Parathyroid Marker Tagging in iPSCs for Directed Differentiation

Michael S. Yao
Mentored by Dr. Shingo Suzuki, Dr. R. Geoffrey Sargent, and (late) Dr. Dieter C. Gruenert
Department of Otolaryngology, University of California, San Francisco, California, USA

Hyper- and hypo-parathyroidism are relatively common diseases in the United States, with over two hundred thousand new cases in the country every year. Regenerative, patient-specific treatment to replace defunct parathyroid glands is a promising method for functional correction of parathyroid activity because these parathyroid diseases, unlike many autosomal disorders, are not typically caused by genetic mutations. In this study, a protocol is described to create a novel iPSC (induced pluripotent stem cell) cell line featuring a reporter system targeting GCM2 (glial cells missing homolog 2), a lineage-specific parathyroid cell marker. Using CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein 9) gene modification technology to enhance homologous recombination through introducing double-strand DNA breaks (DSBs), a fluorescent protein sequence and drug-resistance cassette was installed for intracellular co-expression with GCM2. A modified CF3iPS2 cell line was successfully engineered to be the first iPSCs to direct parathyroid differentiation. The iPSCs were characterized through immunocytochemical analysis, and the modified genomic DNA (gDNA) was analyzed through running multiple analysis PCRs (polymerase chain reactions). The method reported herein opens the door to downstream applications to optimize parathyroid differentiation protocols in culture, using the reporter systems to ensure both genotypic and phenotypic expression of lineage-specific markers for the ultimate purpose of regenerative treatment for parathyroid diseases.

INTRODUCTION

Since the introduction of the microbial CRISPR/Cas9 system (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein 9) for intracellular targeted gene editing, gene therapy has gained increased noticeability as a potential patient-specific treatment for many diseases. With this gene-editing system, a wide variety of applications have since been developed, including specified mutation correction, gene activation, and genome modification. Previous studies have shown the use of CRISPR/Cas9 systems to genotypically and functionally correct mutations in genes that lead to diseases such as cystic fibrosis[1] and sickle cell anemia[2]. However, in regards to certain diseases concerning parathyroid cells, such as hyperparathyroidism and hypoparathyroidism, this traditional gene-editing approach to patient-specific treatment[3] is not applicable because they are not caused by genetic mutations. Hyperparathyroidism is typically caused by tumorous parathyroid cell growth or radioactive iodine treatment for thyroid diseases; on the contrary, familial genetic mutations leading to hyperparathyroidism are extremely rare[4]. The currently accepted treatment for overexpression of the parathyroid hormone (PTH) is parathyroidectomy, the excision of a nonfunctioning parathyroid gland(s), which often leads to hypoparathyroidism and the consequent under-expression of PTH. For these reasons, regenerative therapy promises to be a viable solution for parathyroid diseases, using iPSCs (induced pluripotent stem cells) derived from the patient to generate functionally correct parathyroid tissue and organs.

Although a parathyroid differentiation protocol from embryonic stem cells has been demonstrated[5], a similar differentiation protocol from iPSCs, which is crucial for regenerative treatment[6] because a self-derived iPSC-based treatment eliminates the issue of immune rejection of exogenous parathyroid replacements, remains to be developed. Most screening techniques to assess and isolate a directly differentiated population from iPSCs involve the process of fixation and/or lysis, rendering the screened cells unsuitable to proceed with further analysis[7]. In the experiment described herein, a nondestructive, fluorescence-protein based reporter system is installed as an alternative to screen for cellular clones that express the GCM2 (glial cells missing homolog 2) gene, resulting in live cells post-analysis in the population directed from iPSCs. In order for fluorescent proteins to be linked with the GCM2 protein, a single mRNA transcription needs to code for both proteins. However, translation of the transcript would result in a single chimeric "protein hybrid" of GCM2 and fluorescent protein, likely interfering with GCM2 functional activity. Previous techniques have traditionally used an internal ribosomal entry site (IRES) that allows for translation initiation independently from the protein stop codon, resulting in heterologous gene translations from an mRNA transcript[8]. However, the large sequence size and significant variability between the translations of the "linked" protein sequences are major issues with the IRES system. As a replacement, a smaller, virally derived 2A peptide (P2A) sequence demonstrates high cleavage efficiency of protein sequences immediately flanking the peptide[8]. Thus, through inserting the 2A peptide sequence between the hGCM2 exon 5 and mRFP (monomeric red fluorescent protein) sequences, a fluorescence-based reporter system directly linked with GCM2 expression through proper bicistronic expression can be established without interfering with GCM2 protein function. mRFP was used to visualize GCM2 expression in this experiment because it has been found to be a stably-expressed and relativity bright fluorescent protein with low toxicity levels, largely because of its monomeric nature[9].

The scope of this project is to utilize established methods for gene editing to produce a Sendai-virus-reprogrammed pluripotent cell line[10] tagging the human GCM2 gene expressed specifically in parathyroid cells. Successful generation of the reporter cell line greatly facilitates the analysis and screening of cells expressing the GCM2-coded protein while preserving cell functionality. Furthermore, the targeting of the GCM2 gene is one of the first steps to confirm parathyroid phenotype in differentiated cells. The installation of the mRFP sequence

following the 2A peptide allows for co-transcription of both protein-coding sequences, and provides a reliable, non-destructive method to determine GCM2-expressing cell populations in living culture conditions. The method presented in this report will generate cell lines to provide researchers an easy means to track parathyroid differentiation, and potentially develop a patient-specific method of treatment for parathyroid diseases.

MATERIALS AND METHODS

The overall experimental design and approach are illustrated in (**Fig. 1**), while the detailed materials and methods are described below.

Cells and Culture Conditions

All studies that involve human tissues were approved by the UCSF Committee on Human Research (CHR) and California Pacific Medical Center (CPMC) Institutional Review Board (IRB). The CF3iPS2 pluripotent cell line used in this research has been generated and characterized in concurrent experiments conducted by the same research group (unpublished). iPSCs were fed daily with mTeSR1 (Stemcell Technologies) feeder-free maintenance media, passaged with ReLeSR (Stemcell Technologies) enzyme-free passaging reagent and plated on a cell culture plate coated with Matrigel, a gelatinous protein mixture simulating a standard somatic environment. CFBE41o-, an immortalized human bronchial epithelial cell line[11], was fed every other day using completed Dulbecco's Modified Eagle Medium (Thermo Fisher), a modified solution of basal medium eagle with higher concentrations of amino acids and vitamins. Immortalized cells were cultured on a fibronectin-coated plate and subcultured by trypsinization (detachment from the culturing flask surface) with 0.05% trypsin-versene.

Cell passage number delineates a transfer of a fraction of cells in a confluent culture to new growth media, and is denoted as $Pn_1.n_2.n_3. - n_z$, where n_1 = number of passages as primary cells before reprogramming into iPSCs, n_2 = number of passages since reprogramming, n_3 = number of passages after transfection with a donor DNA, etc. where each period delineates the onset of a specific protocol or treatment that alters the character of the cells[12].

Immunocytochemical Analysis of iPSCs

iPSCs were grown in 24-well plates on Matrigel for immunostaining using primary antibodies and fluorescently-labeled secondary antibodies. To summarize, immunohistochemistry uses of primary antibodies to bind to specific proteinic pluripotent markers (NANOG, SSEA4, TRA1-60, and TRA1-81). Secondary fluorescent antibodies then recognize the fixed primary antibodies and emit color upon UV excitation, which can be analyzed to determine pluripotent protein expression.

Cells were washed with phosphate-buffered saline (PBS) water-based washing solution, fixed in 4% paraformaldehyde (PFA) for 30 min at room temperature (RT) to reduce cell composition degradation during antibody staining, and then washed three times, 5 min each, with ice-cold PBS. The samples were then permeabilized for later antigen detection[13] with 0.25% Triton X-100 non-ionic surfactant for 10 min at RT and incubated for 45 min with 5% normal serum and 1% bovine serum albumin (BSA), both to reduce non-specific antibody binding, in PBS containing 0.1% Tween-20 (Sigma-Aldrich) (PBST), a protein emulsifier, at RT with gentle agitation, followed by overnight incubation at 4°C with the pluripotent marker-targeting primary antibody in 1% BSA in PBST. Next, the samples were subjected to three 5-minute washes with PBS by gentle agitation. After the third wash, the cells were incubated with a 1:200 ratio of secondary antibodies in 1% BSA in PBST for 1 hr at RT in the dark, and again washed three times, 5 min each with gentle agitation. Finally, the samples were sealed with Dapi Fluoromount-G (SouthernBiotech) mounting medium and coverslips, and then examined by fluorescence microscopy. The samples were compared against a DAPI control, staining for cell nuclei, to show specificity of antibody binding to pluripotent-marker proteins.

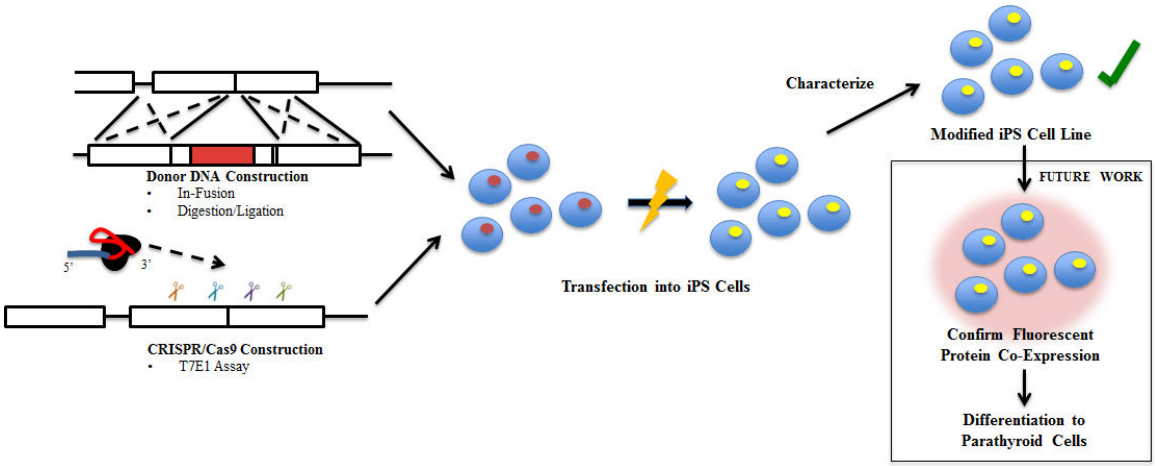


Fig. 1: Experimental Outline An illustration of the overall experimental plan for producing a genetically modified human iPS cell line. The donor DNA construction provided a plasmid with the homology arms and the desired insertion sequence. An optimal gRNA/Cas9 pair targeting hGCM2 was assessed by T7E1 assay. Then, both the optimized CRISPR/Cas9 system and the donor DNA were transfected into an iPSC cell line (CF3iPS2), and the modified cells will be isolated through weeks of drug-screening.

GCM2 Donor DNA Construction

Briefly, a donor DNA plasmid was constructed containing both the mRFP sequence and gene encoding resistance to the Puromycin drug, preceded by a 2A peptide sequence. Together, these sequences were inserted immediately before the GCM2 stop codon, such that they would be transcripted along with GCM2 mRNA for co-expression to "report" GCM2 expression. The donor DNA serves as the template for homology directed repair upon transfection with a DSB-inducing CRISPR/Cas9 system.

The donor DNA targeting the last exon, exon 5, of the hGCM2 gene was created using In-Fusion cloning technology[14] and sub-cloning techniques with restriction enzyme digestions and subsequent ligations. The initial construction was carried out to insert the P2A and mRFP sequences derived from amplifying CFTRex27-mRFP-pUC19 (unpublished) into the GCM2 donor DNA using In-Fusion techniques. This template sequence contained a 2A peptide sequence immediately upstream of the mRFP sequence and was amplified as a part of the mRFP fragment. The 5' homology arm was constructed through amplifying the genomic DNA of immortalized human bronchial epithelial cells, specifically 16HBE14o-[15], immediately before the hGCM2 exon 5 stop codon. The 3' homology arm was constructed through similar amplification of the 3' UTR (untranslated region) in the same cell line (**Fig. 2**). 100 pg of plasmid DNA (pDNA) or 100 ng of gDNA template were each amplified in a 25 µL total PCR (polymerase chain reaction) reaction volume. It is important to note that the forward primer targeting the 3' UTR (NheIHpaI - hGCM2ex5 3'UTR Fw) and the reverse primer targeting the mRFP sequence (mRFP-hGCM2ex5 NheIHpaI Rv) non-seamlessly introduced NheI and HpaI enzyme cut sites immediately downstream of the mRFP stop codon for the eventual insertion of a drug-selectable marker. The final construction inserted the PCR fragments into a pUC19 backbone (Addgene #50005). **Fig. 3** illustrates the In-Fusion process and shows the In-Fusion primer information. This intermediate construct was named hGCM2-P2A-mRFP-NheI/HpaI-pUC19.

Upon completion of this construction, the CAGPuro^{R}ΔTK drug cassette sequence was sub-cloned from CF2B-CAGPuroTK (a plasmid with the cassette insertion) into the newly constructed plasmid by digesting both vectors with NheI and HpaI. After gel purification of the drug cassette from the CF2B-CAGPuroTK construct using NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel), these two fragments were ligated together using T4 DNA Ligase. This new intermediate construct was named hGCM2-P2A-mRFP-PTK-pUC19 (**Fig. 4**). The drug-selection sequence was sub-cloned via restriction enzyme digestion because it is difficult to amplify the CAG promoter sequence using traditional PCR techniques that are required for In-Fusion process due to high GC content that interferes with polymerase activity at the 3' end of the promoter.

However, the insertion of the drug cassette disrupts the availability of the hGCM2 3' UTR region containing a poly-A sequence to appropriately and stably transcribe the mRNA coding the 5' homology arm and the mRFP sequence, possibly leading to poor translation and protein expression. A wPRE-polyA sequence derived from pCXLE-EGFP (Addgene #27082) was therefore sub-cloned into the vector plasmid immediately upstream of the drug cassette by appending an NheI cut site to both the 5' and 3' ends of the wPRE-polyA sequence via PCR. Primers described in **Fig. 4(C)** amplified the appropriate sequence from the pCXLE-EGFP template. The PCR fragment was then inserted into a TA Cloning Vector (Thermo Fisher) and then digested using NheI to isolate the wPRE-polyA fragment. Similarly, hGCM2-P2A-mRFP-PTK-pUC19 was cut using NheI to linearize the pDNA. These two NheI-digested fragments were then ligated together using T4 DNA Ligase. A diagnostic BglII/BspEI double digestion confirmed that the wPRE-polyA sequence was inserted in the correct 5'-3' orientation (data not shown). This final construct was named hGCM2-P2A-mRFP-pA-PTK-pUC19. All ligated plasmids in this experiment were then transformed into DH5α competent cells and isolated using the Pureyield Plasmid Miniprep System (Promega).

CRISPR/Cas9 System Design and Construction

The CRISPR/Cas9 systems tested in this experiment were derived from the Staphylococcus aureus (Sa) and Streptococcus pyogenes (Sp) bacterial species, both of which induce DSBs in mammalian cells. Four sgRNA (single guide RNA) sequences per bacterial species were tested through T7E1 assays for their inductions of DSBs and consequent non-homologous end joining (NHEJ) around the hGCM2 exon 5 stop codon in immortalized human bronchial epithelial cells, specifically CFBE41o-[11]. Sa and Sp sgRNAs were designed using web-based software "CCTop" (http://www.crispr.cos.uni-heidelberg.de/) and "Optimized CRISPR Design – MIT" (http://www.crispr.mit.edu/), respectively. Constructions of Sa and Sp sgRNA were carried out using published protocols by the Joung Lab[16] and Zhang Lab[17],

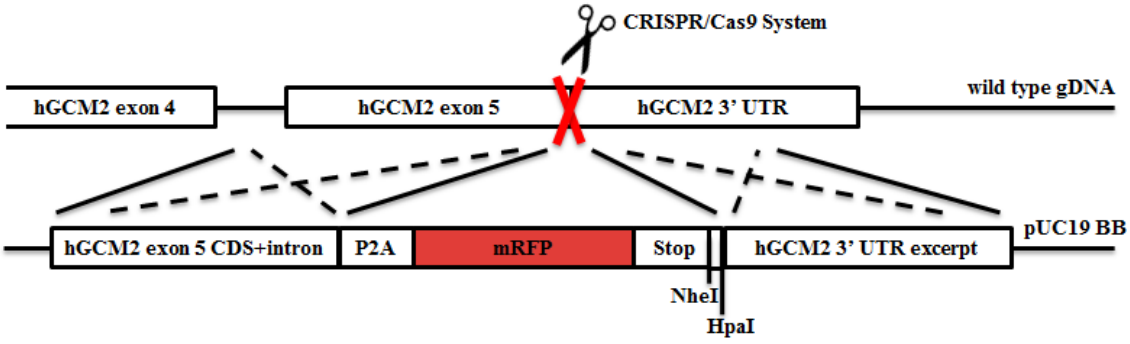
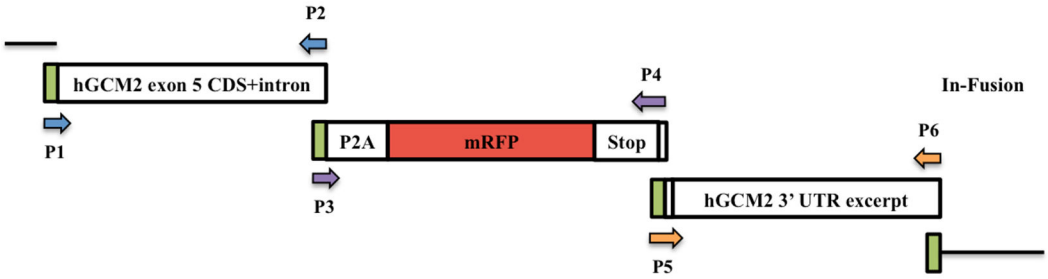
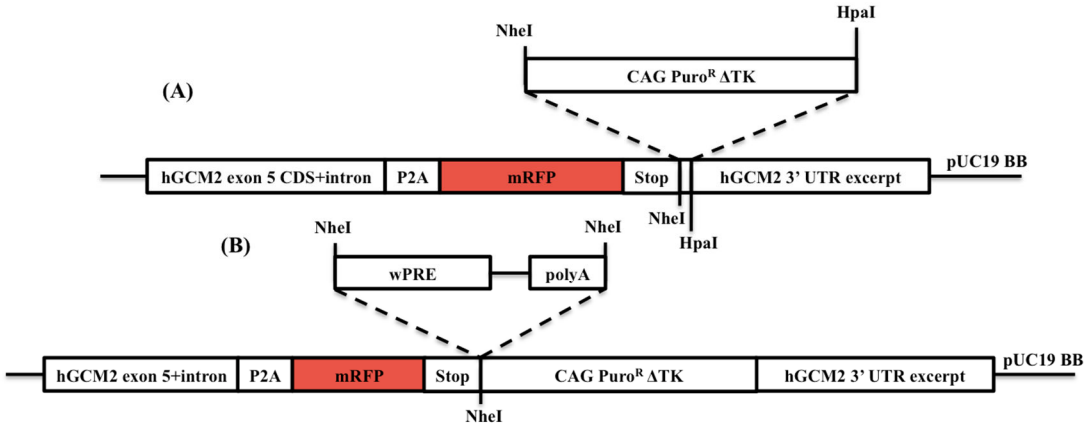


Fig. 2: GCM2 Donor DNA and gDNA Homologous Replacement Enhanced by CRISPR/Cas9 A scheme of homologous recombination between the constructed donor pDNA and wild type hGCM2 gDNA.



hGCM2-P2A-mRFP-NheI/HpaI-pUC19 In-Fusion Primer Table			
Primer Designation	**Primer Length**	**Primer Name**	**Primer Sequence**
P1	39 bp	pUC19-hGCM2ex5 Fw	5'-CGGTACCCGGGGATCGCGTGTGTCTATGTTTGTATCATT-3'
P2	26 bp	hGCM2ex5 CDS Rv	5'-AAAATCCTCATTGTTGTAGGGTAAAG-3'
P3	35 bp	hGCM2ex5 CDS-P2A Fw	5'-AACAATGAGGATTTTGGCTCCGGAGCCACAAACTT-3'
P4	36 bp	mRFP-hGCM2ex5 NheIHpaI Rv	5'-TAACAAGAATAAGCTAGCTTAGGCGCCGGTGGAGTG-3'
P5	39 bp	NheIHpaI - hGCM2ex5 3' UTR Fw	5'-AGCTTATTCTTGTTAACAAGACAATCCAGGGGACATAAT-3'
P6	38 bp	hGCM2 3'UTR – pUC19 Rv	5'-CGACTCTAGAGGATCCTAACAACTGTTTCCCAGGTGAT-3'

Fig. 3: Donor DNA Plasmid Construction using In-Fusion Assembly The In-Fusion process, similar to the Gibson Assembly. The green regions represent the 15 bp homology between successive PCR fragments used in the In-Fusion reaction. Primers P1, P2, P5, and P6 targeted 16HBE14o- P2.61 gDNA to generate the donor DNA homology arms. Primers P3 and P4 targeted CFTRex27-mRFP-pUC19 pDNA.



(C)

NheI-wPREpolyA Insert Construction Primer Table		
Primer Length	**Primer Name**	**Primer Sequence**
34 bp	NheI wPREpolyA Fw	5'-TGAAGCTAGCTTCAAGCTTCGGGGACTAGTCATA-3'
30 bp	NheI wPREpolyA Rv	5'-ACAAGCTAGCTTGTTCATGGCAGCCAGCAT-3'

Fig. 4: Sub-cloning CAGPuroR ΔTK and wPRE-polyA hGCM2 Donor DNA (A) Both CF2B-CAGPuroTK and hGCM2-P2A-mRFP-NheI/HpaI-pUC19 were double digested with NheI and HpaI. (B) The hGCM2-P2A-mRFP-CAGPuroTK-pUC19 backbone was digested with NheI, and the NheI-wPRE-polyA PCR fragment derived from pCXLE-EGFP amplification was ligated into the backbone using T4 DNA Ligase, incubating at 24°C. (C) Primers P7 and P8 targeted pCXLE-eGFP pDNA to generate the wPRE-polyA sequence insertion.

respectively. As an overview, double-strand DNA fragments with overhangs that complement the ligation site of the sgRNA backbone vector were created from sense and anti-sense oligos, denatured, and then slowly annealed, consisting of the desired GCM2 targeting sequence. The annealed fragments were then ligated into the plasmid BPK2660 (Addgene #70709), digesting first with BsmBI and then ligating with T4 DNA Ligase at 24°C for 10 minutes for Sa sgRNA construction, and into the plasmid pX330 (Addgene #42230), digesting first with BbsI and then ligating with T7 DNA Ligase at alternating five-minute cycles of 37°C and 21°C for 1 hour for Sp sgRNA construction. BPK2660 is an optimized vector for Sa targeting gRNA expression[16], while pX330 is an optimized, chimeric vector that expresses both Sp targeting gRNA and the Sp Cas9 protein[18].

Optimization of a Targeting Site via CRISPR/Cas9 System

T7E1 Assays (**Fig. 5**) are commonly used to detect mutations, including those caused by CRISPR/Cas9 activity, using the bacteriophage resolvase T7 endonuclease 1 to cleave heteroduplexes[19]. To perform T7E1 assays to determine CRISPR/Cas9 systems with the highest comparative DNA cutting efficiencies, CFBE41o- cells were nucleofected using SF Cell line solution (Lonza Group) according to their 4D-Nucleofector Protocol with the optimized pulse code, DS120, for this cell line. The cells were trypsinized using 0.05% trypsin-versene and suspended in 20 µL of SF transfection reagent. Transfection conditions are shown in Table 1. The gDNA was harvested 5 days after transfection and analyzed by performing a T7E1 Assay. Agarose gels visualized DNA bands using ethydium bromide and GelDoc 2000 (Biorad).

Percentage of NHEJ incidences (%NHEJ) was determined by the following formula:

$$f_{cut} = \frac{CleavageBand_1 + CleavageBand_2}{CleavageBand_1 + CleavageBand_2 + UncleavedBand} \quad \text{Eq. 1}$$

$$\%NHEJ = 100 * \left(1 - \sqrt{1 - f_{cut}}\right) \quad \text{Eq. 2}$$

PCR Conditions

All PCRs in this experiment were conducted using either 2x MyTaq Hot-Start Reaction Mix (Bioline) or, specifically for generating the aforementioned In-Fusion PCR products, 2x CloneAmp HiFi PCR Premix (Clontech). For MyTaq Reactions, 50-100 ng of gDNA or 100 pg of pDNA was amplified with 1 µL of 10 µM forward and reverse primers each. 12.5 µL of the 2x MyTaq Hot-Start Reaction Mix was added to the reaction mixture, and water was added to bring the final volume up to 25 µL. 35 cycles were run per PCR reaction. CloneAmp reactions followed a similar protocol but instead used 0.625 µL of 10 µM forward and reverse primers each. For MyTaq, the denaturation step occurred at 95°C and the extension step at 72°C, while the CloneAmp denaturation step occurred at 98°C and extension at 72°C. Annealing temperatures were optimized per primer pair.

RESULTS

Sp CRISPR/Cas9 system successfully targets and introduces DSBs in the GCM2 targeting site.

The induction of DSB in a targeting sequence dramatically improves site-specific homologous recombination[3]. Therefore, an optimized CRISPR/Cas9 system was sought to target and introduce DSBs at or near the stop codon of the GCM2 gene. For testing Sp gRNAs, transfection conditions (**Table 1**), as optimized in previous experiments[20], compared the four gRNA sequences with an empty pX330 vector negative control, which demonstrated a small yet detectable level of NHEJ incidences that produced T7E1-digested fragments about 400 bp in length (**Fig. 6**). According to the T7E1 assay results, while Gb11 sgRNA demonstrated the greatest percentage of NHEJ incidences (%NHEJ) at ~19.3%, the band intensity analysis of the gel may have included the negative control's non-specific digestion band patterns to yield a %NHEJ value higher than the actual percentage of DSB on-target occurrences, suggesting that on-target cutting efficiency was roughly the same for all Sp sgRNAs tested in this experiment. Ultimately, Gb11 sgRNA was chosen as the targeting gRNA sequence.

Additionally, four corresponding Sa bacterial CRISPR/Cas9 systems were also tested for their on-target cutting efficiency (**Fig. 6**). Interestingly, none of the four Sa sgRNA sequences successfully enhanced %NHEJ according to a T7 Endonuclease 1 Assay. Other possible candidates for Sa gRNA sequences targeting GCM2 gene must be tested in order to determine if Sa bacterial CRISPR systems are unsuitable for targeting the specific GCM2 gene.

Co-transfection of GCM2 Donor DNA and CRISPR/Cas9 system yields the insertion of a reporter system in iPSCs.

Following the aforementioned T7 Endonuclease 1 Assay data, the CF3iPS2 pluripotent cell line was co-transfected with the constructed GCM2 donor DNA and Gb11 sgRNA in pX330 backbone. 2 x 10^6 cells were electroporated with a 1:1 ratio of donor DNA to gRNA/Cas9 DNA using P3 Primary Cell Solution (Lonza Group) in a 100 µL large cuvette. A large amount of cell death was observed immediately following the transfection, likely due to the introduction of the large GCM2 Donor DNA (10.2 kb). The cells were kept on a 60 mm Matrigel-coated dish with 10 µM Rock inhibitor-containing mTeSR1 feeder-free media for one day post-transfection, after which Rock inhibitor concentration decreased to 5 µM. 0.5 µg/mL of puromycin was added to the mTeSR1 feeding media starting from day 2 for two weeks. On day 9 through 11 post-transfection, individual puromycin resistant ($Puro^R$) clones were picked. A total of 42 clones out of the approximately 100 clones observed were individually isolated and propagated to continue growth.

Figure 7 shows the Analysis PCR reactions conducted to confirm CF3iPS2 clones that successfully inserted the reporter system using the primers described in **Fig. 7(A)**. Briefly, the analysis primers were chosen in such a way such that primer pairs would only amplify junction sites of correctly modified GCM2 DNA, as opposed to amplifying non-modified gDNA, sites of random integration, and the constructed GCM2 donor DNA. PCR fragment signals thus would suggest successful homology directed repair at the GCM2 gene using the GCM2 donor DNA.

After optimization of the analysis PCR conditions, the 42 different clones were analyzed against CF3iPS2 Non-transfected

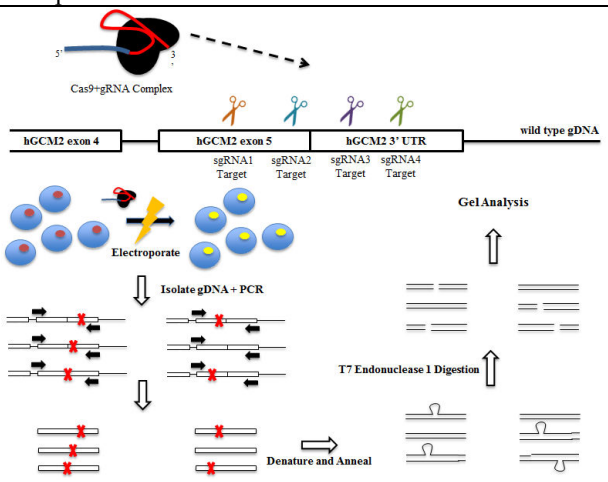


Fig. 5: T7E1 Assay The basic template to determine approximate %NHEJ efficiency for a given CRISPR gRNA sequence. Five days after CRISPR/Cas9 transfection, the gDNA was isolated and then amplified using the T7E1 Primers described in **Fig. 6**. The PCR products were denatured by incubating at 95°C for 10 minutes and then annealed to create heteroduplexes by ramping down at -5°C/min. T7 Endonuclease 1 then cleaved the annealed PCR products at mismatched sites, and the digested products were analyzed on an agarose gel.

Table 1: CFBE41o- Transfection Conditions for T7E1 Assay Both Sp gRNA and Sa gRNA targeting sequences and transfection conditions are shown, following transfection protocol previously established[20].

T7E1 Assay – CFBE41o- Transfection Conditions					
Cell Number	**Cas9**	**gRNA**	**gRNA+PAM Sequence**	**Top Oligo Sequence**	**Bottom Oligo Sequence**
$2x10^5$ per transfection	MSP1830 (800ng)	BPK2660 (200ng)	*	*	*
		GbA SAsgRNA (200ng)	5'-TCCGCAGTCAGTTACTCAGA cagagt-3'	5'-CACCgTCCGCAGTCAGTTACTCAGA-3'	5'-AAACTCTGAGTAACTGACTGCGGAc-3'
		GbB SAsgRNA (200ng)	5'-TGTCCCCTGGATTGTCTTTC aaaaat-3'	5'-CACCgTGTCCCCTGGATTGTCTTTC-3'	5'-AAACGAAAGACAATCCAGGGGACAC-3'
		GbC SAsgRNA (200ng)	5'-GCACACTGCTATTATGTCCC ctggat-3'	5'-CACCgGCACACTGCTATTATGTCCC-3'	5'-AAACGGGACATAATAGCAGTGTGCc-3'
		GbD SAsgRNA (200ng)	5'-ACACAAGGTGAATCTCCCAA ctcaat-3'	5'-CACCgACACAAGGTGAATCTCCCAA-3'	5'-AAACTTGGGAGATTCACCTTGTGTC-3'
$3x10^5$ per transfection	px330 (1.5µg)		*	*	*
	Gb11 SPsgRNA in px330 (1.5µg)		5'-CACACTGCTATTATGTCCCCtgg-3'	5'-CACCgCACACTGCTATTATGTCCCC-3'	5'-AAAcGGGGACATAATAGCAGTGTGc-3'
	Gb12 SPsgRNA in px330 (1.5µg)		5'-GCAATGATCTCTTATTGAGTtgg-3'	5'-CACCgGCAATGATCTCTTATTGAGT-3'	5'-AAAcACTCAATAAGAGATCATTGCc-3'
	Gb13 SPsgRNA in px330 (1.5µg)		5'-CATAATAGCAGTGTGCATGCagg-3'	5'-CACCgCATAATAGCAGTGTGCATGC-3'	5'-AAAcGCATGCACACTGCTATTATGc-3'
	Gb14 SPsgRNA in px330 (1.5µg)		5'-CTCTGCTTCATCTGTCCTAGagg-3'	5'-CACCgCTCTGCTTCATCTGTCCTAG-3'	5'-AAAcCTAGGACAGATGAAGCAGAGc-3'

*Undigested BPK2660 + MSP1830 is the SA negative control. Undigested px330 is the SP negative control. px330 is a chimeric sequence that contains both the gRNA and Cas9 sequence in a single plasmid construct.

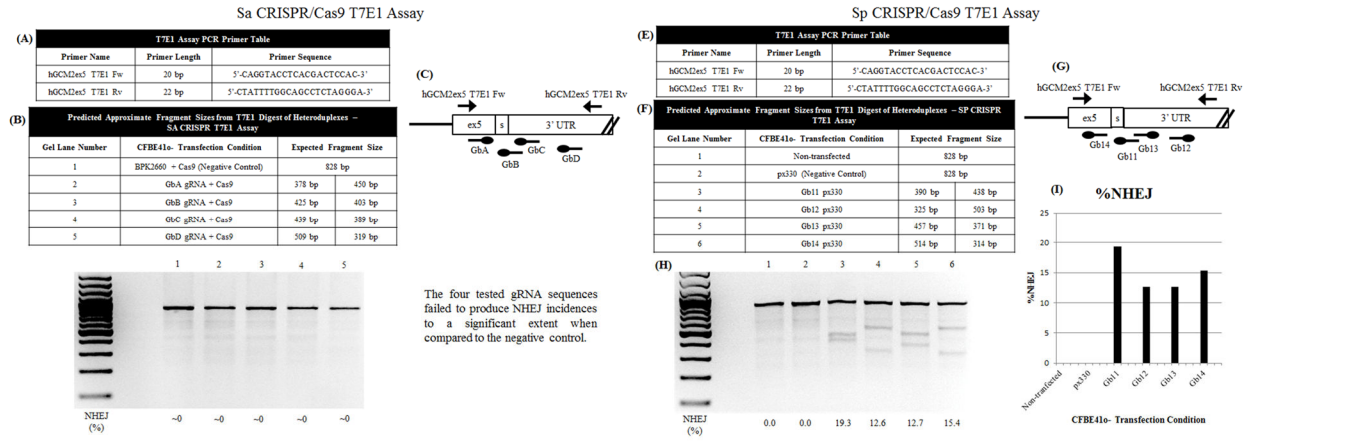
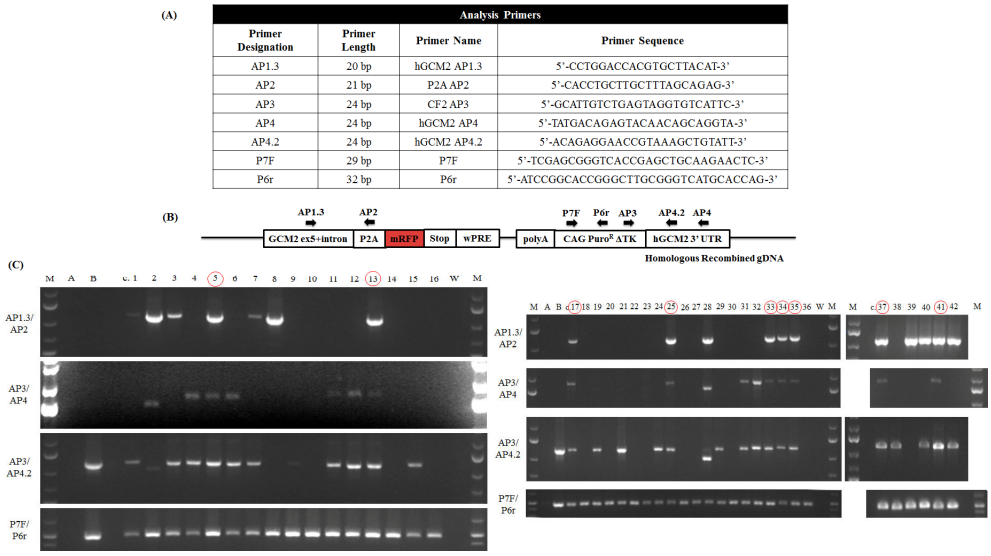


Fig. 6: CRISPR/Cas9 T7E1 Assay Results (A,E) The primers to amplify the modified gDNA are shown. (B,F) The predicted band sizes that result from T7E1 enzymatic activity based on gRNA targeting sequences are shown, assuming on-target DSB induction per gRNA sequence for Sa (left) and Sp (right) systems. (C,G) Relative targeting sites of both Sa (left) and Sp (right) gRNA sequences are shown. (D,H) T7E1-digested fragments run on 2% TBE gels. (I) The results suggested that the tested Sp sgRNAs were significantly more effective at inducing NHEJ incidences than the tested Sa sgRNAs targeting the last exon of the GCM2 gene. Gb11 Sp sgRNA in pX330 was selected to enhance homologous recombination in CF3iPS2 upon co-transfection with the hGCM2 donor DNA constructed earlier.

Fig. 7: CF3iPS2 GCM2 Modification Analysis (A,B) Analysis Primers (AP) 1.3 and 4 targeted genomic DNA outside of the donor DNA homology arms, while AP2, AP3, and AP4.2 targeted regions of the donor DNA specific to the fluorescent protein and drug cassette insertion. P7F/P6r targeted the puromycin resistance gene to confirm that the puromycin selection treatment during cell culture was selective for positive clones. (C) Analysis PCRs for clones 1-42. Lane M was a 1 kb DNA ladder. Lane A was a CF3iPS2 Non-transfected negative control, and lane B was a CF3iPS2 Non-transfected + GCM2 donor DNA negative control.

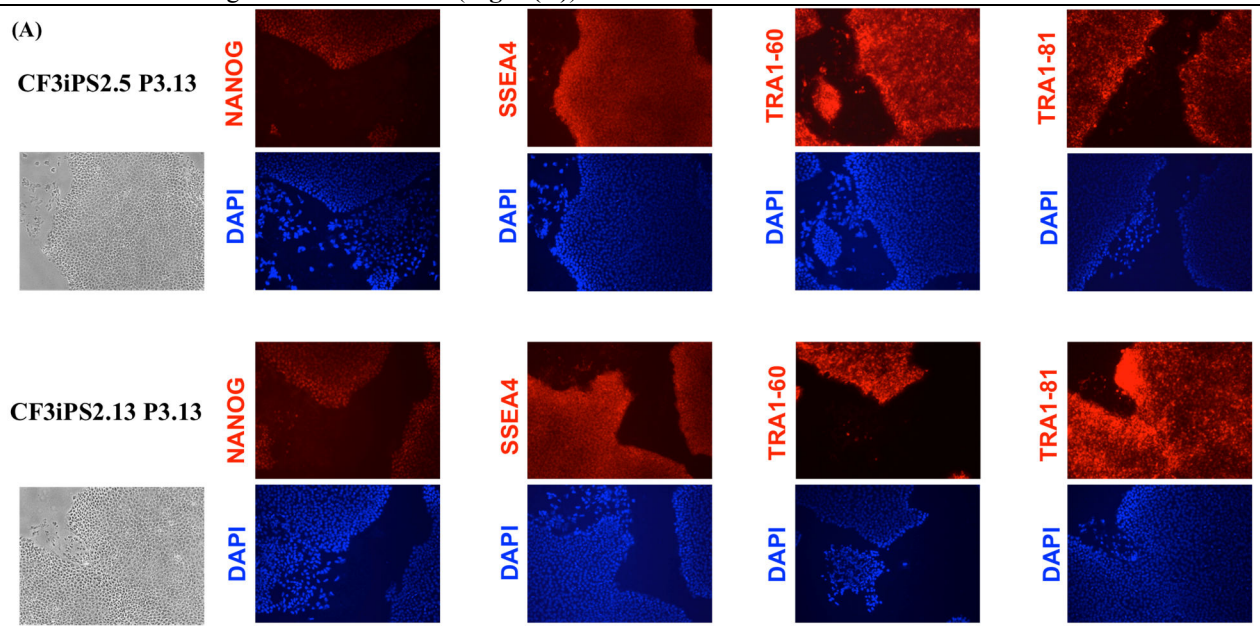


and CF3iPS2 Non-transfected + GCM2 donor DNA negative controls. Positive signals for all 4 analysis PCRs conducted indicated successful insertion of P2A and mRFP sequences immediately downstream of the GCM2 gene. Co-transfection of the Gb11 in pX330 and the GCM2 Donor DNA yielded a targeted modification efficiency of 21.4% (9 out of the 42 clones). Positive clone 13 demonstrated the correct GCM2ex5-P2A junction sequence (**Fig. 8**). CF3iPS2 clones 5 and 13, with positive analysis PCR results, were characterized through immunocytochemical analysis to confirm the retention of pluripotent character after endogenous modification (**Fig. 8(A)**).

Overall, the insertion of mRFP and the puromycin resistance drug cassette demonstrates that site-specific modification of gDNA is a viable procedure to edit pluripotent stem cells using a CRISPR/Cas9 system to enhance the frequency of HDR (homology-directed repair) events.

DISCUSSION

The findings of this experiment report efficient and on-target gene modification, enhanced by the novel use of CRISPR/Cas9 systems to specifically target the GCM2 gene and to increase the



(C)

Immunocytochemical Analysis Antibodies	
PRIMARY ANTIBODY	**VENDOR**
Anti-NANOG antibody, mouse monoclonal IgG1	BD Biosciences 560482
Anti-SSEA-4 antibody, mouse monoclonal IgG3	Abcam Inc. ab16287
Anti-TRA-1-60, mouse monoclonal IgM	BD Biosciences 560071
Anti-TRA-1-81, mouse monoclonal IgM	Millipore MAB4381
SECONDARY ANTIBODY	**VENDOR**
Alexa Fluor 488 goat anti-mouse IgG1	Invitrogen A21121
Alexa Fluor 546 goat anti-mouse IgM	Invitrogen A21045

Fig. 8: CF3iPS2 Characterization: (A) Immunocytochemical analysis for CF3-iPS2 P3.9 pluripotency markers: NANOG, SSEA4, Tra1-60, and Tra1-81. A phase contrast image demonstrated typical pluripotent cell culture morphology in feeder free culture. (B) CF3iPS2 c13 was sequenced and shows the correct junction sequence between GCM2 exon 5 and the 2A Peptide. (C) Antibodies for immunostaining are shown.

number of homologous recombination events between transfected donor DNA and endogenous genomic DNA. This process yields a functional pluripotent cell line that can be used for future work in tracking parathyroid differentiation. Ultimately, this method can find applications in establishing a genotypic and phenotypic differentiation protocol from pluripotent stem cells to parathyroid cells and tissue for regenerative treatment of parathyroid diseases.

An issue encountered during this experiment was the generation of Sa CRISPR gRNA targeting sequences that successfully induced DSBs. Currently, there exists no readily available system for determining predicted on-target cutting efficiencies for Sa gRNA sequences, unlike those systems that exist for Sp CRISPR/Cas9 systems ("Optimized CRISPR Design – MIT"). Rather, present methods base gRNA efficacy predictions solely on off-targeting DNA cutting potential. As a result, all four Sa gRNA sequences tested did not demonstrate the induction of on-target NHEJ, as determined in the T7E1 assay when compared with the negative control. In order to prevent similar occurrences in the future, a Sa gRNA on-target testing system should be investigated to further research work involving the Sa bacterial system, so the prediction step of targeting site will be improved. Unlike the Sa CRISPR/Cas9 system, the Sp CRISPR/Cas9 system has been used since its development as the first working CRISPR/Cas9 system in mammalian cells, and its prediction algorithms of its targeting site were developed and optimized by many research groups[21,22]. In fact, all of the Sp gRNA targeting sequences (Gb11-Gb14) tested using the T7E1 Assay introduced at least 12.6% of %NHEJ, compared with a non-transfected and non-targeting Cas9 transfected negative controls. Therefore, in this stage, the Sp CRISPR/Cas9 system would be the obvious choice to enhance homologous recombination efficiency upon co-transfection with appropriate donor DNA, as prior scientific literature has shown with the Sp bacteria-derived gene editing technology[23,24,25].

Using the optimized CRISPR/Cas9 system, the co-transfection of Gb11 in pX330 with the constructed hGCM2-mRFP-PTK- pUC19 donor DNA provided approximately 21.4% (9 out of 42 clones) targeting efficacy in the puromycin resistant clones, which demonstrated positive band patterns after running various analysis PCRs described in **Fig. 8**. Supplementary literature suggests that the median of DNA repair through HDR is approximately 33%[26]. Clonal isolation during puromycin treatment ensures independence between clone samples, and

$$np = (42)(0.33) = 13.86 > 10$$
$$n(1-p) = (42)(1-0.33) = 28.14 > 10,$$

such that the sample of size n = 42 can be considered normal. Assuming randomness, we have

$$z^* = \frac{\hat{p} - p}{\sigma} = \frac{\frac{9}{42} - 0.33}{\sqrt{\frac{0.33(1-0.33)}{42}}} = -1.595$$
$$p - value = 0.0554$$

which fails to call into question the accepted p = 0.33 proportion of HDR events at an $\alpha = 0.05$ significance level; a standard score of $z^* = -1.595$ standard deviations away from the accepted proportion of p = 0.33 contains greater than 5% of the normal sampling distribution curve. The data collected in this experiment thus corroborates previous findings from prior research and is not considered an anomaly in the scope of other gene-modifying experiments utilizing CRISPR-enhanced HDR. This conclusion assumes that no significant skew exists in the sampling distribution, so that the median and mean of overall DNA repair events via HDR are both approximately 33%.

Furthermore, the band patterns in the Analysis PCRs depicted in **Fig. 7** can be used to demonstrate the nature of the modified clones resistant to puromycin treatment. As stated previously, AP1.3 and AP4 target the genomic DNA around the GCM2 modification site but sufficiently far enough from the point of modification such that the donor DNA homology arms cannot be amplified by AP1.3 and AP4. Similarly, AP2 targets the 2A Peptide sequence, and AP3 targets the drug cassette so that neither primer can amplify unmodified genomic DNA. Therefore, the selection of analysis primer pairs AP1.3/AP2 and AP3/AP4 only amplifies gDNA sequences that correctly integrated the reporter system in the GCM2 site. However, certain clones, such as clone 2 and clone 28, exhibit AP3/AP4 band patterns that are interestingly smaller than expected, suggesting a mutation in the 3' region where the primer pair amplifies. This is likely due to the continued CRISPR/Cas9 cutting activity after the integration of the reporter system occurred, such that Gb11 in pX330 targets the modified 3' arm downstream of the integrated drug cassette and cuts the gDNA again, introducing a noticeable deletion in the 3'UTR region. Although the deletion is in the untranslated region and therefore has no effect on the GCM2 coding region, it is uncertain how large of an impact the deletion has on mRNA transcript stability, as significant deletions may adversely affect GCM2-coded protein expression. Further tests, including sequencing and protein expression analysis, will be necessary in order to determine the nature of the deletion events.

An additional analysis PCR was run using AP3 and AP4.2. AP4.2 targeted a region of the 3'UTR that was part of the donor DNA 3' homology arm. Because the donor DNA only transiently exists after initial electroporation, a positive signal for the AP3/AP4.2 PCR but negative signals for AP1.3/AP2 and AP3/AP4 would suggest random integration(s) of the donor DNA to a different site(s) of the genomic DNA. For example, clones 15 and 29, among others, demonstrate band patterns suggestive of random integration of the donor DNA. Finally, primer pair P7F and P6r targeted the $Puro^R$ gene to confirm that the drug concentration used for selection was specific and killed the cells without the integrated resistance gene.

CONCLUSIONS AND FUTURE WORK

The aforementioned experimental data has shown evidence to positively support the construction of a genetically-modified pluripotent cell line engineered and edited in a way to assist in tracking parathyroid differentiation through modifying GCM2, a lineage-specific marker, to co-express the GCM2 protein and mRFP. A total of 42 CF3iPS2 $Puro^R$ isolated clones were analyzed through various analysis PCRs following puromycin cell culture treatment, 9 clones (21.4%) of which demonstrated expected band patterns of on-target HDR. It has also been shown that CRISPR/Cas9 systems are an effective method of enhancing homologous recombination efficiency; specifically, the Gb11

targeting gRNA inserted into pX330 enhanced %NHEJ to approximately 19.3% in a T7 Endonuclease 1 Assay with CFBE41o- cells, compared with the non-transfected and blank pX330 vector negative controls, both of which showed minimal, if any, detectable on-target cutting efficiency. However, interestingly, none of the Sa CRISPR systems tested in this experiment demonstrated detectable on-target DSB-inducing function, and should be tested further with additional targeting gRNA sequences to determine the nature of gRNA non-targeting, and whether the Sa system simply lacks the ability to target the GCM2 site.

It has been shown that GCM2 is a highly specific genetic marker to determine parathyroid cell expression[27]. However, there are other confirmation genes to target that are also lineage-specific to further confirm successful parathyroid differentiation, such as the parathyroid-hormone-coding gene, as well as other genes expressed earlier in the differentiation pathway from pluripotency to fully-differentiated state. Genes such as Nkx2.1, expressed in the anterior foregut endoderm, and Six1, expressed in the third pharyngeal pouch, both of which are parathyroid precursor markers, could also be marked in a process similar to that shown for GCM2 in this experiment in order to more closely track the differentiation pathway, facilitating parathyroid disease studies.

Furthermore, it is undetermined whether one or both of the alleles in clones exhibiting positive band patterns underwent homologous recombination to insert the reporter system. A non-modified allele may or may not have undergone deletion or insertion events around the Gb11 cutting site due to NHEJ events. However, since the Gb11 targeting site is downstream of the GCM2 stop codon in the 3' UTR region, it is unlikely that NHEJ events were introduced in the GCM2 coding region. Additionally, because the modified allele underwent homologous recombination with the donor DNA and therefore presumably lacks mutagenesis events, the modified cells overall are still able to produce functionally correct GCM2-coded proteins. Future analysis using PCR techniques may be used to determine the number and nature of alleles modified in positively selected clones. However, the issue of allele non-specificity with regards to CRISPR/Cas9 targeting remains an unavoidable obstacle.

Overall, the approach described in this experiment set up the foundations for the establishment of an iPSC differentiation protocol into parathyroid cells, which would be a necessary and important milestone in the development of clinical regenerative treatment for various diseases. Continued use and studies of gene-editing technologies remain at the forefront to progress iPSC-based disease treatments.

1. Suzuki, S., et al. (2016). "TALENs facilitate single-step seamless SDF correction of F508del CFTR in airway epithelial submucosal gland cell-derived CF-iPSCs." Molecular Therapy Nucleic Acids 5, e273. doi: 10.1038/mtna.2015.43
2. Tasan, I., Jain, S. & Zhao, H. (2016). "Use of genome-editing tools to treat sickle cell disease." Human Genetics 135: 1011. doi:10.1007/s00439-016-1688-0
3. Sargent, R. G., Suzuki, S. & Gruenert, D. C. (2014). "Nuclease-mediated Double Strand Break (DSB) Enhancement of Small Fragment Homologous Recombination (SFHR) Gene Modification in Human Induced Pluripotent Stem Cells (hiPSCs)." Methods Molecular Biology 1114: 279-290. doi: 10.1007/978-1-62703-761-7_18
4. Pallan, S. & Khan, A. (2011). "Primary hyperparathyroidism." Can Fam Physician 57(2): 184–189.
5. Bingham, E. L., et al. (2009). "Differentiation of human embryonic stem cells to a parathyroid-like phenotype." Stem Cells and Development 18(7): 1071-1080. doi:10.1089/scd.2008.0337
6. Ozdemir, M., Attar, A. & Kuzu, I. (2012). "Regenerative treatment in spinal cord injury." Current Stem Cell Research and Therapy 7(5): 364-369.
7. Ericson, J. G., et al. (1994). "Effects of whole blood lysis and fixation on the infectivity of human T-lymphotropic virus type 1 (HTLV-I)." Cytometry 18(1): 49-54. doi: 10.1002/cyto.990180110
8. Kim, J. H., et al. (2011). "High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice." PLoS ONE 6(4): e18556. doi: 10.1371/journal.pone.0018556
9. Cranfill, P. J., et al. (2016). "Quantitative assessment of fluorescent proteins." Nature Methods 13: 557-62. doi: 10.1038/nmeth.3891
10. Fusaki, N., et al. (2009). "Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome." Proceedings of the Japan Academy, Series B 85(8): 348-362. doi: 10.2183/pjab.85.348
11. Ehrhardt, C., et al. (2006). "Towards an in vitro model of cystic fibrosis small airway epithelium: characterisation of the human bronchial epithelial cell line CFBE41o-." Cell Tissue Resources 323(3): 405-415. doi: 10.1007/s00441-005-0062-7
12. Gruenert, D. C., et al. (1988). "Characterization of human tracheal epithelial cells transformed by an origin-defective simian virus 40." Proceedings of the National Academy of Sciences USA 85: 5951-5955. doi: 10.1073/pnas.85.16.5951
13. Jamur, M. C. & Oliver, C. (2010). "Permeabilization of cell membranes." Methods in Molecular Biology 588: 63-66. doi: 10.1007/978-1-59745-324-0_9
14. Gibson, D. G., et al. (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases." Nature Methods 6: 343-345. doi: 10.1038/nmeth.1318
15. Cozens, A. L., et al. (1994). "CFTR expression and chloride secretion in polarized immortal human bronchial epithelial cells." American Journal of Respiratory Cell and Molecular Biology 10(1): 38-47. doi: 10.1165/ajrcmb.10.1.7507342
16. Kleinstiver, B. P., et al. (2015). "Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition." Nature Biotechnologies 33(12): 1293-1298. doi: 10.1038/nbt.3404
17. Xie, S., et al. (2014). "sgRNACas9: A software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites." PLoS ONE 9(6): e100448. doi: 10.1371/journal.pone.0100448
18. Cong, L., et al. (2013). "Multiplex Genome Engineering Using CRISPR/Cas Systems." Science 339(6121): 819-23. doi: 10.1126/science.1231143

19. Mashal, R. D., Koontz, J. & Sklar, J. (1995). "Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases." Nature Genetics 9(2): 177-83. doi: 10.1038/ng0295-177
20. Xu, P., et al. (2015). "Both TALENs and CRISPR/Cas9 directly target the HBB IVS2–654 (C > T) mutation in β-thalassemia-derived iPSCs." Scientific Reports 5:12065. Doi: 10.1038/srep12065
21. Ran, F. A., et al. (2013). "Genome engineering using the CRISPR-Cas9 system." Nature Protocols 8(11): 2281-2308. doi: 10.1038/nprot.2013.143
22. Farboud, B. & Meyer, B. J. (2015). "Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design." Genetics 199(4): 959-971. doi: 10.1534/genetics.115.175166
23. Arribere, J. A., et al. (2014). "Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans." Genetics 198(3): 837-846. doi: 10.1534/genetics.114.169730
24. Kim, H., et al. (2014). "A co-CRISPR strategy for efficient genome editing in Caenorhabditis elegans." Genetics 197(4): 1069-1080. doi: 10.1534/genetics.114.166389
25. Ward, J. D. (2015). "Rapid and precise engineering of the Caenorhabditis elegans genome with lethal mutation co-conversion and inactivation of NHEJ repair." Genetics 199(2): 363-377. doi: 10.1534/genetics.114.172361
26. Chu, V. T., et al. (2015). "Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells." Nature Biotechnology 33: 543-548. doi: 10.1038/nbt.3198
27. Zhu, M., Yu, S. & Manley, N. R. (2007). "Gcm2 is required for the differentiation and survival of parathyroid precursor cells in the parathyroid/thymus primordia." Developmental Biology 305(1): 333-346. doi: 10.1016/j.ydbio.2007.02.014