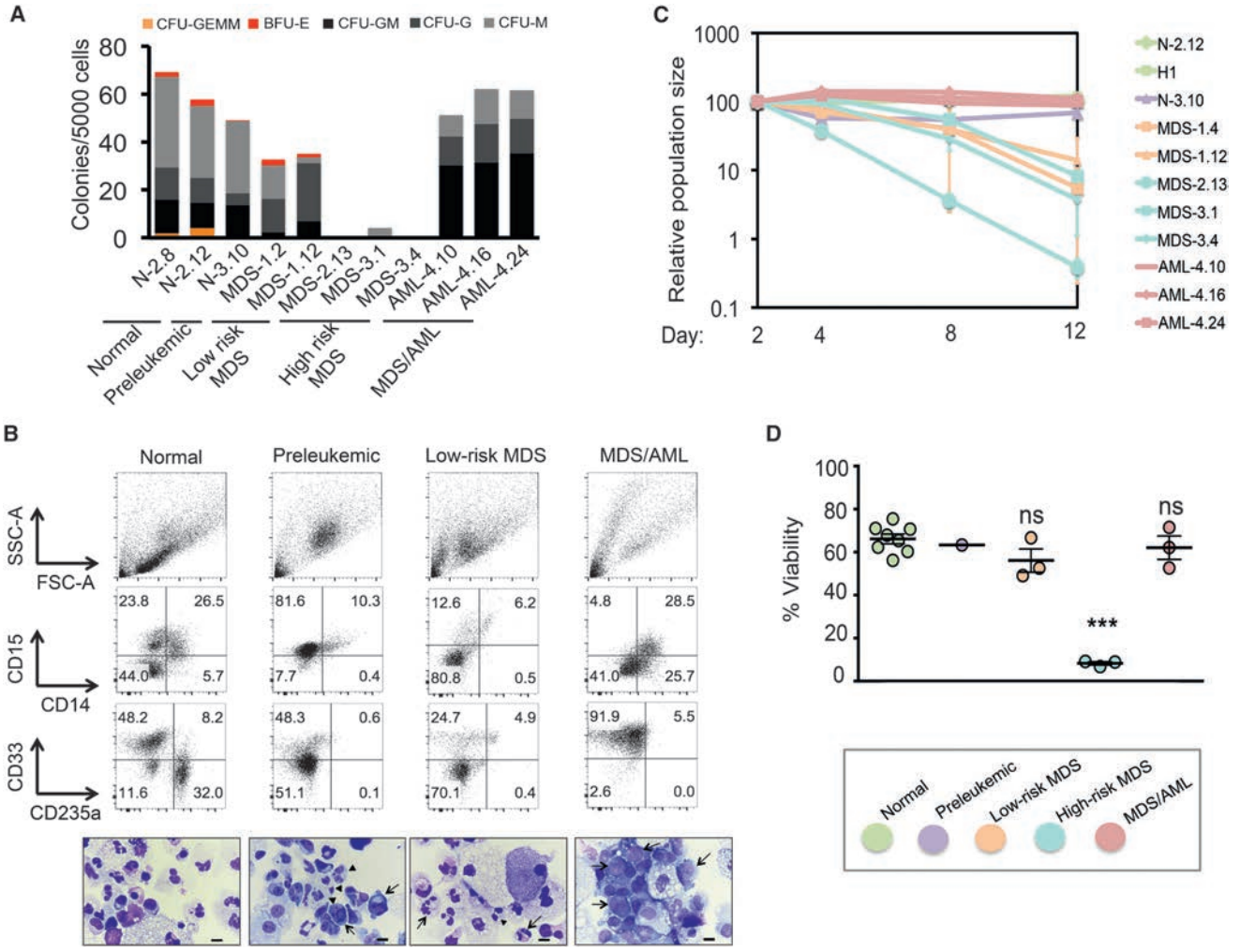


Figure 3. iPSCs from Different Disease Stages Capture Stage-Specific Disease Phenotypes

(A) Methylcellulose assays on day 14 of hematopoietic differentiation. The number of colonies from 5,000 seeded cells is shown. (CFU-GEMM, colony-forming unit-granulocyte, erythrocyte, monocyte, megakaryocyte; CFU-GM: CFU-granulocyte, monocyte; CFU-G: colony-forming unit-granulocyte; CFU-M: CFU-monocyte; BFU-E, burst-forming unit-erythrocyte). Average of three to six independent experiments is shown for each line.

(B) Analysis of lineage markers (top) and morphologic assessment of cells generated in methylcellulose cultures. One iPSC line representative of each disease stage is shown (from left to right: N-2.12, N-3.10, MDS-1.12, AML-4.16). High-risk MDS iPSCs do not give rise to colonies in methylcellulose and are therefore not represented in this panel. Dysplastic changes are observed in preleukemic (arrows point to nuclear blebbing, whereas arrowheads point to pseudo Pelger-Huet cells) and low-risk MDS cells (arrows point to hyper-segmented neutrophils, whereas arrowheads point to pseudo Pelger-Huet cells). Atypical monomorphic myeloid cells (arrows) are the predominant cells observed in methylcellulose cultures from MDS/AML cells. Scale bars, 10 μm.

(C) Growth competition assay. The cells were mixed 1:1 with the N-2.12 line stably expressing GFP at the beginning of hematopoietic differentiation and followed for 12 days by flow cytometry (schematic shown in Figure S5F). The relative population size was calculated as the percentage of GFP^- cells at each time point relative to the population size at day 2. For lines differentiated more than once, the average value is shown.

(D) Cell viability measured by DAPI staining on day 14 of hematopoietic differentiation (extended data are shown in Figure S5E). Mean and SEM of different lines are shown. For lines differentiated more than once (Figure S5E), the average value is shown.

See also Figure S5 and Table S3.

Consistent with this, detailed phenotypic characterization, based on the phenotypic assays defined in the iPSC panel above, confirmed a phenotype corresponding to a preleukemic stage: correction of the emergence of $CD45^+$ cells, loss of CD90 expression, re-emergence of a $CD41a^+/CD45^-$ megakaryocyte progenitor population, partial rescue of clonogenicity, and restored growth and viability (Figures 6B–6F).

Conversely, to model disease progression, we first started with the preleukemic N-3.10 line, harboring a germline *GATA2* mutation. *GATA2* mutations found in patients with familial predisposition syndromes are believed to be loss-of-function mutations (Collin et al., 2015). Since the MDS clone of the same patient from whom this line was derived (patient 3; Figure 1) had acquired a second *GATA2* mutation in the other allele (Figure S1A) together with additional recurrent genetic abnormalities, we sought to model the effects of inactivating the other *GATA2* allele in the disease phenotype. We designed two distinct CRISPR/Cas9-based strategies and isolated two clones

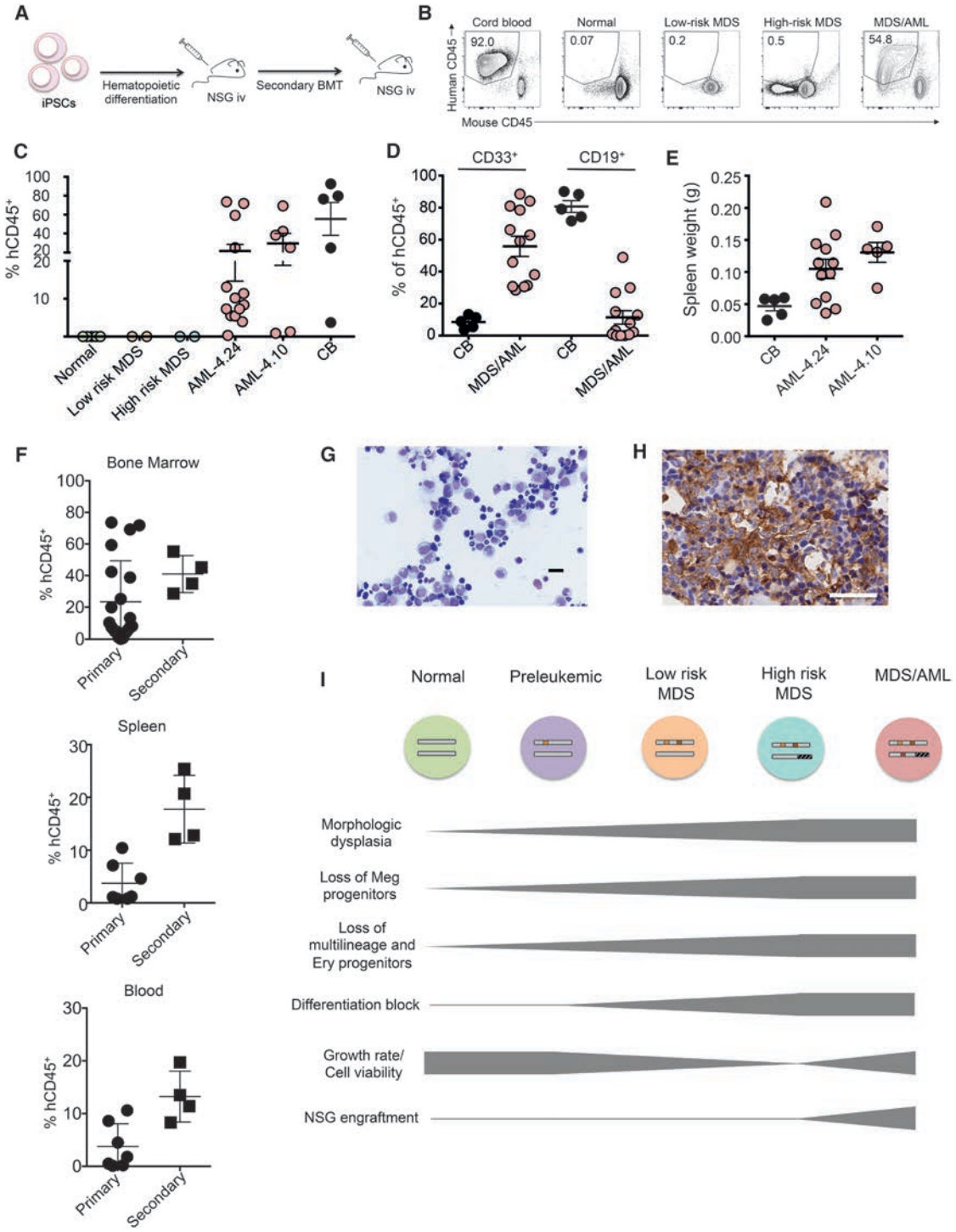


Figure 4. Hematopoietic Cells Derived from MDS/AML-iPSCs Give Rise to Serially Transplantable Leukemia

(A) Scheme of transplantation experiments. Various iPSC lines were differentiated along the hematopoietic lineage for 8–16 days, as shown in Figure 2A and intravenously injected into sub-lethally irradiated or busulfan-treated NSG mice.

(B) Representative flow cytometry panels assessing human cell engraftment in the bone marrow of recipient mice 8–11 weeks post-transplantation.

(C) Engraftment levels in the bone marrow of mice 8–11 weeks after transplantation with HPCs derived from different iPSC lines (normal: N-2.12; low-risk MDS: MDS-1.12; high-risk MDS: MDS-2.13) or with human cord blood CD34$^+$ cells (CB). Error bars show the mean and SEM.

Each data point represents a unique mouse from five independent transplantation experiments.

(D) Fraction of myeloid (CD33$^+$) and lymphoid (CD19$^+$) lineage cells within the hCD45$^+$ population in the BM of mice transplanted with human cord blood CD34$^+$ cells (CB) or MDS/AML-iPSC-derived hematopoietic cells (from lines AML-4.24 and AML-4.10) 8–11 weeks after transplantation. CB engraftment typically gives

(legend continued on next page)

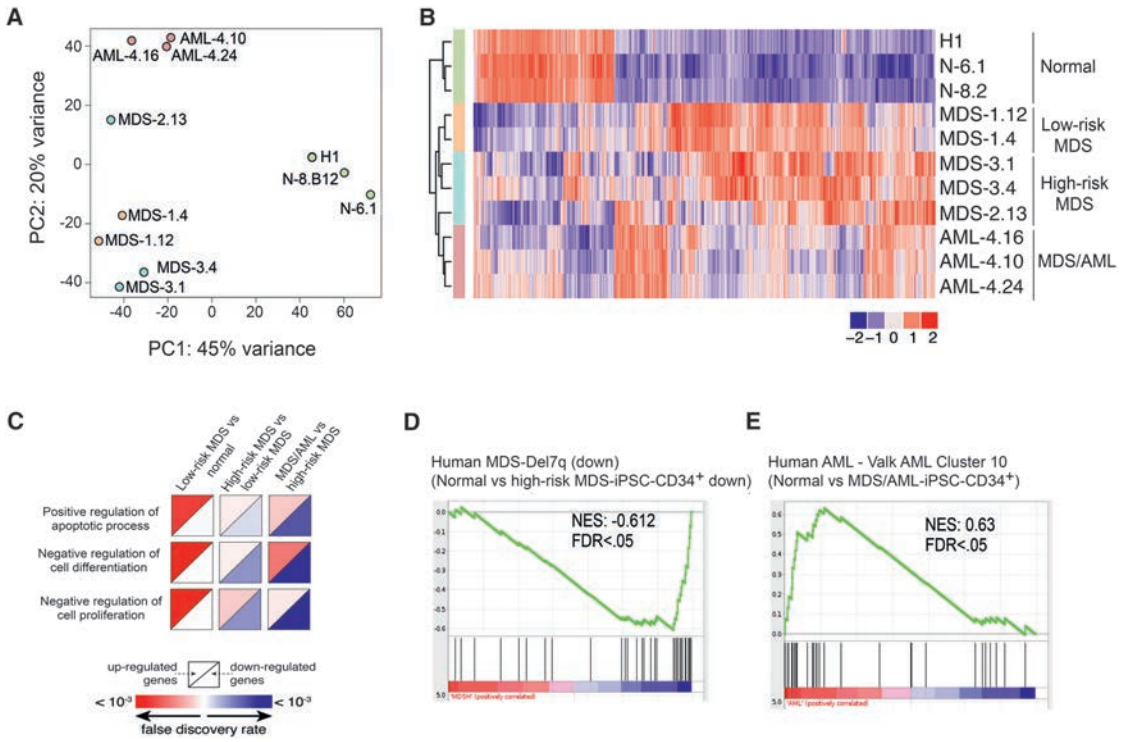


Figure 5. Gene Expression Analysis

(A) Principal-component analysis on regularized log transformed normalized read counts cluster samples by disease status.

(B) Hierarchical clustering of CD34$^+$ HPCs derived from the different iPSC lines based on a total of 2,018 genes differentially expressed between MDS/AML versus normal, high-risk MDS versus normal and low-risk MDS versus normal using scaled log normalized counts.

(C) Gene Ontology (GO) enrichment for genes differentially expressed between low-risk MDS versus normal, high-risk MDS versus low-risk MDS, and high-risk MDS versus MDS/AML.

(D) Gene set enrichment analysis (GSEA) of high-risk MDS-iPSC-CD34$^+$ cells compared to normal iPSC-CD34$^+$ cells shows negative enrichment for down-regulated genes in human high-risk MDS patients harboring a del7q.

(E) GSEA of MDS/AML-iPSC-CD34$^+$ cells compared to normal iPSC-CD34$^+$ cells shows negative enrichment for the human AML Valk cluster 10 patient gene set (group with worse prognosis and chr7q abnormalities) (Valk et al., 2004).

See also Figure S6 and Table S4.

derived from the N-3.10 line, in which both *GATA2* alleles harbored inactivating frameshift indels, as well as a third clone that retained the mutant T357N allele and harbored a deletion of the zinc-finger 2 domain (where many of the mutations found in patients cluster), predicted to abolish DNA binding (Figures 6G, S7A, and S7B) (Collin et al., 2015; Hahn et al., 2011). Since the MDS clone in this patient also had loss of chr7q in the context of a t(1;7) translocation (Figure 1), we also modeled in parallel the contribution of the del(7q) lesion to the disease progression by engineering hemizygous chr7q deletion into the preleukemic N-3.10 line (Figures 6G and S7C). Phenotypic characterization of the three independent *GATA2*-engineered clones and of three independent del(7q) clones, compared to the patient-derived parental preleukemic N-3.10 and MDS lines, revealed a modest reduction in the CD45$^+$ cell population concomitant with a retention of the CD90 marker in the *GATA2*-engineered clones, while the del(7q) clones showed a dramatic decrease in CD45$^+$ cells and a much more prolonged expression of CD90 (Figures 6H

rise to predominantly CD19$^+$ B lymphoid cells. In contrast, MDS/AML-iPSC-derived hematopoietic cells generate predominantly myeloid cells. Each data point represents a unique mouse. Mean and SEM of different mice from three experiments are shown.

(E) Spleen weight in recipient mice 8–11 weeks post-transplantation with human cord blood CD34$^+$ cells (CB) or MDS/AML-iPSC-derived hematopoietic cells, as indicated. Each data point represents a unique mouse. Mean and SEM of different mice from three experiments are shown.

(F) Engraftment levels in the bone marrow, spleen, and peripheral blood of primary and secondary recipient mice transplanted with hematopoietic cells derived from lines AML-4.24 and AML-4.10 (primary) or AML-4.10 (secondary). Each data point represents a unique mouse. Error bars show the mean and SEM.

(G) May-Giemsa-stained cytospin of bone marrow cells of a secondary recipient mouse transplanted with AML-4.10-derived hematopoietic cells. Scale bar, 100 μm.

(H) Human CD45 detection by immunohistochemistry in the bone marrow of a secondary recipient mouse transplanted with AML-4.10-derived hematopoietic cells. Scale bar, 50 μm.

(I) Schematic summary of phenotypic analyses shown in Figures 2, 3, and 4.

See also Table S3.

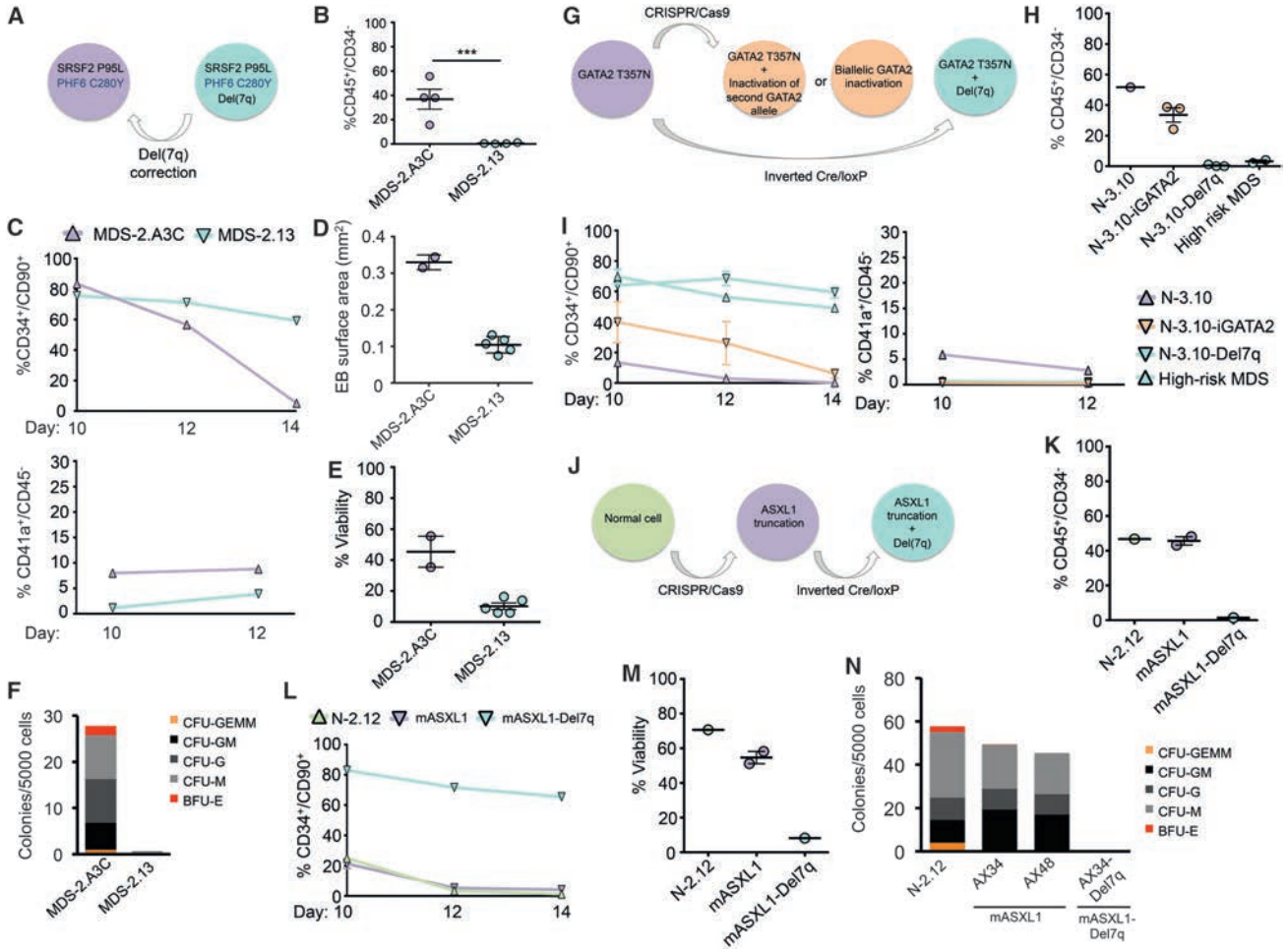


Figure 6. Modeling Disease Stage Transitions

(A) Schematic of reversal of a high-risk MDS line to a stage phenotypically consistent with a preleukemic stage through spontaneous correction of a chr7q deletion.

(B) Fraction of $CD45^+/CD34^-$ cells at day 14 of differentiation. Mean and SEM from independent differentiation experiments are shown.

(C) Fraction of $CD34^+/CD90^+$ cells at the indicated days of differentiation. Average of independent differentiation experiments is shown for each line (top). Fraction of $CD41a^+/CD45^-$ cells at the indicated days of differentiation (Bottom). Average of independent differentiation experiments are shown for each line.

(D) EB surface area at day 8 of hematopoietic differentiation. Mean and SEM from independent differentiation experiments are shown for each line. 10 EBs were measured in each experiment and averaged for each data point.

(E) Cell viability measured by DAPI staining on day 14 of hematopoietic differentiation. Mean and SEM from independent differentiation experiments are shown for each line.

(F) Methylcellulose assays on day 14 of hematopoietic differentiation. The number of colonies from 5,000 seeded cells is shown.

(G) Schematic of progression of a preleukemic line (N-3.10), harboring a heterozygous germline *GATA2* mutation, to a stage corresponding phenotypically to low-risk MDS through CRISPR/Cas9-mediated monoallelic or biallelic *GATA2* inactivation and to a high-risk MDS stage through loss of a copy of chr7q.

(H) Fraction of $CD45^+/CD34^-$ cells on day 14 of differentiation. Mean and SEM of three different *GATA2*-edited and three del7q-engineered clones with average values from two independent differentiations per line are shown.

(I) Fraction of $CD34^+/CD90^+$ cells at the indicated days of differentiation (left). Average values of three different *GATA2*-edited and three del7q-engineered clones with values averaged from two independent differentiation experiments per line are shown. Fraction of $CD41a^+/CD45^-$ cells at the indicated days of differentiation (fight). Average values of three different *GATA2*-edited and three del7q-engineered clones with values averaged from two independent differentiation experiments per line are shown.

(J) Schematic of successive progression of a normal line (N-2.12) to a preleukemic stage through CRISPR/Cas9-mediated *ASXL1* mutation and subsequently to a high-risk MDS stage through engineering of del(7q).

(K) Fraction of $CD45^+/CD34^-$ cells on day 14 of differentiation. Mean and SEM of two different *ASXL1*-edited clones, one *ASXL1*-edited and del7q-engineered clone, and the parental line with average values from three independent differentiations per line are shown.

(L) Fraction of $CD34^+/CD90^+$ cells at the indicated days of differentiation. Average values of two different *ASXL1*-edited clones and one *ASXL1*-edited and del7q-engineered clone with values averaged from three independent differentiation experiments per line are shown.

(M) Cell viability measured by DAPI staining on day 14 of hematopoietic differentiation. Mean and SEM of two different *ASXL1*-edited clones and one *ASXL1*-edited and del7q-engineered clone with average values of three independent differentiations per line are shown.

(N) Methylcellulose assays on day 14 of hematopoietic differentiation. Shown are values averaged from two or three independent differentiation experiments per line.

See also Figure S7.

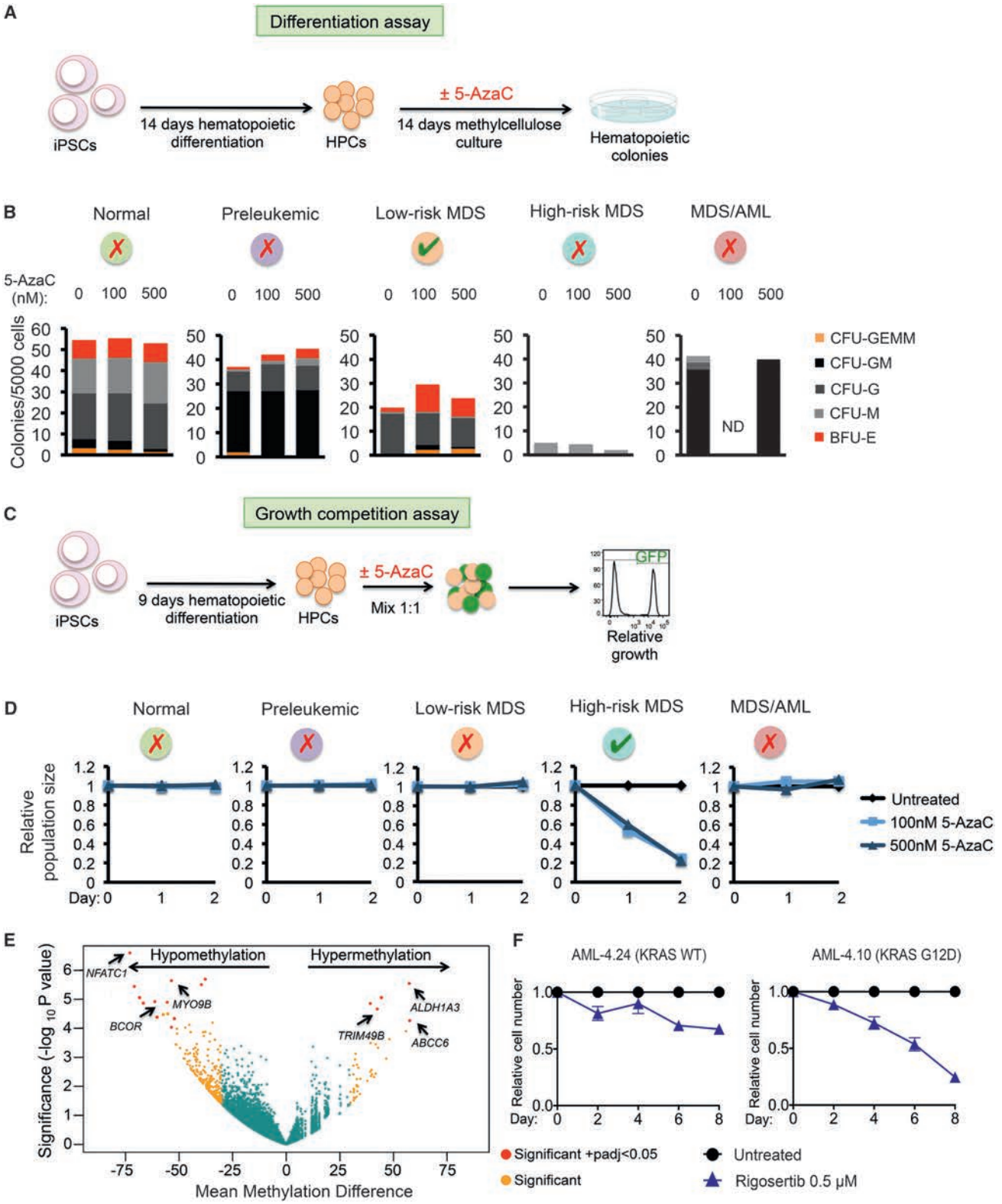


Figure 7. Disease-Stage-Specific Drug Responses

(A) Schematic of experimental design to test the effects of 5-AzaC in differentiation.

(B) Methylcellulose assays at day 14 of hematopoietic differentiation in the presence or absence of 5-AzaC. The number of colonies from 5,000 seeded cells is shown. Normal, average of H1 (two independent experiments) and N-2.12; preleukemic, N-3.10; low-risk MDS, average of MDS-1.2 and MDS-1.12 (four independent experiments); high-risk MDS, MDS-3.4; MDS/AML, AML-4.10 and AML-4.24.

(legend continued on next page)

and 6I). All clones showed a marked decrease of the CD41a$^+$/CD45$^-$ population (Figure 6I). These phenotypes reflect progression to low-risk MDS driven by *GATA2* inactivation and to high-risk MDS driven by the del(7q), suggesting that the del(7q) is a marker of adverse prognosis in the clinic.

We then set to model disease progression along sequential stages driven by the stepwise acquisition of genetic lesions starting from a normal cell. To this end, we first introduced truncating mutations in the *ASXL1* gene and selected two clones with monoallelic truncations (Figures 6J and S7D). *ASXL1* C-terminal truncation is an early event in myeloid malignancies and one of the most common mutations in individuals with CHIP (Link and Walter, 2016; Steensma et al., 2015). We subsequently, in a second step, deleted one copy of chr7q in one of the *ASXL1*-engineered clones (Figures 6J and S7E–S7G). This set of clones recapitulated stepwise progression from normal to preleukemia (*ASXL1* mutation) to high-risk MDS (*ASXL1* mutation + del7q), assessed by CD45 and CD90 marker expression, cell viability, and colony formation (Figures 6K–6N).

These results collectively show that our phenotypic roadmap can be used to model disease stage transitions across the spectrum of myeloid malignancy driven by a variety of genetic lesions and their combinations.

Modeling Disease-Stage-Specific Effects of Therapeutic Interventions

5-Azacytidine (5-AzaC) is a hypomethylating agent that is used as first-line therapy in MDS. 30%–50% of MDS patients show some response, but there are currently limited biomarkers to predict the responders (Bejar and Steensma, 2014). Furthermore, the mechanism by which 5-AzaC exerts its therapeutic effects is not clear. Potential mechanisms may include induction of differentiation or preferential inhibition of the growth of the MDS clone. To first test for potential effects of 5-AzaC in inducing differentiation, we cultured HPCs derived from the different iPSC lines in methylcellulose in the presence or absence of 5-AzaC (Figure 7A). Strikingly, treatment with 5-AzaC resulted in a marked rescue of BFU-E and CFU-GEMM colonies in low-risk MDS-iPSCs (Figures 7B, S7H, and S7I). In contrast, it had no effect in colony growth from normal iPSCs or any other iPSC line from other disease stages. We then tested for selective effects in the growth of the MDS clone using a competitive growth assay. Intriguingly, 5-AzaC had an inhibitory effect in the growth of high-risk MDS-iPSC-derived HPCs, but not of those derived from other disease stage iPSCs or normal iPSCs (Figures 7C and 7D). These results suggest that 5-AzaC may primarily affect differentiation in earlier stages of the disease, whereas its main therapeutic action later on might be mediated through selective inhibition of the MDS clone. DNA methylation analysis of low-risk MDS-iPSC-derived HPCs (MDS-1.12 line) treated with 5-AzaC for 3 days revealed striking genome-wide hypomethylation following 5-AzaC treatment, which included gene promoters, suggesting that hypomethylation may underlie the rescue of colony formation in these cells (Figures 7E and S7J).

To further test for stage-specific drug responses, we treated HPCs derived from two MDS/AML lines from patient 4, capturing a less and a more advanced disease stage, the AML-4.24 line derived from the dominant clone, and the AML-4.10 line derived from the KRAS mutated subclone (Figure 1) with rigosertib, a small-molecule inhibitor of RAS signaling pathways that is currently in clinical trials for high-risk MDS (Athuluri-Divakar et al., 2016). As predicted, AML-4.10 HPCs showed marked sensitivity to rigosertib, whereas AML-4.24 cells were marginally affected (Figure 7F). These results collectively support the use of our disease progression model in drug testing.

DISCUSSION

Here, we used an approach integrating cell reprogramming and cancer genetics to establish iPSC lines representative of distinct stages during the cellular transformation from normal cells to AML through an MDS stage. Detailed genetic and clonal characterization of the starting cell population and the derived iPSC lines allowed us to make additional observations regarding the degree to which the output of reprogramming is representative of the clonal composition of the primary cells. Our results show that the clonal representation of the original cells in the iPSCs is skewed, often in favor of residual normal cells over cells of the premalignant or malignant clone (Table S1). They also show that it is reprogramming per se and not the in vitro stimulation and expansion that accounts for this bias, which seems to be conferred by some MDS- and AML-associated genetic lesions, but not others, while some genetic abnormalities seem to be incompatible with reprogramming (Figures S1B–S1E). Among the ones tested here, del(5q) and monosomy 7 could never be captured in iPSCs, despite cells harboring them comprising over 80% of the starting cell pool. It might be possible to overcome this refractoriness by using alternative reprogramming factor cocktails, which we did not test here. A negative or positive impact of specific cancer-associated gene mutations on the reprogramming "fitness" of the cells would not be surprising given well-studied positive and negative effects, respectively, of TP53 inactivation and Fanconi anemia pathway mutations on reprogramming (Papapetrou, 2016). Importantly, despite the skewed clonal and subclonal representation, we were able to capture normal and preleukemic cells, as well as malignant clones and subclones, and thus compile a panel of lines carrying genomes representative of different disease stages from normal to fully

(C) Schematic of growth competition assay to test the effects of 5-AzaC in cell proliferation relative to normal cells. The cells were mixed 1:1 with the N-2.12 line stably expressing GFP at day 9 of hematopoietic differentiation in the presence or absence of 5-AzaC and followed for an additional 2 days by flow cytometry.
(D) The relative population size was calculated as the percentage of GFP$^-$ cells in the treated cells relative to the percentage of GFP$^-$ cells in the untreated cells at each time point. iPSC lines from left to right: N-2.12, N-3.10, MDS-1.12, MDS-2.13, and AML-4.24.
(E) Volcano plot showing differences in DNA methylation in four HPC samples independently treated with 5-AzaC derived from the MDS-1.12 line in two independent differentiation experiments compared to two untreated controls.
(F) HPCs derived from the AML-4.24 and the AML-4.10 iPSC lines treated with rigosertib. The relative population size was calculated as the number of treated cells relative to the number of untreated cells at each time point. Mean and SEM from triplicate experiments are shown.
See also Figure S7.

transformed states. Here, we applied this strategy to hematologic malignancies, which are particularly amenable to the development of a progression model, because they are relatively genetically simple cancers, and MDS is one of very few well-recognized pre-neoplastic conditions in humans (Martincorena and Campbell, 2015). Although the genetic complexity and low reprogramming efficiency may impose challenges, it is conceivable that similar models can be developed for a variety of other cancers, including solid tumors (Kim et al., 2013; Kim and Zaret, 2015; Papapetrou, 2016).

The phenotypes we characterized in our model bear direct relevance to disease phenotypes at the patient level. Morphologic dysplastic changes are a hallmark and a diagnostic criterion of MDS (Arber et al., 2016). Furthermore, our model presented a pattern of graded severity from unilineage to multilineage dysplasia, similar to what is often observed in the clinic in low-risk versus high-risk MDS cases (Figures 3B and S5B). The impaired differentiation and reduced clonogenicity affecting erythroid and multilineage progenitors first is a very likely correlate of the ineffective hematopoiesis and cytopenias observed in MDS patients, which predominantly affect the erythroid lineage, consistent with findings in primary MDS cells cultured ex vivo (Flores-Figueroa et al., 1999; Sato et al., 1998). The increased cell death is consistent with findings of apoptotic markers in primary patient BM, which has led to the proposition that apoptosis may be another pathophysiologic mechanism accounting for the cytopenias (Kerbauy and Deeg, 2007). The growth and viability defects of MDS cells are abolished upon transformation to full-blown AML, and this is also recapitulated in our model. Our findings are also consistent with previous reports of minimal perturbation of the HSPC compartment in low-risk MDS but a more significant one in higher-risk cases (Elias et al., 2014; Will et al., 2012; Woll et al., 2014). Interestingly, loss of megakaryocyte progenitors is the earliest event in our progression model, which is intriguing in view of recent findings on the close relationship between megakaryocyte progenitors and HSCs (Notta et al., 2016; Sanjuan-Pla et al., 2013; Woolthuis and Park, 2016).

Strikingly, hematopoietic cells derived from our MDS/AML-iPSCs through in vitro differentiation were able to robustly transplant a lethal leukemia when intravenously injected into immunodeficient mice. This is the first demonstration that HSPCs generated from hPSCs through in vitro differentiation possess engraftment ability and is an intriguing finding given the general inability of hematopoiesis derived from human pluripotent stem cells (hPSCs) to engraft (Vo and Daley, 2015). Deeper investigation into the transcriptional programs and cellular processes active in these MDS/AML-iPSC-derived hematopoietic cells may inform ongoing efforts toward the generation of HSPCs with long-term engraftment potential from pluripotent or other cell sources (Vo and Daley, 2015). These cells can also provide an attractive platform for deconstructing and reconstructing clonal evolution in AML and for testing drugs in an in vivo setting.

More than a decade ago, it was proposed that myeloid transformation requires two types of events, one that induces proliferation and one that blocks differentiation, referred to respectively as class I and II mutations (Gilliland and Griffin, 2002; Gilliland and Tallman, 2002). The former would typically involve classic signaling pathways and the latter hematopoietic transcription factors. It was also suggested that class II without class I mutations might result in MDS. Whereas perturbations of proliferation, differentiation, and other processes like self-renewal and cell survival are likely involved in the development of MDS and MDS/AML, it is now obvious that the picture is much more complex and this model can aid future studies in understanding these processes at a cellular and molecular level. However, several limitations need to be noted. MDS is quite heterogeneous genetically and phenotypically, and we only used iPSCs derived from four patients for this study. Our findings that phenotypes of these cells cluster with disease stage supports the well-established observation and long-held idea that diverse genotypes converge to few phenotypes at the cellular and organismal level in myeloid malignancies and cancer more generally. Thus, whereas the derivation of larger collections of MDS and AML iPSC lines in the future can further refine the phenotypic roadmap we delineate here, our findings can already provide a framework to aid investigation into disease mechanisms, drug responses, and the cellular and molecular events driving leukemia progression. Our results align well with the newly emerging view of myeloid malignancy as a spectrum of clinical syndromes encompassing clonal hematopoiesis, MDS, and AML, reflecting disordered hematopoietic processes that can often progress from one to another. However, it is clear that not every patient will necessarily transition through each of these stages. For example, CHIP can progress directly to AML without an MDS stage, whereas MDS and AML can potentially also develop without an antecedent CHIP phase. It is thus conceivable that different routes to myeloid transformation exist and that our findings may not apply to all.

Understanding the cellular events leading to disease stage transitions can help an enhanced understanding of the process of myeloid transformation and cellular transformation more generally and guide drug development targeting specific disease stages or preventing the progression from one stage to another. We provided here proof of principle that transitions between stages (progression or reversal) can be modeled in our system. Our model offers new opportunities to study HSPC populations in MDS and AML, which often cannot be easily obtained at sufficient numbers from primary samples or propagated in patient-derived xenograft models. It also offers the unique opportunity to study disease mechanisms in pure clonal cells devoid of the confounding cellular, genetic, and clonal heterogeneity of primary patient specimens. Mutation of the second *GATA2* allele upon progression to MDS has been described in familial cases of *GATA2* mutation, but its role in disease progression has not been studied before (Collin et al., 2015). Our results suggest that further loss of function of *GATA2* contributes to progression (Figures 6G–6I). This is consistent with a fundamental role of GATA2 in hematopoiesis from studies in mouse models (de Pater et al., 2013). Our findings, however, also suggest that additional events are needed for progression to a more aggressive disease (since *GATA2* knockout [KO] induced rather mild phenotypic changes; Figures 6H and 6I), which is consistent with the finding of additional recurrent MDS-associated somatic lesions in the MDS clone of this patient (Figure 1) and our results showing more dramatic phenotypic changes driven by engineering a del(7q) (Figures 6G–6I). While our results using 5-AzaC and rigosertib treatment warrant further investigation, they demonstrate

the usefulness of this model in testing therapeutic interventions in principle.

Despite the well-established use of iPSCs in disease modeling, their potential to model cancer has barely been explored (Papapetrou, 2016). We show here that integrated patient cell reprogramming and cancer genetics is a powerful way to dissect cancer progression, deconstruct clonal hierarchies, and mimic clonal evolution leveraging CRISPR technology.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENTS AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - iPSC generation
  - Mouse strain used for transplantation experiments
- METHOD DETAILS
  - Mutational analysis
  - Hematopoietic differentiation
  - Globin Gene Expression Analysis
  - Flow cytometry and cell sorting
  - Clonogenic assays
  - Cell growth assays
  - Cytological analyses
  - Transplantation into NSG mice
  - RNA sequencing
  - CRISPR/Cas9-mediated GATA2 inactivation
  - CRISPR/Cas9-mediated ASXL1 mutation
  - Engineering of chr7q deletions
  - Treatment with 5-AzaC
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Genome-wide gene expression analysis
  - DNA methylation analysis by ERRBS
  - Statistical analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2017.01.009.

AUTHOR CONTRIBUTIONS

A.G.K. performed experiments, analyzed data, and assisted with manuscript preparation. C.-J.C., A.C., T.-C.H, T.W., S.V., C H., and M.O. performed experiments and analyzed data. J.T.-F. performed cytological analyses. H.Y., A.S., B.D.G., and C.S.L. performed bioinformatics analyses of RNA-seq data. V.M.K., A.S., and L.S. provided patient samples. R.K.R. and E.P. analyzed gene mutation data. D.P. and S.P. generated and analyzed ERRBS data. E.P.R. provided materials. M.G.K. supervised experiments, analyzed data, and prepared the manuscript. E.P.P. conceived, designed, and supervised the study, analyzed data, and prepared the manuscript.

ACKNOWLEDGMENTS

This work was supported by NIH grants R00DK087923 and R01HL121570, Damon Runyon-Rachleff Innovation Award 30-14 from the Damon Runyon Cancer Research Foundation, the Edward P. Evans Foundation, New Scholar in Aging Award AGS-NS-0984-13 from the Ellison Medical Foundation, and research grants from the Henry and Marilyn Taub Foundation, the Babich Family Foundation, and Alex's Lemonade Stand Foundation (to E.P.P.). This work was also supported by NIH grants R01DK101989, R01CA193842, and P30CA008748, a Kimmel Scholar Award, and a V-Scholar Award (to M.G.K.). H.Y. was supported by the Tri-Institutional Training Program in Computational Biology and Medicine (NIH grant T32GM083937).

Received: June 28, 2016
Revised: December 18, 2016
Accepted: January 26, 2017
Published: February 16, 2017

REFERENCES

Akalin, A., Garrett-Bakelman, F.E., Kormaksson, M., Busuttil, J., Zhang, L., Khrebtukova, I., Milne, T.A., Huang, Y., Biswas, D., Hess, J.L., et al. (2012a). Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. PLoS Genet. *8*, e1002781.

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012c). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. *13*, R87.

Arber, D.A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M.J., Le Beau, M.M., Bloomfield, C.D., Cazzola, M., and Vardiman, J.W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood *127*, 2391–2405.

Athuluri-Divakar, S.K., Vasquez-Del Carpio, R., Dutta, K., Baker, S.J., Cosenza, S.C., Basu, I., Gupta, Y.K., Reddy, M.V., Ueno, L., Hart, J.R., et al. (2016). A Small Molecule RAS-Mimetic Disrupts RAS Association with Effector Proteins to Block Signaling. Cell *165*, 643–655.

Bejar, R., and Steensma, D.P. (2014). Recent developments in myelodysplastic syndromes. Blood *124*, 2793–2803.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219.

Collin, M., Dickinson, R., and Bigley, V. (2015). Haematopoietic and immune defects associated with GATA2 mutation. Br. J. Haematol. *169*, 173–187.

de Pater, E., Kaimakis, P., Vink, C.S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P.S., Camper, S.A., Speck, N., and Dzierzak, E. (2013). Gata2 is required for HSC generation and survival. J. Exp. Med. *210*, 2843–2850.

Elias, H.K., Schinke, C., Bhattacharyya, S., Will, B., Verma, A., and Steidl, U. (2014). Stem cell origin of myelodysplastic syndromes. Oncogene *33*, 5139–5150.

Flores-Figueroa, E., Gutiérrez-Espindola, G., Guerrero-Rivera, S., Pizzuto-Chavez, J., and Mayani, H. (1999). Hematopoietic progenitor cells from patients with myelodysplastic syndromes: in vitro colony growth and long-term proliferation. Leuk. Res. *23*, 385–394.

Genovese, G., Kähler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M., et al. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N. Engl. J. Med. *371*, 2477–2487.

Gilliland, D.G., and Griffin, J.D. (2002). The roles of FLT3 in hematopoiesis and leukemia. Blood *100*, 1532–1542.

Gilliland, D.G., and Tallman, M.S. (2002). Focus on acute leukemias. Cancer Cell *1*, 417–420.

Hahn, C.N., Chong, C.E., Carmichael, C.L., Wilkins, E.J., Brautigan, P.J., Li, X.C., Babic, M., Lin, M., Carmagnac, A., Lee, Y.K., et al. (2011). Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. Nat. Genet. *43*, 1012–1017.

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A., et al. (2014).

Age-related clonal hematopoiesis associated with adverse outcomes. N. Engl. J. Med. *371*, 2488–2498.

Jan, M., Snyder, T.M., Corces-Zimmerman, M.R., Vyas, P., Weissman, I.L., Quake, S.R., and Majeti, R. (2012). Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. Sci. Transl. Med. *4*, 149ra118.

Kerbauy, D.B., and Deeg, H.J. (2007). Apoptosis and antiapoptotic mechanisms in the progression of myelodysplastic syndrome. Exp. Hematol. *35*, 1739–1746.

Kim, J., and Zaret, K.S. (2015). Reprogramming of human cancer cells to pluripotency for models of cancer progression. EMBO J. *34*, 739–747.

Kim, J., Hoffman, J.P., Alpaugh, R.K., Rhim, A.D., Reichert, M., Stanger, B.Z., Furth, E.E., Sepulveda, A.R., Yuan, C.X., Won, K.J., et al. (2013). An iPSC line from human pancreatic ductal adenocarcinoma undergoes early to invasive stages of pancreatic cancer progression. Cell Rep. *3*, 2088–2099.

Kotini, A.G., Chang, C.J., Boussaad, I., Delrow, J.J., Dolezal, E.K., Nagulapally, A.B., Perna, F., Fishbein, G.A., Klimek, V.M., Hawkins, R.D., et al. (2015). Functional analysis of a chromosomal deletion associated with myelodysplastic syndromes using isogenic human induced pluripotent stem cells. Nat. Biotechnol. *33*, 646–655.

Link, D.C., and Walter, M.J. (2016). 'CHIP'ping away at clonal hematopoiesis. Leukemia *30*, 1633–1635.

Martincorena, I., and Campbell, P.J. (2015). Somatic mutation in cancer and normal cells. Science *349*, 1483–1489.

Ng, E.S., Davis, R., Stanley, E.G., and Elefanty, A.G. (2008). A protocol describing the use of a recombinant protein-based, animal product-free medium (APEL) for human embryonic stem cell differentiation as spin embryoid bodies. Nat. Protoc. *3*, 768–776.

Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O.I., Wilson, G., Kaufmann, K.B., McLeod, J., Laurenti, E., Dunant, C.F., et al. (2016). Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science *351*, aab2116.

Pandolfi, A., Barreyro, L., and Steidl, U. (2013). Concise review: preleukemic stem cells: molecular biology and clinical implications of the precursors to leukemia stem cells. Stem Cells Transl. Med. *2*, 143–150.

Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C.J., Ellis, P., Wedge, D.C., Pellagatti, A., et al.; Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood *122*, 3616–3627, quiz 3699.

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V.I., Paschka, P., Roberts, N.D., Potter, N.E., Heuser, M., Thol, F., Bolli, N., et al. (2016). Genomic classification and prognosis in acute myeloid leukemia. N. Engl. J. Med. *374*, 2209–2221.

Papapetrou, E.P. (2016). Patient-derived induced pluripotent stem cells in cancer research and precision oncology. Nat. Med. *22*, 1392–1401.

Papapetrou, E.P., and Sadelain, M. (2011). Generation of transgene-free human induced pluripotent stem cells with an excisable single polycistronic vector. Nat. Protoc. *6*, 1251–1273.

Papapetrou, E.P., Tomishima, M.J., Chambers, S.M., Mica, Y., Reed, E., Menon, J., Tabar, V., Mo, Q., Studer, L., and Sadelain, M. (2009). Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. Proc. Natl. Acad. Sci. USA *106*, 12759–12764.

Papapetrou, E.P., Lee, G., Malani, N., Setty, M., Riviere, I., Tirunagari, L.M., Kadota, K., Roth, S.L., Giardina, P., Viale, A., et al. (2011). Genomic safe harbors permit high β-globin transgene expression in thalassemia induced pluripotent stem cells. Nat. Biotechnol. *29*, 73–78.

Pellagatti, A., Cazzola, M., Giagounidis, A., Perry, J., Malcovati, L., Della Porta, M.G., Jädersten, M., Killick, S., Verma, A., Norbury, C.J., et al. (2010). Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. Leukemia *24*, 756–764.

Sanjuan-Pla, A., Macaulay, I.C., Jensen, C.T., Woll, P.S., Luis, T.C., Mead, A., Moore, S., Carella, C., Matsuoka, S., Bouriez Jones, T., et al. (2013). Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. Nature *502*, 232–236.

Sato, T., Kim, S., Selleri, C., Young, N.S., and Maciejewski, J.P. (1998). Measurement of secondary colony formation after 5 weeks in long-term cultures in patients with myelodysplastic syndrome. Leukemia *12*, 1187–1194.

Shlush, L.I., Zandi, S., Mitchell, A., Chen, W.C., Brandwein, J.M., Gupta, V., Kennedy, J.A., Schimmer, A.D., Schuh, A.C., Yee, K.W., et al.; HALT Pan-Leukemia Gene Panel Consortium (2014). Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature *506*, 328–333.

Sperling, A.S., Gibson, C.J., and Ebert, B.L. (2017). The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. Nat. Rev. Cancer *17*, 5–19.

Steensma, D.P., Bejar, R., Jaiswal, S., Lindsley, R.C., Sekeres, M.A., Hasserjian, R.P., and Ebert, B.L. (2015). Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood *126*, 9–16.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Themeli, M., Kloss, C.C., Ciriello, G., Fedorov, V.D., Perna, F., Gonen, M., and Sadelain, M. (2013). Generation of tumor-targeted human T lymphocytes from induced pluripotent stem cells for cancer therapy. Nat. Biotechnol. *31*, 928–933.

Valk, P.J., Verhaak, R.G., Beijen, M.A., Erpelinck, C.A., Barjesteh van Waalwijk van Doorn-Khosrovani, S., Boer, J.M., Beverloo, H.B., Moorhouse, M.J., van der Spek, P.J., Löwenberg, B., and Delwel, R. (2004). Prognostically useful gene-expression profiles in acute myeloid leukemia. N. Engl. J. Med. *350*, 1617–1628.

Vo, L.T., and Daley, G.Q. (2015). De novo generation of HSCs from somatic and pluripotent stem cell sources. Blood *125*, 2641–2648.

Walter, M.J., Shen, D., Ding, L., Shao, J., Koboldt, D.C., Chen, K., Larson, D.E., McLellan, M.D., Dooling, D., Abbott, R., et al. (2012). Clonal architecture of secondary acute myeloid leukemia. N. Engl. J. Med. *366*, 1090–1098.

Walter, M.J., Shen, D., Shao, J., Ding, L., White, B.S., Kandoth, C., Miller, C.A., Niu, B., McLellan, M.D., Dees, N.D., et al. (2013). Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. Leukemia *27*, 1275–1282.

Will, B., Zhou, L., Vogler, T.O., Ben-Neriah, S., Schinke, C., Tamari, R., Yu, Y., Bhagat, T.D., Bhattacharyya, S., Barreyro, L., et al. (2012). Stem and progenitor cells in myelodysplastic syndromes show aberrant stage-specific expansion and harbor genetic and epigenetic alterations. Blood *120*, 2076–2086.

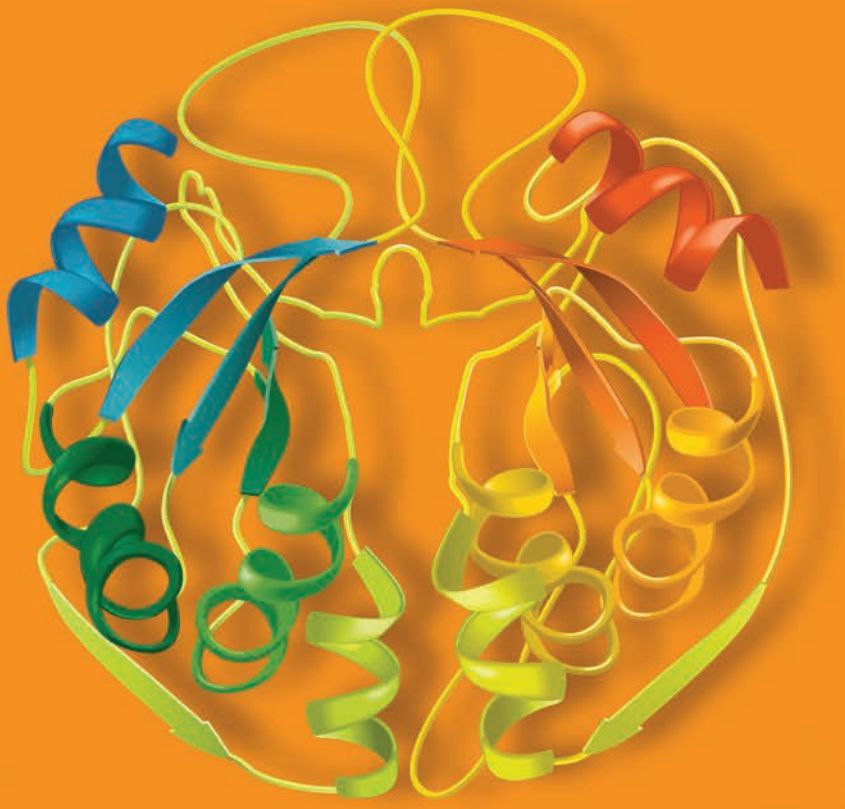
Wlodarski, M.W., Hirabayashi, S., Pastor, V., Starý, J., Hasle, H., Masetti, R., Dworzak, M., Schmugge, M., van den Heuvel-Eibrink, M., Ussowicz, M., et al.; EWOG-MDS (2016). Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents. Blood *127*, 1387–1397, quiz 1518.

Woll, P.S., Kjällquist, U., Chowdhury, O., Doolittle, H., Wedge, D.C., Thongjuea, S., Erlandsson, R., Ngara, M., Anderson, K., Deng, Q., et al. (2014). Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo. Cancer Cell *25*, 794–808.

Woolthuis, C.M., and Park, C.Y. (2016). Hematopoietic stem/progenitor cell commitment to the megakaryocyte lineage. Blood *127*, 1242–1248.

Xie, M., Lu, C., Wang, J., McLellan, M.D., Johnson, K.J., Wendl, M.C., McMichael, J.F., Schmidt, H.K., Yellapantula, V., Miller, C.A., et al. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat. Med. *20*, 1472–1478.

Over 50% of the Human Proteome Made From Human Cells



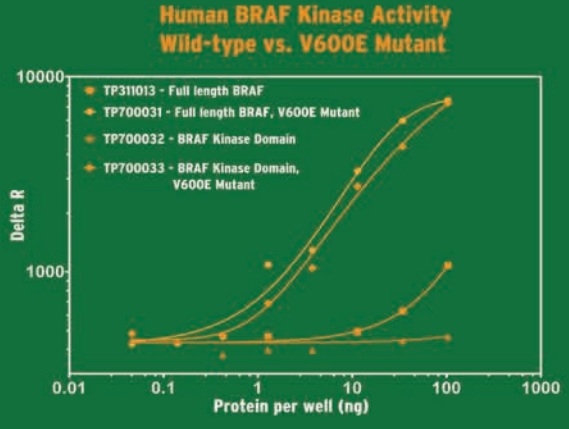
10,000 Proteins Mammalian Cell Produced

Protein functional studies

Compound screening

Immunogens for antibody production

Assay standards: ELISA, Mass Spec, and more



origene.com/protein

ORIGENE

Knock Out Any Gene!

CRISPR/CAS 9 Genome Editing Kits

OriGene's system for genome disruption and gene replacement delivers pre-designed plasmids and all the vectors needed to knock out any human or mouse gene. Knockout is as simple as 1-2-3:

1. Search the gene symbol on origene.com and order
2. Follow the simple protocol for transfection and Puro selection
3. Validate the knockout

Scan to view CRISPR/Cas-9 Video

Kit Components Include:

- 2 guide RNA vectors to ensure efficient cleavage
- Donor vector with predesigned homologous arms
- Knockin GFP-Puro for selection
- Scramble gRNA as negative control included

Come to OriGene, the trusted molecular biology expert, for your CRISPR needs.

ORIGENE

www.origene.com/CRISPR-CAS9