ADVANCED GENOME ENGINEERING OF INDUCED PLURIPOTENT STEM CELLS FOR THERAPEUTIC PURPOSES

A DISSERTATION
SUBMITTED TO THE PROGRAM IN GENETICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jonathan M. Geisinger
June 2015
I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Michele Calos, Primary Adviser

Stuart Kim

Matthew Porteus

Tim Stearns

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.
Abstract

The ability to precisely modify the genome with high efficiency in a directed manner is dependent on generating and exploiting a double-strand break in DNA. The tools available to generate double-strand breaks include the recombinases and resolvases, the phage integrases, the homing endonucleases, zinc finger nucleases, transcription activator-like effector nucleases, and the CRISPR/Cas9 system, the most recently developed tool. These tools allow for in-depth studies of the function of genes and genome structure, the generation of organisms bearing economically important bio-synthetic pathways, and the development of novel gene and cell therapies for addressing previously untreatable diseases. Taking full advantage of each tool and knowing when to use each one requires a thorough understanding of how the tool functions. For most of the genome engineering toolkit, this understanding has been achieved. The CRISPR/Cas9 system, however, has been assumed to function in the same manner as the zinc finger and transcription activator-like effector nucleases despite generating blunt-end double-strand breaks rather than staggered breaks. This difference is important to understanding how Cas9-mediated double-strand breaks are repaired, and, thus, how best to exploit these breaks for genome engineering. This thesis demonstrates the use of different genome engineering tools for uses appropriate to how they function. Chapters 2 through 4 demonstrate the exploitation of blunt-ended double-strand breaks generated by the CRISPR/Cas9 system to remove endogenous sequences and insert exogenous sequences with a high level of precision into the genome in a human immortalized cell line and in induced pluripotent stem cells. Chapter 5 presents preliminary results using phage integrases and Cre recombinase to generate an induced pluripotent stem cell-derived therapy for skeletal muscle. Together, these chapters illustrate the importance understanding how each genome engineering tool functions in developing advantageous methods for gene and cell therapies.
Acknowledgments

I would first like to thank my advisor, Michele Calos, for allowing me to pursue the studies detailed in this thesis. Being able to pursue these studies greatly facilitated my personal growth as a scientist and for that I am thankful.

I would also like to thank the members of the Calos lab, in particular, Chris Bjornson, Sören Turan, Alfonso Farruggio, and Tawny Neal. Chris has been an invaluable source of information on skeletal muscle and satellite cells. Sören has provided insightful comments that always improved the quality of the science they were directed at, and Alfonso provided invaluable lessons on cloning vectors. Tawny was the best lab manager we had during my Ph.D.

I also thank my adopted lab, that of Stuart Kim’s. By attending their lab meetings, I learned a great deal about statistical analysis, genomics, aging, and C. elegans. Through Stuart’s lab, I was able to attend the Bay Area Aging Meeting throughout my Ph.D., which further expanded my scientific horizons. For these experiences and the lessons learned from them, I am in debt to Stuart, Biff Mann, Stephanie Zimmerman, Zach O’Brown, and Kristen Fortney, as well as the previous members Hinco Gierman, Dror Sagi, and Ari Friedland.

I thank the rest of my committee as well: Tim Stearns, Anne Brunet, and Matt Porteus. They have all been invaluable sources of feedback and helpful career advice.

My class and the rest of people I have met at Stanford have all been helpful in their own ways, whether just having fun or studying early on. A good support network is very helpful for Ph.D. studies. I particularly thank Nandita Garud, Mei-hsin Cheng, Shuo Han, Lisl Esherick, Stephanie, Shaila Musharoff, and Justin Smith.

I am forever indebted to Erin Mitsunaga for being my best friend. She is a wonderful person and her friendship has been and is incredibly valuable to me.
# Table of Contents

**Abstract**

Acknowledgments

## 1  Introduction

## 2  Investigation of the Repair Pattern of CRISPR/Cas9-Mediated Double-Strand Breaks

2.1 Introduction  
2.2 Results  
2.3 Discussion  
2.4 Methods  
   2.4.1 Choice of sgRNAs and Vector Construction  
   2.4.2 Cell Culture  
   2.4.3 Detection of CRISPR/Cas9-Facilitated Excision  
   2.4.4 Statistics

## 3  Knock-in Blunt Cloning: Exploiting CRISPR/Cas9-Mediated Non-Homologous End-Joining

3.1 Introduction  
3.2 Results  
   3.2.1 Knock-in Blunt Ligation: CRISPR/Cas9-Mediated in vivo Blunt-End Cloning  
   3.2.2 Stabilized Cas9-DD Combined with Nuclease Protection Facilitates a Higher Degree of Precise Repair at the Genome-Cassette Junction  
3.3 Discussion  
3.4 Methods  
   3.4.1 Choice of sgRNAs and Vector Construction  
   3.4.2 Knock-In Cassette Construction
### 3.4.3 Detection of Knock-in Blunt Ligation

### 3.4.4 Flow Cytometry

### 3.4.5 Western Blotting

### 3.4.6 Statistics

#### 4 Knock-in Blunt Cloning in Human Induced Pluripotent Stem Cells

##### 4.1 Introduction

##### 4.2 Results

- **4.2.1** KiBL results in efficient, precise knock-ins in hiPSCs
- **4.2.2** Analysis of Off-Target KiBL Events
- **4.2.3** HYE-STCC: High-Yield Efficient Single Transmigrant Cell Cloning

##### 4.3 Discussion

##### 4.4 Methods

- **4.4.1** Knock-In Cassette Construction
- **4.4.2** Cell Culture
- **4.4.3** Detection of Knock-in Blunt Ligation and Homologous Recombination
- **4.4.4** Analysis of Off-Target KiBL Events
- **4.4.5** HYE-STCC
- **4.4.6** Statistics

#### 5 Engineering and Identification of a Multi-Purpose iPSC-Derived Myogenic Progenitor Cell

##### 5.1 Introduction

##### 5.2 Results

- **5.2.1** Therapeutic Engineering of Murine iPSCs
- **5.2.2** Identification of a Suitable Myogenic Progenitor Cell
- **5.2.3** Engraftment of iPSC-Derived Myogenic Progenitor Cells

##### 5.3 Discussion

##### 5.4 Methods

- **5.4.1** Cell Culture
- **5.4.2** Retargeting via Bxb1 and Cre Excision
5.4.3 Myogenic Differentiation and FACS 76
5.4.4 Genomic Copy Number QPCR and qrt-PCR 76
5.4.5 Engraftment of iPSC-Derived Myogenic Progenitor Cells 77
5.4.6 Immunofluorochemistry 78
5.4.7 Statistics 79

Appendix 80

Bibliography 87
List of Tables

A.1 Interrogated regions used for design of guide RNAs. 80
A.2 Guide RNAs and cloning oligos 81
A.3 Excision detection primers 83
A.4 Primers used in cassette amplification 84
A.5 Primers for Amplifying Cassette Junctions 85
A.6 Off-target sites and primers for H11-r1-2 and H11-r2-3 86
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Deletion repair generated by paired sgRNAs</td>
<td>6</td>
</tr>
<tr>
<td>2.2</td>
<td>Outcomes of precise repair by PAM orientation</td>
<td>7</td>
</tr>
<tr>
<td>2.3</td>
<td>Patterns of Repair at the MYOD1 3’UTR</td>
<td>8</td>
</tr>
<tr>
<td>2.4</td>
<td>Sequences of CRISPR/Cas9-mediated deletions at the MYOD1 3’UTR</td>
<td>9</td>
</tr>
<tr>
<td>2.5</td>
<td>Patterns of Repair at the PAX3 3’UTR</td>
<td>10</td>
</tr>
<tr>
<td>2.6</td>
<td>Sequences of CRISPR/Cas9-mediated deletions at the PAX3 3’UTR</td>
<td>11</td>
</tr>
<tr>
<td>2.7</td>
<td>Patterns of Repair at the PAX7 3’UTR</td>
<td>12</td>
</tr>
<tr>
<td>2.8</td>
<td>Sequences of CRISPR/Cas9-mediated deletions at the PAX7 3’UTR in HEK293</td>
<td>13</td>
</tr>
<tr>
<td>2.9</td>
<td>Sequences of CRISPR/Cas9-mediated deletions at the PAX7 3’UTR in H9</td>
<td>13</td>
</tr>
<tr>
<td>2.10</td>
<td>Patterns of Repair at the mUtrn 3’UTR</td>
<td>14</td>
</tr>
<tr>
<td>2.11</td>
<td>Sequences of CRISPR/Cas9-mediated deletions at the mUtrn 3’UTR</td>
<td>15</td>
</tr>
<tr>
<td>3.1</td>
<td>Knock-in blunt ligation</td>
<td>22</td>
</tr>
<tr>
<td>3.2</td>
<td>Schematic of the PGK-PTKmChR and DICE-EPv2 cassettes.</td>
<td>22</td>
</tr>
<tr>
<td>3.3</td>
<td>Initial scheme targeting KiBL to the PAX7 3’UTR.</td>
<td>23</td>
</tr>
<tr>
<td>3.4</td>
<td>Representative FACS plot of DICE-EPv2 KiBL</td>
<td>24</td>
</tr>
<tr>
<td>3.5</td>
<td>Quantification of initial KiBL experiments via flow cytometry</td>
<td>24</td>
</tr>
<tr>
<td>3.6</td>
<td>KiBL can generate perfect genome-cassette junctions</td>
<td>25</td>
</tr>
<tr>
<td>3.7</td>
<td>Analysis of PAX7-PTKmChR/DICE-EPv2.0 junctions</td>
<td>26</td>
</tr>
<tr>
<td>3.8</td>
<td>Schematic diagrams of the pKER cassette series</td>
<td>26</td>
</tr>
<tr>
<td>3.9</td>
<td>Targeting the H11 locus with CRISPR/Cas9</td>
<td>27</td>
</tr>
<tr>
<td>3.10</td>
<td>Structure and conservation of the H11 safe harbor locus</td>
<td>28</td>
</tr>
<tr>
<td>3.11</td>
<td>Suitability of the pKER series for KiBL</td>
<td>28</td>
</tr>
<tr>
<td>3.12</td>
<td>Schematics of destabilized Cas9</td>
<td>29</td>
</tr>
<tr>
<td>3.13</td>
<td>Stability analysis of Cas9-DD and DD-Cas9 in HEK293 cells</td>
<td>30</td>
</tr>
<tr>
<td>3.14</td>
<td>KiBL at the PAX7 3’UTR with pKER-Clover and Cas9-DD</td>
<td>31</td>
</tr>
<tr>
<td>3.15</td>
<td>Molecular analysis of KiBL with Cas9-DD and nuclease protection</td>
<td>32</td>
</tr>
<tr>
<td>3.16</td>
<td>Comparison of WT-Cas9 with destabilized Cas9-DD and nuclease protection</td>
<td>33</td>
</tr>
<tr>
<td>4.1</td>
<td>Diagram of the targeted H11 locus in JF10 hiPSCs</td>
<td>43</td>
</tr>
<tr>
<td>4.2</td>
<td>Schematic for examining the efficiency KiBL in hiPSCs</td>
<td>43</td>
</tr>
</tbody>
</table>
4.3 Comparison of WT-Cas9- and Cas9-DD-mediated KiBL at the H11 locus via fluorescence microscopy
4.4 Examining KiBL efficiency with Cas9-DD in hiPSCs
4.5 Sequence analysis of KiBL in hiPSCs with Cas9-DD
4.6 Sequence analysis of KiBL in hiPSCs with DD-Cas9
4.7 Destabilized DD-Cas9 does not facilitate KiBL
4.8 Single-cell PCR analysis of KiBL
4.9 KiBL clonality analysis through multicolor FACS
4.10 Clonality analysis through barcoded cassettes
4.11 Targeted analysis of off-target KiBL events in hiPSCs
4.12 Schematic diagram of HYE-STCC
4.13 Survivability analysis of Clover+ cells subjected to HYE-STCC
5.1 Sex Determination of MEF-derived iPSC clones
5.2 Schematics of therapeutic vectors for Bxb1-based retargeting
5.3 Confirmation of reprogramming cassette excision in PCX cells
5.4 Excised clones retain markers of pluripotency
5.5 FACS plots of fully stained cells from a 20-day myogenic differentiation
5.6 CD106+ cells are enriched for myogenic markers early in differentiation
5.7 CD106+CD140a- cells are more stem-like than other iPSC-derived CD106+ subpopulations
5.8 CD106+ CD140a- CD140b+ cells are potentially more myoblastic than progenitor-like
5.9 CD106+ CD34-/low CD140a-b- cells are enriched for Pax7 expression
5.10 iPSC-derived myogenic progenitors contribute to muscle repair
5.11 mCherry+ cells are relatively rare in engrafted muscle
5.12 Treatment with PCX2-derived myogenic progenitors results in greater fiber CSA than injury alone
Chapter 1

Introduction

Jonathan M. Geisinger

One of the major challenges in the field of genome engineering is the application of genome engineering approaches to developing viable gene and cell therapies. Addressing this challenge involves not only understanding the genome engineering toolkit, but also understanding the disease in question. Understanding the disease may require additional knowledge of cell biology, stem cell biology, immunology, systems biology, or even population genetics. Without this complementary knowledge, it is extremely hard, if not outright impossible, not only to design a viable therapy, but to be able to interpret negative results, should they arise. Taking this multidisciplinary approach is even more important now that the CRISPR/Cas9 system has become widely available.

Until fairly recently, genome engineering could be considered a specialist field, as constructing functional tools for manipulating the genome through the generation of double-strand breaks (DSBs) had a fair amount of difficulty associated with it. However, with the rise of the CRISPR/Cas9 system (Mali, Yang et al., 2013; Cong et al., 2013; Jinek et al., 2013), construction of genome engineering tools is now technically easy. This ease of construction has led to application of the CRISPR/Cas9 system in numerous organisms at a greatly accelerated pace compared to that of zinc finger and transcription activator-like effector nucleases (ZFNs and TALENs). In the rush to apply this new, deceptively simple technology, the development of a considered, nuanced understanding of how best to utilize the CRISPR/Cas9 system has been neglected. The aspect of the system that has been focused on is that Cas9 generates DSBs and that these can be exploited in the same way that DSBs generated by ZFNs and TALENs have been. However, ZFNs and TALENs share the same nuclease domain, FokI, which generates DSBs with overhangs, whereas Cas9 generates blunt-
ended DSBs (Jinek et al., 2012). This underappreciated difference explains much of the difference between these tools and should have led to the rapid development of methods to best exploit how the eukaryotic cell would repair this blunt DSB. Instead, the methods and knowledge gained from the ZFNs and TALENs were applied, especially that a DSB would be repaired through the non-homologous end-joining pathway in a predominantly error-prone manner. For the type of DSBs generated by Cas9, this error-prone repair should generally be bypassed in favor of direct-end ligation, i.e., precise repair (see Chapman, Taylor, and Boulton, 2012). Additionally, the vast majority of work to improve the CRISPR/Cas9 system has focused on suppression of non-homologous end-joining to promote homologous recombination (Lin, Staahl et al., 2014; Yu, Liu et al., 2015; Maruyama et al., 2015; Chu et al., 2015). These approaches may have unforeseen consequences on the genome and do not best exploit the CRISPR/Cas9 system for genome engineering.

In considering how to apply genome engineering to the development of disease therapeutics, the importance of understanding how each component of the genome engineering toolkit functions and the etiology of disease when developing a therapy is exemplified by Duchenne muscular dystrophy. Duchenne muscular dystrophy is a progressive, degenerative disease with a prevalence of 1 in 3,500 boys, and is caused by mutations in the dystrophin gene (Hoffman, Brown, and Kunkel, 1987). Most efforts to address this disease have centered on viral transduction of the protein or cell therapy using wild-type myoblasts (see Miller et al., 1997 for an example of cell therapy). These approaches have not been particularly successful, and thus genome engineering strategies have begun to be developed. In examining the spectrum of mutations in dystrophin, a hotspot of deletions spanning exons 45 through 55 was observed (White and den Dunnen, 2006). Targeting this region would result in a treatment for approximately 60% of all Duchenne muscular dystrophy cases. Recently, a CRISPR/Cas9 strategy was developed targeting this hotspot (Ousterout et al., 2015). This strategy targets two sites, one near exon 45 and one near exon 55, resulting in the excision of the intervening sequence and the restoration of an open reading frame. Such a strategy, however, does not cure Duchenne; rather, the patients would be left with a milder form, as dystrophin would still be truncated. Additionally, this strategy relied on injecting a pool of cells enriched for corrected cells for determining therapeutic viability. While this strategy led to human dystrophin-positive fibers in mouse muscle, the number of human-mouse chimeric human dystrophin-negative fibers significantly outnumbered the positive fraction. Thus, it remains
unknown if this strategy would actually be efficacious if modified. One can certainly imagine that transfecting the fibers directly with CRISPR/Cas9 may be more efficacious, but such a strategy would almost certainly fail to target the muscle stem cells at any significant level. This study did consider how their genome engineering tool of choice functions, but did not consider how to make this a viable therapy. It remains, for now, simply a proof-of-principle that excising the deletion hotspot region of dystrophin can result in the generation of truncated dystrophin protein.

In this thesis, I will demonstrate the importance of understanding how two components of the genome engineering toolkit, the CRISPR/Cas9 system and the phage integrases, function in order to design genome engineering strategies best suited to their characteristics. Chapters 2 through 4 focus on CRISPR/Cas9 and Chapter 5 focuses on the use of phage integrases. In Chapter 2, I demonstrate that the repair pattern of CRISPR/Cas9-mediated DSBs is consistent with that of precise repair through the non-homologous end-joining pathway. In Chapter 3, I present a method for exploiting this high level of repair precision to precisely place exogenous sequences into the genome of a human immortalized cell line in a homology-free manner. Chapter 4 details the application of this knock-in blunt ligation method to human induced pluripotent stem cells and presents preliminary results for a novel method of single-cell cloning. Lastly, Chapter 5 describes the application of a triple-recombinase-based genome engineering strategy to developing a multi-purpose cell therapy for skeletal muscle. This chapter underscores the challenges faced in developing a viable therapy.

Additionally, significant portions of Chapters 2 through 4 appear verbatim from a manuscript submitted on May 22, 2015.
Chapter 2

Investigating the Repair Pattern of CRISPR/Cas9-Mediated Double-Strand Breaks

*Jonathan M. Geisinger, Sophia Hernandez, Laura P. Spector, Michele P. Calos*

**2.1 Introduction**

The clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 system is one of several components of the RNA-mediated adaptive immune system in bacteria (Barrangou et al., 2007). CRISPRs are derived from fragments of phage and plasmid DNA that have been integrated into the host bacterium’s genome (Barrangou et al., 2007). The arrays formed by such fragments are transcribed and processed into short CRISPR RNAs (crRNAs) that are then used to direct the Cas9 enzyme to specific targets (hence their description as guide RNAs) by first base-pairing to trans-activating crRNA and can function *in vitro* (Jinek et al., 2012). Cas9 itself is an RNA-guided DNA endonuclease that recognizes potential DNA targets first on the basis of a protospacer adjacent motif (PAM) downstream of the crRNA target sequence (Sternberg, Redding *et al*., 2014). DNA-RNA base pairing then facilitates the binding of Cas9 to the target DNA, followed by Cas9 making a blunt, double-stranded break (DSB) three or four bases upstream of the PAM through the use of two nuclease domains, a RuvC-like domain and an HNH domain (Jinek *et al*., 2012).

When this system was first adapted from *Streptococcus pyogenes* for genome editing in eukaryotic cells, the initial applications focused on making targeted, random mutations through error-prone NHEJ or modifying the genome through homology-directed repair or homologous recombination (Mali, Yang *et al*., 2013; Cong *et al*., 2013; Jinek *et al*., 2013).
Interestingly, the use of two guide RNAs at once was observed to be capable of generating precise deletions with no loss of additional nucleotides (Mali, Yang et al., 2013; Cong et al., 2013). Despite this observation, the main repair mechanism of CRISPR/Cas9-induced DSBs has been proposed to be error-prone NHEJ. This conclusion appears to be somewhat at odds with what is known about the repair of blunt-end DSBs. First, use of the Tn5 transposon system in eukaryotic cells has demonstrated that blunt, chemically unmodified (i.e., 5’-phosphorylated, 3’hydroxy) DSBs can be repaired precisely through the canonical NHEJ pathway, which includes the proteins Ku70, Ku80, DNA ligase IV (LIG4), DNA-PKcs, and XRCC4 (van Heemst et al., 2004; for extensive review of the NHEJ pathway see Chapman, Taylor, and Boulton, 2012; also see Bétermier, Bertrand, and Lopez, 2014). Resolving the crystal structure of Cas9 revealed that the DSBs generated by the cleavage of Cas9 should be chemically unmodified (Nishimasu et al., 2014). Second, experiments in Saccharomyces cerevisiae utilizing the blunt-cutting restriction enzymes PvuII and EcoRV demonstrated very poor survivability upon induction of expression of these endonucleases (Westmoreland et al., 2010). The authors concluded that such blunt DSBs were poor substrates for repair, but the alternative explanation is that these DSBs were repaired precisely and repeatedly cleaved, effectively preventing progression through the cell cycle as DSBs had previously been demonstrated to do in S. cerevisiae (Demeter et al., 2000). Third, rather large differences in efficiency as measured by the level of induced error-prone NHEJ have been observed between cell types when using the same guide RNA, with use in human iPSCs generating among the lowest efficiencies (Mali, Yang et al., 2013). Taken together, these observations suggest that CRISPR/Cas9-induced DSBs may not be repaired predominantly through the error-prone NHEJ pathway.

Thus, we hypothesized that the majority of CRISPR/Cas9-mediated DSBs should be resolved by canonical NHEJ with a high degree of precision. Here, we demonstrate that large and small precise deletions can be generated through the simultaneous use of two guide RNAs to generate two DSBs, and show that the repair pattern of these breaks is consistent with the predominance of canonical NHEJ.
2.2 Results

Because Cas9 generates blunt DSBs, we hypothesized that these blunt ends would be preferentially repaired by precise NHEJ. To test this hypothesis, we chose to generate two DSBs within the same locus rather than one because precise NHEJ mediated by one individual sgRNA would be indistinguishable from uncleaved DNA at the sequence level. However, if a cell received two sgRNA vectors expressing two different sgRNAs, there would be a high probability of excision of the intervening sequence between the paired sgRNA targets, allowing for religation of the genomic junctions and subsequent detection and analysis via PCR and Sanger sequencing (Figure 2.1).

![Figure 2.1. Deletion repair generated by paired sgRNAs.](image)

Others (Cong et al., 2013; Mali, Yang, et al., 2013; Canver, Bauer et al., 2014) have also observed this resulting excision and religation. This excision may occur because Cas9 appears to bind the PAM side of the DSB less strongly after cleavage (Sternberg, Redding et al., 2014). A high level of precision has also been observed in translocations generated by two sgRNAs (Ghezraoui et al., 2014). However, the repair consequences of paired PAM orientations resulting from using two sgRNAs simultaneously have not been explored. For pairs of sgRNAs, there are four possible orientations: 1) PAMs Out, 2) PAMs In, 3) PAMs 5’, and 4) PAMs 3’ (Figure 2.2).

It is important to note that PAMs 5’ and PAMs 3’ are functionally equivalent at first glance. Because Cas9 can cleave at either the third or fourth base upstream of the PAM, precise rejoining can occur with 0, +1, or -1 bases on each of the repaired genomic ends, depending on which PAM orientation is used. As such, each orientation has four possible outcomes for precise repair (detailed in Figure 2.2 and shown as 5’sgRNA bases/ 3’ sgRNA bases). To this end, we generated multiple sgRNAs against the 3’UTRs of human MYOD1, PAX3, PAX7,
and mouse Utrophin (Utrn) to examine the repair patterns and degree of precision utilized by paired PAM orientations.

Figure 2.2. Outcomes of precise repair by PAM orientation. Each paired PAM orientation has four possible outcomes for precise repair (depicted below each orientation) based on SpCas9’s ability to cleave at the third or the fourth base upstream of the PAM.

<table>
<thead>
<tr>
<th>PAMs Out</th>
<th>PAMs In</th>
<th>PAMs 5’</th>
<th>PAMs 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precise Repair Outcomes</td>
<td>Precise Repair Outcomes</td>
<td>Precise Repair Outcomes</td>
<td>Precise Repair Outcomes</td>
</tr>
<tr>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>+1/0</td>
<td>-1/0</td>
<td>+1/0</td>
<td>-1/0</td>
</tr>
<tr>
<td>0/+1</td>
<td>0/-1</td>
<td>0/-1</td>
<td>0/+1</td>
</tr>
<tr>
<td>+1/+1</td>
<td>-1/-1</td>
<td>+1/-1</td>
<td>-1/+1</td>
</tr>
</tbody>
</table>

To date, the largest study of repair patterns of paired PAM-mediated excisions focused only on the PAMs 3’ orientation (Canver, Bauer et al., 2014). In our study, we first sought to examine if there were any large differences in repair patterns between all four orientations. Each pair was transfected separately into HEK293 cells. In examining all four paired PAM orientations at the MYOD1 3’UTR in HEK293 cells, we observed a remarkably high level of precision of NHEJ repair as indicated by the high frequency (64.71-96.15% of observed amplicons) of precise repair, regardless of orientation (Figure 2.3A-B; for sequences, see Figure 2.4). PAMs 5’ showed the highest frequency of precise repair, reaching 96.15%, whereas PAMs 3’ showed the lowest frequency of precise repair (Figure 2.3B). Imprecise repair was mainly limited to small deletions and occasional inclusion of bases that should have been excised. PAMs Out and PAMs 3’ displayed a deviation from 0/0 precise repair, with both possessing a mostly 0/+1 precise repair pattern, which may be the result of them sharing the same 3’ sgRNA (Figure 2.3C). Interestingly, the predominance of 0/+1 precise repair observed with PAMs 3’ appears to be due to 0/0 repair generating a perfect target site for the 5’ sgRNA in this pair. Comparing the deviations from 0/0 precise repair between PAM orientations revealed that there was a significant difference between distributions ($P < 0.0001$; Kruskal-Wallis test). Post hoc multiple comparisons revealed no significant difference between PAMs Out and PAMs 3’ or between PAMs In and PAMs 5’ ($P > 0.9999$ for both comparisons), but that all other comparisons between orientations were significantly different ($P < 0.0001$; Dunn’s multiple comparisons test for all comparisons). Additionally, we observed only one stereotypical error-prone event (Figure 2.4). These findings suggest that
there are no inherent differences in repair precision between paired PAM orientations, the differences are dependent on individual sgRNAs, and NHEJ repair of CRISPR/Cas9-induced DSBs is not inherently error-prone.

Figure 2.3. Patterns of repair at the MYOD1 3’UTR. (A) Schematic diagrams of the targeted locus with sgRNA targets depicted as sgRNA/Cas9 complexes. Additionally, the largest deletion size for each locus is depicted. Red indicates exonic sequence and purple indicates 3’UTR. (B) Graphs depicting the cumulative percentage of precise (green), excision+insertion (orange), and imprecise (red) repair observed in sequenced amplicons for each tested sgRNA pair. (C) Dot plots of deletion-containing amplicons described in sub-panel ii. Each dot represents the sum of bases beyond 0/0 precise repair. The numbers of amplicons analyzed for each sgRNA pair appear on the plot. Data from (C) were analyzed with the Kruskal-Wallis test followed by post-hoc Dunn’s multiple comparisons test. All sgRNA pairs were compared against all pairs. **** = $P < 0.0001$. 
Figure 2.4. Sequences of CRISPR/Cas9-mediated deletions at the MYOD1 3'UTR. Underlined blue sequence denotes sgRNA targets. Red denotes PAMs. ~ = Not sequenced.
Having observed that all four orientations were capable of facilitating precise repair at one locus, we next asked if different paired PAMs in the same orientation would display the same degree of repair precision at the same locus. Thus, we utilized four pairs of sgRNAs in the PAMs In orientation at the PAX3 3’UTR in HEK293 cells (Figure 2.5A). We observed that the vast majority of sequenced deletion-containing amplicons for all four pairs were

**Figure 2.5. Patterns of repair at the PAX3 3’UTR.** (A) Schematic diagrams of the targeted locus with sgRNA targets depicted as sgRNA/Cas9 complexes. Additionally, the largest deletion size for each locus is depicted. Red indicates exonic sequence and purple indicates 3’UTR. (B) Graphs depicting the cumulative percentage of precise (green), excision+insertion (orange), and imprecise (red) repair observed in sequenced amplicons for each tested sgRNA pair. (C) Dot plots of deletion-containing amplicons described in sub-panel ii. Each dot represents the sum of bases beyond 0/0 precise repair. The numbers of amplicons analyzed for each sgRNA pair appear on the plot. Data from (C) were analyzed with the Kruskal-Wallis test followed by post-hoc Dunn’s multiple comparisons test. All sgRNA pairs were compared against all pairs. **** = P < 0.0001; ** = P < 0.004.
remarkably precise, with the frequency of precise repair ranging from 87.5% to 100% (Figure 2.5B; Figure 2.6). Interestingly, we did not observe insertions in any of these sequenced amplicons. We observed that there was a significant difference between the distributions of deviations ($P < 0.0001$; Kruskal-Wallis test), but that only PAMs In #1 displayed a noticeable deviation from 0/0 precise repair. The majority of these sequenced amplicons were of the 0/-1 precise repair variety, which was found to be significantly different than that of the other three pairs ($P < 0.004$ for PAMs In #1 versus each other pair; post hoc Dunn’s multiple comparisons test for all comparisons; Figure 2.5C). These findings provide additional evidence for both the high degree of precision involved in CRISPR/Cas9-induced NHEJ repair, and that differences in repair are dependent on individual sgRNAs, possibly based on target sequence.

Figure 2.6. Sequences of CRISPR/Cas9-mediated deletions at the PAX3 3’UTR. Underlined blue sequence denotes sgRNA targets. Red denotes PAMs. Dark blue and pink denote overlapping sgRNA target sequence and PAMs respectively.

Having observed that CRISPR/Cas9-induced NHEJ repair does not appear to be inherently error-prone, we next sought to examine whether a given pair of sgRNAs generates similar repair patterns in different cell types. Thus, we designed pairs of sgRNAs against the PAX7 3’UTR, which were subsequently transfected into HEK293 cells and H9 human ESCs (hESCs) (Figure 2.7A). We observed a mostly high level of precise repair in the H9 hESCs, with only one pair resulting in less than 75% precise repair (Figure 2.7B; Figure 2.8 for HEK293; Figure 2.9 for H9 hESCs). However, we observed what appeared to be a more varied level of precise repair in HEK293 cells with the same PAM pairs, ranging from 44.44% to 90.48% (Figure 2.7B). Additionally, we only observed two typical error-prone events for HEK293, one resulting from the use of PAX7r1-4 and a double event for PAX7r1-4 and PAX7r1-1 (Figure 2.8). For H9 hESCs, we observed two typical error-prone repair events: one for PAX7r1-2 and one for PAX7r1-3 (Figure 2.9). In spite of this variation, deviation from 0/0 precise repair was not significant for any given pair between cell types. In fact,
Figure 2.7. Patterns of repair at the PAX7 3’UTR. (A) Schematic diagrams of the targeted locus with sgRNA targets depicted as sgRNA/Cas9 complexes. Additionally, the largest deletion size for each locus is depicted. Red indicates exonic sequence and purple indicates 3’UTR. (B) Graphs depicting the cumulative percentage of precise (green), excision+insertion (orange), and imprecise (red) repair observed in sequenced amplicons for each tested sgRNA pair. (C) Dot plots of deletion-containing amplicons described in sub-panel ii. Each dot represents the sum of bases beyond 0/0 precise repair. The numbers of amplicons analyzed for each sgRNA pair appear on the plot. Data from (C) were analyzed with the Kruskal-Wallis test followed by post-hoc Dunn’s multiple comparisons test. Each pair was only compared between cell types.

for each pair, the deviation from 0/0 precise repair was not significantly different between cell types ($P > 0.8$ for all four comparisons; Dunn’s multiple comparisons test; Figure 2.7C). These data illustrate that a given pair of sgRNAs generates similar patterns of repair with similar degrees of precision in different cell types of the same species.
Given that we observed a high degree of precise repair in human cells across three loci and 16 pairs of sgRNAs, we sought to examine if a high level of repair precision would also be observed in mouse cells. Paired sgRNAs have been previously utilized to generate deletions in murine embryos, although there was significant variability in deletion size between embryos (Zhou, Wang et al., 2014). To examine repair patterns in mouse cells, we transfected C2C12 murine immortalized myoblast cells with three pairs of sgRNAs in different orientations against the mUtrn 3’UTR (Figure 2.10A). We observed not only a more varied pattern of repair than with human cells (precise repair ranged from 41.67% to 100% of amplicons; Figure 2.10B; Figure 2.11), but also a higher degree of repetition in the imprecise excision repair amplicons in their deviation from 0/0 precise repair (Figure 2.10C).

Interestingly, we found that there was no significant difference between the distributions of deviance from 0/0 precise repair for the three orientations ($P = 0.1354$; Kruskal-Wallis test).
Figure 2.10. Patterns of repair at the mUtrn 3’UTR. (A) Schematic diagrams of the targeted locus with sgRNA targets depicted as sgRNA/Cas9 complexes. Additionally, the largest deletion size for each locus is depicted. Red indicates exonic sequence and purple indicates 3’UTR. (B) Graphs depicting the cumulative percentage of precise (green), excision+insertion (orange), and imprecise (red) repair observed in sequenced amplicons for each tested sgRNA pair. (C) Dot plots of deletion-containing amplicons described in sub-panel ii. Each dot represents the sum of bases beyond 0/0 precise repair. The numbers of amplicons analyzed for each sgRNA pair appear on the plot. Data from (C) were analyzed with the Kruskal-Wallis test followed by post-hoc Dunn’s multiple comparisons test. All sgRNA pairs were compared against all pairs.

These results indicated that CRISPR/Cas9 facilitates a high level of repair in murine cells as well, but may potentially have generated a more varied distribution of imprecise repair due to species-specific differences in NHEJ repair.
Figure 2.11. Sequences of CRISPR/Cas9-mediated deletions at the mUtrn 3’UTR. Underlined blue sequence denotes sgRNA targets. Red denotes PAMs.

2.3 Discussion

These results demonstrate that CRISPR/Cas9-mediated DSBs are repaired with a high degree of precision in human and murine cells (confirming reports by others), and that the individual guide RNAs have more influence over the pattern of repair than the orientation of the paired PAMs. As such, these results not only provide evidence for precise canonical NHEJ being the major repair pathway for CRISPR/Cas9-induced DSBs, they also suggest that the choice of PAM orientation when using paired guide RNAs does not greatly affect the outcome of repair, which had not been previously known.

From our results, the use of paired guide RNAs appears to facilitate excision of the intervening sequence followed by end-joining with a high degree of precision. Others (Cong et al., 2013; Mali, Yang, et al., 2013; Canver, Bauer et al., 2014) have also observed this resulting excision and religation. A high level of precision has also been observed in translocations generated using two guide RNAs in human immortalized cell lines (Ghezraoui et al., 2014). At least for the PAMs Out orientation, this phenomenon may occur because Cas9 appears to bind the PAM side of the DSB less strongly after cleavage (Sternberg, Redding et al., 2014). However, based on the results of Sternberg, Redding et al., (2014) we initially thought the PAMs Out orientation would display the highest level of precise repair due to Cas9 remaining bound to intervening sequence while simultaneously releasing the PAMs on each side. Such mechanism would allow for precise canonical NHEJ to take place with relative ease. As such, we predicted that the other orientations would display a more imprecise and stereotypical error-prone pattern of repair due to Cas9 remaining bound to at least one side of the deletion junction. Because all orientations unexpectedly displayed similar
levels of precise repair, the mechanism by which the CRISPR/Cas9 complex is removed from the DNA remains unilluminated. We speculate that it is possible that Cas9 is degraded by a nuclear protease, potentially in a ubiquitin-dependent, proteasome-independent manner as has been recently described for replication-coupled proteolysis-based repair of DNA-protein crosslinks (Duxin et al., 2014). It is possible that Cas9 is somehow displaced after binding, although this displacement would require relaxation of the strong binding Cas9 displays on the guide target DNA after cleaving, which appears to not occur based on in vitro experiments (Sternberg, Redding et al., 2014).

In our data we also observed that a given pair of guide RNAs will generate patterns of repair that are not significantly different from in each other in different cell types of the same species. This finding combined with previous observations that gene editing of hiPSCs with one guide RNA is less efficient as measured by error-prone NHEJ than that of HEK293 and K562 cells (Mali, Yang et al., 2013) suggests not only that NHEJ is not as error-prone as previously believed, but also that iPSCs and ESCs may be more poised to repair through the canonical precise NHEJ pathway rather than the error-prone pathway. This preference for the canonical, precise repair may also be possessed by primary stem and somatic cells. For example, hematopoietic and muscle stem cells from aged mice appear to retain the capacity for repairing DSBs through NHEJ in a precise manner (Flach et al., 2014; Beerman et al., 2014; Vahidi Ferdousi et al., 2014). Regarding somatic cells, CRISPR/Cas9-mediated in vivo gene editing of the murine liver through tail-vein injection resulted in a relatively low level of gene editing as assessed by the measurement of error-prone repair (Xue, Chen, Yin et al., 2014). This observation suggests that somatic cells in vivo may also prefer the canonical precise NHEJ pathway for DSB repair.

While our work demonstrates that CRISPR/Cas9-induced DSBs are repaired with a high degree of precision when using two guide RNAs regardless of paired PAM orientation, the mechanism behind the apparent activation of the error-prone NHEJ pathway when using one guide RNA remains unknown. One possibility is that this error-prone repair is dependent on low levels of LIG4. Inhibition of LIG4 through the small molecule SCR7 has been reported to lead to an increase in the efficiency of CRISPR/Cas9-facilitated homology-directed repair, but the effect on NHEJ was not directly examined (Maruyama et al., 2015). Another possibility is that the use of only one guide RNA leads to a cycle of cleavage and repair until the
error-prone pathway is activated. To address this possibility, future studies should examine the overexpression and repression of components of the error-prone NHEJ repair pathway, such as Mre11, Rad50, Nbs1, and CtlP, in combination with the CRISPR/Cas9 system.

2.4 Methods

2.4.1 Choice of sgRNAs and Vector Construction

All sgRNAs were chosen using the MIT CRISPR Design tool (http://crispr.mit.edu/). Briefly, genomic regions consisting of up to 250 bp were chosen for each locus, and the highest quality guides were chosen for cloning into the pX330 (Addgene #42230) backbone following the protocol developed by Feng Zhang’s lab. These regions appear in Table A.1. All oligos, primers, and gBlocks were ordered from Integrated DNA Technologies (IDT, Coralville, IA). A description of the sgRNAs chosen, as well as the oligos used for cloning, appear in Table A.2. sgRNA/Cas9 vectors, as well as all additional plasmids in this study, were isolated using the Nucleobond Midi Plus EF kit (Machrey-Nagel, Bethlehem, PA).

2.4.2 Cell Culture

HEK293 cells were maintained in 6-well tissue culture plates coated with Poly-L-Lysine (Sigma-Aldrich, St. Louis, MO) in high-glucose DMEM (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS; Gemini Bio-Products, West Sacramento, CA), 1x non-essential amino acids (Life Technologies), and 1x GlutaMAX (Life Technologies). Cells were passaged with 0.05% Trypsin-EDTA (Life Technologies) as needed at a 1:12 split. C2C12 cells were maintained in the same media as HEK293 cells and passaged in the same manner. C2C12 cells were grown on 6-well tissue culture plates. H9 hESCs were maintained in 0.1% gelatin-coated 6-well tissue culture plates on γ-irradiated mouse embryonic fibroblast feeder cells in human embryonic stem cell media consisting of DMEM/F12 (Thermo Fisher Scientific), 20% Knockout Serum Replacement (Life Technologies), 1x non-essential amino acids (Life Technologies), 1x GlutaMAX (Life Technologies), 0.1mM β-mercaptoethanol (Life Technologies), and 8ng/mL bFGF
(PeproTech, Rocky Hill, NJ). Cells were passaged as clumps using collagenase IV (Stem Cell Technologies, Vancouver, British Columbia, Canada).

2.4.3 Detection of CRISPR/Cas9-Facilitated Excision

HEK293 and C2C12 cells were transfected using FuGene HD (Promega, Madison, WI). Briefly, 150,000 cells were plated one to two days prior on poly-L-lysine-coated (HEK293) or standard (C2C12) 24-well tissue culture plates and were transfected with 5 μg (HEK293) or 3 μg (C2C12) of each sgRNA/Cas9 vector for a total of 10 μg or 6 μg per transfection, respectively, in Opti-MEM I (Life Technologies) media with FuGene HD being used at a 3:1 ratio. Cells were harvested four days post-transfection and genomic DNA was isolated using the DNeasy Blood and Tissue Mini Kit (Qiagen). Each transfection was separately performed at least twice. H9 hESCs were transfected using the Amaxa Nucleofector (Lonza, Basel, Switzerland) with the Human Embryonic Stem Cell Nucleofection Kit 2 (Lonza). Briefly, cells were passaged normally into Matrigel-coated 6-well plates three days prior to nucleofection. Before nucleofection, media was replaced at least 1 hr beforehand with fresh media containing 10μM ROCK inhibitor Y-27632 (R&D Systems, Minneapolis, MN). Cells were dissociated with collagenase IV, and subsequently electroporated with 1.5 μg of each sgRNA/vector for a total of 3 μg using program B-16 according to the manufacturer’s instructions, and plated on 0.1% gelatin-coated 12-well plates containing at one reaction per well in human stem cell media containing 10μM ROCK inhibitor Y-27632. Each reaction was carried out at least twice. Genomic DNA was isolated from the cells 2 days later with the DNeasy Blood and Tissue Mini Kit.

PCR amplification of the deletion-containing regions was carried out using OneTaq (NEB) and Q5 high fidelity polymerase using 150 ng of gDNA according to the manufacturer’s instructions. Amplicons were subjected agarose gel electrophoresis, and subsequently the deletion-containing bands, as well as the full-length bands where applicable, were excised and purified using the MinElute Gel Extraction Kit (Qiagen). Purified amplicons were then subcloned using the CloneJet PCR Cloning Kit (Thermo-Fisher Scientific), and subsequently transformed into α-Select electrocompetent cells (Bioline, Taunton, MA), 10-β chemicompetent cells (NEB), or STBL3 chemicompetent cells according to their manufacturers’ instructions. Colonies were inoculated and grown overnight. Plasmids were
isolated the following day using the Miniprep Spin Kit (Qiagen), and subsequently subjected to Sanger sequencing of the inserts. Primers used for amplification of excision events appear in Table A.3.

2.4.4 Statistics

All statistical analysis was carried out using GraphPad Prism 6.0 (GraphPad Software, La Jolla, CA). The Kruskal-Wallis test followed by post hoc Dunn’s multiple comparisons test was used to analyze all distributions of deviation from 0/0 precise repair.
Chapter 3

Knock-in Blunt Cloning: Exploiting CRISPR/Cas9-Mediated Non-Homologous End-Joining

Jonathan M. Geisinger, Sophia Hernandez, Michele P. Calos

3.1 Introduction

The ability to make precise double-strand breaks (DSBs) in the genome is extremely useful for genome engineering, as it can facilitate directed changes in the genome. These changes can be used for applications ranging from studying a gene to engineering an entire biosynthetic pathway. These DSBs greatly increase the rate of gene targeting above homologous recombination alone. The most popular tools for making DBSs are currently zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the CRISPR/Cas9 system. Of these three, the CRISPR/Cas9 system offers the greatest ease of use.

This ease of use has greatly facilitated the application of the CRISPR/Cas9 system to virtually any organism. As such, there have been several proof-of-principle demonstrations of the CRISPR/Cas9 system in various eukaryotes, including *C. elegans*, *Drosophila*, *S. cerevisiae*, and the turquoise killifish *N. furzeri* (Friedland et al., 2013; Kim, Ishidate et al., 2014; Zhang, Koolhass, and Schnorrer, 2014; Gratz, Ukken et al., 2014; DiCarlo et al., 2013; Harel et al., 2015). Additionally, the system has been used for disease-modeling with a focus on cancer (Xue, Chen, Yin et al., 2014; Heckl et al., 2014; Platt, Chen et al., 2014). There are also already efforts underway for using the CRISPR/Cas9 system for therapeutic development, including Duchenne muscular dystrophy, HIV, herpes virus infection, and hereditary
tyrosemia (Ousterout et al., 2015; Hu, Kaminski et al., 2014; Wang and Quake, 2014; Yin, Xue et al., 2014). All of these applications rely on generating insertions and deletions through error-prone NHEJ or the precise insertion of exogenous sequences through homologous recombination (HR) or homology-directed repair (HDR). Initially, the efficiencies of CRISPR/Cas9-mediated HR and HDR were relatively low (Mali, Yang et al., 2013; Cong et al., 2013). However, these efficiencies have been increased through the use of cell cycle synchronization, small molecule enhancers of HDR and inhibitors of NHEJ, and suppression of the expression of members of the canonical NHEJ repair pathway (Lin, Staahl et al., 2014; Yu, Liu et al., 2015; Maruyama et al., 2015; Chu et al., 2015).

Alternative methods have also been developed for precisely knocking exogenous sequences into the genome. One method, termed CRIS-PITCh, exploits microhomology-mediated repair through the use of three guide RNAs (Nakade et al., 2014). One guide targets the genome, while the others target the donor vector in such a way that the cassette to be knocked-in possesses short microhomologies on both ends. Another method has been developed in zebrafish using plasmid DNA in which one guide RNA is used to cleave both the genome and the plasmid, ultimately resulting in homology-independent knock-in of the plasmid, albeit with reduced levels of perfect ligation (Auer et al., 2014).

Similarly to Auer and colleagues (2014), we hypothesized that we could exploit the canonical precise NHEJ pathway to precisely insert exogenous sequences into the genome. Here, we demonstrate the use and optimization of such a method we call knock-in blunt ligation (KiBL) in HEK293 cells to achieve efficiencies of up to 35.8% and precise genome-cassette junctions at levels up to 60%.

3.2 Results

3.2.1 Knock-in Blunt Ligation: CRISPR/Cas9-Mediated in vivo Blunt-End Cloning

Because of the high level of precise NHEJ repair we observed in Chapter 2, we asked if we could exploit this apparent in vivo blunt-end ligation to replace endogenous sequences
precisely with linear exogenous sequences in a homology-independent manner using CRISPR/Cas9 (Figure 3.1).

Figure 3.1. Knock-in blunt ligation. The use of two sgRNAs at the same locus should facilitate ligation of exogenous sequence in place of the excised sequence.

To test this idea and examine the level of repair precision, we constructed an expression cassette consisting of the mouse PGK promoter driving expression of the PuroΔTK fusion gene linked via a P2A skipping peptide to mCherry, followed by the rabbit beta globin terminator sequence (PTKmChR). We also constructed another cassette containing the same sequences, but flanked by a phiC31 attP site upstream of the PGK promoter and a Bxb1 attP site downstream of the terminator sequence for the initial purpose of generating a selectable landing pad that did not rely on homology arms for integration (DICE-EPv2; Figure 3.2).

Figure 3.2. Schematic of the PGK-PTKmChR and DICE-EPv2 cassettes.

Each cassette was amplified via PCR with 5’ phosphorylated primers, purified, and subsequently co-transfected into HEK293 cells with various guide RNA/Cas9 constructs, singly or in pairs, before being subjected to flow cytometric analysis and junction PCR,
followed by Sanger sequencing (Figure 3.3A). In this set of experiments, we utilized sgRNAs targeting the PAX7 3’UTR (Figure 3.3B).

For the DICE-EPv2 cassette, flow cytometric analysis revealed that the percentage of mCherry+ cells was increased above cassette alone (0.20 ± 0.15%) for sgRNAs PAX7r2-1 and PAX7r3-2, both alone and paired (referred to as PAMs Out Large) with a range 1.61-2.67% two days post-transfection, and that the percentage of mCherry+ cells could be further increased to 17.80-24.50% with four days of puromycin selection, albeit at the risk of increasing selection for random integration (Figures 3.4 and 3.5A; 14.30 ± 0.20% for cassette alone). For the PTKmChR cassette after two days, we also observed that the percentage of mCherry+ cells was increased above cassette alone (0.56 ± 0.32%) in the presence of all guide RNAs tested, which consisted of two pairs of PAMs 5’, two pairs of PAMs Out (the four pairs previously characterized in Figure 2.7), and the PAMs Out Large pair, together and singly for a range of 2.63-5.74% (Figure 3.5B). Interestingly, four days of puromycin selection resulted in a lower percentage of mCherry+ cells with cassette alone (6.57 ± 1.64%) than with the DICE-EPv2 cassette. Puromycin selection also resulted in an increase in the percentage of mCherry+ cells for the PAMs 5’ and PAMs Out pairs, ranging from 17.10% to 32.60%.

Figure 3.3. Initial scheme targeting KiBL to the PAX7 3’UTR. (A) Workflow for KiBL experiments: HEK293 cells are transfected with sgRNA/Cas9 vectors and cassettes before flow cytometric analysis and sequence analysis. (B) Schematic diagram of the PAX7 3’UTR region along with sgRNAs used in these experiments.
Figure 3.4. Representative FACS plot of DICE-EPv2 KiBL. This data displaying the difference in the percentage of mCherry+ cells between the DICE-EPv2.0 cassette transfected alone and with sgRNA/Cas9 vectors after two days.

Figure 3.5. Quantification of initial KiBL experiments via flow cytometry. (A) Quantification of the DICE-EPv2 transfection experiments in terms of percentage of mCherry+ cells using sgRNAs PAX7r2-1 and PAX7r3-2. (B) Quantification of the PGK-PTKmChR cassette transfections. Same experimental conditions as in panel F. n = at least 2 independent experiments consisting of 3 replicates each for each transfection. Data is shown as the mean ± SEM.

Subsequent PCR analysis of bulk unsorted cells for all possible junctions of genomic and cassette sequences revealed that knock-in blunt ligation of the DICE-EPv2 cassette led to imprecise repair of the genomic junction (60% of sequenced amplicons) and an absence of precise cassette junction repair, whereas knock-in blunt ligation of PTKmChR cassette resulted in a high level of precise genomic junction repair (93.75%) and an appreciable degree
of precise cassette junction repair (37.5%; Figure 3.6A-B). We speculated that the lack of precise cassette junction repair with DICE-EPv2 was attributable to the formation of hairpin structures by the palindromic \textit{AttP} sites. Such hairpins could be unfavorable to direct end-

![Graph of cumulative percentage of sequenced amplicons categorized by repair status of the genome-cassette junction.](image)

**Figure 3.6.** KiBL can generate perfect genome-cassette junctions. (A) Graph of cumulative percentage of sequenced amplicons categorized by repair status of the genome-cassette junction. \(n\) for each is displayed on the graph. (B) An example chromatogram of the precise genomic/precise cassette junction for the PTKmChR cassette. Boxed sequence denotes PGK promoter sequence, PAM is underlined, and the first four bases of sgRNA target are overlined.

joining. Thus, these hairpins would be more subject to removal by nucleases. Additionally, we observed one instance of an insertion of part of the pX330 plasmid (similarly to that observed in Canver, Bauer \textit{et al.}, 2014). These results indicated that the blunt DSBs generated by the CRISPR/Cas9 system could be exploited in mammalian cells to knock-in PCR cassettes in a homology-independent manner. We observed that cassettes were knocked into the genome in both orientations, indicating that ligation by precise NHEJ lacks directionality, as expected due to lack of homology arms on our cassettes (Figure 3.7).

Because of the relatively large size of the DICE-EPv2 and PTKmChR cassettes and the relative dimness of mCherry, we developed a second series of smaller, brighter cassettes. These cassettes consist of the \textit{Ef1\(\alpha\)} promoter driving the expression of a fluorescent protein, which is followed by the rabbit \(\beta\)-globin terminator sequence (Figure 3.8). Our fluorescent proteins of choice were Clover, mRuby2 (both from Lam \textit{et al.}, 2012), mKO\(\kappa\) (Tsutsui \textit{et al.}, 2014).
Figure 3.7. Analysis of PAX7-PTKmChR/DICE-EPv2.0 junctions. PCRs were carried out to analyze the PAX7 5'-5' cassette and PAX7 3'-5' cassette junctions in transfected HEK293 cells through agarose gel electrophoresis. The PAX7-X2 primers were used along with the PGK-rev primer for amplification. Marker ladder bands are denoted for reference. Expected band sizes for 5'-5' junctions: r1-1 + r1-2 = 1188 bp; r1-1 + r1-4 = 1233 bp; r1-2 + r1-3 = 1188 bp; r1-3 + r1-4 = 1233 bp; r2-1 = 1595 bp; r3-2 = 808 bp; r2-1 + r3-2 = 808 bp. Expected band sizes for 3'-5' junctions: r1-1 + r1-2 = 1028 bp; r1-1 + r1-4 = 1028 bp; r1-2 + r1-3 = 1036 bp; r1-3 + r1-4 = 1036 bp; r2-1 = 929 bp; r3-2 = 1716 bp; r2-1 + r3-2 = 929 bp.

2008), mCardinal (Chu et al., 2014), and mCerulean (Rizzo et al., 2004) due to minimal spectral overlap and (for the most part) brightness. These cassettes were less than 2 kb in size and could be readily amplified from template plasmid and easily purified. We called these cassettes the pKER (polymerase-chain-reaction-based Knock-in EF1α-RBG terminator) series, and they are depicted by brightness in Figure 3.8.

Figure 3.8. Schematic diagrams of the pKER cassette series. Cassettes are organized by brightness.
To quickly assess the suitability of the pKER series for KiBL, we used pKER-mKOκ to investigate the effects of the orientation of the paired PAMs as well as the effects of the cassette’s phosphorylation status on the efficiency of KiBL, using the percentage of positive cells as a proxy. We reasoned that unphosphorylated cassettes may have a selective advantage for KiBL as they may not be detected as free, broken ends of DNA as readily as phosphorylated cassettes may be, and they may be decrease the possibility of amplicon concatemerization. For these experiments, the sgRNA pairs consisted of all four possible PAM orientations directed against the same region of the human H11 locus on chromosome 22 (Figure 3.9), which was previously identified as a safe harbor site in the human genome (Zhu et al., 2014).

Figure 3.9. Targeting the H11 locus with CRISPR/Cas9. Schematic diagram of the 5’ side of the H11 locus on human chromosome 22 depicting the four sgRNAs positioned relative to the sense strand. 548 bp separates the 5’-most and the 3’-most sgRNAs from each other.

H11 is an intergenic locus between *DRG1* and *EIF4ENIF1* on human chromosome 22 (Figure 3.10). There is a potential non-coding RNA at this locus, but it has only been observed in human lung and only from a cDNA screen. We transfected HEK293 cells with phosphorylated or unphosphorylated pKER-mKOκ cassettes and pairs of sgRNA/Cas9 vectors, before subjecting the cells to flow cytometric analysis two days later (Figure 3.11A). Two days post-transfection, we observed a higher percentage of mKOκ+ cells in the cultures co-transfected with CRISPR/Cas9 vectors (ranging from 8.15-11.64%) than in those transfected with cassette alone (1.34-1.49%; Figure 3.11B). Surprisingly, flow cytometry did not reveal any significant differences between PAM orientations or cassette phosphorylation status in terms of percentage of mKOκ+ cells or relative normalized median fluorescence intensity (MFI), other than that the presence of Cas9 increased both measures (Figure 3.11C; two-way ANOVA with repeated measures followed by post-hoc Tukey’s multiple comparisons test, P < 0.003 for all comparisons when comparing PAMs-treated to cassette alone). These results suggested that the pKER cassettes were well-suited for use in KiBL: positive cells were bright, and treatment with Cas9 resulted in a large increase in positive cells.
Figure 3.10. Structure and conservation of the H11 safe harbor locus. Schematic from the UCSC Genome Browser (http://genome.ucsc.edu; used Feb. 2009 (GRCh37/hg19) assembly) displaying H11 locus in the human genome on chromosome 22. The flanking genes are DRG1 and EIF4ENIF1. AK074476 is a potential non-coding RNA isolated from human lung. Below are depicted alignments of the region for several vertebrates.

Figure 3.11. Suitability of the pKER series for KiBL. (A) Diagram depicting the workflow for analyzing pKER expression. pKER cassettes and pairs of sgRNA/Cas9 vectors are co-transfected into HEK293 cells and analyzed by flow cytometry after two days. (B) Representative flow cytometric data generated with pKER-mKOx and the H11 sgRNAs. (C) Quantification of the percentage of mKOx+ cells and their relative median fluorescence intensity for phosphorylated and unphosphorylated cassettes for all four possible PAM orientations. n = 3 independent experiments each consisting of 3 technical replicates. Data is displayed as the mean ± SEM of the averages of each experiment and were analyzed using a two-way ANOVA with repeated measures followed by a post hoc Tukey’s multiple comparisons test. ** = P < 0.003.
3.2.3 Stabilized Cas9-DD Combined with Nuclease Protection Facilitates a Higher Degree of Precise Repair at the Genome-Cassette Junction

We and others have observed that, in utilizing the CRISPR/Cas9 system, successful modifications of one allele appeared to coincide with undesired mutations in the sister allele (unpublished data; Canver, Bauer et al., 2014). The duration of Cas9 activity may be responsible for these additional modifications. Thus, we initially sought to generate a more temporally controllable Cas9. We chose to utilize the FKBP12 L106P destabilization domain because it was relatively small, has been shown to be effective at destabilizing a wide range of proteins, and its stabilizing ligand, the small molecule Shield-1, was relatively inexpensive and had virtually no effect on cell viability, as well as good intracellular availability and potency (Banaszynski et al., 2006; Maynard-Smith et al., 2007). We initially chose to fuse the destabilization domain to the C-terminus of SpCas9 because we desired a version of Cas9 that possessed some residual stability in the absence of Shield-1 (Banaszynski et al., 2006). The resulting fused domain was in close proximity to the PAM-binding domain of SpCas9 (Anders et al., 2014). We called this version of Cas9 “Cas9-DD” (Figure 3.12A-B). Additionally, we also constructed a version of Cas9 with the destabilization domain fused to the N-terminus, which we called “DD-Cas9” (Figure 3.12A and C).

![Figure 3.12. Schematics of destabilized Cas9.](image)

(A) Representation at the abstract sequence level. (B) and (C) Representation at the structural level for Cas9-DD and DD-Cas9 respectively.
We then examined the relative stability of Cas9-DD and DD-Cas9 via Western blotting of HEK293 cells transfected with constructs encoding the H11-r1-2 sgRNA and WT-Cas9, Cas9-DD, or DD-Cas9 (Figure 3.13).

After 29 hours, we observed a higher level of Cas9-DD protein present in cells treated with 0.5 µM Shield-1, compared to cells treated with Cas9-DD vector alone (destabilized Cas9-DD). Similarly, for DD-Cas9, treatment with 0.5 µM Shield-1 appeared to increase the level of DD-Cas9 protein relative to DD-Cas9 vector alone. These results indicated that the addition of the destabilization domain to Cas9 resulted in a destabilized protein that could be stabilized by the addition of Shield-1.

To characterize how Cas9-DD would affect KiBL efficiency, we transfected our PAX7 PAMs Out Large sgRNA pairs on vectors encoding the sgRNAs and either wild-type Cas9 (WT-Cas9) or Cas9-DD, along with phosphorylated or unphosphorylated pKER-Clover cassettes into HEK293 cells, which were subjected to flow cytometric analysis and harvested for
genomic DNA after two days (Figure 3.14A). We chose PAX7 because we had already successfully carried out KiBL and analyzed the pattern of repair at this locus (Figure 3.6).

**Figure 3.14. KiBL at the PAX7 3’UTR with pKER-Clover and Cas9-DD.** (A) Schematic diagram displaying the workflow for comparing WT-Cas9 to Cas9-DD. HEK293 cells are transfected with pKER-Clover cassette and the PAX7 PAMs Out-Large pair of vectors encoding either WT-Cas9 or Cas9-DD. Two days post-transfection, a portion of the cells are subjected to flow cytometric analysis and the remainder are reserved for genomic DNA isolation for sequence analysis. (B) Representative flow cytometric data for the pKER-Clover cassette and the PAX7 PAMs Out-Large sgRNAs with destabilized Cas9-DD. (C) Quantification of the percentage of Clover+ cells and their relative median fluorescence intensity for phosphorylated and unphosphorylated cassettes for WT-Cas9, destabilized Cas9-DD, and stabilized Cas9-DD (1 µM Shield-1 for 24 hours). n = 5 independent experiments each consisting of 3 technical replicates. Data is displayed as the mean ± SEM of the averages of each experiment and were analyzed using a two-way ANOVA with repeated measures followed by a *post hoc* Tukey’s multiple comparisons test. ** = P < 0.001.

Flow cytometric analysis confirmed that co-transfection with the sgRNA/Cas9 vectors increased the percentage of Clover+ cells (18.12-35.8%) over that of cassette alone (2.36-2.64%; Figure 2.14B-C; P < 0.002 for all comparisons; Tukey’s multiple comparisons test). Interestingly, we did not detect any significant difference between Cas9-DD and WT-Cas9 for the percentage of Clover+ cells or for normalized MFI (Figure 3.14C). Additionally, we did not detect any significant difference between the presence and absence of 1 µM Shield-1 with Cas9-DD.
Having already observed that WT-Cas9 facilitated knock-in blunt ligation of linear cassettes, we sought to determine the effect of Cas9-DD at the molecular level. To this end, we examined the 5’ PAX7 genomic-5’ pKER-Clover junction resulting from KiBL using unphosphorylated cassettes with WT-Cas9 or destabilized Cas9-DD with our sgRNAs (Figure 3.15A-B).

**Figure 3.15. Molecular analysis of KiBL with Cas9-DD and nuclease protection.** (A) Diagram of the modified PAX7 3’UTR locus that displays pKER-Clover placed at the locus in the 5’ to 3’ orientation. Arrows indicate the primers used to generate amplicons for (B). (B) Analysis of the effect of phosphorothioate bonds on KiBL. PCR was carried out to analyze the PAX7 5’-5’ pKER cassette junction in transfected HEK293 cells through agarose gel electrophoresis. The PAX7-x2-fwd and pKER-detector-5’ primers were used in all reactions. The expected band size is 750 bp. Each lane represents one technical replicate for the given condition from the same experiment.

We did not examine junctions using stabilized Cas9-DD at this time as we reasoned that its junctions would resemble those of WT-Cas9. Additionally, we examined whether the addition of three phosphorothioate bonds in the 5’ ends of the primers used to amplify the cassettes affected the precision of the cassette side of the junction by protecting the cassette from nuclease degradation. We observed that unprotected cassettes co-transfected with WT-Cas9 resulted in precise genomic junctions, but imprecise cassette junctions in 100% of observed amplicons (Figure 3.16). Interestingly, protecting the 5’ ends of the cassette only resulted in a relatively modest increase in precise genomic/precise cassette junctions (40% of amplicons) when using WT-Cas9 (Figure 3.16). When destabilized Cas9-DD was used in conjunction with phosphorothioate bonds, we observed a greater increase in precise genomic/precise cassette junctions (60% of amplicons), but, surprisingly, we also observed a slight increase in imprecise genomic/imprecise cassette junctions (20% of amplicons; Figure 3.16). These results appeared to demonstrate that the use of destabilized Cas9-DD, along with the addition
of nuclease protection to the pKER cassettes, facilitated a higher degree of cassette junction precision in KiBL.

![Figure 3.16. Comparison of WT-Cas9 with destabilized Cas9-DD and nuclease protection.](image)

Numbers of amplicons analyzed appear beneath their respective treatments. Also depicted are the sequences of the amplicons from the destabilized Cas9-DD junction analysis. Black denotes genomic sequence, turquoise denotes cassette sequence, PAM is underlined, and dashes denote unobserved sequence.

### 3.3 Discussion

These results demonstrate that CRISPR/Cas9-generated DSBs efficiently facilitate the uptake of multi-kilobase PCR amplicons in a targeted, homology-independent manner. These results also illustrated that using two guide RNAs simultaneously allows for replacement of the endogenous sequences with the amplicon cassette. We also constructed a destabilized version of Cas9, the use of which appeared to result a higher degree of precise junction repair between the genome and the cassette. Lastly, we found that the addition of phosphorothioate bonds in the 5’ termini of the primers used to amplify the cassette also increased the frequency of precise junction repair.

KiBL has several advantages compared to other non-HR methods. Regarding CRIS-PITCh, KiBL has the advantage of possessing less complexity. CRIS-PITCh requires three different guide RNAs in addition to the donor vector and requires all three to be cleaved at roughly the same time for microhomology-mediated end-joining to occur (Nakade et al., 2014), whereas KiBL only requires one to two guide RNAs and the cassette. Additionally, there is the potential concern for the donor vector backbone to integrate into the genome. Compared to the method of Auer and colleagues (2014), KiBL results in a higher level of precise genome-
cassette junctions: the Auer method resulted in up to 17% precise junctions, whereas KiBL reaches up to 60% in HEK293 cells. Generally, the plasmid-based methods depend on in vivo cleavage of a donor to mediate integration, which was previously demonstrated to increase targeting efficiency with ZFNs (Cristea et al., 2012), whereas KiBL has no such restriction. It is worth mentioning that the Cas9 nickase has been used to knock-in a double-stranded oligonucleotide (dsODN) with overhangs in a strategy similar to the ObLiGaRe method used with ZFNs, albeit with relatively low efficiency compared to KiBL as well (Ran, Hsu et al., 2014; Maresca et al., 2013).

Another set of methods for knocking sequences into the genome has been developed around exploiting HR and HDR. In cultured Drosophila cells, the use of PCR amplicons containing homology arms has been used to great effect (Böttcher et al., 2014), which is similar to what has been observed with human immortalized cell lines and mouse ESCs (Zheng et al., 2014; Li et al., 2014). A subset of these methods has focused on increasing HDR frequency through cell cycle synchronization (Lin, Staahl et al., 2014) and by inhibition of the NHEJ machinery (Chu et al., 2015; Maruyama et al., 2015). Compared to these methods, KiBL offers the advantage of functioning during the G1, S, and G2 phases of the cell cycle, whereas the HR and HDR strategies are restricted to late S and G2 phases. Thus, our method may be particularly useful in non-dividing cells, such as neurons where the application of Cas9 has been limited in vivo to generating mutations through error-prone repair (Swiech, Heidenreich et al., 2014).

While being able to place exogenous sequences at any genomic location is useful, being able to knock-in sequences at loci where transgenes do not disrupt endogenous gene regulation (known as safe harbor sites) is also useful. We chose the H11 as one of our target loci because it has previously been identified as a safe harbor site in the human and mouse genome (Zhu et al., 2014; Tasic et al., 2011). The use of a safe harbor site is of great importance for the placement of transgenes in cell therapies because of the potential for the integration event itself to perturb endogenous genes in a possibly oncogenic manner (for further information see Sadelain, Papapetrou, and Bushman et al., 2011). Additionally, the gene order of the H11 locus appears to be highly conserved, particularly with respect to mammals and through vertebrates in general. For example, the order appears even to be conserved in the coelacanth. This conservation makes H11 particularly attractive as a genome engineering site.
The CRISPR/Cas9 system has the advantages of being relatively inexpensive and easy to use, but may have larger off-target effects than TALENs or ZFNs. Several approaches have been put forward to restrict off-target cleavage, including simply using less guide RNA and Cas9, using truncated guide RNAs for increased specificity, using paired nickases, controlling Cas9 expression through the inducible Tet-ON system, and splitting Cas9 in half (Hsu et al., 2013; Fu et al., 2013; Ran, Hsu et al., 2013; Zhu, Gonzalez, and Huangfu, 2014; Wright, Sternberg et al., 2015; Zetsche, Volz, and Zhang, 2015). While the first three approaches appear to lack undesired effects, Tet-ON induction, however, generates a large amount of mRNA, which could lead to an increase in Cas9 off-target activity. The split-Cas9 variants, while potentially being more controllable, have substantially reduced cleavage activity, relying on either the sgRNA or rapamycin to bind the two halves together (Wright, Sternberg et al., 2015; Zetsche, Volz, and Zhang, 2015). Our destabilized Cas9 retains the simplicity of requiring only two expression cassettes, rather than three or four. Additionally, Shield-1 itself does not induce any undesired responses when applied to cells in culture or in animals, unlike rapamycin, (Banaszynski et al., 2006; Maynard-Smith et al., 2007; Banaszynski et al., 2008). Recently, a fifth variant was developed using an evolved ligand-dependent intein to restrict the activity of Cas9 until the binding of the ligand (in the form of 4-hydroxytamoxifen) results in the cleavage of the intein and activation of Cas9 (Davis et al., 2015). Such a system is elegant, but possesses the drawback of being unable to restrict Cas9’s activity once the intein is cleaved. It would be interesting to combine the destabilization domain with this intein because such a combination could allow for an unprecedented level of control of Cas9 activity.

Our observation that unphosphorylated cassettes are capable of ligating into CRISPR/Cas9-induced DSBs at relatively high frequency in HEK293 cells appears to contradict observations in the GUIDE-seq method (Tsai et al., 2014b). Tsai and colleagues observed that phosphorothioate bonds were not only necessary for integration, but that they were required in both the 5’ and 3’ termini. However, our finding can be explained by our much larger cassettes: our unprotected, unphosphorylated cassettes may be much less negatively affected by endogenous nucleases than the 34-bp GUIDE-seq dsODN. Additionally, our cassettes are generated through PCR amplification whereas the GUIDE-seq dsODN was generated through annealing. This difference may explain the requirement of the 5’ and 3’ phosphorothioate
bonds in the dsODN as the annealing may not generate a fully double-stranded cassette, i.e., the termini may not be completely annealed to each other.

In this chapter, we have presented a method for utilizing the CRISPR/Cas9 system to facilitate in vivo blunt-end cloning in precise, homology-independent manner through NHEJ. This method, KiBL, is made possible by the ability of Cas9 to make blunt DSBs, which we and others have exploited using two sgRNAs at once to facilitate precise excision of the intervening sequence (Cong et al., 2013; Mali, Yang et al., 2013; Canver, Bauer et al., 2014; Zheng, Cai et al., 2014). The main advantages of our method are its relatively high frequency of precise genomic-cassette junctions, its lack of incorporation of additional exogenous sequences, and its independence from sgRNA efficiency. Additionally, KiBL may prove amenable to high-throughput applications, whereas HR is not, due to the necessity of constructing homology arms for each targeted locus. This construction is of particular concern for the concept of saturation editing, which could become very expensive for targeting large numbers of loci (Findlay, Boyle et al., 2014). The main drawbacks of our method are generating large amounts of cassette, which can be easily overcome by increasing the number of cassette amplification reactions in parallel. Future development of the KiBL will involve examining its efficiency in generating fluorescently-tagged reporter genes at their endogenous locus with minimal disruption of their regulatory sequences. Such reporter alleles would permit the interrogation of the response of the gene or pathway to a wide range of genetic and environmental perturbations.

3.4 Methods

3.4.1 Choice of sgRNAs and Vector Construction

The choice of and design of guide RNAs for targeting the H11 locus followed the same procedure as in section 2.4.1. The regions used and the sequences of the guides appear in Tables A.1 and A.2, respectively.

The C-terminal L106P destabilization domain was obtained as a gBlock (IDT), and was digested with FseI (New England Biolabs (NEB), Ipswich, MA) and BsaI-HF (NEB). pX330 was digested with FseI and EcoRI-HF (NEB), and dephosphorylated with Antarctic
phosphatase (NEB). The digested C-terminal DD and the pX330 backbone were ligated with T4 DNA ligase (NEB), and transformed into STBL3 chemicompetent cells (Life Technologies, Grand Island, NY). In order to clone sgRNAs into pX330-Cas9-DD, the BbsI cloning site was replaced with a BsaI x2 cloning site using a gBlock containing the entire sgRNA expression cassette. Both pX330-Cas9-DD and the gBlock were sequentially digested with PciI (NEB) and KpnI-HF (NEB) before dephosphorylation of the backbone and subsequent ligation and transformation to generate pX330-BsaIx2-Cas9-DD. This new cloning site uses FastDigest Eco31I (isochizomer of BsaI; Thermo Fisher Scientific, Waltham, MA) and the two following general primers:

- Sense oligo: 5’- accg-NNNNNNNNNNNNNNNNNNNNN-3’
- Antisense oligo: 5’-aaac-NNNNNNNNNNNNNNNNNNNNN-3’

Additionally, pX330-Cas9-DD versions of the H11 sgRNAs were generated by cloning Cas9-DD directly into their pX330 vectors through the use of FastDigest BshTI (Thermo Fisher Scientific) and FastDigest NotI (Thermo Fisher Scientific).

To generate pX330-BsaIx2-DD-Cas9, the N-terminal DD was obtained as a gBlock and digested with FastDigest BshTI and FastDigest BglII (Thermo Fisher Scientific) before being cloned into the pX330 backbone that had been digested with the same enzymes and dephosphorylated with FastAP (Thermo Fisher Scientific). This vector was subsequently digested with FastDigest KpnI (Thermo Fisher Scientific) and FastDigest NotI to isolate the DD-Cas9 insert. pX330-BsaIx2-Cas9-DD was digested with the same enzymes and dephosphorylated to remove Cas9-DD. The DD-Cas9 insert was cloned into the pX330-BsaIx2 backbone to generate pX330-BsaIx2-DD-Cas9.

### 3.4.2 Knock-In Cassette Construction

The DICE-EPv2.0 and PTKmChR cassettes were amplified from a common vector containing the Puro$_r$ΔTK fusion gene (from Addgene # 22733) followed by a P2A ribosomal skipping element and mCherry followed by the rabbit beta-globin terminator. This cassette is under the control of the mouse phosphoglycerol kinase (PGK) promoter. Cassettes were amplified via PCR with Q5 high-fidelity polymerase, subjected to gel electrophoresis, excised, and purified
with the MinElute Gel Extraction Kit. For the DICE-EPv2.0 cassettes, the phiC31 and Bxb1 attP sites were included in the primers used for amplification. All primers used for cassette amplification appear in Table A.4.

The pKER cassettes were assembled by first obtaining phosphorylated gBlocks for Clover, mRuby2, and mCerulean in the form of EF1α promoter-fluorescent protein-rabbit beta globin terminator. These three gBlocks were cloned separately into a kanamycin resistance backbone with blunt ends. mCardinal and mKOκ were obtained as gBlocks, digested with FastDigest NheI (Thermo Fisher Scientific) and FastDigest NotI, and cloned into a dephosphorylated pKER-Clover backbone that had been digested with the same enzymes to remove Clover. Cassettes were amplified from 1-3 ng of plasmid using Q5 high fidelity polymerase in eight individual reactions. After amplification, the reactions were pooled and purified using either the MinElute PCR Purification kit (Qiagen) or the GeneJet PCR Purification kit (Thermo Fisher Scientific). The purified reaction was then treated with FastDigest DpnI to digest residual plasmid. After digestion, the reaction was brought to 100 µL with nuclease-free water (Life Technologies) and purified with a CHROMASpin-1000+TE chromatography column (Clontech, Mountain View, CA) to remove digested plasmid and restriction enzyme.

### 3.4.3 Detection of Knock-in Blunt Ligation

For HEK293 cells, FuGene HD was used to transfect cells as described in section 2.4.3, using 100 ng of cassette and 1.5 µg of each sgRNA/Cas9 vector for a total of 3.1 µg of DNA. Reactions were carried out in triplicate. Shield-1 (Clontech) was added to the cells immediately prior to transfection at a concentration of 1 µM. Puromycin (Life Technologies) selection was started two days after transfection at a concentration of 1µg/mL and carried out for four days with fresh media and antibiotic being replaced every two days. Cells were analyzed by flow cytometry two days after transfection or six days in the case of puromycin-selected cells. Briefly, cells were trypsinized with 0.05% trypsin, resuspended in PBS (Thermo Fisher Scientific) with 2% FBS, and were filtered before analysis to remove clumps and debris. gDNA was isolated from the remaining cells with the DNeasy Blood and Tissue Mini Kit. Genome-cassette junctions were amplified with Q5 Hi-Fidelity polymerase using at least 100 ng of DNA per reaction. Amplicons were subjected to gel electrophoresis, excised,
purified, subcloned, transformed, and sequenced as described above. Primers used appear in Table A.5.

### 3.4.4 Flow Cytometry

Flow cytometric analysis was carried out on LSRII- and FACScan-class analyzers (BD Biosciences, San Jose, CA). Sorting was carried out on FACSemiaII-class sorters (BD Biosciences). Live cells were discriminated on the basis of DAPI exclusion using the NucBlue Fixed Cell Stain ReadyProbes reagent (Life Technologies). mCherry was detected via excitation with a 561 nm yellow/green 100 mW laser with a 571 nm LP filter in the optical path, a 725 nm SP splitter, and a 690±40 nm BP filter. Clover was detected via excitation with a 488 nm 50mW blue laser, a 505 nm LP splitter, and 525±50 nm BP filter. mKOκ was detected via excitation with a 532 nm 150mW green laser and a 575±25 nm BP filter. mCardinal was detected with a 640 nm 40mW red laser and a 670±30 nm BP filter. mRuby2 was detected with a 532 nm 150mW green laser, a 600 nm LP splitter, and a 610±20 nm BP filter. mCerulean was detected with a 405 nm 50mW violet laser and a 450±50 nm BP filter in the absence of DAPI. All flow cytometric data were analyzed using FlowJo software (Tree Star, Ashland, OR).

### 3.4.5 Western Blotting

HEK293 cells in 6-well plates that had been plated two days prior were transfected using FuGene HD with 5 µg of plasmid encoding the H11-r1-2 sgRNA and WT-Cas9, Cas9-DD, or DD-Cas9 as described above. Cas9-DD and DD-Cas9 were transfected in duplicate. All wells received 0.5 µM Shield-1 except one replicate each for Cas9-DD and DD-Cas9. The following day, media was aspirated and replaced with fresh media containing 0.5 µM Shield-1 for all wells except for the replicates that did not receive Shield-1 the previous day. 5 hours later, media was aspirated, cells were washed twice with cold PBS, and protein was extracted with RIPA buffer (Thermo Fisher Scientific) containing 1x HALT protease inhibitor cocktail (Thermo Fisher Scientific). Cell lysate was snap-frozen for later analysis. Protein concentration was determined with the Bradford assay.
For Western blotting, 50µg of protein was denatured in the presence of 1x Laemmlie buffer (Bio-Rad Laboratories, Hercules, CA) and 0.1 M DTT (Life Technologies) at 100°C for 5 minutes. Samples were then resolved on a 10% SDS-PAGE gel (Bio-Rad) at 100 V in 1x Tris/Glycine/SDS buffer (Bio-Rad). Protein was transferred to an Immuno-Blot PVDF membrane (Bio-Rad) under wet transfer conditions at 200 mA for 1 hour with constant current at 4°C in 1x Tris/Glycine/SDS buffer + 10% methanol. Following transfer, the membrane was blocked for 1 hour at room temperature with agitation with Protein-Free T20 blocking buffer (Thermo Fisher Scientific) and incubated overnight with agitation at 4°C with 1:10,000 monoclonal ANTI-FLAG M2 antibody (Sigma-Aldrich) and 1:10,000 monoclonal anti-GAPDH antibody (clone GAPDH-71.1; Sigma-Aldrich). The following day, the membrane was washed three times for 5 minutes each with PBS + 0.05% TWEEN-20 (Sigma-Aldrich). Secondary incubation was carried out at room temperature with agitation for 1 hour with 1:5000 goat anti-mouse IgG conjugated to horseradish peroxidase (Thermo Fisher Scientific). The membrane was washed again as described above and then incubated for 5 minutes at room temperature with Clarity Western ECL substrate (Bio-Rad). Amersham Hyperfilm ECL (GE Healthcare, Buckhamshire, UK) was exposed to the membrane and subsequently developed. The blot was then quantified using ImageJ and normalized to the WT-Cas9 sample.

3.4.6 Statistics

All statistical analysis was carried out using GraphPad Prism 6.0.
Chapter 4

Knock-in Blunt Cloning in Human Induced Pluripotent Stem Cells

Jonathan M. Geisinger, Sören Turan, Michele P. Calos

4.1 Introduction

Induced pluripotent stem cells (iPSCs) have the potential to be of great therapeutic use. The greatest advantages of iPSCs are their capability to differentiate to virtually any cell type and that they can be derived directly from a patient making them immunologically matched. However, obtaining a physiologically relevant, patient-matched therapeutic cell often requires genome engineering of the iPSCs.

Genome engineering of human iPSCs (hiPSCs) has been greatly facilitated by the development of methods involving targeted nucleases. Traditional HR has been employed as well, but it is less efficient in general and may not be favorable in cells derived from patients with certain diseases (Zhu et al., 2014). TALENs and ZFNs have been used to successfully modify hiPSCs, but these nucleases do not provide the ease of use that the CRISPR/Cas9 system does. Demonstrating this point, the CRISPR/Cas9 system has been used to generate targeted, multi-kilobase, homozygous gene replacements in hiPSCs through HR mediated by one or two guide RNAs (Byrne et al., 2014).

However, the spectre of off-target cleavage may present significant challenges to the use of CRISPR/Cas9 in therapeutic applications. The guide RNA not only tolerates mismatches, insertions, and deletions throughout its target sequence, particularly outside of the seed sequence, but may also display cell type-specific off-target effects (Fu et al., 2013; Hsu et al.,
2013; Pattanayak et al., 2013; Cho et al., 2014; Lin et al., 2014; Fu et al., 2014; Wu et al., 2014; Kuscu, Arslan et al., 2014; Cencic, Miura et al., 2014). Contrary to these observations, off-target cleavage in murine embryos and human iPSCs and embryonic stem cells (ESCs) has been reported to be extremely rare (Yang, Wang et al., 2013; Veres et al., 2013; Xie et al., 2014; Yang, Grishin, Wang et al., 2014).

To counter off-target cleavage at the cellular level, cells are usually subjected to single-cell cloning. Several cell lines, including murine iPSCs, are amenable to this cloning step, but hiPSCs are not, leading to the necessity of cloning many single cells to obtain a few surviving colonies. This cloning step essentially becomes a bottleneck effect, raising concerns about not only the overall efficiency of the editing event, but also concerns about the integrity of the cells that survived single cell cloning. Including antibiotic resistance in the donor vector is one possible solution, but this inclusion can lead to genetic scars and may not be desirable if minimal perturbation of genome is a goal for the therapeutic cell. Further compounding the difficulty of isolating pure clones is the relatively low efficiency of HDR-based editing in hiPSCs (1% or less; Yang et al., 2013; Miyaoka et al., 2014). Recently, the sib-selection method for isolating rare cell types in yeast genetics has been adapted to isolating gene edited hiPSCs with the purpose of enriching gene editing events, but unfortunately does not address the issue of single-cell cloning (Miyaoka et al., 2014).

Having observed that the CRISPR/Cas9 system could facilitate the precise knock-in of cassettes in an immortalized cell line in Chapter 3, we sought to apply KiBL to human iPSCs (hiPSCs). The successful application of such a method would provide a viable alternative to traditional nuclease-mediated homologous recombination (HR). We also sought to develop a novel single-cell cloning method that could reduce the bottleneck effect typically associated with single-cell cloning of hiPSCs. Here, we demonstrate that KiBL works in hiPSC with an average efficiency of 0.233% and with a high level of precise joining between the genome and the cassette. Additionally, we present preliminary data for novel method of single-cell cloning.
4.2 Results

4.2.1 KiBL results in efficient, precise knock-ins in hiPSCs

To test the efficiency of KiBL in hiPSCs, we chose a line of hiPSCs generated from a healthy donor that we refer to as JF10, using H11 as our target locus with the PAMs Out guide RNAs. The JF10 cells fortuitously possessed two allele-specific SNPs, one flanking each guide RNA target site, which allowed us to determine which allele was targeted for genome editing (Figure 4.1). In order to minimize the effect of untransfected cells on downstream analyses,
we isolated fluorescent-protein+ cells via FACS four to seven days post-electroporation with unphosphorylated pKER cassettes containing phosphorothioate bonds and the guide RNA/Cas9 vector pairs (Figure 4.2). To our surprise, we observed rather low cell viability and relatively low numbers of Clover+ cells when using the WT-Cas9 vectors, whereas we observed greater cell viability and higher numbers of Clover+ cells when using the Cas9-DD vectors in the absence of Shield-1 (Figure 4.3). 0.5 μM and 1 μM of Shield-1 resulted in a moderately increased level of cell viability compared with WT-Cas9. FACS analysis revealed that the Clover+ cells formed a mostly discrete population in these hiPSCs (Figure 4.4). Additionally, the percentage of Clover+ cells was always observed to be less than 0.5% of live cells (Figure 4.4). This low percentage may be due both to the relatively low amount of transfected cassette (200ng) and the lower electroporation efficiency of linear double-stranded DNA. However, we found that there was a significant difference between the mean percentages of Clover+ cells ($P = 0.0193$; ordinary one-way ANOVA), but that only the mean percentages of cassette alone (0.069%) and Cas9-DD + 0.5 μM Shield-1 (0.147%) were significantly different from each other ($P = 0.019$; post hoc Tukey’s multiple comparisons test). In order to analyze precise end-joining at the junctions, we carried out targeted genomic

![Figure 4.3. Comparison of WT-Cas9- and Cas9-DD-mediated KiBL at the H11 locus via fluorescence microscopy. Scale bar = 50 μm.](image-url)
amplification on pooled Clover+ cells by amplifying across the H11 locus in order to capture

**Figure 4.4. Examining KiBL efficiency with Cas9-DD in hiPSCs.** Examples of FACS plots from KiBL using UPT-Clover. Plots are of singlet live cells. Quantification of the percentage of Clover+ cells via FACS for KiBL using UPT-Clover and Cas9-DD. *n* for each condition = at least 2 independent experiments with at least 2 technical replicates each. Each data point represents the mean percentage of Clover+ cells for one experiment. The mean of the experiments is graphed as a horizontal line; the error bars are ± SEM. Data were analyzed using an ordinary one-way ANOVA with a post-hoc Tukey’s multiple comparisons test. * = *P* < 0.02.

all of the KiBL events (Figure 4.5A). Subsequent analysis of the H11 5’- Clover 5’ junction revealed that, in the presence of Shield-1, Cas9-DD-induced DSBs resulted in the precise joining of both the genomic and cassette sequences, but that the absence of Shield-1 resulted in the loss of the PAM and the three bases preceding it on the genomic side, but no loss of bases on the cassette side (Figure 4.5B). Additionally, the use of destabilized Cas9-DD appeared to result in a C-to-T mutation nine bases 5’ of the PAM. We have only observed this mutation with Cas9-DD in the absence of Shield-1. Bulk sequencing of the Clover 3’-H11 3’ junction for stabilized Cas9-DD also showed precise joining of the genome with the cassette (Figure 4.5C). When utilizing stabilized DD-Cas9, we observed precise genomic repair and mostly precise cassette repair, but we also observed the same genomic deletion and C-to-T mutation at the 5’ junction as we had observed with destabilized Cas9-DD (Figure 4.6). When analyzing destabilized DD-Cas9, we did not observe the presence of 5’ H11-5’ cassette junctions (Figure 4.7). These results demonstrated that KiBL is feasible in hiPSCs and that Cas9-DD may be preferable to WT-Cas9 due to increased
Figure 4.5. Sequence analysis of KiBL in hiPSCs with Cas9-DD. (A) Schematic diagram showing targeted amplification strategy of knock-in blunt ligation events at the H11 locus. Full arrowheads denote the primers used to amplify across the locus in the primary PCR. Half arrowheads denote the primers used in conjunction with the full arrowheads to amplify the genome-cassette junctions via nested PCR. (B) Sequence analysis of the 5'-genome 5'-cassette junctions of the KiBL event depicted in panel (A) for various degrees of stability of Cas9-DD. 5' allele-specific SNP is denoted in red. The 5' side of the pKER cassette is denoted in turquoise. Nucleotides in green correspond to Cas9 cleaving at the fourth base upstream of the PAM (underlined) instead of the third. The cytosine colored violet indicates the reference base for JF10 at that position; the thymine in orange indicates a mutation not present in JF10. Dashes indicate unobserved bases. (C) Chromatogram of bulk sequenced 3’pKER-Clover cassette-3’genome junctions resulting from the use of stabilized Cas9-DD. Boxed sequence is the 3’ end of the cassette. Overlined sequence is the expected first three bases of the 3’ sgRNA target site. Underlined sequence is the PAM. Arrowhead indicates the 3’ allele-specific SNP at this locus.

Cell viability. Additionally, these results demonstrated a difference between the stabilized and destabilized forms of Cas9-DD in that stabilized Cas9-DD may result in higher precision KiBL whereas destabilized Cas9-DD may be more useful for probing repair dynamics of Cas9-induced DSBs.

We recognized that there was the possibility that we were repeatedly sampling and sequencing the same clonal population in our initial hiPSC KiBL experiments. We attempted to rule out this possibility in three ways. First, we carried out single-cell PCR across the H11 locus.
utilizing FACS to sort single cells into wells of 96-well plates. From this experiment, we identified two cells that underwent KiBL: one originated from treatment with stabilized Cas9-DD (cell I), the other from destabilized Cas9-DD (cell II) (Figure 4.8). Cell I, interestingly, appeared to have only received the 5' guide RNA/Cas9-DD vector and had precise genomic and cassette junctions for both the 5'-5' and 3'-3' junctions. Sequence analysis revealed that cell II received both guide RNA vectors and possessed a precise 5' genomic junction, but was lacking four bases of the 5' cassette junction. Interestingly, the 3' cassette junction was precise, but the 3' genomic junction appeared to have been cut 16 bases downstream of the actual PAM site for the 3' guide RNA (Figure 4.8). In both cells, the other allele appeared be completely unmodified except for an additional adenosine present at the cleavage site of the 5' guide RNA, potentially as the result of error-prone NHEJ.

To further examine the level of clonality generated by KiBL, we took advantage of the pKER cassettes’ limited spectral overlap and co-transfected 1) pKER-Clover and pKER-mKOκ, and 2) pKER-Clover, -mKOκ, and –mCardinal with our H11 PAMs Out guide RNA pair into JF10 hiPSCs, which were subsequently analyzed via FACS. In the first condition, we observed that Clover+mKOκ- and Clover-mKOκ+ cells were roughly equivalent in relative frequency and that Clover+mKOκ+ cells made up a smaller frequency of the population (Figure 4.9A). Interestingly, in the second condition, we observed all three possible single-positive populations, and no double- or triple-positive populations (Figure 4.9B). These data illustrated that there were at least three clones in each experiment.
Figure 4.7. Destabilized DD-Cas9 does not facilitate KiBL. PCR was performed to analyze the H11 5’-5’ pKER cassette junction and the H11 locus of sorted JF10 hiPSCs transfected with pKER-mKOk cassettes and the H11 PAMs Out sgRNA pair with Cas9-DD or DD-Cas9 in the presence or absence of Shield-1. The primers H11-x1-fwd and mKOk-AS-N2 were used to amplify the H11 5’-5’ cassette junction and the primers H11-X1-fwd and H11-X1-rev were used to amplify the H11 locus. Gel electrophoresis was performed to visualize bands. The expected band size of the H11 5’-5’ mKOk cassette junction is 898 bp. The expected band size of the unmodified H11 locus is 992 bp and the expected size of the KiBL allele is 2233 bp.

Figure 4.8. Single-cell PCR analysis of KiBL. Single-cell PCR-based analysis of two cells from independent transfections. Red denotes allele-specific SNPs and de novo mutations, turquoise denotes cassette sequence, black denotes genomic, green denotes wobble-cleaved bases, violet denotes the wild type SNP for JF10 that differs from reference, dashes denote unobserved sequence, and orange denotes an insertion.
Figure 4.9. **KiBL clonality analysis through multicolor FACS.** (A) Representative FACS plot of singlet live hiPSCs transfected with equal amounts (100 ng each) UPT-Clover and UPT-mKoκ cassettes with stabilized Cas9-DD the H11 PAMs Out sgRNA pair. Data is representative of two independent experiments. (B) Representative FACS plot of singlet live hiPSCs transfected with equal amounts (66.67 ng each) UPT-Clover, -mKoκ, and -mCardinal cassettes with stabilized Cas9-DD the H11 PAMs Out sgRNA pair.

To address clonality at the molecular level, we employed an approach using pKER-Clover cassettes containing 6-base barcodes and 50-base homology arms. Homology arms of this size have been previously used with the CRISPR/Cas9 system to facilitate homologous recombination in human immortalized cell lines (Zheng, Cai et al., 2014) and mouse ESCs (Li et al., 2014). After co-transfection with the H11 PAMs Out sgRNA pairs and Cas9-DD,

Predicted KiBL Junction

| ATAGGCT...57 bp...TAAGTAACTGATGATGTCCTCCGCG | ATAGGCT...57 bp...TAAGTAACTGATGATGTCCTCCGCG |

HR

| ATAGGCT...57 bp...TAAGTAACTGATGATGTCCTCCGCG | ATAGGCT...57 bp...TAAGTAACTGATGATGTCCTCCGCG |

Barcoded KiBL

| ATAGGCT...57 bp...TAAGTAACTGATGATGTCCTCCGCG | ATAGGCT...57 bp...TAAGTAACTGATGATGTCCTCCGCG |

Figure 4.10. **Clonality analysis through barcoded cassettes.** Sequence analysis KiBL events at 5’ H11-5’ pKER junction resulting from the use of stabilized Cas9-DD and barcoded UPT-Clover cassettes possessing short 50 bp homology arms. Red denotes allele-specific SNPs (also indicated by an arrow), violet denotes HR-mediated conversion of a base (also indicated by an arrow), blue denotes a base HR-mediated insertion, orange denotes a common index sequence, green denotes the barcode sequence, turquoise denotes the 5’ cassette sequence, dashes indicated unobserved sequence. PAM is underlined.
followed by isolation with FACS, sequence analysis of the 5’ genomic-5’ Clover junction revealed that HR occurred several times (Figure 4.10). After isolation with FACS, PCR amplification of the 5’ genomic-5’ Clover junction, and subsequent subcloning, we observed 11 different barcodes in 12 sequenced amplicons (Figure 4.10). Out of these, 10 had a thymine (the human reference allele, which we had placed in the 5’ homology arm) instead of cytosine (the allele possessed by JF10) in the position prior to the PAM that differed from reference, indicating that homologous recombination did indeed occur. We also observed that both alleles underwent HR, as indicated by detection of both alleles of our allele-specific T/C SNP on the 5’ side of the locus. Additionally, we observed one non-barcoded KiBL event and a barcoded KiBL event where the cassette appeared to lack at least the 5’ homology arm. We did not observe any KiBL events where the homology arm was incorporated via NHEJ rather than HR. These results indicated that several independent KiBL events occurred in the Clover+ populations, ruling out concerns over clonal amplification affecting our analyses. Additionally, these data demonstrated that PCR amplicons possessing short homology arms combined with the CRISPR/Cas9 system facilitated homologous recombination in hiPSCs, which had not previously been demonstrated.

4.2.2 Analysis of Off-Target KiBL Events

Because KiBL utilizes the CRISPR/Cas9 system, there was a concern that off-target cleavage would result in the uptake of cassettes at sites other than the targeted locus. Indeed, such events have been observed in HEK293T cells using the GUIDE-seq method for unbiased identification of CRISPR/Cas9 off-target cleavage events (Tsai et al., 2014b). We chose to examine the top six off-target sites for both guideRNAs in the H11 PAMs Out pair as determined by the MIT CRISPR Design tool. For each off-target site, we attempted to amplify the 5’ genomic-5’ cassette, 3’ genomic-5’ cassette, 5’ genomic-3’ cassette, and 3’ genomic-3’ cassette junctions, which led to 48 off-target reactions per sample. In order to perform such a large number of assays, we used the multiple displacement amplification variation of whole genome amplification to ensure that there would be enough DNA. We analyzed seven pools of Clover+ cells: one pool treated with WT-Cas9, two pools treated with destabilized Cas9-DD, two pools treated with Cas9-DD and 0.5 μM Shield-1, and two pools treated with Cas9-DD and 1 μM Shield-1. We were unable to detect off-target KiBL events at the vast majority of predicted junctions, with the exception of faint detection of the 3’H11
sgRNA OT4 genomic reverse primer-3’pKER cassette junction (Figure 4.11). These data suggest that, at least in hiPSCs, KiBL occurred predominantly at the on-target locus. It is also possible that the frequency of KiBL off-target events fell below the sensitivity of our assay.

Figure 4.11. Targeted analysis of off-target KiBL events in hiPSCs. Genomic DNA from various CRISPR/Cas9-treated UPT-Clover+ cells was subjected to whole genome amplification and used for analysis of the top six off-target sites (denoted OT1-6) for both the sgRNAs of the H11 PAMs Out pair. Each off-target site was analyzed using a sense (denoted F) and an antisense (denoted R) primer designed to amplify the off-target locus. Each F and R primer were used in two amplifications, one with the 5’ pKER cassette detection primer and the other with the 3’ pKER cassette detection primer. H11 F and H11 R correspond to the primers H11-X1-Fwd and H11-X1-Rev respectively and act as on-target controls. White indicates lack of detection, grey indicates faint detection, and black indicates strong detection. Each Cas9 treatment indicates an independent pool of cells.
4.2.3 HYE-STCC: High-Yield Efficient Single Transmigrant Cell Cloning

Because traditional single-cell cloning could potentially lead to bottlenecking effects and distort the efficiency of genome editing, we sought to develop a more unbiased method of cloning. We reasoned that hiPSCs grow best in the presence of other cells, so to maintain xeno-free conditions, we utilized unmodified JF10 hiPSCs as feeder cells. We initially planned to expand the modified cells in the presence of the feeders for only a short time before isolating the modified cells via FACS, and thus called our method high-yield efficient single transmigrant cell cloning, or HYE-STCC (Figure 4.12). However, we realized that to apply this method for non-fluorescent genome editing, the method would have to be modified, which will be described in the Discussion.

Figure 4.12. Schematic diagram of HYE-STCC.

We thus sought to examine the feasibility of cloning single hiPSCs onto hiPSC feeders. We performed KiBL on a separate population of JF10 hiPSCs at the H11 locus using our unphosphorylated, nuclease-protected pKER-Clover cassette and both destabilized and stabilized (1 µM Shield-1) Cas9-DD. These cells were single-cell cloned via FACS of Clover+ cells onto the feeder JF10 hiPSCs in 48-well plates eight days after electroporation.
and plates were scored after two days for the presence of Clover+ cells. In examining plate-wise survivability, we observed a range from 20.8% to 41.7% with an average of 28.1 ± 4.7% (Figure 4.13A). In examining treatment-wise survivability, we observed an average of 40.2 ± 11.1% for destabilized Cas9-DD and an average of 22.2 ± 3.7% for stabilized Cas9-DD (Figure 4.13B). These preliminary results suggest that, at least, the single-cell cloning step is feasible.

![Figure 4.13. Survivability analysis of Clover+ cells subjected to HYE-STCC. (A) Plate-wise survivability two days after cloning. (B) Treatment-wise survivability of the plates analyzed in (A).](image)

### 4.3 Discussion

These results demonstrate that KiBL can be used to modify hiPSCs in a targeted, homology-independent manner with a high level of precision. Additionally, we demonstrated differences between destabilized and stabilized Cas9-DD and DD-Cas9 at the sequence level. We also illustrated that our analyses were not confounded by the expansion of a clonal population. We observed that short homology arms were also capable of facilitating targeted knock-in of exogenous sequences with the CRISPR/Cas9 system in hiPSCs. We also observed minimal off-target KiBL events, consistent with previous reports of CRISPR/Cas9-based editing in hiPSCs (Veres et al., 2013; Xie et al., 2014; Yang, Grishin, Wang et al., 2014). Lastly, we presented preliminary results of a novel single-cell cloning method for hiPSCs.
Intriguingly, we observed differences between stabilized and destabilized Cas9 in hiPSCs when we did not in HEK293 cells. Our observation that the use of Cas9-DD in the absence of Shield-1 leads to the loss of nucleotides on the genomic side of the genome-cassette junction is particularly interesting. One possibility is that these deletions are the result of a transient binding event by Cas9 as it searches for its programmed target site. Cas9 has been found to probe potential binding sites by finding PAMs (Sternberg, Redding et al., 2014), and unbiased interrogation of Cas9 cleavage sites in HEK293 cells has revealed that SpCas9 can recognize non-canonical PAMs (Tsai et al., 2014b). Additionally, experiments in C. elegans have demonstrated that choosing sgRNAs that target loci enriched with potential PAMs increases successful editing (Farboud and Meyer, 2015), which was motivated by the observation that Cas9 can be sequestered by competitor DNA containing a high density of PAM sequences (Sternberg, Redding et al., 2014). Those findings, coupled with our observations, suggest that there may be several low-affinity binding events at such target sites that can result in cleavage by Cas9. Of further interest is the co-occurrence of a single C-to-T transition upstream of the PAM with the nucleotide loss event. Because neither allele of JF10 possesses a T at this position and we only observed this mutation when using destabilized Cas9, we speculate that this T could be the result of a deamination of a methylated C, which may occur asymmetrically, as we have only observed this occurrence in the T-C-T allele and not the C-C-C allele. Embryonic stem cells are known to possess non-CpG methylation at CpA and CpT sites, which appears to be mediated in part by Dnmt3a (Ramsahoye et al., 2000). Additionally, homologous recombination of direct repeats of GFP facilitated by I-SceI in mouse ESCs has been found to stimulate de novo methylation through Dnmt1 that leads to silencing of the recombined cassette (Cuozzo, Porcellini et al., 2007). These findings combined with the consistent presence of the mutation when using destabilized Cas9-DD lead us to speculate that DSBs induced by destabilized Cas9-DD could be repaired in a different manner than those of WT-Cas9 and stabilized Cas9-DD. Additionally, the lack of knock-in events when using destabilized DD-Cas9 suggests that DD-Cas9 may be preferable for restricting Cas9 activity. This observation is in accordance with the initial work of Banaszynski et al. (2006) describing N-terminal DD fusions as inherently less stable than C-terminal fusions.

Analyzing cells in bulk is informative, but it lacks the resolution only possible at the single-cell (or –clone) level. When working with hiPSCs, single-cell analysis is possible, albeit
rather difficult, as the current method for obtaining pure clones is to single-cell sort cells into numerous 96-well plates and analyze the survivors once colonies have formed. The use of a sib-selection-based strategy for enrichment of modified cells is useful, but this strategy is dependent on subsequent subcloning for isolating a pure population as well (Miyaoka et al., 2014). Thus, in our analyses, we investigated the consequences of KiBL via single-cell PCR. Cell I was found to have only received the 5’sgRNA vector, but the cassette was knocked-in precisely without loss of any nucleotides, underscoring that KiBL is compatible with single sgRNAs as well. Cell II received both sgRNA vectors, and the modified allele possessed a precise 5’ genomic-imprecise 5’ cassette junction and a 3’ precise cassette-imprecise 3’ genomic junction, which is somewhat consistent with destabilized Cas9-DD targeting a nearby incorrect PAM. Interestingly, in both cells, the other allele was wholly correct except for an additional A at the cleavage site of the 5’ sgRNA. It is possible that this additional A is the result of error-prone NHEJ or that there is a small subpopulation of cells in culture that acquired an A at this position undergoing KiBL. Further analysis involving deep sequencing and more single-cell PCR would have to be carried out to determine which actually occurred. It is worth noting that single-cell cloning is the major obstacle to the application of the CRISPR/Cas9 system to hiPSCs. Additionally, our analysis suggests that multi-sgRNA KiBL may benefit from having both sgRNAs present within the same vector, similar to what has been developed for multiplexed editing using lentiviral vectors (Kabadi et al., 2014).

In this work, we chose to identify off-target KiBL events through a targeted, candidate-based method utilizing off-target sites possessing the fewest number of mismatches. While this method provides a good starting point, it does not necessarily take into account the finding that SpCas9 can tolerate bulges in the target DNA and the sgRNA, as well as cleave at non-canonical PAMs and tolerate more than four mismatches (Lin et al., 2014; Tsai et al., 2014b). Whole genome sequencing would identify the off-target KiBL events, in addition to error-prone off-target cleavage repair, but its cost makes it prohibitive for hiPSCs, and the additional information may not be worth the increased cost, as error-prone off-target repair appears not to occur at high frequency in hiPSCs and primary stem cells (Mandal et al., 2014; Yang, Grishin, Wang et al., 2014). There are currently three unbiased methods for identifying off-target cleavage events: GUIDE-seq, Digenome-seq, and identification through integrase-deficient lentiviral vectors (Tsai et al., 2014b; Kim et al., 2015; Wang et al., 2015). GUIDE-seq and identification through integrase-deficient lentiviral vectors are functionally similar, and
it does not escape us that KiBL could be used in a similar approach. Ideally, this would be combined with a linear amplification-based method to maximize sensitivity. Such a strategy has been previously used to examine translocations generated by TALENs and CRISPR/Cas9 and to measure the frequency of promoterless gene targeting using adeno-associated virus vectors (Frock et al., 2014; Barzel et al., 2015).

In developing HYE-STCC, we first carefully considered the sib-selection method described by Miyaoka and colleagues (2014). In sib-selection, the treated cells are split into pools, which are then analyzed for the event of interest and divided until the event is enriched. At this point, single-cell cloning or clump-based cloning is performed to eventually isolate a pure line. Our initial strategy for HYE-STCC would also rely on a similar final subcloning step, which would increase the time needed to obtain a pure clone and would not be compatible with non-fluorescent genome engineering approaches. As such, we postulate combining the sib-selection method with HYE-STCC as a first optimization. This combination permits the initial enrichment of a rare editing event. The second optimization involves generating a line of modified feeder hiPSCs. Through genome engineering, these cells could be modified to possess a negative selection marker, such as thymidine kinase. Thus, when the hiPSCs of interest are single-cell cloned onto the feeders, the feeders can be negatively selected against once the clone has expanded. This optimization would facilitate the generation of a pure clone.

One disadvantage that we observed with KiBL in hiPSCs is the lower electroporation efficiency of linear dsDNA. The limitation of the amount of cassette generated greatly affects directly comparing KiBL to CRISPR/Cas9-mediated HR, because HR uses at least 10-fold more donor vector than our method (≥2 μg versus 200 ng; Byrne et al., 2014). Thus, there is more donor vector available for incorporation in HR relative to KiBL. This drawback can be most easily addressed by increasing the amount of cassette transfected.

KiBL offers a viable alternative to HR and may be a better choice for targeting some cell types. For example, aged hematopoietic stem cells and aged skeletal muscle stem cells retain the capacity to precisely repair DSBs through NHEJ, making KiBL the method of choice for genome engineering in them, particularly in vivo or in their quiescent, i.e., non-dividing, state (Flach et al., 2014; Beerman et al., 2014; Vahidi Ferdousi et al., 2014). In summary, KiBL is
a versatile method capable of facilitating advanced genome engineering strategies and providing new insights into how Cas9-induced DSBs are repaired. Future work involving KiBL in hiPSCs will involve building reporter alleles for differentiation. Of particular interest are the genes involved in muscle differentiation and development, such as \textit{PAX7}, \textit{MYOD1}, and \textit{PAX3}. The generation of such reporters will greatly facilitate high-throughput approaches for directing differentiation to specific lineages and cell types.

4.4 Methods

4.4.1 Knock-In Cassette Construction

Barcoded cassettes containing homology arms were first amplified with primers containing, in the 5’ direction, a 20-base secondary priming sequence, a 6-base randomized barcode, and the priming sequence for the pKER cassette. This reaction was then purified with the GeneJet PCR Purification kit and used as the template for secondary PCR using primers containing 50-base homology arms and the secondary priming sequence, which was purified and digested in the same manner as described above. All primers used for cassette amplification appear in Table A.4.

4.4.2 Cell Culture

JF10 hiPSCs were maintained in 6-well tissue culture plates on recombinant human vitronectin (Life Technologies) in Essential 8 media (Life Technologies). Cells were passaged as clumps every three to four days with Versene (Life Technologies) at a 1:6 split.

4.4.3 Detection of Knock-in Blunt Ligation and Homologous Recombination

JF10 cells were plated three to four days prior to electroporation on vitronectin-coated 6-well plates. Cells were incubated at least one hour beforehand with Essential 8 media containing
10 µM ROCK inhibitor Y-27632. Cells were then dissociated into single cells with Versene. Cells were resuspended in the homemade nucleofector buffer described in Zhang, Vanoli et al., 2014, with 500ng of each sgRNA/Cas9 vector and 200ng of cassette per reaction (for a total of 1.2 µg of DNA) and electroporated with the Amaxa nucleofector using program B-16 in triplicate. Electroporated cells were plated on vitronectin-coated 12-well plates in Essential 8 media containing 10 µM ROCK inhibitor Y-27632 and 0, 0.5, or 1 µM of Shield-1. ROCK inhibitor was added each day for four days after nucleofection. Modified cells were subjected to FACS four to seven days after nucleofection. Briefly, cells were treated 1 hour prior with 10 µM ROCK inhibitor Y-27632 before being trypsinized with TrypLE (Life Technologies) for 10 minutes. Harvested cells were resuspended in DPBS (Life Technologies) with 2% AlbuMAX I (Life Technologies), 2 mM EDTA (Life Technologies), NucBlue Fixed Cell Stain ReadyProbes reagent, and 10 µM ROCK inhibitor Y-27632, and subsequently filtered to remove clumps and debris. Genomic DNA was isolated from sorted cells using the ZymoBead Genomic DNA kit, (Zymo Research, Irvine, CA). gDNA was then subjected to targeted amplification of the H11 locus via PCR amplification with Q5 high-fidelity polymerase or multiple displacement amplification (MDA; a variant of whole genome amplification) following Dean et al., 2002, with the addition of inorganic (yeast) pyrophosphatase (NEB). PCR-amplified products were purified with the MinElute PCR Purification kit (Qiagen), diluted to 100 µL with Buffer EB (Qiagen), and used for further genome-cassette junction PCR amplifications. MDA reactions were diluted to 200 µL with TE-EF buffer (Machrey-Nagel) and also used for further genome-cassette junction PCR amplifications. Primers used appear in Table A.5.

4.4.4 Analysis of Off-Target KiBL Events

For each sgRNA, forward and reverse primers were designed to each of the top six off-target sites as predicted by the MIT CRISPR Design Tool. Each forward and reverse primer was used in conjunction with the 5’ and 3’ pKER-Clover detection primers, for a total of 48 reactions per sample. 1 µL of 1:200 diluted WGA of Clover+ cells was used in each PCR amplification. PCR was carried out with Q5 high-fidelity polymerase. Off-target primers appear in Table A.6.
4.4.5 HYE-STCC

To generate feeder plates, JF10 hiPSCs were treated for at least 1 hour with 10 µM ROCK inhibitor Y-27632 prior dissociation with TrypLE. Cells were centrifuged at 300 x g for 5 minutes and resuspended in Essential 8 media with 10 µM ROCK inhibitor Y-27632. Cells were then plated onto vitronectin-coated 48-well tissue culture plates in a volume of 250 µL of media. For reference, the cells of one well of a 6-well plate was used for each 48-well plate.

After one to two days, the feeder cells were treated with 10 µM ROCK inhibitor Y-27632 for at least one hour. CRISPR/Cas9-treated JF10 hiPSCs were then single-cell cloned onto the feeders using a FACSAria II sorter. Following cloning, plates were briefly centrifuged and placed in a 37°C incubator at atmospheric oxygen levels overnight. The next day, media was replaced with fresh Essential 8 media containing 10 µM ROCK inhibitor Y-27632. Wells were scored for the presence of Clover+ cells the following day.

4.4.6 Statistics

All statistical analysis was carried out using GraphPad Prism 6.0.
Chapter 5

Engineering and Identification of a Multi-Purpose iPSC-Derived Myogenic Progenitor Cell

Jonathan M. Geisinger, Tawny L. Neal, Michele P. Calos

5.1 Introduction

Induced pluripotent stem cells (iPSCs) possess great therapeutic potential through their ability to be differentiated to nearly any cell type in the body. This ability facilitates their use in targeted drug screening as well as in cell-based therapies. iPSCs possess another advantage for both therapeutic approaches in their capacity to theoretically generate large numbers of cells, which is an attractive feature for addressing diseases and disorders of skeletal muscle.

In considering skeletal muscle, the genetic muscle wasting diseases, i.e., the dystrophies, the sarcoglycanopathies, etc., and aging present somewhat similar pathologies. Both are characterized by increased fibrosis and adipogenesis, negatively altered muscle function leading muscle weakness through muscle atrophy and sarcopenia, respectively, and decreased mitochondrial function (Hoffman, Brown, and Kunkel, 1987; Dillon, Rebelo, and Moraes, 2012; Uezumi et al., 2011; Godin, Daussin et al., 2012). Because of these shared pathologies, it is possible that therapy targeting one category may also benefit the other. Indeed, expression of the transcriptional coactivator peroxisome proliferator-activated receptor γ coactivator 1α (PGC-1α), which originally was found to reduce muscle wasting and protect against sarcopenia with age and is considered a master regulator of mitochondrial biogenesis (Sandri et al., 2006; Wenz et al., 2009), has been found to improve dystrophic skeletal muscle in the mdx mouse model of Duchenne muscular dystrophy through increased mitochondrial biomass and a fast-to-slow fiber type shift (Godin, Daussin et al., 2012; Selsby et al., 2012).
Interestingly, the overexpression of dystrophin, mutations in which cause Duchenne muscular dystrophy, rescues the age-dependent decrease in heart ejection fraction (Townsend et al., 2011). Thus, a cellular therapy for aging or genetic muscle diseases has the potential to be multifunctional.

Much of the efforts to derive a cell-based therapy for skeletal muscle from iPSCs has focused on obtaining a myogenic progenitor cell. The approaches taken thus far can be broadly divided into three categories: 1) Small molecule-based and media supplementation-based differentiation, 2) transgene-directed differentiation, and 3) standard embryoid body-based differentiation. In the small molecule/media supplementation category, most differentiation protocols rely on inhibition of GSK3β (Borchin, Chen, and Barberi, 2013; Shelton et al., 2014; Xu, Tabebordbar et al., 2013), but one protocol utilizes EGF and FGF-2 instead (Hosoyama et al., 2014). These methods, while they produce cells suitable for disease modeling, generate little to no engraftment in immunocompromised mice. The transgene-based methods utilize genes from the myogenic differentiation pathway, with the use of Pax7 and MyoD leading to the generation of engraftable myogenic progenitor cells and cell suitable for modeling myopathies in vitro (Darabi et al., 2012; Goudenège et al., 2012; Tanaka et al., 2012). However, the integration of non-therapeutic transgenes in these methods makes them less desirable for an actual therapy. The third category is derived from early experiments demonstrating the expression of myogenic genes through embryoid body differentiation of mouse ESCs (Rohwedel et al., 1994). The protocols in this category have utilized FACS for isolating the myogenic progenitors arising in such a culture, although they have the drawback of relying on additional culturing or using a non-commercially available antibody (Hwang, Suk et al., 2013; Chang et al., 2009; Mizuno, Chang et al., 2010). Regardless of the method of differentiating the cells, characterization of the potential myogenic progenitor cells has relied on intracellular staining of a few transcription factors or limited immunophenotyping through the use of FACS. These characterizations have been based almost entirely on adult markers of myogenic progenitor cells, also known as satellite cells.

In this study, we sought to engineer and identify a murine iPSC-derived multi-purpose myogenic progenitor cell. We chose two separate genes as our therapeutics, dystrophin and PGC-1α, so that we could examine the efficacy of our cells in ameliorating Duchenne muscular dystrophy and aging. To engineer our cells, we used our previously published triple
recombinase system consisting of phiC31 integrase, Bxb1 integrase, and Cre recombinase (Zhao et al., 2014). In order to generate and isolate myogenic progenitor cells, we utilized embryoid body-based differentiation couple with FACS-based isolation. To identify a suitable myogenic progenitor, we used a combination of surface markers more consistent with fetal myogenic cells rather than adult myogenic cells, as we reasoned iPSC-derived cells would be more developmentally immature. These preliminary results not only represent the first steps towards a viable, multi-purpose cellular therapy for skeletal muscle, but also the characterization of baseline myogenic differentiation in murine iPSCs.

5.2 Results

5.2.1 Therapeutic Engineering of Murine iPSCs

We previously generated several iPSC lines from mdx mouse adult and embryonic fibroblasts (Zhao et al., 2014). Each of these lines contained one reprogramming cassette consisting of the mouse Oct4, Sox2, Klf4, and c-Myc genes and EGFP linked via 2A skipping peptides under the control of the CAG promoter. The cassette also contained a Bxb1 AttP site. These cassettes were integrated into the genome through the use of phiC31 integrase. The genomic integration sites were previously identified through linker-mediated PCR. Additionally, we determined the gender of the embryonic fibroblast-derived lines through genomic copy-

![Figure 5.1. Sex Determination of MEF-derived iPSC clones.](image)

Il2rg was used as an X chromosome marker while Kdm5d was used as a Y chromosome marker.
number QPCR (Figure 5.1). We chose two lines for further engineering: the female adult-fibroblast-derived line 1:20#9 and the male embryonic fibroblast-derived line W8-1.

Our previous triple-recombinase strategy relied on selection with puromycin to identify clones successfully retargeted with Bxb1, but identifying successfully Cre-excised clones relied solely on loss of EGFP expression (Zhao et al., 2014). We sought to improve upon this third step through the inclusion of a negatively selectable thymidine kinase cassette in the Bxb1 retargeting step. We constructed therapeutic retargeting vectors containing this negative selection cassette and either mouse dystrophin under the control of the muscle-specific CK6 promoter or mouse PGC-1α linked via a P2A skipping peptide to mCherry under the control of the CAG promoter (Figure 5.2). We transfected 1:20#9 cells with the mouse dystrophin construct and W8-1 cells with the PGC-1α construct along with a Bxb1 integrase expression vector. Following selection with puromycin, we picked surviving colonies and expanded them before transfection with a Cre recombinase expression vector to excise the reprogramming cassette. Post-transfection, we treated the cells with ganciclovir to negatively select against unexcised cells. After selection, we picked colonies, expanded them, and confirmed excision of the reprogramming cassette and retention of the therapeutic cassette via PCR of the LoxP junction and genomic copy number QPCR for EGFP (Figure 5.3). We then referred to the dystrophin iPSC clone as DXA1 and the PGC-1α iPSC clone as PCX2. After

Figure 5.2. Schematics of therapeutic vectors for Bxb1-based retargeting.

We then referred to the dystrophin iPSC clone as DXA1 and the PGC-1α iPSC clone as PCX2.
confirming excision, we validated that the lines were still pluripotent through immunofluorocytochemistry for Oct4, Sox2, SSEA-1, and Nanog (Figure 5.4).

**Figure 5.3. Confirmation of reprogramming cassette excision in PCX cells.** The top panel depicts the result of a PCR amplification bridging the *LoxP* site in the reprogramming vector and the *LoxP* site in the retargeting vector. The presence of a band indicates that these recombined and removed the intervening reprogramming cassette. The bottom panel depicts the results of a genomic copy number QPCR against EGFP. All PCX clones were confirmed to lack EGFP. Similar results were obtained for the Cre-treated 1:20#9 + pKTCmD clones.

<table>
<thead>
<tr>
<th></th>
<th>Oct3/4</th>
<th>Sox2</th>
<th>Nanog</th>
<th>SSEA-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCX2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 5.4. Excised clones retain markers of pluripotency.** iPSC clones that successfully excised the excision cassette still express Oct3/4, Sox2, Nanog, and SSEA-1, suggesting retention of pluripotency and activation of endogenous pluripotency genes. Note that DXA1 appears to express Oct3/4 lower than PCX2. DAPI (blue) stains nuclei.
5.2.2 Identification of a Suitable Myogenic Progenitor Cell

For our differentiation protocol, we utilized a modified version of the protocol of Mizuno, Chang, and colleagues (2010). This embryoid body-based protocol uses three days of hanging drop culture followed by three days of suspension culture before 14-24 days of adherent culture on collagen-coated plates. We chose to isolate our cell of interest via FACS.

To identify a potentially suitable myogenic progenitor cell, we took an informed candidate-based approach. Even though we reasoned that myogenic progenitors derived from iPSCs would be developmentally immature, we first examined what surface markers are used to isolate adult satellite cells. In general, CD31, CD45, and Sca-1 are used to exclude non-satellite cells during FACS-based isolation, so we included these as negative markers in our panel as well (Fukada et al., 2004; Sherwood et al., 2004; Cheung et al., 2012). CD106 (Vcam-1) is used as a reliable positive marker of adult satellite cells in combination with the previous three negative markers, and has been found to be prevalent on fetal myogenic progenitor cells (Cheung et al., 2012; Fukada et al., 2007; Rosen et al., 1992; Biressi et al., 2007). Another reliable negative marker is CD140a (platelet-derived growth factor receptor α), which is not found on adult satellite cells nor fetal myogenic progenitors (Bosnakovski et al., 2008; Yablonka-Reuveni and Seifert, 1993). CD34 has been identified as a marker of quiescent adult satellite cells in mouse, but its absence marks activated satellite cells that can revert back to CD34+ following injury (Bosnakovski et al., 2008; Ieronimakis et al., 2010). We reasoned that the myogenic progenitor cells in our differentiation culture would be proliferative, so we used CD34 as a negative marker. Lastly, we considered CD140b (platelet-derived growth factor receptor β). Fetal myogenic progenitor cells have been reported to be both enriched and depleted for this marker (Yablonka-Reuveni et al., 1992; Biressi et al., 2007). Thus, we decided that further characterization using this marker would be necessary.

Our candidate combination of surface markers was thus CD31- CD45- Sca-1- CD106+ CD140a- CD140b+/- CD34-.

Having defined our markers of interest, we used FACS to isolate different subpopulations from our differentiated cultures of iPSCs (Figure 5.5). Initially, we isolated CD31- CD45- Sca-1- CD106+ and - cells from DXA1 cultures at day 20, 25 and 33 of differentiation to
determine if this population expressed the transcription factor Pax7, a reliable intracellular marker of myogenic stem cells in skeletal muscle, via quantitative real-time PCR (qrtPCR) (Figure 5.6). We observed that the CD106+ fraction was enriched for Pax7 expression above both unsorted cells and the CD106- fraction. Additionally, we found that 20 days of differentiation resulted in the highest levels of Pax7 expression among the timepoints examined. We also observed that the CD106+ population was enriched over unsorted and CD106- cells for Myogenin expression, a myogenic transcription factor more indicative of committed myoblasts (Figure 5.6). We also examined the relative levels of Dystrophin and found a large increase expression in our CD106+ relative to unsorted cells, consistent with expression from the CK6 promoter (Figure 5.6). Interestingly, this level of expression was extremely strong on day 25 amongst all populations analyzed.

**Figure 5.5.** Representative FACS plots of fully stained cells from a 20-day myogenic differentiation. Depicted here are singlet live PCX2 cells. This diagram also displays the sorting order of the markers in our panel.

**Figure 5.6.** CD106+ cells are enriched for myogenic markers early in differentiation. Depicted are the results of qrt-PCR on sorted and unsorted populations. Pax7, a marker of myogenic progenitors is most strongly enriched at day 20 of differentiation. Myogenin (Myog) is also enriched in day-20 CD106+ cells. Dystrophin (Dmd) increases in expression through day 25. DXA1 cells were used.
The previous results confirmed to us that the CD106+ subpopulation of our differentiated cultures contained myogenic progenitors as early as day 20 of differentiation. Additionally, these results demonstrated that the CD106+ contained more differentiated myogenic cells as well. Thus, we began interrogating further subpopulations using CD140a and CD140b. First, we examined the CD140a and CD140b positive and negative subpopulations separately. For DXA1, we examined 6 populations of cells: unsorted, CD106- CD140a-, CD106+ CD140a-, CD106+ CD140a+, CD106+ CD140b-, and CD106+ CD140b+. In this analysis, CD140a and CD140b were examined separately. Upon examination of Pax7 expression, we observed high Pax7 levels in the CD106+ CD140a- and CD106+ CD140b- subpopulations, with CD106+ CD140a- cells possessing a higher level of expression (Figure 5.7). Interestingly, examination of Notch1 expression, the activation of which promotes satellite cell self-renewal in adult skeletal muscle (Conboy and Rando, 2002), revealed a concurrently high level of expression in the same subpopulations as for Pax7, but also decreased expression level in the CD106+ CD140b+ cells relative to CD106+ CD140b- (Figure 5.7). These results suggested that any CD140a+ subpopulation would be depleted for myogenic progenitor cells and that more committed myogenic cells may be present in CD140b+ subpopulations.

**Figure 5.7.** CD106+CD140a- cells are more progenitor-like than other iPSC-derived CD106+ subpopulations. qrt-PCR demonstrates an enrichment for expression of the myogenic progenitor markers Pax7 and Notch1 in the CD106+ CD140a- subpopulation of differentiated DXA1 iPSCs after 20 days.
With the knowledge gained from these findings, we then chose to examine the CD140a and CD140b subpopulations of CD106+ cells together using the PCX2 line. For these experiments, we analyzed unsorted, CD106+ CD140a+ CD140b-, CD106+ CD140a- CD140b-, and CD106- CD140a- CD140b- cells. For Pax7, we found that CD106+ CD140a- CD140b+ cells displayed the highest level of expression (Figure 5.8). Similarly, this subpopulation was enriched for Notch1 expression (Figure 5.8). Intriguingly, these cells were also enriched for Myogenin expression (Figure 5.8). These results suggested that the CD106+ CD140a- CD140b+ cells were actually more differentiated and more similar to myoblasts rather than myogenic progenitors.

Figure 5.8. CD106+ CD140a- CD140b+ cells are potentially more myoblastic than progenitor-like. qrt-PCR reveals the expression of the myogenic progenitor markers Pax7 and Notch1 in the CD106+ CD140a- CD140b+ subpopulation of differentiated PCX2 iPSCs after 20 days, but a relatively high enrichment of Myogenin-expressing cells.

Because of this finding, we chose to examine a subpopulation that we had neglected: CD106+ CD140a- CD140b-. Additionally, we included CD34 in this analysis, but it appeared to be a fairly sparse marker with the vast majority of cells appearing to be negative, so we will refer to these cells as CD34-/low. We first examined Pax7 expression for this subpopulation relative to the CD106- CD34-/low population for DXA1. Through this analysis, we observed
an enrichment of *Pax7* expression in the CD106+ CD34-/low CD140a- CD140b- cells (Figure 5.9). This trend was also observed in PCX2 cells (Figure 5.9). Thus, we decided to use the CD31- CD45- Sca-1- CD106+ CD34-/low CD140a-b- cells as our putative myogenic progenitor cells.

![Figure 5.9. CD106+ CD34-/low CD140a-b- cells are enriched for Pax7 expression. Depicted are the results of qrt-PCR for Pax7 expression after 20 days of differentiation on collagen.](image)

### 5.2.3 Engraftment of iPSC-Derived Myogenic Progenitor Cells

To test the suitability of our myogenic progenitor cells to contribute to muscle regeneration and repair, we first analyzed the engraftment of cells derived from DXA1 in *mdx/SCID* mice, an immunocompromised model of Duchenne muscular dystrophy. For this experiment, we injured the tibialis anterior (TA) muscle of these mice with an injection of 1.2% BaCl₂ three days prior to cell treatment. Only one leg of each mouse was injured, so that the other could be used as a contralateral control. For cell treatments, two mice received ~83,000 putative myogenic progenitor cells each, two mice received 400,000 CD106- CD34- cells each as negative cell control, and one mouse received no cells as an injury-only control. Injections were carried out intramuscularly. 21 days later, the mice were sacrificed and both TA muscles were harvested for sectioning and immunofluorochemistry. At least four distinct regions of each TA were sectioned with at least four consecutive sections per region. All sections were stained for dystrophin and laminin. Following staining, each TA was scored for dystrophin+ fibers, the results of which appear in Figure 5.10. We had planned to compare each treated TA to each other treated TA and each control leg to the contralateral treated TA. These comparisons revealed that the CD106- CD34- cells treatments did not generate significantly
different numbers of dystrophin+ fibers relative to contralateral controls, but that treatment with putative myogenic progenitors did ($P < 0.002$). Additionally, treatment with myogenic progenitors tended to increase the average number of dystrophin+ fibers versus CD106-CD34- cells, which we found to trend towards significant, possibly reflecting mouse-specific differences in revertant fibers (two comparisons involving the same CD106-CD34- treated mouse: $P < 0.0001$; one comparison involving the other CD106-CD34- treated mouse: $P < 0.05$; remaining comparison: $P = 0.0581$). Interestingly, treatment with putative myogenic progenitors increased average dystrophin+ fiber number over that of injury alone ($P > 0.03$). These results suggested that our iPSC-derived putative myogenic progenitor cells were capable of engraftment and contribution to muscle regeneration and repair.

Figure 5.10. iPSC-derived myogenic progenitors contribute to muscle repair. Left panel: TA muscle section demonstrating repair in mdx/SCID 21 days after injection with DXA1-derived myogenic progenitors. White arrows denote dystrophin+ fibers, blue (DAPI) indicates nuclei, green indicates laminin staining, yellow indicates overlapping dystrophin and laminin staining. Right panel: summary of dystrophin+ fibers across all sections analyzed per muscle per mouse.

We then sought to examine the ability of our PCX2-derived progenitor cells to engraft in aged skeletal muscle. For this study, we used seven aged C57BL/6J male mice (~21-23 months old at the start of the study, born and aged at Stanford). This group consisted of three groups
of littermates and one mouse from a separate litter. These mice were the offspring of two sets of parents. The mice were divided into two groups, keeping littermates together, for carrying out

![Figure 5.11. mCherry+ cells are relatively rare in engrafted muscle.](image)

**Figure 5.11. mCherry+ cells are relatively rare in engrafted muscle.** Depicted is a TA muscle section from an aged C57BL/6J male mouse injected 21 days prior with PCX2-derived myogenic progenitor cells. White arrows indicate potentially two mCherry+ cells. Nuclei are blue and the laminin is stained green. These cells appear rarely in across sections.

injections due to limiting numbers of cells. Thus, two sets of injections were carried out one week apart. One TA muscle for each mouse was injured three days before injection with 1.2% BaCl₂. On the day of injection, two mice were injected intramuscularly with 30,000 cells each, whereas the remaining one to two mice received no cells. Where possible, one littermate received cells whereas the other did not. All in all, 4 mice were treated with cells and 3 mice were not. Mice were sacrificed after 21 days and TA muscles were harvested, sectioned in the manner described above, and stained with antibodies against mCherry and laminin. We initially had planned on counting mCherry+ fibers and cells, but examination of the sections revealed relatively few mCherry+ individual cells (Figure 5.11) and very few, if any, mCherry+ fibers. We reasoned that our cells may have still contributed to muscle regeneration, given that the regenerative capacity of endogenous satellite cells declines with age. Thus, we examined the cross sectional area (CSA) of TA fibers for one pair of littermates for both the injured muscle and the contralateral control. We observed no significant difference between mean fiber CSA for the contralateral control TA muscles between these two mice, but we did observe a significant difference between cell treatment versus injury alone (1713 ± 81.13 μm² versus 1228 ± 103.1 μm², P = 0.0004; Figure 5.12). These
preliminary results suggested that our myogenic progenitor cells were contributing to the repair of damaged fibers in aged mouse muscle.

![Graph](image)

**Figure 5.12.** Treatment with PCX2-derived myogenic progenitors results in greater fiber CSA than injury alone. Left panel: mean fiber CSA for each TA muscle of two aged C57BL/6J male littermates, one treated with PCX2-derived myogenic progenitors (positive) and one receiving only injury (negative). LTTA injected/injured in both mice. Right panel: histogram of the CSA of individual fibers of the LTTA for the treated and injured mice.

### 5.3 Discussion

These preliminary results demonstrated that an engineered iPSC-derived cell therapy may be feasible for the treatment of genetic diseases of skeletal muscle and potentially for muscle aging. Additionally, these results illustrated the challenges and benefits of using a candidate-marker-based approach to identifying a suitable target cell type for developing an iPSC-derived cell therapy in the absence of genome-engineering-based screening methods. These results also showed the challenges associated with iPSC-derived therapies in general and in measuring relevant outcomes of such therapies.

When considering how to engineer iPSC-based therapies, it is important to consider what amount of genome disruption would be considered permissible. This study made use of murine iPSC lines chosen following the safe harbor criteria (Sadelain, Papapetrou, and Bushman, 2011). These criteria minimize potentially oncogenic and epigenomic disruption,
and, thus, the targeted insertion of transgenes is permitted under these criteria. The use of the Bxb1 phage integrase to target a therapeutic cassette to the safe harbor site is also favorable, as in our hands, Bxb1 appears incapable of recognizing pseudo-attP sites. Our Cre-mediated removal of the reprogramming cassette is also favorable, as the cells no longer possess the exogenous copies of the reprogramming genes. This removal minimizes the risk of cancer resulting from constitutive expression of these genes. Certainly, it may be more beneficial to target a relatively safe locus directly with programmable nucleases directly, rather than relying on the spectrum of phiC31 pseudo-attP sites in the genome. Indeed, this approach was taken in generating the landing pad for dual integrase cassette exchange, a method that uses nuclease-mediated homologous recombination to place a landing pad, which is then exchanged for a cassette of interest through simultaneous use of the phiC31 and Bxb1 integrases (Zhu et al., 2014).

In this work, we demonstrated that it is possible to use a candidate-based approach for identifying a cell type of interest solely from embryoid body-based differentiation. Because we were interested in identifying a pure a population of myogenic progenitor cells as possible, we followed a testing strategy similar to the design-build-test strategy used in synthetic biology. This strategy led us to identify candidate surface markers, purify cells with those markers, and examine these cells for expression of intracellular markers via qrt-PCR before undertaking another differentiation experiment to further subdivide the previous population in search of generating the most pure population we could. This strategy was time-consuming and limited in its throughput. It certainly would not have been feasible if we had chosen to carry out a screen for directed differentiation. This strategy in general would benefit from the use of reporter alleles. Such alleles could consist of fluorescent proteins linked via skipping peptides to Pax7 and Myogenin. These alleles could be generated through the CRISPR/Cas9 system via KiBL. Having a cell line possessing multiple reporter alleles would greatly decrease the amount of time spent differentiating cells and would be amenable to high-throughput approaches for directed differentiation.

One of the major challenges of cell-based therapies is obtaining a large enough number of cells for efficacious treatment. The use of iPSCs addresses this challenge somewhat as one is no longer limited by obtaining a large enough number of cells, but raises the new challenge of obtaining that cell number efficiently. Our differentiation strategy relied on identifying what
myogenic progenitors would be produced without the addition of exogenous factors. While this strategy enabled us to obtain the cells we desired, obtaining large numbers of cells meant setting up large numbers of differentiation cultures, and, by extension, sorting large numbers of cells, which is not very efficient and could be detrimental to obtaining an efficacious number of cells. Conversely, the directed muscle differentiation protocols generate a large number of cells that are positive for one or two specific markers very efficiently, but, puzzlingly, generally appear to have little ability to engraft even into immunocompromised muscle (specifically see Xu, Tabebordbar et al., 2013). One interpretation is that iPSC-derived myogenic progenitor cells are not as good at engraftment as bona fide satellite cells. Another more likely interpretation is that the use of small molecules, proteins, and transgenes to direct differentiation of iPSCs to myogenic progenitor cells leads to a rather large percentage of false positive cells in the population, i.e., cells that superficially resemble myogenic progenitors, but do not actually possess the ability to behave as such in a cell therapy scenario. In such a scenario, the population of true myogenic progenitors may not actually be that different in number from the number generated through undirected, embryoid body-based differentiation. Future work in the field of iPSC differentiation should address this possibility.

In this study, we demonstrated that our iPSC-derived myogenic progenitor cells were capable of engraftment in both dystrophic and aged skeletal muscle. In dystrophic muscle, however, we did not detect a particularly high level of engraftment as measured by the number dystrophin-positive fibers. This low level of detection may indicate that our cells are not as potent as satellite cells or that dystrophin expression was limited to the vicinity of the engrafted cell and was simply not included in the analyzed sections. When we injected our PCX2-derived myogenic progenitor cells into aged skeletal muscle, we observed relatively little mCherry staining at both the fiber and the mononuclear cell level. This low level of staining at the fiber level may be due to the amount of mCherry produced by the PCX2 cells being too low to be detected in something as large as a muscle fiber. At the mononuclear cell level, our PCX2-derived cells may be outcompeted for the satellite niche by the endogenous satellite cells, thus leading to the rarity of mCherry-positive cells. If we had only relied on mCherry, we would have concluded that the cells failed to have any effect. Fortunately, we examined fiber CSA and determined that treatment with our myogenic progenitor cells lead to larger fibers on average compared with injury alone. These results are preliminary, but they
warrant further investigation into whether these cells are capable of rejuvenating aged muscle. Additionally, it is worth examining if the PCX2-derived cells can have a positive benefit on dystrophic muscle as well.

5.4 Methods

5.4.1 Cell Culture

iPSCs were maintained on tissue culture plates coated with 0.2% gelatin and γ-irradiated CF1 mouse embryonic fibroblast feeder cells in high-glucose DMEM supplemented with 20% FBS, 1x non-essential amino acids, 1x GlutaMAX, 0.1mM β-mercaptoethanol, and 10 ng/mL mouse LIF (Millipore). Cells were passaged with 0.05% Trypsin-EDTA every 4 to 5 days at a 1:12 split.

5.4.2 Retargeting via Bxb1 and Cre Excision

For nucleofections, iPSCs were first preplated on gelatin-coated plates for 30 minutes to remove feeder cells. For each nucleofection, 2 million iPSCs were transfected using the Mouse ES Cell Nucleofection Kit (Lonza) using program A-30 with 1.5 or 3 µg of pCSBxb1 (a Bxb1 integrase-expressing plasmid) and 1.5 or 3 µg of either pKTCmd or pKTCfimPCR. Nucleofected cells were plated on DR4 feeder cells and allowed to recover for 2 days. Puromycin selection was carried out for 6 days at a concentration of 1 µg/mL. Resistant colonies were picked and expanded before being subjected to nucleofection as described above with 3 µg of a Cre-expressing vector. 2 days post nucleofection, cells were subjected to negative selection with 10 µM ganciclovir for 6 days. GFP-negative colonies were picked and expanded before being subjected to LoxP junction analysis via PCR. Briefly, 150 ng of genomic DNA was used as a template for nested PCR using HotStarTaq Plus. The primers used in the first round were 5'-GCGGGGGTTCGTAAGCCGTCAG*C*C-3' and 5'-CCACCCACCCTGCCCACTGAG*G-3'. The second round primers were 5'-GCGGCGCTCGAGAAGCTTTAG-3' and 5'-GCCCGACCCTCCCTCTGCAAC-3'.
5.4.3 Myogenic Differentiation and FACS

Undifferentiated cells were preplated for at least 30 minutes to remove feeder cells before being resuspended in differentiation media consisting of high-glucose DMEM supplemented with 10% FBS, 5% horse serum, 1x non-essential amino acids, 1x GlutaMAX, and 0.1mM β-mercaptoethanol. Cells were then subjected to three days in hanging drop culture (1000 cells/20 µL), followed by 3 days in suspension culture on tissue plates treated with 20 mg/mL poly-HEMA (Sigma-Aldrich). The resulting embryoid bodies were then transferred to collagen-coated tissue-culture dishes (Thermo-Fisher Scientific) with media changes every 3-5 days. On days of interest, cells were dissociated with TrypLE and centrifuged at 300 x g with a 100% FBS underlay to remove dead cells. Cells were then resuspended in Ham’s F10 (Thermo-Fisher Scientific) with 10% horse serum and filtered through a 70 µm cell strainer. Cells were blocked with mouse BD Fc Block (BD Biosciences) before staining with the following primary antibodies for FACS: PerCP-Cy5.5 rat anti-mouse CD31 (1 µg/mL; BD Biosciences), PerCP-Cy5.5 rat anti-mouse CD45 (1 µg/mL; BD Biosciences), BV605 rat anti-mouse Sca-1 (1 µg/mL; BD Biosciences), FITC rat anti-mouse CD34 (3 µg/mL; BD Biosciences), PE-Cy7 rat anti-mouse CD106 (2.5 µg/mL; BioLegend), APC rat anti-mouse CD140a (3 µg/mL; BD Biosciences), and Biotin rat anti-mouse CD140b (3 µg/mL; BioLegend). Cells were stained for 45 minutes at 4°C on a nutator. Secondary staining consisted of SAv-APC-Cy7 (3 µg/mL; BD Biosciences) for 15 minutes at 4°C on a nutator. Between staining, cells were washed with an excess of 1x PBS and centrifuged at 300 x g at 4°C for 5 minutes. Following secondary staining and washing, cells were resuspended at a concentration of 20 million/mL in Ham’s F10 with 10% horse serum and NucBlue Fixed Cell Stain ReadyProbes reagent. Cells were sorted on FACSariaII-class sorters. The fluorescence minus one method was used to set gates. Post-sort analysis was carried out using FlowJo software.

5.4.4 Genomic Copy Number QPCR and qrt-PCR

Genomic copy number QPCR was carried out using 20 ng of genomic DNA per reaction. Each sample was carried out in quadruplicate. The 2x TaqMan Genotyping Master Mix (Life Technologies) was used with the TaqMan copy number reference assay for Trfc (Life Technologies) as a reference control. For the examination of iPSC clone gender, the Iil2rg
TaqMan copy number assay was used to detect the X chromosome and the Kdm5d TaqMan copy number assay was to detect the Y chromosome. For analysis of reprogramming cassette presence, the EGFP TaqMan copy number assay was used to detect the reprogramming cassette.

For qrt-PCR, mRNA was isolated from cells using the RNEasy mini kit (Qiagen). cDNA was generated using SuperScript III first strand synthesis kit (Life Technologies) using random hexamers according to the manufacturer’s instructions. 1 µL of cDNA synthesis reaction was used per qrt-PCR reaction. All reactions were carried out in quadruplicate using TaqMan Universal Master Mix II with UNG (Life Technologies). The mouse Gapdh TaqMan expression reference assay (Life Technologies) was used as a loading control. TaqMan expression assays (Life Technologies) for Pax7, Myogenin, Notch1, and Dystrophin were used for analysis.

All reactions were carried out on a CF1000 thermal cycler equipped with a CFX96 real-time system (Bio-Rad).

5.4.5 Engraftment of iPSC-Derived Myogenic Progenitor Cells

Animal maintenance, surgical procedures and husbandry were carried out at the Veterinary Service Center of Stanford University. All procedures were approved by the Institutional Animal Care and Use Committee. Three days prior to engraftment, the left TA muscles of 8-week-old male \textit{mdx}/\textit{SCID} mice or 21-23 month old C57BL/6J mice were injected with 1.2% BaCl\textsubscript{2}. On the day of engraftment, myogenic progenitor cells were isolated by FACS, aliquoted and resuspended in 30 µL of Ham’s F10 + 10% horse serum, loaded into an insulin syringe (BD Biosciences), and injected into the left TA in three sites for 10 µL per site. 21 days post-engraftment, TA muscles were harvested, fixed in 5% paraformaldehyde for 4 hours, dehydrated in 20% sucrose overnight and frozen in OCT the following day using liquid nitrogen-cooled methyl-butane.
5.4.6 Immunofluorochemistry

Reprogramming cassette-excised iPSCs were plated on 4-well chamber slides (Millipore) that had been coated with 0.2% gelatin and γ-irradiated CF1 feeders. After two days, slides were washed with cold PBS and fixed with 4% paraformaldehyde for 10 minutes. Slides were washed and rehydrated with PBS before blocking and permeabilization with PBS-BT (PBS containing 5% bovine serum albumin (Sigma-Aldrich) and 0.02% Tween-20 (Sigma-Aldrich)) for 30 minutes. The following primary antibodies were incubated overnight at 4°C: anti-Oct3/4 (1:200, Santa Cruz Biotechnology), anti-SSEA-1 (1:100, Santa Cruz Biotechnology), anti-Sox2 (1:200, Rockland) and anti-Nanog (1:200, Rockland). The following day, cells were washed 3 times with PBS-BT, and incubated for 1 hour at room temperature with the following antibodies: AlexaFluor 488 donkey anti-rabbit, AlexaFluor 488 goat anti-mouse, and NucBlue Fixed Cell Stain ReadyProbes reagent (all Life Technologies). Cells were washed 3 times with PBS-BT before treatment with FluoroGel and were sealed with a coverslip.

TA muscles embedded in OCT blocks were cryosectioned and processed for immunofluorohistochemical analysis with a 10 minute fixation with 4% paraformaldehyde followed by washing with PBS and blocking with PBS-BT for 20 minutes. Sections were incubated overnight at 4°C with the following primary antibodies in PBS-BT: rat anti-laminin (1:1000, Sigma-Aldrich) and either rabbit anti-dystrophin (1:100, Abcam) or goat anti-mCherry (1:100, Acris). The following day, sections were washed three times with PBS-BT and were incubated for 1 hour at room temperature with the following secondary antibodies: AlexaFluor 488 goat or donkey anti-rat (1:500, Life Technologies), AlexaFluor 594 donkey anti-rabbit or anti-goat (1:500, Life Technologies), and NucBlue Fixed Cell Stain ReadyProbes reagent. Cells were then washed 3 times with PBS-BT and coverslips were mounted through FluoroGel.

Fiber CSA was measured using ImageJ analysis software.
5.4.7 Statistics

All statistical analysis was carried out using GraphPad Prism 6.0. Planned comparisons for the number of dystrophin+ fibers used the Student’s t test with $\alpha = 0.05$. All comparisons involving fiber CSA in aged mice used Student’s t test with $\alpha = 0.05$, as well.
Appendix

Table A.1. Interrogated regions used for design of sgRNAs. The GRCh37/hg19 assembly was used for human regions and the GRC38/mm10 assembly was used for murine regions.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Region #</th>
<th>Chromosomal Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYOD1</td>
<td>1</td>
<td>Hs chr11: 17742664-913</td>
</tr>
<tr>
<td>MYOD1</td>
<td>2</td>
<td>Hs chr11: 17743679-928</td>
</tr>
<tr>
<td>PAX3</td>
<td>1</td>
<td>Hs chr2: 223066487-736</td>
</tr>
<tr>
<td>PAX7</td>
<td>1</td>
<td>Hs chr1: 19062359-608</td>
</tr>
<tr>
<td>PAX7</td>
<td>2</td>
<td>Hs chr1: 19062633-845</td>
</tr>
<tr>
<td>PAX7</td>
<td>3</td>
<td>Hs chr1: 19061876-2125</td>
</tr>
<tr>
<td>mUrtn</td>
<td>2</td>
<td>Mm chr10: 12381712-961</td>
</tr>
<tr>
<td>mUrtn</td>
<td>4</td>
<td>Mm chr10: 12384485-734</td>
</tr>
<tr>
<td>H11 locus</td>
<td>1</td>
<td>Hs chr22: 31830203-452</td>
</tr>
<tr>
<td>H11 locus</td>
<td>2</td>
<td>Hs chr22: 31830847-1096</td>
</tr>
</tbody>
</table>
Table A.2. **sgRNAs and cloning oligos used.** All sequences are in the 5’ to 3’ direction and relative to the sense strand where appropriate. Dashes separate PAM from target. Lower case indicates overhang sequence for cloning.

<table>
<thead>
<tr>
<th>sgRNA Name</th>
<th>Sequence</th>
<th>Sense oligo</th>
<th>Antisense oligo</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYOD-r1-1</td>
<td>CCT-</td>
<td>caccgGGAGGGCTCGGGTATGTAACG</td>
<td>aaaGCTTTACTACGGGAGCC</td>
</tr>
<tr>
<td></td>
<td>GCTTACCTAACG</td>
<td></td>
<td>CTCCc</td>
</tr>
<tr>
<td></td>
<td>AGCCCTCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYOD-r1-2</td>
<td>GTCTCAAAGTAC</td>
<td>accgGTCTCAAGTACTGGGCCC</td>
<td>aaaCGGGGCAGTACTTTG</td>
</tr>
<tr>
<td></td>
<td>TGGGCAGGGG</td>
<td></td>
<td>AGACc</td>
</tr>
<tr>
<td>MYOD-r2-1</td>
<td>TACATTACTACAC</td>
<td>caccgTAACATTACTACACACGTCG</td>
<td>aaaCGGAGCTGAGTGAGTA</td>
</tr>
<tr>
<td></td>
<td>AGCTCCG-TGG</td>
<td></td>
<td>ATGTAac</td>
</tr>
<tr>
<td>MYOD-r2-2</td>
<td>CCC-</td>
<td>caccgCCGCTTAGGCTCATGGAACGTACG</td>
<td>aaaTCAGCGACGCTC</td>
</tr>
<tr>
<td></td>
<td>TCAGCGACGCTT</td>
<td></td>
<td>GCGGGe</td>
</tr>
<tr>
<td></td>
<td>GTAGGCGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAX3-1</td>
<td>CCA-</td>
<td>caccgCCTGTACAGGCTACCAATA</td>
<td>aaaTATTGAGCTAGCTATG</td>
</tr>
<tr>
<td></td>
<td>TATTGATAGGTG</td>
<td></td>
<td>CAGGe</td>
</tr>
<tr>
<td></td>
<td>TGGACGGGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAX3-2</td>
<td>AAAATTTGATACC</td>
<td>caccgAAAATTTGATACGGCAGGTATG</td>
<td>aaaCACATGCGGAGTAC</td>
</tr>
<tr>
<td></td>
<td>GGATGATGG-TGG</td>
<td></td>
<td>ATTTTc</td>
</tr>
<tr>
<td>PAX3-3</td>
<td>CCT-</td>
<td>caccgTACAGTGATGGACCTCCTGAC</td>
<td>aaaGTCAGAGGCTCATA</td>
</tr>
<tr>
<td></td>
<td>GTGACGGTGTC</td>
<td></td>
<td>CTGTAc</td>
</tr>
<tr>
<td></td>
<td>ATACGGTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAX3-4</td>
<td>TTACACAAGGAA</td>
<td>caccgTTACACAAGGACAGCCCTG</td>
<td>aaaGCAGGAGGCTCCTTGT</td>
</tr>
<tr>
<td></td>
<td>GGGCCTGC-TGG</td>
<td></td>
<td>GTAAc</td>
</tr>
<tr>
<td>PAX7r1-1</td>
<td>AAGCCTCTGCTTC</td>
<td>caccgAAGCCTAGCTGCTGACTTA</td>
<td>aaaTAGTACGGAAGCAG</td>
</tr>
<tr>
<td></td>
<td>CGTACCTGGAGGTG</td>
<td></td>
<td>GGGTCat</td>
</tr>
<tr>
<td>PAX7r1-2</td>
<td>CCT-</td>
<td>caccgAATGGGAGCAGGGAGGACG</td>
<td>aaaGGGCGTCCCCCGTCCC</td>
</tr>
<tr>
<td></td>
<td>GGGCGCTCCCCCG</td>
<td></td>
<td>CATTC</td>
</tr>
<tr>
<td>PAX7r1-3</td>
<td>CCT-</td>
<td>caccgGAGCCATAGTACGGAAGCAG</td>
<td>aaaCGTTCTCCCCTGACTATG</td>
</tr>
<tr>
<td></td>
<td>CTGTTTCTACGCTACT</td>
<td></td>
<td>GCTTC</td>
</tr>
<tr>
<td>PAX7r1-4</td>
<td>CCT-</td>
<td>caccgGTAGCTTCCATAAACAGA</td>
<td>aaaGCTTTATTGTAGAG</td>
</tr>
</tbody>
</table>

81
<table>
<thead>
<tr>
<th>Region</th>
<th>Primer A</th>
<th>Primer B</th>
<th>Primer C</th>
<th>Primer D</th>
<th>Primer E</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAX7r2-1</td>
<td>CAACAGAATACG</td>
<td>caaccGAAACGAATCACCTTTGAGCCC</td>
<td>aaaacGGGCTCAAAGTGATTC</td>
<td>GCTACe</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TTGAGCCC-AGG</td>
<td></td>
<td></td>
<td></td>
<td>GTTTGe</td>
</tr>
<tr>
<td>PAX7r3-2</td>
<td>CCC-</td>
<td>caaccGATTTCGTCGCCCTCATGCAT</td>
<td>aaaacATGCAATGAGGCGACG</td>
<td>CAAATc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATGCATGAGGCG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGGCAAAAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mUtrn-r2-1</td>
<td>CCC-</td>
<td>caaccGTACATTTCTAAAGCCCATT</td>
<td>aaaacATGGGCTTTAGGAA</td>
<td>TGTAc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AATGGGCTTTAG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAATGTCAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mUtrn-r2-3</td>
<td>ATGAAGGCCCAAT</td>
<td>caaccATGAAGGCCCATTGGCTTTT</td>
<td>aaaacAAAGCCCATTGGGCC</td>
<td>TTCATc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGGGTCTTTAG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mUtrn-r4-1</td>
<td>CCA-</td>
<td>caaccTAGCCCGATTTAGACCCCAT</td>
<td>aaaacATGGGTCTAAATACG</td>
<td>GGCTAc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATGGGTCTAAAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACGGCTA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mUtrn-r4-3</td>
<td>ATGCAATGGGTG</td>
<td>caaccATGCAATGGGTCTTTTAAATAC</td>
<td>aaaacGTATTTAGACCCATTG</td>
<td>GCATc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTAAATACGGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mUtrn-r4-3</td>
<td>ATGCAATGGGTG</td>
<td>caaccATGCAATGGGTCTTTTAAATAC</td>
<td>aaaacGTATTTAGACCCATTG</td>
<td>GCATc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTAAATACGGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H11-r1-1</td>
<td>TTAATTGATCAGT</td>
<td>caaccTTAAATTGATCACTCCCGCA</td>
<td>aaaacTGCGGGGAGTGATCAGCAT</td>
<td>ATTAAc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCCGCA-AGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H11-r1-2</td>
<td>CCC-</td>
<td>caaccCAAATAGGTTAGCCTTTGCG</td>
<td>aaaacCGCAAGGCTAACCTA</td>
<td>TTTTGe</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGCAAGGCCTAAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTATTTTGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H11-r2-1</td>
<td>CCT-</td>
<td>caaccGAGAATTCGCTTGAACCCGGG</td>
<td>aaaacCCCGGGTTCAAGCGA</td>
<td>TCTCc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCCGGGTCAAG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGATTTCTC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H11-r2-3</td>
<td>GTGTCAGCCTTCCT</td>
<td>caaccGTGTCAGCCTTCCTGATTAGC</td>
<td>aaaacGCTAATCAGGAGGCT</td>
<td>GACAc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GATTAGC-TGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table A.3. **Primers used in cassette amplification**. All oligos are in the 5’ to 3’ direction.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>DICE-EP-v2-PCR-fwd</td>
<td>/5phos/-GAAGAA-CTCGAGGTAGTGCCCCAACTGGGGTAACTTTTGAGTTCTCTCACAGTTGGGCGGCT- GCTAGCTACCCGGGTAGGGGAGGC</td>
</tr>
<tr>
<td>DICE-EP-v2-PCR-rev</td>
<td>/5phos/-GAAGAA-GGTACCGGGTTTGTACCGTACCCACTGAGACCGCGTGGTTGACCAGACAAACCA-CGCAGAGAGAGACAGCTATG</td>
</tr>
<tr>
<td>PTKPmChR-PCR-fwd</td>
<td>/5phos/-TACCGGGTAGGGGAGGCGCCTTCTT</td>
</tr>
<tr>
<td>PTKPmChR-PCR-rev</td>
<td>/5phos/-GAGAAGAGGGACAGCTATGAGGGGAGTAGT</td>
</tr>
<tr>
<td>phospho-pKER-fwd</td>
<td>/5phos/-GATCTGCGATCGCTCCGGTG</td>
</tr>
<tr>
<td>phospho-pKER-rev</td>
<td>/5phos/-GAGAAGAGGGACAGCTATGACTGGG</td>
</tr>
<tr>
<td>pKER-fwd</td>
<td>gatctcgccagtcccggtg</td>
</tr>
<tr>
<td>pKER-rev</td>
<td>GAGAAGAGGGACAGCTATGACTGGG</td>
</tr>
<tr>
<td>UPT-pKER-fwd</td>
<td>G<em>A</em>T*CTCGGATCGCTCCGGTG</td>
</tr>
<tr>
<td>UPT-pKER-rev</td>
<td>G<em>A</em>G*AAGAGGACAGCTATGACTGGG</td>
</tr>
<tr>
<td>BarUniv-pKER-fwd</td>
<td>AATGATACGGGCAGCGATGT-NNNNNN-GATCTGCGATCGCTCCGGTG</td>
</tr>
<tr>
<td>BarUniv-pKER-rev</td>
<td>CTGCCGACGGAGCATGAAGTCGCACTATGAGCTGGG</td>
</tr>
<tr>
<td>UPT-H11PO-</td>
<td>G<em>A</em>T*ACTGGAGAGGAAGGACTTTATGTAAGTTATATGACTCCTCCGCG-</td>
</tr>
<tr>
<td>BarCommon-fwd</td>
<td>AATGATACGGGGACAGCTATGAGCTGGT</td>
</tr>
<tr>
<td>UPT-H11PO-</td>
<td>A<em>A</em>A*AATACAAAATTAGCGGGCATGGTGGTACCTGCCTGAATCCCAGCT-</td>
</tr>
<tr>
<td>BarCommon-rev</td>
<td>CTGCCGACGAGTCTAGTCTT</td>
</tr>
</tbody>
</table>
Table A.4. Excision detection primers. All oligos are in the 5’ to 3’ direction.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYOD1-X1-fwd</td>
<td>CTTTTGGATTGTCCCCGGACTCTA</td>
</tr>
<tr>
<td>MYOD1-X1-rev</td>
<td>CACATCCTCTATACCAACGAGG</td>
</tr>
<tr>
<td>PAX3-X1-fwd</td>
<td>GAAAGGCACCTGTAAGGAAACAC</td>
</tr>
<tr>
<td>PAX3-X1-rev</td>
<td>TGCAATCTGGATCTCTCTGCATT</td>
</tr>
<tr>
<td>PAX7-X1-fwd</td>
<td>AGGGCAGCAAAATCAGGTAA</td>
</tr>
<tr>
<td>PAX7-X1-rev</td>
<td>AAATGTGCCCAGTAAATGTTAAG</td>
</tr>
<tr>
<td>PAX7-X2-fwd</td>
<td>CAAGTGAACATCCACGTAACGTAAC</td>
</tr>
<tr>
<td>PAX7-X2-rev</td>
<td>GAGAGAGTGGAAACACAGCCTTTC</td>
</tr>
<tr>
<td>mUtrn-X2-fwd</td>
<td>GCACCAATGAAACGTAATGAGAG</td>
</tr>
<tr>
<td>mUtrn-X2-rev</td>
<td>TTTAGCCCAGTAGCTTTTCAGTTTG</td>
</tr>
<tr>
<td>H11-X1-fwd</td>
<td>CAGAATTTGTTTGAAACGATGGCT</td>
</tr>
<tr>
<td>H11-X1-rev</td>
<td>TCATGGGATTCAAGCAGAAAGT</td>
</tr>
</tbody>
</table>
Table A.5. Primers for Amplifying Cassette Junctions. All oligos are in the 5’ to 3’ direction and are to be used in conjunction with the genomic primers in Table A.4.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>DICE-EP-detection-mch-fwd</td>
<td>GACATCCCGACTACTTGAAGC</td>
</tr>
<tr>
<td>PGK-rev1</td>
<td>AAAGAACGGAGCCGGTTGG</td>
</tr>
<tr>
<td>pKER-detector-5’</td>
<td>TACACGACATCACTTTCCCAATTTA</td>
</tr>
<tr>
<td>pKER-clover-detector-3’</td>
<td>CAACCACCTACCTAGCCACATAGTC</td>
</tr>
<tr>
<td>pKER-detector-5’-NP</td>
<td>TACACGACATCACTTTCCAGTTTA</td>
</tr>
<tr>
<td>pKER-clover-detector-3’-NP</td>
<td>TAACCACCTACCTAGCCATAGTC</td>
</tr>
<tr>
<td>mKOk-AS-N2</td>
<td>CACCTTCAATTGTAATCATCGGCTT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site Name</th>
<th>Off-target Sequence</th>
<th>Chromosomal Location</th>
<th>F Primer</th>
<th>R Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>H11-r1-2-OT1</td>
<td>CAAAACAGGGTTAG</td>
<td>chr1: 1-4565443</td>
<td>TCACCTACCCAGAA</td>
<td>GCTGTACCCTCCCTCA</td>
</tr>
<tr>
<td></td>
<td>CCTTGCTAGG</td>
<td></td>
<td>TGTTATCCAGT</td>
<td>TTCC</td>
</tr>
<tr>
<td>H11-r1-2-OT2</td>
<td>CAAAATGTGGTAG</td>
<td>chr14: 1-95753361</td>
<td>GAAAGTGACGGGG</td>
<td>CACGGTAGATCGTCA</td>
</tr>
<tr>
<td></td>
<td>CCTTGCGGGG</td>
<td></td>
<td>ACAGTTCC</td>
<td>GGGA</td>
</tr>
<tr>
<td>H11-r1-2-OT3</td>
<td>AAATATAGGGTTAG</td>
<td>chr8: 1-15580041</td>
<td>ACAATTGCTATCCG</td>
<td>TTAAGAGTGGCTCATCAA</td>
</tr>
<tr>
<td></td>
<td>CCTTGCGGGG</td>
<td></td>
<td>TTCAATTGCTCTCTCA</td>
<td>TTTAGCTGTCC</td>
</tr>
<tr>
<td>H11-r1-2-OT4</td>
<td>GAAACTAGGGTTAG</td>
<td>chr9: 1-1961769</td>
<td>GTGAGGTACTGTCT</td>
<td>GTTATTTTTGGATGCG</td>
</tr>
<tr>
<td></td>
<td>CCATGCGTTG</td>
<td></td>
<td>GGAACTAGG</td>
<td>AGTTTTCCTCC</td>
</tr>
<tr>
<td>H11-r1-2-OT5</td>
<td>CAAAATAGGGTTAG</td>
<td>chr3: 1-54074337</td>
<td>CGAGAGCTATTTTC</td>
<td>TCCAAGGATGGAAG</td>
</tr>
<tr>
<td></td>
<td>CCTTGAGGTGG</td>
<td></td>
<td>TGTTATTTCTCC</td>
<td>AACCCCTATGCC</td>
</tr>
<tr>
<td>H11-r1-2-OT6</td>
<td>TCAAATAGGGTATG</td>
<td>chr10: 1-75935789</td>
<td>TCATAAAAGCATTATA</td>
<td>CGGCAGCACGACGTAC</td>
</tr>
<tr>
<td></td>
<td>CCTTGAGTGG</td>
<td></td>
<td>CTGATTCCCCC</td>
<td>TAAT</td>
</tr>
<tr>
<td>H11-r2-3-OT1</td>
<td>CTCTCAGCTTCTCTG</td>
<td>chr1: 1-7585363</td>
<td>TCCTGCTTTTTTCA</td>
<td>GAATCTGAAGCTCCCCCTT</td>
</tr>
<tr>
<td></td>
<td>ATTAGCTGG</td>
<td></td>
<td>GAGTGTGC</td>
<td>TCTCAG</td>
</tr>
<tr>
<td>H11-r2-3-OT2</td>
<td>CTGACAGCTCTTCT</td>
<td>chr2: 1-143987130</td>
<td>AGGAGGCCCTACA</td>
<td>CAACTAGTGCGCTGGAG</td>
</tr>
<tr>
<td></td>
<td>GATTAGGAAG</td>
<td></td>
<td>TATACCTTA</td>
<td>AGCAA</td>
</tr>
<tr>
<td>H11-r2-3-OT3</td>
<td>GTCAGAGCTCTCTG</td>
<td>chr7: 1-18216913</td>
<td>GTTGGGGAACCTGA</td>
<td>AGCCTGAAAAGGGGTAC</td>
</tr>
<tr>
<td></td>
<td>ATTAGCGAAG</td>
<td></td>
<td>TAATCCGC</td>
<td>GGGT</td>
</tr>
<tr>
<td>H11-r2-3-OT4</td>
<td>ATGTCAGACTCAT</td>
<td>chr13: 1-68310653</td>
<td>TGATCTGCTAGGTCA</td>
<td>CCCTTGTGGCAGGATA</td>
</tr>
<tr>
<td></td>
<td>GATTAGCGAAG</td>
<td></td>
<td>TATTAGCTTCA</td>
<td>AGGT</td>
</tr>
<tr>
<td>H11-r2-3-OT5</td>
<td>GTGTCAGCTCTTCA</td>
<td>chr15: 1-27274159</td>
<td>CACTGCTAGCATT</td>
<td>GTAAGACTGCTAGTCA</td>
</tr>
<tr>
<td></td>
<td>GATGAGCAGG</td>
<td></td>
<td>CGGCTTC</td>
<td>GACCCC</td>
</tr>
<tr>
<td>H11-r2-3-OT6</td>
<td>TTGTCAGCTCTTTG</td>
<td>chr4: 1-38229570</td>
<td>ACATGAGAATCCCC</td>
<td>TCAGTGCTCATCAAGG</td>
</tr>
<tr>
<td></td>
<td>ATTAGAGG</td>
<td></td>
<td>AAACAAACG</td>
<td>TAGTGT</td>
</tr>
</tbody>
</table>
Bibliography


Fukada S, Higuchi S, Segawa M, Koda K, Yamamoto Y, Tsujikawa K et al. 2004. Purification and cell-surface marker characterization of quiescent satellite cells from murine skeletal


Sandri M, Lin J, Handschin C, Yang W, Arany ZP, Lecker SH, Goldberg AL, Spiegelman BM. 2006. PGC-1α protects skeletal muscle from atrophy by suppressing FoxO3 action and


