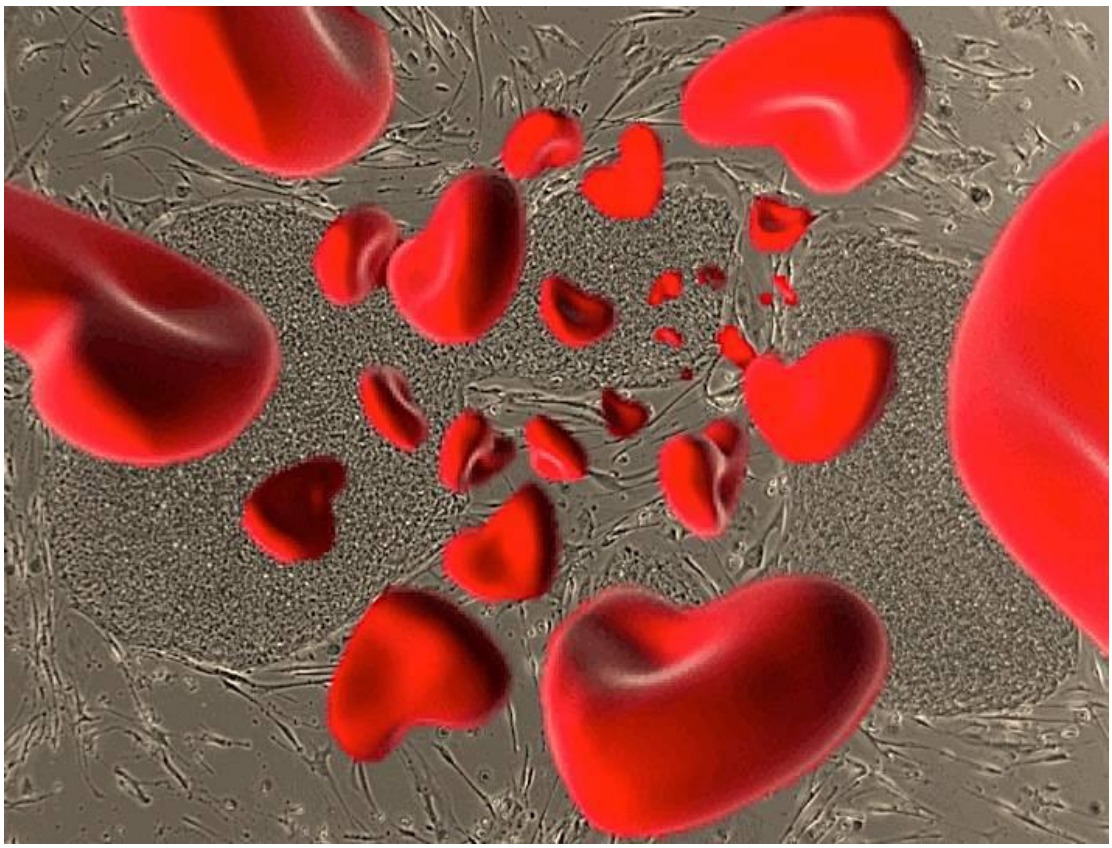
UNIVERSIDAD AUTONOMA DE MADRID

Knock-In Gene Correction of induced Pluripotent Stem Cells from Pyruvate Kinase Deficient Patients



Universidad Autónoma de Madrid,
Departamento Biología Molecular

Zita Garate Mutiloa

Madrid 2013

UNIVERSIDAD AUTÓNOMA DE MADRID

Facultad de Ciencias

Departamento Biología molecular

Zita Garate Mutiloa

Knock-In Gene Correction of induced Pluripotent Stem Cells from Pyruvate Kinase Deficient patients

Memoria presentada por **ZITA GARATE MUTILOA,** licenciada en Biotecnología, para optar al grado de doctor por la Universidad Autónoma de Madrid

Directores de Tesis:

José Carlos Segovia Sanz

Oscar Quintana Bustamante

CENTRO DE INVESTIGACIONES ENERGÉTICAS, MEDIOAMBIENTALES Y TECNOLÓGICAS (CIEMAT) y

CENTRO DE INVESTIGACIONES BIOMÉDICAS EN RED DE ENFERMEDADES RARAS (CIBERER)

División de Terapias Innovadoras del Sistema Hematopoyético

MADRID

2013

Knock-In Gene Correction of induced Pluripotent Stem Cells from Pyruvate Kinase Deficient patients

José Carlos Segovia Sanz, Jefe de la Unidad de Diferenciación y Citometría de la División de Terapias Innovadoras del Sistema Hematopoyético del Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, y **Oscar Quintana Bustamante,** investigador post doctoral de la misma unidad, certifican que la memoria adjunta, titulada *"*Knock-In Gene Correction of induced Pluripotent Stem Cells from Pyruvate Kinase Deficient patients*"* ha sido realizada por Zita Garate Mutiloa bajo la dirección conjunta de los que suscriben, y cumple las condiciones exigidas para optar al título de Doctor por la Universidad Autónoma de Madrid.

José Carlos Segovia Sanz **Oscar Quintana Bustamante**

El presente trabajo de investigación ha sido realizado en la División de Terapias Innovadoras del Sistema Hematopoyético del **CIEMAT** y Centro de Investigaciones Biomédicas en Red de Enfermedades Raras (**CIBERER**), del Instituto de Salud Carlos III, con la colaboración del proyecto **PERSIST** del programa de salud de la Unión Europea, la Red Española de Terapia Celular (**TERCEL**) y el Ministerio de Economía y Competitividad (**MINECO**), Secretaría de Estado de Investigación, Desarrollo e Innovación.

Zita Garate Mutiloa ha disfrutado de una beca del programa PREDOC del departamento de Educación, Política Lingüística y Cultura del **Gobierno Vasco**.

I. Summary

Pyruvate Kinase Deficiency (PKD) is a rare disease caused by mutations in the *PKLR* gene that leads to Chronic Non-Spherocytic Hemolytic Anemia (CNSHA). The only definitive treatment for severe cases of PKD is allogeneic Bone Marrow Transplantation (BMT). The risks associated to BMT, such as graft versus host disease, together with the low availability of suitable donors, make autologous cell therapy desirable for this disease. Patient specific induced Pluripotent Stem Cells (hiPSC) coupled with targeted gene correction via Homologous Recombination (HR), is a promising alternative for the treatment of hematopoietic inherited disorders. In order to prove the feasibility of this therapeutic alternative for PKD, we have generated integration free iPSCs from Peripheral Blood Mononuclear Cells (PB-MNC) of PKD patients (PKDiPSCs) using Sendai based viral vectors, and have corrected them through a Knock-In approach in the *PKLR* locus by using two types of DNA nucleases, Meganucleases (MG) and TALE nucleases (TALEN™). Different hiPSC clones were obtained from two patients and one healthy donor; these hiPSC clones showed pluripotent characteristics even after the disappearance of reprogramming vectors. A strategy to avoid lymphoid cells reprogramming within PB-MNC was successfully applied as neither T nor B cell receptor rearrangements were found in any of the analyzed hiPSC lines. Interestingly, erythroid differentiation of PKDiPSC was impaired as occurs in PKD patients, showing an accumulation of immature CD71/CD235a double positive erythroid cells and assessing the use of hiPSC for disease modeling. To restore the genetic defect, specific MG and TALEN™ were used to facilitate the Knock-In of a codon optimized *RPK* cDNA in the second intron of the *PKLR* gene. Whereas the MG generated DSB with very low specificity, after using the PKLR1 TALEN™, correct integration in *PKLR* locus was confirmed by PCR and southern blot, and the presence of the recombinant therapeutic RPK was assessed at the protein level. Surprisingly, allele specific integration due to the presence of a single nucleotide polymorphism was identified in one of the patients, pointing out its potential use in specific allele substitution. Genome integrity was examined by analyzing the appearance of *de novo* somatic mutations and Copy Number Variations (CNVs), detecting three single nucleotide variants and six CNVs in the corrected PKD2iPSC. The majority of them were already present before correction but not in the PB-MNC. These modifications did not include genes that were clearly associated either to a selective advantage or to the use of the nucleases. More importantly, gene corrected coPKDiPSCs displayed a normal erythroid maturation profile, similar to the one observed in wild-type hiPSCs. Overall, we show the feasibility of *PKLR* locus specific gene correction in patient specific iPSCs.

I Resumen

La Deficiencia en Piruvato Quinasa (DPQ) es una enfermedad rara causada por mutaciones en el gen *PKLR* que provoca Anemia Hemolítica no Esferocítica Crónica (AHNEC). El único tratamiento definitivo para los casos graves de DPQ es el Trasplante Alogénico de Médula Ósea (TAMO). Debido a los riesgos asociados a TAMO, como la enfermedad de injerto contra huésped, y la baja disponibilidad de donantes adecuados, hacen que la terapia celular autóloga sea una alternativa atractiva para el tratamiento de los casos graves de DPQ. Una alternativa prometedora para el tratamiento de trastornos hereditarios hematopoyéticos es la combinación de la generación de células madre pluripotentes inducidas (*iPSCs*) a partir de muestras de pacientes, junto con la corrección especifica de sitio mediada por Recombinación Homóloga (RH). Con el fin de evaluar esta posibilidad como alternativa terapéutica para DPQ, hemos generado *iPSC* derivadas de Células Mononucleares de sangre periférica (CMN) de pacientes con DPQ (DPQiPSCs) mediante vectores virales basados en el virus Sendai y las hemos corregido mediante RH específica de sitio asistida por dos tipos de nucleasas de ADN, Meganucleasas (MG) y TALE nucleasas (TALEN^{TM}). Para ello, hemos seguido una estrategia de *Knock-in* en el locus *PKLR*. Se obtuvieron diferentes clones de *hiPSC* a partir de dos pacientes de DPQ y un donante sano, de los cuales se confirmó la pluripotencia y la desaparición de los vectores de reprogramación. Con el fin de evitar la reprogramación de células linfoides, se llevó a cabo una estrategia basada en el uso de citoquinas, verificándose mediante el análisis de reordenamientos somáticos genómicos característicos de linfocitos B y T. Es importante señalar que la diferenciación eritroide de DPQiPSCs es deficiente tal como ocurre en los pacientes con DPQ, manifestando una acumulación de células eritroides inmaduras positivas para los marcadores CD71 y CD235a, lo que confirma el potencial uso de *hiPSC* para el modelado de enfermedades. Con el fin de restaurar el defecto genético, se han utilizado MG y TALEN^{TM} para facilitar la introducción de parte del ADNc de *RPK* terapéutico en el intrón 2 del gen *PKLR*. La MG presentó una muy baja especificidad para el gen *PKLR*, sin embargo, después del uso de PKLR1 TALEN^{TM}, se consiguió una correcta integración en un alto porcentaje de clones analizados mediante *PCR* y *Southern Blot*, así como la presencia de la proteína terapéutica RPK recombinante. Sorprendentemente, debido a la presencia de un polimorfismo en un único nucleótido en el ADN genómico de uno de los pacientes, se detectó exclusivamente integración específica en uno de alelos. Se examinó la integridad del genoma mediante el análisis de mutaciones somáticas y de Variaciones de Número de Copia (VNC), detectándose tres mutaciones y seis VNC en la muestra de DPQiPSC corregida, que no estaban presentes en la muestra de CMN de sangre periférica original. Estas modificaciones no afectan a ningún gen que confiera ninguna ventaja selectiva durante el

cultivo de *hiPSC* ni al uso de nucleasas. Por último, cabe destacar que las células DPQiPSC corregidas, una vez diferenciadas a al linaje eritroide, muestran un perfil de maduración normal, similar al observado en las hiPSCs de CMN de un donante sano. Como conclusión, los resultados presentados muestran la posibilidad de generar eritrocitos genéticamente corregidos específicamente en el locus *PKLR,* a partir de *hiPSC* de pacientes de DPQ.

II. Index

I. Summary *9*

I Resumen *10*

II. Index *13*

III. Abbreviations *17*

IV. Introduction *21*

1. Pyruvate Kinase Deficiency (PKD) **21**

1.1 Pyruvate Kinase (PK) and erythroid differentiation 22

1.2 Mutations, clinical manifestations and treatment 23

1.3 New treatments development 25

1.3.1 Cell and Gene Therapy 25

1.3.2 In situ gene correction 26

2. Induced Pluripotent Stem Cells (iPSCs) **27**

2.1 Choice of Reprogramming platform 29

2.2.1 Sendai Viral Vectors 30

2.2 Choice of cell source for reprogramming 31

2.3 Erythroid differentiation of iPSCs 32

3. Gene correction of disease specific iPSCs **33**

3.1 Homologous recombination mediated gene editing 34

3.2 DNA Nucleases 36

3.2.1 Meganucleases 36

3.2.2 ZFN 37

3.2.3 TALE nucleases 37

3.2.4 CRISPR/Cas system 38

3.3 Limitations of the use of nuclease mediated gene editing, DSBs 38

3.4 Reported patient specific hiPSC gene correction 39

3.5 Limitation of the therapeutic use of hiPSCs: genomic instability 40

V. Aims *43*

VI. Materials and Methods *45*

1. Patient and healthy donor samples **45**

1.1 Patient specific mutations detection 45

2. Induced Pluripotent Stem Cells **45**

2.1 Generation ... 45
2.2 Culture conditions ... 46
2.2.1 Freezing/Thawing ... 46
2.3 hiPSC characterization ... 47
2.3.1 Alkaline Phosphatase ... 47
2.3.2 RT-PCR gene expression array ... 47
2.3.3 Teratomas ... 47
2.3.4 Promoter demethylation ... 48
2.3.5 Karyotyping ... 48
2.3.6 Immunoglobulin Chain and T cell receptor rearrangement analysis ... 48
2.3.7 Comparative Genomic Hybridization (CGH) ... 48
2.3.8 Exome sequencing ... 48
2.4 Erythroid Differentiation ... 49

3 Homologous Recombination enhanced by Nucleases ... 49
3.1 Plasmids ... 49
3.2 mRNA synthesis ... 50
3.3 Nuclease cleavage activity ... 50
3.3.1 MG PKLR-specific DSBs generation analysis by deep sequencing ... 50
3.3.2 Surveyor assay ... 51
3.4 Generation of the targeting matrix ... 51
3.5 Homologous recombination in K-562 cell line ... 53
3.5.1 PCR assessment of Homologous Recombination ... 53
3.5.2 Erythroid differentiation of K-562 ... 53
3.6 Homologous recombination in hiPSCs ... 53
3.6.1 HR analysis ... 53
3.6.1.1 Colony PCR ... 53
3.6.1.2 Southern Blot ... 54
3.6.2 Western Blot ... 54

4 General techniques ... 55
4.1 Immunohistochemistry ... 55
4.2 Flow cytometry ... 56
4.3 Gene expression analysis ... 56

5 Primers Table ... 58

6 Antibodies Table ... 59

VII Results ... 61

1. Cell reprogramming ... 61
1.1. Generation of hiPSCs from Peripheral Blood Monuclear Cells (PB-MNC) ... 61

1.1.1 Targeting non-lymphoid cells for reprogramming 62
1.1.2 Pyruvate Kinase Deficient (PKD) patient and healthy donor reprogramming by SeV 63
1.1.2.1 hiPSC characterization 65
D) Reprogramming vectors disappearance 70
1.2 Erythroid differentiation of PKDiPSC 70
1.2.1 RPK expression during in vitro erythroid differentiation 70
1.2.2 Pyruvate Kinase deficiency phenotype in PKDiPSC derived erythrocytes 72

2 Nuclease mediated *Knock In* gene correction in *PKLR* locus 74
2.1 Homologous recombination (HR) strategy design 74
2.2 PKLR Meganuclease Efficiency 75
2.2.1 PKL01 MG versions comparison in 293T cell line by deep sequencing 75
2.2.2 PKLR Meganuclease dose optimization 76
2.3 HR analysis in *PKLR* locus using K-562 cell line 77
2.3.1 PCR assessment of Homologous Recombination 78
2.3.2 Correct splicing assessment 78
2.4 HR in PKD2iPSC facilitated by PKLR MG 79
2.5 HR by Transcription activator-like effector nucleases (TALEN™) 81
2.5.1 DSB generation efficiency 81
2.5.2 HR mediated gene correction in PKDiPSC from two different patients 82
2.5.2.1 Non Homologous End Joining (NHEJ) in untargeted *PKLR* locus 83
2.5.2.2 Off target analysis 85
2.5.2.3 Allele specific targeting 86
2.5.2.4 Pluripotency maintenance after HR 87
2.5.2.6 Flagged RPK expression in differentiated coPKD2iPSC 88
2.5.2.5 Erythroid differentiation in corrected PKD2iPSC 88
2.6 Genome Stability analysis 89

***VIII Discussion* 93**

1 Cell Reprogramming 93
1.1 Generation of hiPSCs from PKD patients for gene correction 93
1.1.1 Reprogramming on a safe manner 94
1.1.2 Use of PB-MNC as cell source for reprogramming 95
1.1.3 PKD and healthy samples reprogramming 96
1.1.4 PKD phenotype recapitulation during PKDiPSC erythroid differentiation 97

2 PKDiPSC Knock-In gene correction 98
2.1 Meganuclease mediated homologous recombination 100
2.2 TALEN™ mediated homologous recombination 101
2.2.1 Allele specific targeting 103
2.2.2 Transgene expression and phenotype restoration in corrected PKDiPSC 104

2.3 Genome stability analysis.....104

3 Future perspectives 106

***IX Conclusions*.....109**

***IX Conclusiones*.....110**

***X References*.....111**

***XI Annex*.....123**

III. Abbreviations

A1ATD-hiPSC: Alpha 1 Antitrypsin Deficiency induced Pluripotent Stem Cells

Aa: Amino Acids AAVS1: Adeno-Associated Virus Integration Site 1

aCGH: array based Comparative Genomic Hybridization

ADN: Ácido Desoxirribonucleico

ADNc: ADN copia

AHEC: Anemia Hemolítica No-Esferocítica Crónica

AP: Alkaline Phosphatase

APEL: Albumin Polyvinylalcohol Essential Lipids

ATP: Adenosine Triphosphate

bFGF: basic Fibroblast Growth Factor

bGH: bovine Growth Hormone

BGM: Blast Growing Media

BM: Bone Marrow

BM-MSC: Bone Marrow Mesenchymal Stem Cells

BMP4: Bone Morphogenetic Protein 4

BMT: Bone Marrow Transplantation

bp: base pairs

BSA: Bovine Serum Albumin

Cas: CRISPR associated system

CMN: Células Mononucleadas de Sangre Periférica

CNSHA: Chronic Non-Spherocytic Hemolytic Anemia

CNV: Copy Number Variation

coPKDiPSC: corrected Pyruvate Kinase Deficiency induced Pluripotent Stem Cells

coRPK: codon optimized RPK

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

DAPI: 4',6-Diamidino-2-Phenylindole, Dihydrochloride

DHA: Docosahexaenoic Acid

DM: Donor Matrix

DMEM: Dulbecco's modified Eagle's medium

DMSO: Dimethyl sulfoxide

DNA: Deoxyribonucleic acid

DPQ: Deficiencia en Piruvato Quinasa

DPQiPSC: Células Madre Pluripotentes inducidas de Deficiencia en Piruvato Quinasa

DSB: Double Strand Breaks

EBs: Embryoid Bodies

EC: Expression Cassette

ECL: Enhanced Chemiluminescence

EDTA: Ethylenediaminetetraacetic Acid

EF1 α: Elongation Factor 1 α

FACS: Fluorescence Activated Cell Sorting

FA-hiPSC: Fanconi Anemia Human induced Pluripotent Stem Cells

FBS: Fetal Bovine Serum

FLT3L: Fms-Related Tyrosine Kinase 3 Ligand

FV: Foamy Virus

GAPDH: Glyceraldehyde 3-phosphate dehydrogenase

G-CSF: Granulocyte Colony-Stimulating Factor

gDNA: genomic DNA

GFP: Green Fluorescent Protein

GT: Gene Therapy

GVHD: Graft Versus Host Disease

HA: Homology Arms

HE: Homing Endonucleases

hES: human Embryonic Stem

hESCs: human Embryonic Stem Cells

HFF: Human Foreskin Fibroblasts

hiPSCs: Human induced Pluripotent Stem Cells

HIV: Human Immunodeficiency Virus

hPSCs: human Pluripotent Stem Cells

HR: Homologous Recombination

HSCs: Hematopoietic Stem Cells

IgH: Immunoglobulin Heavy Chain

IL3: Interleukin 3

IL-6: Interleukin 6

ISCA: International Standards for Cytogenomics Array

lncRNA: Long non-coding RNA

LOH: Loss of Heterozygosity

LPK: liver pyruvate kinase

mCpGs: Methylated CpG dinucleotide

MEF: Mouse Embryonic Fibroblasts

MG: Meganucleases

mRNA: messenger RNA

NHEJ: Non-Homologous End Joining

NSG: Non-obese diabetic SCID gamma

Nts: nucleotides

OMIM: Online Mendelian Inheritance in Man

OpPCR: Overlapping PCR

PB: Peripheral Blood

PB1: Peripheral Blood from healthy donor number 1

PB2: Peripheral Blood from healthy donor number 2

PBA: BSA and Sodium Azide in PBS

PB-MNC: Peripheral Blood Mononuclear Cells

PBS: Phosphate Buffer Saline

PCR: Polymerase Chain Reaction

PFA: Paraformaldehyde

PK: Pyruvate Kinase

PKD: Pyruvate Kinase Deficiency

PKD1: Pyruvate Kinase Deficient patient number 1

PKD2: Pyruvate Kinase Deficient patient number 2

PKD3: Pyruvate Kinase Deficient patient number 3

PKDiPSCs: Pyruvate Kinase Deficiency induced Pluripotent Stem Cells

PKLR: Pyruvate Kinase, Liver and Red Blood Cells

PKM: Pyruvate Kinase, Muscle

Puro-MEF: Puromycin resistant mouse embryonic fibroblasts

qRT-PCR: quantitative Real Time PCR

RBCs: Red Blood Cells

RH: Recombinación Homóloga

RNA: Ribonucleic acid

RPK: R type Pyruvate Kinase

RT: Room Temperature

SA: Splicing Acceptor

SC: Selection Cassette

SCF: Stem Cell Factor

SDS: sodium dodecyl sulfate

SELEX: Systematic Evolution of Ligands by Exponential Enrichment

SeV: Sendai Viral Vectors

SNP: Single Nucleotide Polymorphism

SNV: Single Nucleotide Variants

SR-RNA: Self-Replicating RNAs

SSC: saline-sodium citrate

TA: TALENTM: Transcription Activator-Like Effector Nucleases

TAMO: Trasplante Alogénico de Médula Ósea

TBS: Tris-Buffered Saline

TBS-T: Tris-Buffered Saline Tween

TCR: T Cell Receptor

TF: Transcription Factors

TPO: Thrombopoietin

UV: Ultraviolet

VCC: Variaciones de Número de Copia

VEGF: Vascular Endothelial Growth Factor

YS: Yolk Sac

ZFNs: Zinc Finger Nucleases

IV. Introduction

1. Pyruvate Kinase Deficiency (PKD)

PKD (OMIM 266200) is a rare erythroid metabolic disease diagnosed for the first time in the early 1960s (Valentine, Tanaka et al. 1961), caused by mutations in *PKLR* gene, which codes the R-type Pyruvate Kinase (RPK) in erythrocytes. It shows a worldwide geographical distribution and the most common type of genetic inheritance is autosomal recessive. Its incidence is around 51 cases per million in the North American white population (Beutler and Gelbart 2000) and together with glucose-6-phosphate deficiency (G6PD) are the most common causes of Chronic Non-spherocytic Hemolytic Anemia (CNSHA). The majority of the diagnosed patients are compound heterozygotes, as homozygotes are rare but very severe (van Wijk, Huizinga et al. 2009).

The human genome encodes four Pyruvate Kinase (PK) Isotypes: M1, M2, R and L and its expression is tissue specific. M1 and M2 are encoded by the *PKM* gene and are expressed in Muscle and Brain (M1) and lung and many types of highly proliferating cells (M2) (Tsutsumi, Tani et al. 1988). RPK and LPK are expressed from the *PKLR* gene, the first one in erythrocytes and the second in liver, kidney and intestine. Two different tissue specific promoters drive the transcription of these two isoforms. The length of the *PKLR* gene is 9.5 Kilobases (Kb) and is located in Chromosome 1 (1q21) (Fig. 1). *RPK* and *LPK* transcripts differ in the first exon, which is different in each transcript, sharing all the downstream exons (Kanno, Fujii et al. 1992). *RPK* cDNA is 2060 base pairs (bp) long and codes for 574 amino acids (aa), while *LPK*s cDNA and protein length are 1632 bp and 543 aa respectively.

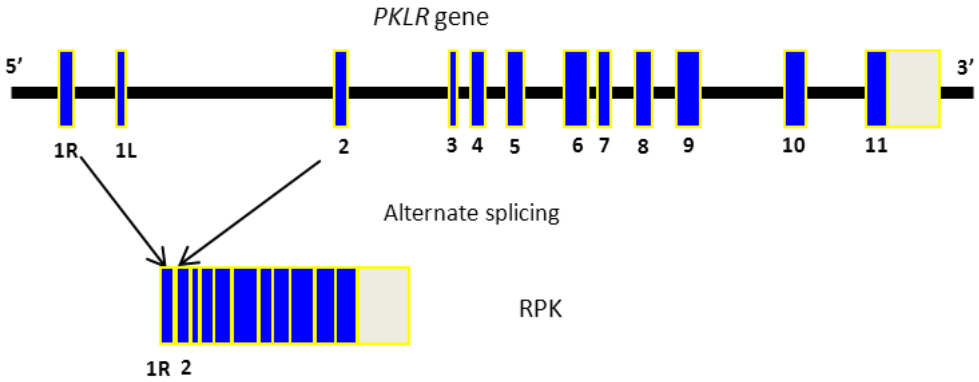


Figure 1: *PKLR* gene structure. Exons 1 to 11 are represented in blue and untranslated region (UTR) in grey. There are two exons 1 that by alternative splicing are going to generate the R-isoform or the L-isoform respectively.

PK catalyzes the conversion of phosphoenolpyruvate to pyruvate in Embden-Meyerhof pathway of anaerobic glycolysis (Zanella, Bianchi et al. 2007) (Fig. 2), which is highly important in mature

erythrocytes. The disease becomes clinically relevant when, due to *PKLR* mutations, the protein activity decreases below 25% of the normal activity in erythrocytes (Certo, Gwiazda et al. 2012). ATP depletion leads to ATP-dependent cation pumps failure disrupting the adequate cell electrolyte concentration. This causes damage in erythrocyte cell membrane leading to early erythrocyte destruction in liver and spleen (Mentzer, Baehner et al. 1971).

1.1 Pyruvate Kinase (PK) and erythroid differentiation

During erythroid differentiation there is a switch in the presence of the different PK isoenzymes (Fig 2). M2PK is expressed in immature proerythroblasts but at the stage of polychromatic erythroblast it decreases while RPK levels rise, being the only PK enzyme expressed in mature erythrocytes. Mature erythrocytes are then highly affected by alterations in RPK function. As erythroid cells lose the nucleus and mitochondria, they become highly dependent on glycolysis for ATP production. Mature erythrocytes do not have residual M2PK and as they do not have nucleus and cannot compensate RPK defect by increasing RPK levels. Although the majority of *PKLR* mutations affect both isoforms, no hepatic affectation is observed in PKD patients, mainly because hepatocytes can compensate the defect in two ways: 1) when the genetic defect leads to a loss of activity, the level of enzyme generated can be increased by increasing gene expression, and 2) residual levels of M2PK could be present (Zanella, Bianchi et al. 2007).

RPK protein structure varies during erythrocyte differentiation. Whereas eritroblastic RPK is a homotetramer of 4 L subunits (unmodified), RPK isolated from mature erythrocytes is a heterotetramer composed by 2 L and 2 Lc subunits, which are generated by proteolysis of L in its N terminal region. L molecular weight is 62 kDa whereas Lc is 58 kDa and the catalytic constant of the heterotetramer is lower than the one from the homotetramer (van Dongen, Langerak et al. 2003, Ju, Kim et al. 2011).

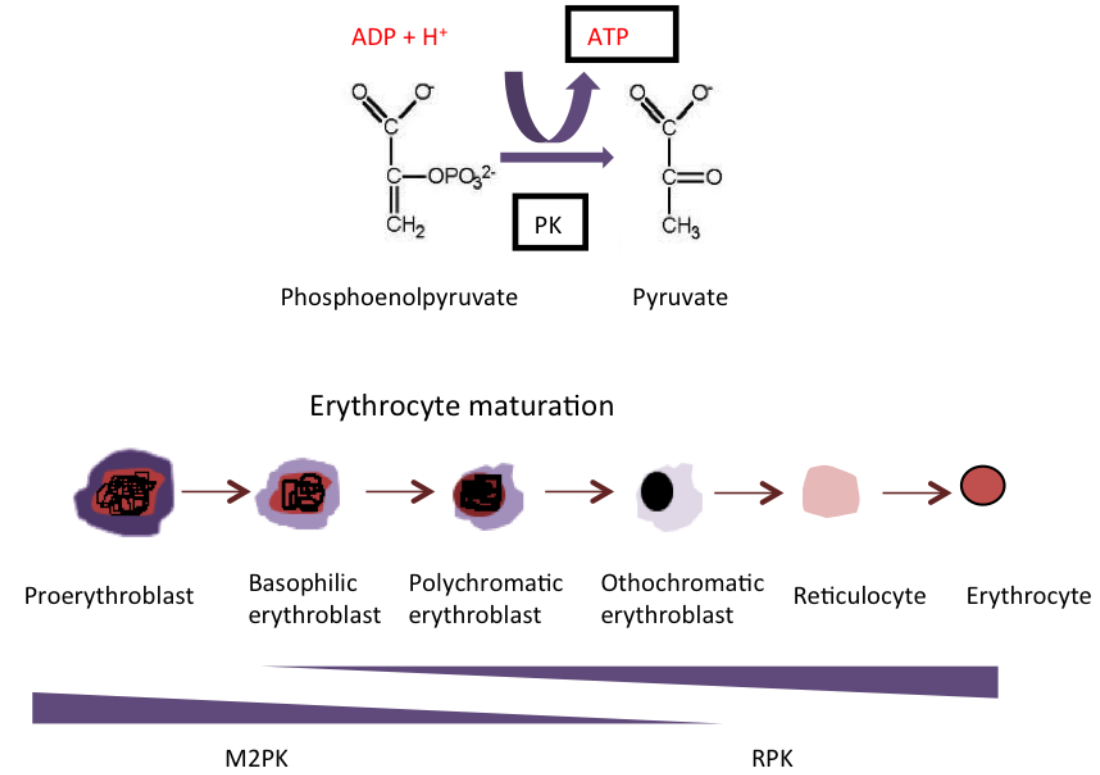


Figure 2: Energy generation during erythrocyte maturation. Top: Pyruvate Kinase (PK) catalyzes the conversion of phosphoenolpyruvate into pyruvate and its consequent generation of ATP. Bottom: The presence of the different PKs varies during erythropoiesis. In early progenitors, M2PK isoform catalyzes this reaction whereas in mature erythrocytes RPK is the responsible one.

1.2 Mutations, clinical manifestations and treatment

Up to now, more than 190 PKD-associated mutations have been identified throughout the whole *PKLR* gene (Zanella, Fermo et al. 2007, Kedar, Hamada et al. 2009, van Wijk, Huizinga et al. 2009, Machado, Manco et al. 2012). The number of mutations described in exons 8 to 11 is higher than in the rest of exons, and mutations in exon 1R and 1L have never been described. The mutations have been designated using the *RPK* cDNA sequence, being the A from the ATG the +1 (Genebank Accession number: NM_000298.5). The majority are missense mutations (69%), but splice sites (11%) and stop codon (5%) mutations have also been identified. Small deletions, insertions and frameshift mutations are rare (12%). Two large deletions have been reported: the "Gypsy" deletion of the whole exon 11 and the PK "Viet", which leads to the loss of exons 4 to 10. A few mutations at the promoter region have also been described (-72A→G and -83G→C) and have been associated to defects in gene regulation. The most common ones are 1529A, 1456T and 1468T, being the first one the most common in USA and in Northern and central Europe. (Zanella, Fermo et al. 2005).

Alberto Zanella and colleagues established a correlation between mutations and clinical phenotype in 54 PKD patients (Zanella, Fermo et al. 2007). They identified three phenotypes: 1) Severe phenotype, in which patients presented severe neonatal jaundice and most of them required blood

transfusions, splenectomy and one case of death *in utero*. The associated mutations included disruptive mutations (stop codons, frame shifts, splicing and large deletions) and missense mutations where amino acids from the active site or crucial for protein stability were affected. 2) Moderate phenotype, usually diagnosed around the age of 25 and the majority showing 1529A mutation. 3) Mild phenotype: showing mild or fully compensated hemolysis and diagnosed in adulthood. The predominant mutation in this group is: 1456T. Although not included in this study because the known incidence is very low, patients suffering homozygous ¨null mutations¨ would also be included in the first category, as they usually result in a total loss of protein activity leading to the reported symptoms and in some cases intrauterine growth retardation (Zanella and Bianchi 2000, Sedano, Rothlisberger et al. 2004).

Being its incidence between 5 and 51 cases per million in the North American white population, PKD prevalence might be influenced by the facts that its diagnosis is mainly performed in some population groups and that it is based on observed mutation frequencies (Beutler and Gelbart 2000). Therefore, the frequencies of unreported PKD cases due to *hydrops fetalis* and early neonatal death may increase its incidence (Gilsanz, Vega et al. 1993, Certo, Gwiazda et al. 2012). Based on the fact that heterozygous PKD mutants show a decreased RPK activity, its frequency was studied in Spanish population (Garcia, Moragon et al. 1979) and it was determined that the frequency of a mutated allele in the healthy population is 0.24%.

Overall, PKD patients show a highly variable phenotype. Depending on the severity of the mutation, the symptomatology can range from fully compensated hemolysis with no apparent anemia to fetal death *in utero* (Gilsanz, Vega et al. 1993, Zanella, Fermo et al. 2005). The most common symptoms are CNSHA and jaundice in new-borns. Patients suffering from severe anemia may need periodical blood transfusions. Splenectomy is also sometimes recommended, as it decreases erythrocyte destruction. Nevertheless, the only definitive cure for PKD is allogeneic bone marrow transplantation (BMT), which was successfully carried out in a 5-year-old boy who developed neonatal jaundice and severe transfusion-dependent hemolytic anemia (Suvatte, Tanphaichitr et al. 1998, Tanphaichitr, Suvatte et al. 2000). Three years after the BMT, he showed normal hemoglobin levels and normal red blood cell pyruvate kinase activity without evidence of hemolysis. This treatment usually requires the use of immunosuppressive drugs or regimes that can induce toxicity and increase the risk of infections and cancer. This, together with the low availability of suitable donors makes the development of new treatment options a need to treat severe patients.

1.3 New treatments development

1.3.1 Cell and Gene Therapy

As an alternative to avoid BMT complications, other approaches such as gene therapy of autologous hematopoietic stem cells (HSCs) or the transfusion of *in vitro* generated erythrocytes have also been proposed. Cell therapy and related applications should be developed and tested in the very valuable PKD animal models before they are used in the clinics. Natural mutations in homologous genes of the human *PKLR* have been described in several animal models including mice (Morimoto, Kanno et al. 1995, Min-Oo, Fortin et al. 2004), cats (Kohn and Fumi 2008, Grahn, Grahn et al. 2012) and dogs (Whitney and Lothrop 1995, Beard and Kiem 2009). These models are of enormous importance for the study of the physiopathological characteristics of the disease and to explore new treatment choices. They have already been used in several reports, especially in the gene and cell therapy field. The first attempt of cell therapy was performed in the PKD canine model, in which the feasibility of non-myeloablative BMT was tested as PKD treatment. The achieved level of chimerism was not enough to accomplish a long-term cure (Zaucha, Yu et al. 2001). Nevertheless, similar levels of chimerism in an equivalent approach carried out in a PKD mouse model (CBA-Pk-1slc/Pk-1slc) were successful in PKD phenotype correction (Morimoto, Kanno et al. 1995). In order to use autologous HSCs for PKD treatment, previous work from our group explored the use of human RPK expressing γ-retroviral vectors to treat the disease in a mouse model of PKD, by transplanting autologous BM hematopoietic progenitors and HSC transduced with the developed retroviral vectors into myeloablated PKD mice (Meza, Alonso-Ferrero et al. 2009). In this study, the PKD mouse model used showed a phenotype very similar to the one seen in PKD patients (Min-Oo, Fortin et al. 2004). Long-term expression of RPK was achieved in red blood cells of primary and secondary recipients and PKD associated hematological manifestations were reverted without showing any adverse effects. The latest reported PKD gene therapy (GT) approach used a foamy virus (FV) engineered vector to express the wild type *RPK* cDNA. For this study, one PKD dog was transplanted with autologous $CD34^+$ hematopoietic cells transduced with the *RPK* FV resulting in a long-term functional cure (Trobridge, Beard et al. 2012).

Periodical blood transfusion is a routine therapeutic procedure for PKD treatment. However, adverse effects have been observed in some patients treated with frequent blood transfusions, such as alloimmunization against donor blood cells. This and others side effects made them refractory to future additional blood transfusions. Consequently, autologous blood transfusions after gene correction might overcome this problem, which could be obtained by the *in vitro* generation of terminally differentiated and functional red blood cells (RBC) (Migliaccio, Whitsett et al. 2012,

Zeuner, Martelli et al. 2012). Many attempts have been carried out by using different sources of stem cells such as hematopoietic progenitors (Giarratana, Rouard et al. 2011) and induced pluripotent stem cells (Lu, Feng et al. 2008). By using peripheral blood isolated $CD34^+$ cells, Giarratana et al generated and expanded RBC *in vitro* and *in vivo* which showed all the biochemical and functional characteristics of normal RBCs. More importantly, they were capable of transfusing the same donor with the *in vitro* generated chromium labeled autologous erythrocytes and showed that after 26 days, between 41% and 63% of the total infused cells survived.

Another disease model that could also be used to study PKD and to develop new therapeutic options is patient specific induced Pluripotent Stem Cells (iPSCs). Human iPSCs show unlimited self-renewal capacity while retaining the ability, in principle, to differentiate into any cell type of the human body (Takahashi, Tanabe et al. 2007). This includes the hematopoietic system and which is more important for PKD study, the opportunity to study the whole erythropoietic differentiation process (Lu, Feng et al. 2007, Lu, Feng et al. 2008, Hatzistavrou, Micallef et al. 2009). Due to this enormous potential, many disease-specific iPSC lines have been generated and in some cases genetically corrected (Garate, Davis et al. 2013). These aspects will be deeply analyzed in depth through-out this thesis manuscript.

1.3.2 In situ gene correction

The viral vectors used in conventional GT integrate randomly in the genome and this fact could cause insertional mutagenesis and promote the activation of surrounding oncogenes, which may result in oncogenic events (Hacein-Bey-Abina, Von Kalle et al. 2003, Hacein-Bey-Abina, Garrigue et al. 2008, Fischer, Hacein-Bey-Abina et al. 2010). Newly designed vectors are exploring the possibility of using safer strategies but until now no definitive ideal vector has been reported. The identification of integrations in safe harbor genomic sites (e.g., far away from genes or coding information) in clonal cultures could represent an alternative. However, the selection of cells with safe integrations requires long culture periods, incompatible with the maintenance of stem cell properties of HSCs. The clonal nature of pluripotent cultures could allow this type of selection and the concomitant differentiation of hiPSCs to HSCs could provide the proper cells for safe hematopoietic transplantation. In two different approaches performing lentiviral gene correction in thalassemia and erythropoietic porphyria patient iPSC (Papapetrou, Lee et al. 2011, Bedel, Taillepierre et al. 2012), the clonal expansion of corrected patient hiPSC allowed the analysis of the integration of these vectors at a clonal level a and the selection of the ones integrated in safe harbors. However, the definition of a safe harbor is complex and might be influenced by unknown genomic functions. Special attention should be taken to the newly generated information from ENCODE project, which is showing unannotated genomic functions (http://www.encodeproject.org/ENCODE/pubs.html or

http://www.nature.com/ENCODE/). Future therapeutic applications of hiPSCs for cell therapy would benefit from a site-specific gene correction approach. The cooperation between hiPSC technology and gene editing, which is the introduction of a piece of genetic information in a specific site of the genome on a controlled manner via homologous recombination (HR), has been extensively explored. HR is presented as an exciting and novel alternative, to avoid insertional mutagenesis associated with integrative vector-mediated correction. Its possibilities will be extensively reviewed later.

2. Induced Pluripotent Stem Cells (iPSCs)

Human Pluripotent Stem Cells (human Embryonic Stem Cells (hESCs) and hiPSCs) are a powerful source of cells for regenerative medicine, as they have the ability to differentiate into any cell type of the human body. Thomson's group described the isolation of hESCs from the Inner Cell mass of human blastocysts in 1998 (Thomson, Itskovitz-Eldor et al. 1998) establishing the pluripotent culture conditions that were later used for the derivation of human induced Pluripotent Stem Cells. Human iPSC technology combines the potential of hESCs in terms of self-renewal and pluripotency with the absence of problems related to hESC lines establishment and clinical application, (i.e. ethical issues associated with embryo disruption and immuno-incompatibility with the recipient of the cells). In addition to their potential as a pluripotent autologous cell source, iPSCs clonal nature makes them an ideal platform to perform gene editing mediated gene correction. iPSCs generation was reported for the first time in 2006 in the mouse system by the transduction of fibroblasts with the following transcription factors (TFs): *Oct4, Sox2, Klf4 and C-Myc* (Takahashi and Yamanaka 2006). One year later the groups of Yamanaka (Takahashi, Tanabe et al. 2007) and Thomson (Yu, Vodyanik et al. 2007) achieved the same result in human cells revolutionizing the stem cells field. Many laboratories have been able to reprogram a large variety of somatic cells into pluripotent stem cells, from neural stem cells (Kim, Zaehres et al. 2009) to terminally differentiated B-lymphocytes (Hanna, Markoulaki et al. 2008). The reproducibility and potentiality of the iPSC technology has caused this field to advance very rapidly worldwide.

Human iPSCs derived from patients with specific disorders are presented as a very valuable material to study human diseases *in vitro* and also *in vivo*. From a clinical point of view, two main applications can be pursued (Fig. 3):

1) Disease modeling and drug development. Patient specific hiPSC could be a model of the patient pathology. As they have the ability to differentiate into any cell type of the human body, they could resemble the disease phenotype in the desired tissue and specific drugs could be tested. This strategy has been used to test drugs used for arrhythmias in hiPSC derived beating cardiomyocytes (Braam, Tertoolen et al. 2013), drugs for Niemann Pick type C disease in hiPSC derived neurons

(Swaroop, Thorne et al. 2012) or to study the responsiveness to DHA (Docosahexaenoic Acid) in Alzheimer's disease iPSC derived neurons (Kondo, Asai et al. 2013). Although yet not applied, many authors have also proposed the use of iPSC-derived hepatocytes for drug toxicity testing (Greenhough, Medine et al. 2010, Nakamura, Saeki et al. 2012)

2) Cell source for regenerative medicine: As iPSCs can be generated from a concrete patient, they could serve as an unlimited source of autologous cells for replacement (Qin, Song et al. 2013, Yuan, Liao et al. 2013). In the case of a genetic disease, the mutation could be corrected on a safe manner before differentiating the cells towards the tissue of interest. This has not been applied in clinics yet, but many proof of principles have been reported in literature (Raya, Rodriguez-Piza et al. 2009, Howden, Gore et al. 2011, Liu, Suzuki et al. 2011, Sebastiano, Maeder et al. 2011, Soldner, Laganiere et al. 2011, Yusa, Rashid et al. 2011, Zou, Mali et al. 2011, Zou, Sweeney et al. 2011, An, Zhang et al. 2012, Bedel, Taillepierre et al. 2012, Wang, Zheng et al. 2012)

There is an additional application for hiPSCs under development, which is its use as an immunotherapy tool by *in vitro* generation of T lymphocytes with reactivity against a specific antigen. This application has been explored for HIV (Nishimura, Kaneko et al. 2013), for cancer (Vizcardo, Masuda et al. 2013) and for pathogen defense (Wakao, Yoshikiyo et al. 2013) with promising results.

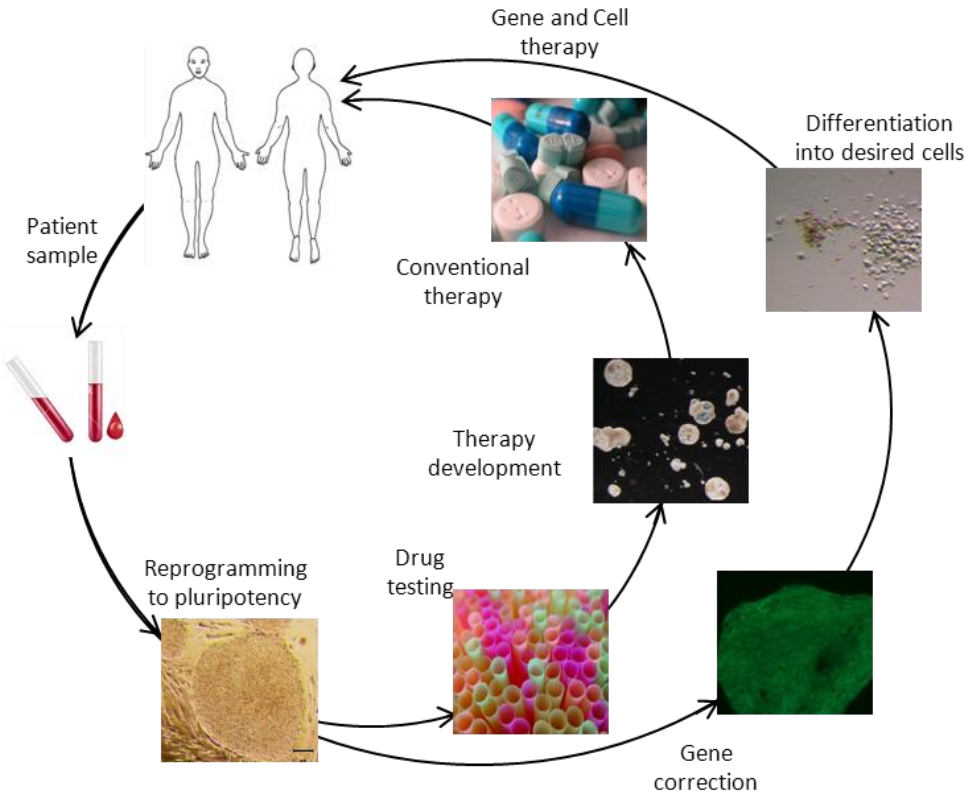


Figure 3. Biomedical potential of hiPSCs. Cells obtained from a patient are reprogrammed to pluripotency. Due to their acquired capacity to differentiate into any cell type, these cells could serve for disease modeling and drug screening and also could be genetically corrected to serve as source for regenerative medicine, Modified from (Garcia-Gomez 2013)

2.1 Choice of Reprogramming platform

Since their initial description, many different procedures have been described to generate iPSCs. Viral vectors have been the preferred system because of their ability to express a gene of interest in a certain cell type due to their natural cell entry and replication mechanisms. That is why first reports of hiPSC generation used retroviruses to express the four defined TFs required for reprogramming (Takahashi, Tanabe et al. 2007, Yu, Vodyanik et al. 2007). Nevertheless, a future clinical application of hiPSC will ideally involve the use of non-integrative systems, in which the risk of insertional mutagenesis due to random integration of the reprogramming genes is avoided. Many different integrative, removable and non-integrative systems have been developed with its strengths and limitations (Table 1). First reported approaches for reprogramming were carried out by delivering each TF separately in different viral vectors whereas very soon, other groups reported the inclusion of all of them in a unique polycistronic lentiviral vector, which improved reprogramming efficiency (Carey, Markoulaki et al. 2009).

Two of the four Yamanaka TFs, *C-MYC* and *KLF-4*, have been designated as oncogenes and their presence in the reprogramming cocktail was seen as a risk factor. It was shown that it was possible to remove them from the reprogramming cocktail but just for certain cell types, like hematopoietic progenitors (Liu, Zou et al. 2012, Meng, Neises et al. 2012) or neural stem cells (Kim, Zaehres et al. 2009). Even reprogramming with *OCT4* alone was achieved (Thier, Munst et al. 2010). Unfortunately, this kind of approaches usually decreased reprogramming efficiency.

Another explored integrative possibility is the use of Cre-mediated excisable polycistronic lentiviral vectors (Somers, Jean et al. 2010, Papapetrou and Sadelain 2011) or transposon-based reprogramming systems (Woltjen, Hamalainen et al.), which could be removed after obtaining the hiPSC clones. In the first case just a LoxP signal will remain after excision by Cre recombinase and in the second the genome should end up, in principle, completely unmodified.

However, the safest reprogramming procedure will eventually involve non-integrative systems. These systems are less genotoxic as the chance of integration is low (plasmids, adenovirus) or even null (Sendai viral vectors (SeV), mRNAs or proteins). Thomson's group reported the first strictly non-integrative system for hiPSCs generation in 2009 by using episomal plasmids for the expression of the four Yamanaka TF plus *NANOG*, *LIN28* and SV40 large T antigen (*SVLT*) (Yu, Hu et al. 2009). In the same year, reprogramming by recombinant proteins (Kim, Kim et al. 2009), synthetic mRNAs (Warren, Manos et al. 2010), and non-integrating RNA SeV (Fusaki, Ban et al. 2009) was also reported. Reprogramming with recombinant proteins or mRNAs, although ideal, shows a reduced efficiency and the need for multiple rounds of transfection/protein delivery. SeV, which combine the safety of non-integrating systems with a high efficiency, are a very promising choice for cell

reprograming. Very recently another very promising non-integrative approach came out based on synthetic self-replicating RNAs expressing reprogramming factors (Yoshioka, Gros et al. 2013). Non integrative systems are not only safer but also generate iPSC showing a gene expression profile more similar to ESC than iPSC generated by integrative methods (Liu, Cheng et al. 2012).

TF delivery method	Efficiency/ Rounds	Integrative	Cell type
Lenti/ retroviruses	0.01-1/ single	Yes	Fibroblasts (Takahashi, Tanabe et al. 2007) Keratinocytes(Aasen and Izpisua Belmonte 2010) B-Lymphocytes(Hanna, Markoulaki et al. 2008) Human Adipose Stem cells (Sun, Panetta et al. 2009) BM-MSCs (Zou, Sweeney et al. 2011) CD34 Progenitors (Ye, Zhan et al. 2009) BM-MNC (Kunisato, Wakatsuki et al. 2010) Hepatocytes(Hansel, Gramignoli et al. 2013) Pancreatic Islet Beta Cells(Bar-Nur, Russ et al. 2011) Neural Stem Cells (Kim, Zaehres et al. 2009)
Adenoviruses	≈ 0.0002/ multiple	No	Fibroblast (Zhou and Freed 2009)
Sendai virus	≈ 1/single	No	Fibroblasts (Fusaki, Ban et al. 2009) T cells (Seki, Yuasa et al. 2011) CD34 progenitors (Ban, Nishishita et al. 2011)
Excisable Lentivirus	0.01-1/single	Removable	Fibroblasts (Sommer, Stadtfeld et al. 2009, Papapetrou and Sadelain 2011, Sebastiano, Maeder et al. 2011, Soldner, Laganiere et al. 2011) CD34 Progenitors (Ramos-Mejia, Montes et al. 2012)
Excisable Transposons	≈ 0.1/single	Removable	Fibroblasts (Kaji, Norrby et al. 2009) BM-MSCs (Zou, Mali et al. 2011)
Minicircles	≈ 0.005/ multiple	No	Adipose derived Stromal cells (Jia, Wilson et al. 2010)
Episomal Vectors	0,001/multiple	No	Fibroblasts (Yu, Hu et al. 2009) BM and CB MNCs (Hu, Yu et al. 2011)
Proteins	0,001/multiple	No	Fibroblasts (Kim, Kim et al. 2009)
miRNAs	0,1/single	Yes	Fibroblasts (Anokye-Danso, Trivedi et al. 2011)
mRNAs	1-4/multiple	No	Fibroblasts (Warren, Manos et al. 2010)
Self-replicating RNAs	0.01-1/single	No	Fibroblasts (Yoshioka, Gros et al. 2013)

Table1. Technological platforms and cell sources for reprogramming, modified from (Garate, Davis et al. 2013)

2.2.1 Sendai Viral Vectors

SeV are presented as a very promising tool for gene transfer in mammalian cells and therefore a very good tool for iPSC generation. These virus were isolated for the first time in Japan in 1950 (Kuroya and Ishida 1953) and have been associated with respiratory tract infections in many rodents and occasionally in pigs, being neither pathogenic nor tumorogenic for humans. As seen in figure 4, one of the major advantages of this vector is the fact that the whole replication process takes place as RNA and exclusively in the cytoplasm of the transduced cell without any risk of chromosomal integration. Additionally, as it is cell replication-independent, SeV do not need a specific cell cycle status. SeV also show a higher transgene expression rate than other common viral vectors (Yu,

Shioda et al. 1997). Several clinical trials have been carried out using SeV showing no adverse effects (Masaki, Yonemitsu et al. 2001, Masaki, Yonemitsu et al. 2002, Slobod, Shenep et al. 2004). All the characteristics that made the vectors based in this virus attractive for clinical purposes made them also an ideal tool for hiPSC generation. They can be used for many starting cell types with a high transduction efficiency leading to high reprogramming efficiency (≈1%). Their ability to replicate during the beginning of reprogramming avoids multiple transductions, but when iPSCs are established and divide very rapidly, the viral vectors tend to disappear, increasing the safety of the method (Fusaki, Ban et al. 2009, Ban, Nishishita et al. 2011, Ye, Muench et al. 2013).

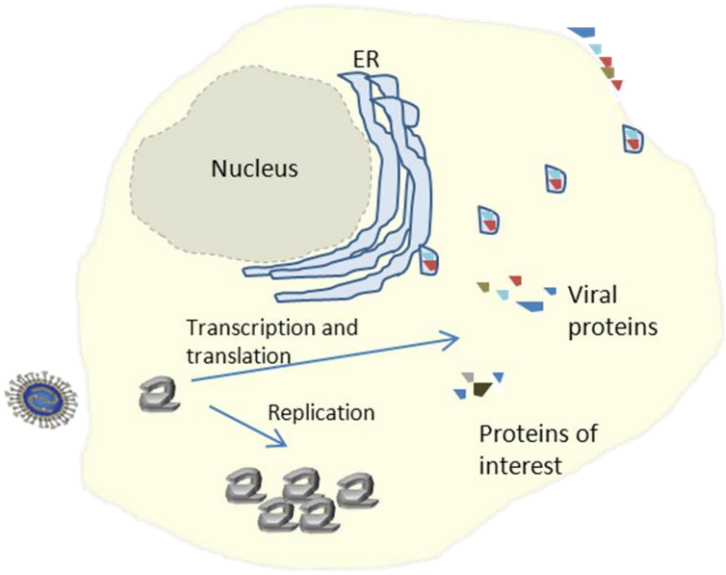


Figure 4: Sendai viral vector cycle. Once the cell is transduced by SeV, the viral RNA genetic material penetrates the cell and replicates. The RNA is transcribed and translated generating viral proteins together with the protein(s) of interest. The virus is genetically modified to avoid the generation of infectious particles.

2.2 Choice of cell source for reprogramming

Many different cell sources have been successfully used for induction of pluripotency using different reprogramming platforms (Table1). According to Hanna et al, as long as the reprogramming genes are expressed at an adequate level and the cell does not show any abnormality, any cell type is likely capable of giving rise to an iPSC line (Hanna, Saha et al. 2009). In fact, the *in vivo* induction of reprogramming genes induced the generation of iPSC from any tissue (Abad, Mosteiro et al. 2013). Therefore, the desired cell source is the most easily accessible and the one in which the reprogramming factors can be delivered more efficiently. This is the reason why fibroblasts have been the preferred option by many groups (Takahashi, Tanabe et al. 2007, Park, Arora et al. 2008, Fusaki, Ban et al. 2009, Carvajal-Vergara, Sevilla et al. 2010, Liu, Sumer et al. 2010, Howden, Gore et al. 2011, Papapetrou, Lee et al. 2011, Sebastiano, Maeder et al. 2011, Tanaka, Takahashi et al. 2012), as they can be easily grown from a small harmless biopsy and efficiently be transduced and transfected. Another cell source presented as an alternative to fibroblasts is peripheral blood mononuclear cells (PB-MNC), which can be also easily obtained from routine blood tests. PB-MNC

present several additional advantages (Kunisato, Wakatsuki et al. 2010) among other sources. They are a heterogeneous cell population in which stimulation of the preferred cell type by cytokines is possible. Staerk, J. et al revealed that by stimulation by G-CSF, GM-CSF, IL-3, and IL-6 before and during the first days of reprogramming, the induction of pluripotency could be encouraged in progenitors and myeloid origin cells, avoiding reprogramming of genetically rearranged B or T lymphoid cells (Staerk, Dawlaty et al. 2010). Additionally, PB-MNC in enough numbers for reprogramming can be directly obtained without any need of culture expansion, which has been shown to induce genomic abnormalities. Also, they could be used to reprogram cells from any disease, even from patients in which skin biopsies are not recommended (i.e. aberrant scaring) (Sommer, Rozelle et al. 2012). However, bulk PB-MNCs cannot be nucleofected and its transduction with lentiviral vectors is not as efficient as it is in fibroblast.

2.3 Erythroid differentiation of iPSCs

hESCs are isolated from the Inner Cell Mass of the developing human embryos before gastrulation and therefore have the ability to differentiate into any cell type from the embryo (Thomson, Itskovitz-Eldor et al. 1998). hiPSCs acquire all the characteristics of hESCs in terms of pluripotency. *In vitro* differentiation of hESC and iPSC has been revealed as a very promising model that mimics gastrulation and in which all the events that take place during human embryonic development can be examined (Fig. 5). One of the developmental pathways that have been studied is the emergence of Hematopoietic Stem Cells (HSC) from mesodermal tissues. Hematopoietic cells arise in two waves during human embryogenesis: the primitive wave, generated in the Yolk Sac (YS) (extraembryonic mesoderm) just after gastrulation and the definitive one, which is the adult type hematopoiesis, which starts after 6-8 weeks in the aorta-gonad-mesonephros region (AGM) of the developing embryo. The erythropoiesis that takes place in the YS shows some differences with the one in adult hematopoietic tissues (Palis, Malik et al. 2010); the primitive erythroblast (EryP) generated are nucleated and produce mainly embryonic (ζ and ε) and fetal (γ) globin chains. Once heartbeat is established, they begin to circulate while maturing, and after several days in circulation they enucleate. Even after enucleation, embryonic erythrocytes show a larger size than erythrocytes generated from adult tissues (Baron 2013) (Fig. 6). Trying to reproduce embryonic development, several authors have shown the generation of erythroid progenitors by culturing hESC/iPSC in different platforms in which the presence of erythroid promoting cytokines (hSCF, hTPO, hEPO, hIL3, hIL6), iron saturated transferrin and insulin are common among them (Lu, Feng et al. 2008, Hatzistavrou, Micallef et al. 2009, Lapillonne, Kobari et al. 2010, Dias, Gumenyuk et al. 2011). Some of the cell markers used for the identification of erythroid differentiation are: Glycophrin A (CD235a), a glycoprotein present in erythrocyte membrane, and transferrin receptor (CD71),

implicated in iron transfer into the red blood cell. Erythroid cells derived from hESC/hiPSC are phenotypically similar to the ones that appear during the first waves of embryonic/fetal development, being mainly nucleated and expressing embryonic and fetal globins. Some *in vitro* protocols involving feeder supporting cells (Lu, Feng et al. 2008, Dias, Gumenyuk et al. 2011) or prolonged culture in the presence of human plasma (Lapillonne, Kobari et al. 2010) showed some adult β globin expression and some level of enucleation (10-60%). In an *in vivo* maturation approach carried out by Kobari L et al (Kobari, Yates et al. 2012), in which early differentiated erythroid cells were injected into the retro-orbital vein of immunodeficient mice, the modulation of the human globin chains was complete, being β and α chains each approx. 50% of the total globins.

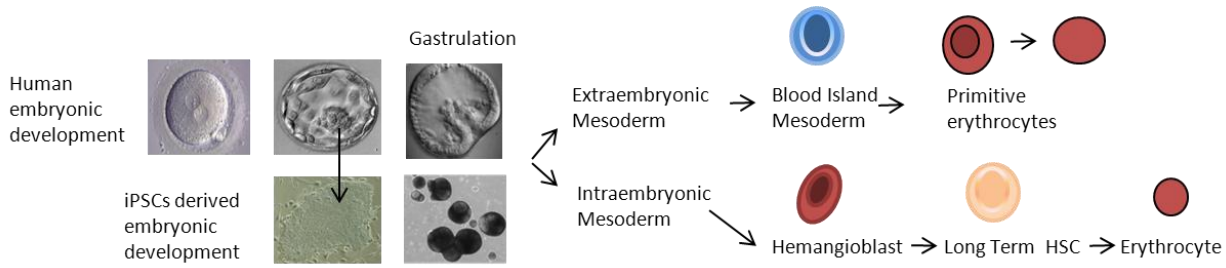


Figure 5: Erythropoiesis during human development. Two waves of hematopoiesis have been reported during development: the primitive that arises from the Yolk Sac and generates primary primitive erythroblast, and the definitive one, which is the adult type and through the Long Term Hematopoietic Stem Cell, generates totally mature erythrocytes. Embryoid bodies generated from pluripotent stem cells resemble human gastrulation *in vitro.*

3. Gene correction of disease specific iPSCs

The possibility of generating and correcting patient specific iPSC has opened up new avenues for regenerative medicine. Up until now, different strategies have been carried out for patient-specific hiPSC gene correction (reviewed in Garate, Davis et al. 2013). Although many of the reported approaches performed genetic correction directly in hiPSC, this is not always possible, as some disease causing mutations also interfere with the reprogramming process. For example, this is the case of Fanconi Anemia (FA), in which the generation of FA derived hiPSC was mainly possible after correction by an integrative lentiviral vector (Raya, Rodriguez-Piza et al. 2009). The use of integrative vectors for correction of the diseased cells, before or after cell reprogramming, has important risks already mentioned, mainly insertional mutagenesis. Other strategies for non-integrative gene correction are being studied, as in the case of integrative defective lentiviral vectors (Natale, Frangipane di Regalbono et al. 2007, Weidauer, Vatter et al. 2008), adenoviruses or adenoassociated viruses (Alberghini, Pasquinelli et al. 2007). But when the target cells are proliferative, no method to maintain the exogenous DNA has been developed yet. Having the opportunity of unlimited self-renewal and clonal selection, we believe that the best genomic correction approach would be

homologous recombination (HR) mediated direct site integration, which could be fully explored in human pluripotent cultures (hiPSCs and hESCs).

3.1 Homologous recombination mediated gene editing

During gene editing, a desired DNA sequence is inserted into a specific place at the genome very accurately, at a single base resolution. HR is the cellular mechanism used to carry out this type of genetic modification and is part of the DNA repair machinery that under normal conditions is responsible for the maintenance of cell genome integrity. HR and Non Homologous End Joining (NHEJ) are the two main mechanisms activated to repair a DNA Double Strand Break (DSB) in the genome. HR is a very accurate reparation procedure as the undamaged DNA strand (sister chromatid) is used as a template to generate the new DNA sequence. In contrast, during NHEJ, DSB ends are directly ligated without the need for a homologous template, being highly error prone. HR is a very secure process to repair DSB. The fidelity of HR gives the specificity and accuracy that gene editing requires and therefore, researchers have developed molecular biology tools to introduce exogenous DNA into a specific spot of the genome by using the endogenous HR machinery. Part of the DNA of interest needs to be homologous or surrounded by homologous sequences to the locus where it will be inserted, the so-called Homology Arms (HA). Once a DSB is generated in the desired locus, the HR machinery uses the homology region of the exogenous DNA as a template, copying and inserting all genomic elements surrounded by the HAs. These techniques have been widely used for the generation of Knock-Out, and Knock-In transgenic animals (Robbins 1993) and their application was awarded by the Nobel price to Mario Capecchi, Oliver Smithies and Martin Evans in 2007, for being pioneers in disrupting or altering mouse genes (Smithies 2001).

As targeted integration mediated by HR shows a low efficiency in mammalian cells (1 HR event per 10^6 cells) (Cathomen and Joung 2008), positive and negative selection markers (i.e. drug resistance gene between HAs and suicide genes outside HAs) can be included in the targeting vector (also called Donor Matrix (DM)) in order to select those cells in which the desired insertion has happened correctly. The final structure and complexity of the DM will vary according to the targeting strategy and the place of the genome to be targeted. There are three different gene editing strategies that could be considered to correct or insert/express a transgene by HR (Fig. 6).

1. Gene correction: One or more bases are exchanged from the original strand by integrating them within the homology sequence; this strategy is the one to be chosen when the goal is to introduce or repair point or small mutations.
2. Safe harbor integration: Insertion of a whole expression cassette (promoter, transgene and regulatory signals) in a harmless place of the genome. Examples of this type of genomic locus

reported in literature are: *AASV1*, *CCR5* or *ROSA26* locus (Irion, Luche et al. 2007, Torres, Garcia et al. 2011, Yao, Nashun et al. 2011).

3. Knock-In: a partial cDNA of the gene of interest is introduced in its endogenous locus. Commonly, a splicing acceptor is located before the cDNA to anchor the previous endogenous splicing donor. The endogenous elements of the locus will regulate the expression of the endogenous/exogenous resulting chimeric gene.

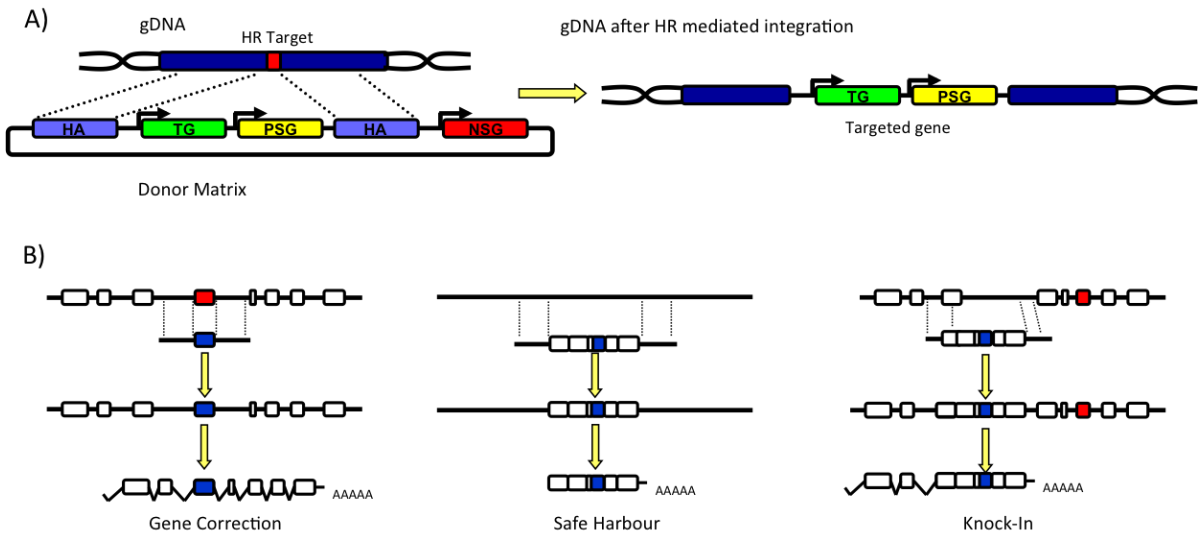


Figure 6: Homologous recombination matrix elements and different strategies for gene editing. A) Figure showing all the elements that a donor matrix could be composed of in order to integrate in a particular locus and to allow correctly targeted cells to be selected. HA: Homology Arm, PSG: Positive selection gene, TG: Therapeutic gene, NSG: Negative selection gene. B) Three different possibilities for gene editing can be considered. Left: a mutation (in red) is corrected leaving no other exogenous DNA. Middle: the DNA is introduced into a harmless place in the genome. Right: a piece of the cDNA is introduced in an intron, taking advantage of endogenous splicing machinery. Modified from (Garate, Davis et al. 2013)

Independently of the chosen strategy, gene editing process can be separated in two different steps; generation of DSB and HR, being both crucial for gene editing efficiency. In order to activate HR machinery in the intended locus, a DSB should be generated, that as previously mentioned could be resolved by HR or by NHEJ. Unfortunately NHEJ is the preferred mechanism to solve this type of lesions increasing the risk of genomic deletions/insertions. Cell cycle status is also a critical parameter for gene editing as it requires transit through S-G2 phase of the cell cycle to take place (Delacote and Lopez 2008). These limitations make gene editing in human cells difficult to achieve, especially in hESC and hiPSC, which are difficult to clone from single cells (Amit, Carpenter et al. 2000) and in which transduction efficiency with common techniques used for mouse ESCs is low (10-50%) (Braam, Denning et al. 2008). In order to solve this bottleneck, different approaches have been reported to improve gene editing efficiency in hPSCs. The use of very long HAs (27.7 and 8.8 kb) in a Bacterial Artificial Chromosome (BAC) was carried out by Howden et al. (Howden, Gore et al. 2011) to correct gyrate atrophy hiPSCs showing an efficiency of 10%. Different reports have shown that by

using Helper-dependent adenoviral vectors (HDAdV), which have the ability to transduce hPSC very efficiently, the overall gene editing efficiency is increased (7-80%) (Liu, Suzuki et al. 2011, Aizawa, Hirabayashi et al. 2012). Another explored possibility is the induction of DSBs specifically in the locus to be targeted by using engineered DNA nucleases (Fig. 7), that have been shown to increase HR efficiency by up to 10^3 fold (Porteus and Carroll 2005). This strategy has been deeply studied and applied for correction of patient specific hiPSCs and has been reviewed in Garate, Davis et al. 2013 and in Simara, Motl et al. 2013.

3.2 DNA Nucleases

The discovery of DSB generation in a specific site of the genome promoting HR in this specific locus, lead to the development of the first DNA nucleases with specific targeting, the meganucleases (Colleaux, D'Auriol et al. 1988). After that, very versatile engineered types of nucleases came out, first the Zinc Finger Nucleases (ZFNs) (Kim, Cha et al. 1996) and then the transcription activator-like effector nucleases (TALENs) (Cermak, Doyle et al. 2011) . Lately, a new approach has been explored, the CRISPR/Cas system (Mali, Yang et al. 2013). All of them are intended to generate DSB in a specific locus but each of them has its own particularities. A scheme of their structures and the strategy to use them for HR is depicted in figure 7.

3.2.1 Meganucleases

Rare cutting DNA nucleases or homing endonucleases (HE) were firstly explored in lower eukaryotes, being responsible for a process called intron homing, during which an intron is copied from one allele to another of the same gene after a DSB generation (Jasin 1996). One of the best characterized HE is I-SceI, which was isolated from the yeast *Saccharomyces cerevisiae* and recognizes an intron in the 21S Ribosomal RNA gene. The main advantage of this type of nucleases is that their recognition sites are much longer than the ones of bacterial restriction enzymes, and that they could be engineered to recognize other sequences. Their characteristics make these enzymes cut; in principle, only once in a particular site of the genome, which could be our site of interest. They recognize a sequence around 14-44 bp and assuming a random organization of the human genome, the frequency of occurrence of this sequence more than once is expected to be very rare (Colleaux, D'Auriol et al. 1988). They are composed of a DNA recognition domain, which is responsible for the HE specificity, and a cutting domain, which is the one that will catalyze the DSB generation. There are different families and each of them show different structures, some having two binding modules and working as homodimers, and some working as monomers (Stoddard 2005). The specificity of natural HE can be altered by mutagenesis of the DNA binding domain followed by functional screening (Chen and Zhao 2005, Arnould, Chames et al. 2006, Doyon, Pattanayak et al. 2006, Silva,

Belfort et al. 2006, Delacote, Perez et al. 2013), and they are called customized Meganucleases (MG). Regarding their DNA binding domain design, a unique and specific MN recognition sequence can be found approximately every 300 bp (Paques and Duchateau 2007). One of the main advantages of this type of nucleases is their reduced cytotoxicity in mammalian cells (Rouet, Smih et al. 1994).

3.2.2 ZFN

Zinc Finger Nucleases (ZFNs) are artificial endonucleases firstly generated through a very fruitful collaboration between the groups of doctors S. Chandrasegaran and J. Berg. In order to generate a DNA nuclease with a specific targeting, they combined the DNA binding domain of Zinc Finger (ZF) proteins and the type IIS nuclease domain of the restriction enzyme Fok-I (Kim, Cha et al. 1996). As cleavage activity of FokI requires dimerization, ZFNs were designed to work as pairs of two monomers of ZFN in reverse orientation (Fig. 7). Each ZF modules is composed of a tandem repeat of Cys_2-His_2 that recognizes three nucleotides. Therefore an array of 3 or 4 ZF recognizes and binds a sequence of 9 or 12 nucleotides, leading to the targeting of a genomic sequence of 18-24 nucleotides long (Porteus and Carroll 2005, Carroll 2011). Several *in vitro* studies in oocytes were performed in order to study and confirm the specific cutting, but the first genomic locus that was targeted *in vivo* by a ZFN was the yellow gene of *Drosophila Melanogaster* in the early 2000's (Bibikova, Golic et al. 2002, Bibikova, Beumer et al. 2003). Later on, its use has been applied for the generation of transgenic plants, zebrafish, mouse, rat and xenopus and in several human somatic cell types, including HSCs, hESCs and hiPSCs. Lombardo et al. showed for the first time the insertion of GFP into the *CCR5* safe harbor locus in hESC after inducing HR by ZFN expression; targeted hESC were able to differentiate into neurons keeping GFP expression (Lombardo, Genovese et al. 2007). Many reports have now shown ZNF mediated gene correction in disease specific hiPSCs, being until now the preferred type of DNA nuclease for this purpose. ZFN can be designed to target any gene but not any sequence, as not all the combinations of three nucleotides are feasible.

3.2.3 TALE nucleases

Transcription activator-like effector nucleases (TALENs) are also artificial endonucleases similar to ZFN but with a different DNA binding domain, sharing the nuclease domain. In this case the specific DNA binding is obtained using TAL effector (TALE) proteins, which were discovered in a plant pathogen from the species Xantomonas as plant genes regulators (Hopkins, White et al. 1992). The code for TALEs DNA recognition was elucidated by Adam Bogdanove (Bogdanove and Voytas 2011) and Jens Boch (Boch, Scholze et al. 2009) leading to a very versatile gene editing tool. TALENs (TAs) DNA recognition domain consist of tandem arrays of 33-35 aa (Fig. 7), being each of these units able

to bind one nucleotide. Thus, the DNA binding domain can be designed to recognize any desired genomic sequence. Best TAs cleavage conditions were found to be similar to ZFN, as dimers in reverse orientation (Li, Huang et al. 2011) and recently they have been described and licensed as single monomers (Beurdeley, Bietz et al. 2013).

3.2.4 CRISPR/Cas system

CRISPR/Cas system is a sequence specific RNA mediated DNA nuclease that is part of the bacterial and archeal immune system, aiming the destruction of invading viral and plasmid DNAs. The first reported adaptation of this prokaryotic system in eukaryotic cells was described this February (Mali, Yang et al. 2013) and, in half a year more than 20 studies have been reported in which this type of system was used in eukaryotic cells. It was efficiently used for the generation of transgenic models of many organisms including plants (Jiang, Zhou et al. 2013, Miao, Guo et al. 2013), mice (Wang, Yang et al. 2013, Yang, Wang et al. 2013) and zebrafish (Blackburn, Campbell et al. 2013, Hwang, Fu et al. 2013, Hwang, Fu et al. 2013, Jao, Wente et al. 2013, Xiao, Wang et al. 2013) and in many cell lines, including hPSCs (Hou, Zhang et al. 2013). The versatility of this system relies on the fact that the specificity is determined by a guide RNA that targets a DNA sequence of 20 nt (crRNA) and that is delivered separately from the protein that generates the DSBs, the Cas9. This also allows the targeting of multiple loci at the same time. The other needed element, the tracrRNA, is responsible for the delivery of the Cas9 to the targeted sequence and can be expressed as a fusion of the crRNA, also called single guide RNA (sgRNA). The technique development is so recent that there is limited data on efficiency and specificity, but there are two reports so far in which it is shown that its low specificity might be one of its main limitations (Cradick, Fine et al. 2013, Fu, Foden et al. 2013), being sequences with few mismatches highly cut.

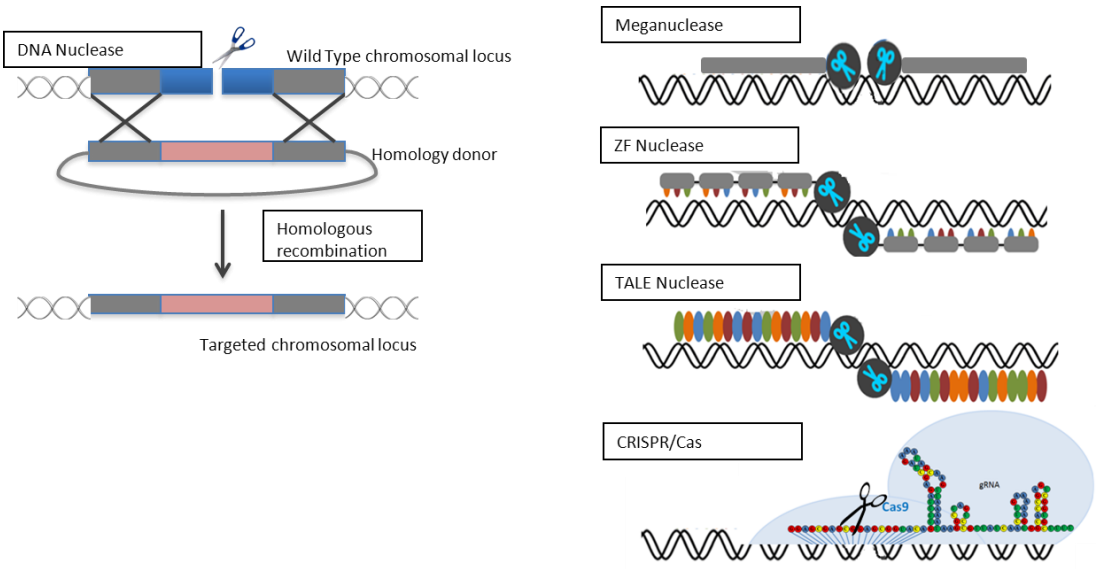


Figure 7: DNA-Nuclease enhanced Homologous Recombination (HR). On the left, a schematic diagram of nuclease mediated HR. After Double Strand generation, the Homology donor is served as a repair template and the piece in pink is introduced in the intended locus. On the right, the structure and DNA binding of the 4 different used nucleases. CRISPR/Cas system image modified from http://www.addgene.org/crispr/jounglab/. ZFN and TALEN images modified from a commercial brochure of PNA Bio.

Although it is possible to perform nuclease mediated targeting in any site of the genome, including transcriptionally inactive loci (Hockemeyer, Soldner et al. 2009, Aizawa, Hirabayashi et al. 2012), the epigenetic status of the locus to be targeted influences targeting frequency, as it has been shown by using MGs (Daboussi, Zaslavskiy et al. 2012), ZFNs (Hockemeyer, Soldner et al. 2009) and TAs (Hockemeyer, Wang et al. 2011). The mechanism by which it is affected is poorly understood; it might be the recognition and accessibility of the nuclease to the target or the accessibility of the repair machinery elements to the locus. It has been shown that the presence of methylated CpGs dinucleotides (mCpG) in TAs recognition site can dramatically decrease its efficiency (Shijia Chen 2013) and in the case of MN this also happens if the mCpG is in the central tetrabase of the recognition site (Valton, Daboussi et al. 2012).

The main concern when using DNA nucleases is the off target cutting (Gabriel, Lombardo et al. 2011), which is the recognition and DSB generation in other genome sequences different from the intended one. These unexpected DSB will be mainly solved by NHEJ, which is highly error prone and will lead to genomic modifications. Consequently, this type of cutting usually increases cell toxicity, decreases the desired DSB generation efficiency and can lead to undesired chromosomal alterations (Richardson and Jasin 2000). Theoretically, a DNA nuclease that recognizes a sequence of 18 bp will find its target once in every $7x10^{10}$pb, which means once in the human genome. The problem is that the recognition might tolerate a variable number of mismatches and recognize and cut similar sequences to the original. This has been mainly studied in ZFN and by *in vitro* and *in vivo* assays, identifying off targeted sequences that were not recognized by bioinformatics approaches (Gabriel, Lombardo et al. 2011, Pattanayak, Ramirez et al. 2011). Conclusions from both studies matched with the ones generated by the biochemical assay SELEX, which is a very popular off target analysis tool based on affinity binding of immobilized ZF molecules to DNA molecules and subsequent analysis of the sequences captured by each monomer (Perez, Wang et al. 2008, Gabriel, Lombardo et al. 2011). The fact that ZFNs and TAs cut as heterodimers increases the possibility of off target cutting, as homodimers with a different target could be formed. There are several strategies under study to force their correct conformation in order to be able to cut and reduce the derived associations cutting in undesired sites, the so called forced heterodimers.

3.4 Reported patient specific hiPSC gene correction

The combination of patient specific hiPSCs and gene editing is a powerful source for regenerative medicine and therefore many proofs of principle have been described from different disea showing that they could be successfully corrected using nuclease mediated HR (Garate, Davis et al. 2013). So far, the majority of them used ZFN and gene correction strategy has been the most popular one. Table 2 shows all the reported strategies and its characteristics.

Disease	Strategy	Reference
β-Thalassemia	Safe Harbor integration	(Macarthur, Fontes et al. 2012)
Diamond Blackfan anemia	Safe Harbor integration	(Garcon, Ge et al. 2013)
Xl-inked Chronic Granulomatosis	Safe Harbor integration	(Zou, Sweeney et al. 2011)
Sickle Cell Anemia	Gene correction	(Sebastiano, Maeder et al. 2011) and (Zou, Mali et al. 2011)
Huntington's disease	Gene correction	(An, Zhang et al. 2012)
β-Thalassemia	Gene correction	(Wang, Zheng et al. 2012)
Parkinson's disease	Gene correction	(Soldner, Laganiere et al. 2011) and (Ban, Nishishita et al. 2011)
a1-antitrypsin deficiency	Gene correction	(Yusa, Rashid et al. 2011)
Tauopathy	Gene correction	(Fong, Wang et al. 2013)
Diabetes	Knock-In	(Hua, Shang et al. 2013)

Table 2. Corrected iPSC reported in literature by ZFN mediated HR. Disease, correction strategy and reference are shown

3.5 Limitation of the therapeutic use of hiPSCs: genomic instability

As pointed out by Blasco et al (Blasco, Serrano et al. 2011), there are several facts related to genomic instability in cell reprogramming that should be taken into consideration: two of the four reprogramming genes are oncogenes (*C-MYC* and *KLF4*), reprogramming efficiency is increased by including SV40 large T antigen or by the suppression of some tumor suppressor genes (*P53* or *ARF*) and, most importantly, the fact that mice generated from iPSCs developed tumors and showed developmental problems. As a consequence, many groups have studied the genomic integrity of pluripotent cultures making clear that genetic abnormalities may be introduced into the hiPSCs either through the reprogramming process, the tissue culture expansion, and/or the gene correction process (Gore, Li et al. 2011, Hussein, Batada et al. 2011, Laurent, Ulitsky et al. 2011, Martins-Taylor, Nisler et al. 2011). The risks of genetic correction are not only related to DSB generation but also to the fact that there are several steps of the procedure in which clonal cultures are selected and, due to the low replating efficiencies, cells with a proliferative advantage could be selected. There are many techniques that can be used for interrogating genomic integrity of hiPSCs. The most popular one is G-banded Karyotyping, by which it has been shown that there are common aneuploidies that occur during PSCs culture or reprogramming, like trisomy of chromosomes 12, 8, 20q and X, detected in both hESC and hiPSC and of Chr 17 in hESCs (Martins-Taylor and Xu 2012). However, chromosomal aberrations smaller than 5 MB in size cannot be detected by karyotyping and, therefore, many other methodologies have been developed for this purpose, like whole genome and whole exome sequencing, array-based Comparative Genomic Hybridization (aCGH) and high resolution SNP genotyping. Throughout the whole year 2011, 4 studies came out in which by using one or more of these techniques several cell lines were analyzed showing the appearance of many genetic abnormalities (Gore, Li et al. 2011, Hussein, Batada et al. 2011, Laurent, Ulitsky et al. 2011,

Martins-Taylor, Nisler et al. 2011). By comparing early and late passage PSC two types of CNV could be distinguished, the ones generated during reprogramming and the ones generated during culture. It was also reported that some of the variations proceeded from the population over long-term culture pointing out that they were positively selected during reprogramming and negatively selected during long-term passage. Regarding replication stress and appearance of somatic coding mutations, Gore at al found an average of five coding mutations in each cell line being sometimes present in cancer related genes (Gore, Li et al. 2011).

Regarding genome modifications during gene correction, several studies have looked at it carefully. In the study of gene correction of A1ATD-hiPSC, a complete genome integrity study was performed (Yusa, Rashid et al. 2011). The authors concluded that more genetic alterations were generated during the reprogramming process and the extensive culture of the hiPSC than during the HR correction. Thus, genome integrity should be carefully studied in gene-edited hiPSCs in the case they are intended to be used in clinical settings.

As it was the case of erythroid diseases, such as Sickle cell anemia or β-Thalassemia, and taking into consideration the possible genomic modifications that could happen during hiPSC culture, we believe that Pyruvate Kinase Deficiency (PKD) patients could also benefit from this powerful technology and therefore, we have explored its use on the development of Knock-In gene corrected erythrocyte

V. Aims

PKD is a rare erythroid metabolic genetic disease, being the most common cause of Non-Spherocytic Hemolytic Anemia (CNSHA). Although in many cases the symptomatology is moderate, there are some severe cases in which the only definitive treatment is allogeneic bone marrow transplantation. In order to avoid transplant-associated complications, an autologous cell therapy in combination with gene correction is presented as a promising therapeutic approach. As cell source we consider that patient specific hiPSCs, having the ability to differentiate into any cell type and unlimited self-renewal, are a very powerful cell type. Regarding the gene correction approach, we believe that nuclease mediated site-specific gene correction is also a very accurate method. With the purpose of studying the feasibility and safety of this kind of therapies, we wanted to develop strategies to reprogram and genetically correct PKD samples. Our overall goal was the design of an innovative therapy that could be used in a future for severe cases of PKD and to develop tools that could also be applied to other genetic hematological diseases. In order to reach our goal we had to achieve the following specific aims, also outlined in figure 8:

1. Generation of transgene-free human induced Pluripotent Stem Cells from Pyruvate Kinase Deficient patients (PKDiPSCs) and evaluation of their potential for disease modeling.

2. Knock-In gene correction of the *PKLR* gene by DNA-nuclease assisted homologous recombination in *PKLR* locus and study of the correction efficacy.

3. Evaluation of genomic integrity of corrected PKDiPSC lines.

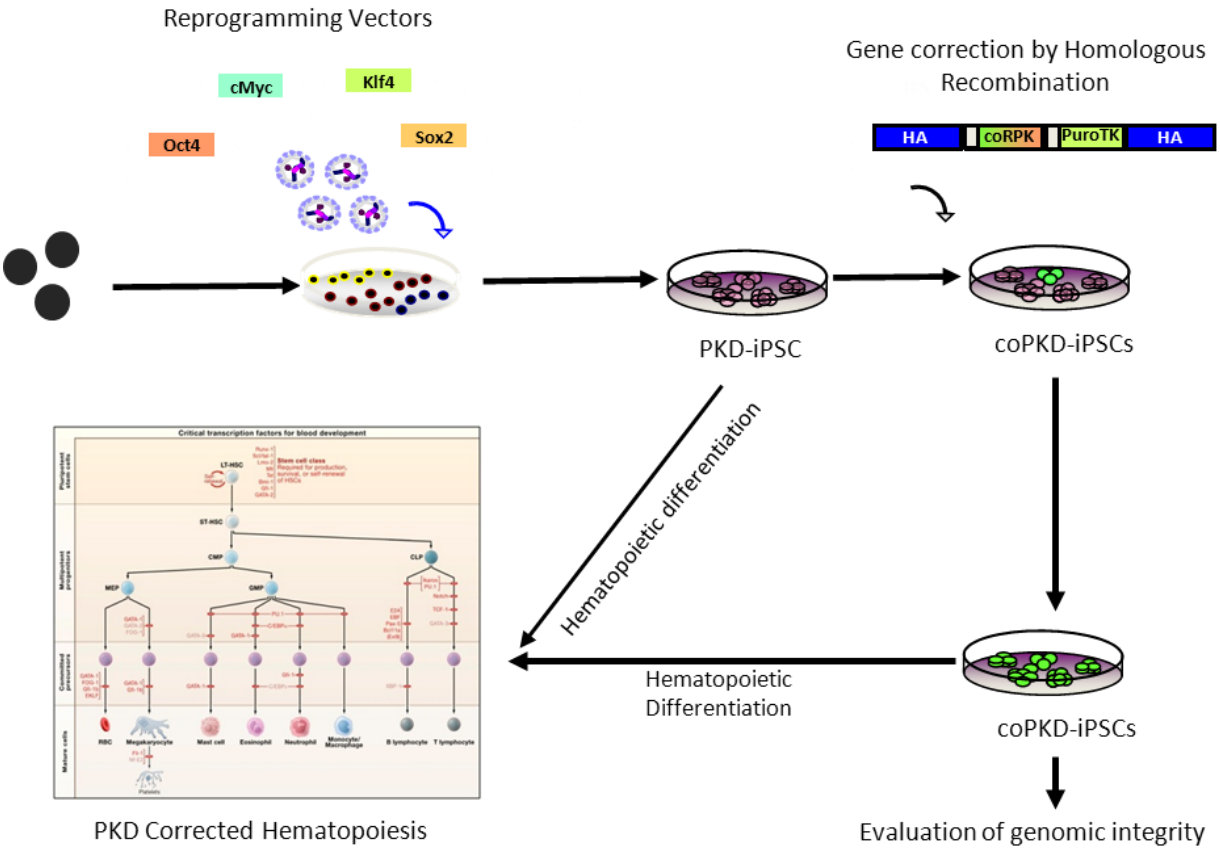


Figure 8: Gene editing of induced pluripotent stem cells from Pyruvate Kinase Deficient patients. Specific aims of the present thesis manuscript are summarized in this diagram. First, patient specific cells are reprogrammed to pluripotency (aim 1) and then genetically corrected by DNA nuclease mediated HR (aim 2). Uncorrected and corrected PKDiPS are going to be differentiated into erythroid cells in order to evaluate PKD phenotype before and after correction (aims 1 and 2). Genomic integrity of corrected PKDiPSC is also going to be evaluated (aim 3).

VI. Materials and Methods

1. Patient and healthy donor samples

Peripheral blood (PB) from Pyruvate Kinase Deficient (PKD) patients was obtained from Hospital 12 de Octubre (Madrid, Spain) (patient PKD1), Fundación Jimenez Díaz (Madrid, Spain) (PKD2) and Centro Hospitalario de Coimbra (Coimbra, Portugal) (PKD3). In addition, PB from healthy donors was also collected in the Medical Care Service of CIEMAT (Madrid, Spain). All the patient samples were collected under written consent.

PB mononuclear cells (PBMNC) were isolated by Ficoll-Paque (GE Healthcare, WI, USA: density 1.077) following manufacturer recommendations and were kept frozen or used fresh in further experiments.

1.1 Patient specific mutations detection

Patient specific mutations were confirmed by DNA Sequencing. Genomic DNA was extracted from the different patient and healthy donor cells by QIAamp DNA Blood kit (Qiagen, Hilden, Germany) and the sequences surrounding the mutations were amplified by high fidelity Pfx50 polymerase (Invitrogen, Life technologies, CA, USA) using primers number 1, 2 and 3, 4 for PKD2, and primers number 5 and 6 for PKD3 (see primers table). PCR products were purified from agarose gel by QIAquick Gel Extraction Kit (Qiagen) and sequenced with the primers used for amplification (Stabvida, Portugal; Seqwrite, Houston Texas, USA). Sequences analysis was performed by Vector NTI software (Invitrogen) or by Lasergene 10 (DNAstar, WI, USA).

2. Induced Pluripotent Stem Cells

2.1 Generation

Reprogramming was performed right after isolation (PKD2 and PB2) or from frozen samples (PKD1, PKD3 and PB1). PB-MNCs were pre-stimulated for four days in StemSpan (StemCell Technologies, British Columbia, Canada) plus 100 ng/mL hSCF, 100 ng/mL hFLT3L, 20 ng/mL hTPO, 10 ng/mL G-CSF, and 2 ng/mL hIL3 (Peprotech, NJ, USA). The cells were collected and transduced with either a mix of Sendai viral vectors (SeV) expressing OCT3/4, KLF4, SOX2, c-MYC and Azami Green or with STEMCCA excisable polycistronic lentiviral vector expressing mouse OCT3/4, KLF4, SOX2, c-MYC. Both vector systems were used at multiplicity of infection 3. The Sendai Viral vectors were kindly provided by the company DNAvec (Ibaraki, Japan) in form of viral particles. The STEMCCA lentiviral vector was kindly provided by Gustavo Mostoslavsy and the viral particles were generated at CIEMAT as

described (Sommer, Stadtfeld et al. 2009). Transduced cells were maintained with the described cytokines for four more days and then supplemented with 10 ng/ml of bFGF. Five days after transduction cells were collected, counted and seeded on irradiated HFF-coated 100 mm culture plates with hES media (DMEM/F12, 20% knockout serum replacement [KO-SR], 1 mM L-glutamine, 1% nonessential Amino Acids [All from Life technologies], 0.1 mM β-Mercaptoethanol (Sigma-Aldrich, MO, USA) and 10 ng/ml basic hFGF (Peprotech, NY, USA or Invitrogen)) at density 150-800 cells/cm^2. hES media was changed every other day. When hES-like colonies appeared, they were selected under the stereoscope (Olympus, Tokio, Japan) and a clonal culture from each colony was established. All the established iPSC lines were generated by SeV.

2.2 Culture conditions

hiPSC lines were maintained over a layer of irradiated fibroblast, which were either irradiated Human Foreskin Fibroblast (HFF-1, ATCC), irradiated Mouse Embryonic Fibroblasts (MEF, ATCC) or irradiated puromycin resistant MEFs (Puro-MEF, STEMCELL Technologies). Fibroblasts were cultured and expanded in DMEM medium, 1mM L-glutamine, 1% nonessential Amino Acids (All from Life technologies), 20% Hyclone FBS (Thermo Scientific, Thermo Fischer Scientific, MA, USA), irradiated at 45 Gy and frozen in viability at 2.5 x 10^6 cells/vial. The day before iPSC splitting, feeders were seeded at 19,000 cells/cm^2 in gelatin pre-coated plates. hiPSC lines were cultured in hES media, which was changed daily. When iPSCs reached 80-90% confluence and the colonies looked dense (usually once a week) they were passed in a 1:4-6 ratio using 1 mg/ml Collagenase IV (Life technologies) for 3 minutes and then scraping.

For electroporation experiments, a single cell suspension was generated. Cultures ready to split (70-80% confluent) were washed twice with PBS and treated with StemPro Accutase (Life Technologies) for 3 minutes. Cells already at single cell stage but still forming the colony were very gently washed with PBS and hES media was added. Cells were finally detached by pipetting up and down several times. When a higher cell survival was needed, like higher amplification ratio or after generating single cell suspensions, the cells were cultured for 1 day in hES media supplemented with 10 μM of the Rock Inhibitor Y-27632 (Sigma-Aldrich).

2.2.1 Freezing/Thawing

On average, one 80% confluent well of a 6 well plate was used for generating 2 vials of frozen iPSCs. Cells were Collagenase IV treated, scraped and resuspended in hES media. An equal amount of 2X freezing media (20% DMSO, 20% hES media and 60% Hyclone FBS) was gently added while shaking the mixture. Cells were frozen overnight in a controlled freezing container at -80 °C. After 24 hours frozen vials were transferred to liquid nitrogen cryogenic tank. For thawing, the cryotube was quickly

transferred from the cryogenic tank to the water bath at 37°C. Before the cells were completely thawed they were taken out, drop wise mixed with hES media and spun down for 5 minutes at 200 g. The cell pellet was resuspended in hES media with 10 μM of Rock Inhibitor Y-27632 and transferred into a feeder coated 6 well plate.

2.3 hiPSC characterization

2.3.1 Alkaline Phosphatase

hiPSC lines growing on feeders were fixed for 10 minutes with 2% PFA, washed twice with PBS and Alkaline Phosphatase staining was performed with VECTOR Blue Alkaline Phosphatase Substrate Kit, following manufacturer specifications (Vector Laboratories, CA, USA). The stained samples were analyzed and counted under an Olympus IX6 microscope (Olympus, Tokio, Japan)

2.3.2 RT-PCR gene expression array

Gene expression of 84 key genes involved in the maintenance of pluripotency and the self-renewal status of human embryonic stem cells (hESC) was analyzed by using hESC RT2 Profiler PCR Array as described by manufacturer (SaBiosciences, Qiagen). First, RNA was isolated from 1 confluent 6 wells plate of hiPSC by RNeasy Plus Mini Kit (Qiagen) and then cDNA was generated using 5 μg of RNA/ reaction with RT2 First Strand Kit (SaBiosciences). cDNA was added to the quantitative PCR array plate and the rest of the protocol was followed as recommended by the manufacturer. The machine used was an ABI 7900 (Applied Biosystems, Life Technologies). cDNA extracted from the hESC line H9, obtained from Wicell (Wisconsin, USA) was used as control.

2.3.3 Teratomas

Teratomas were generated by subcutaneous injection of a whole confluent 6 well plate into immunocompromised NOD.Cg-PrkdcscidIL2rgtm/Wjl/SzJ (NOD scid gamma null (or NSG)) mice (Jackson Laboratories, Maine, USA). The culture was treated with collagenase IV, the colonies were cut in small pieces, spun down for 5 minutes at 200 g and resuspended in 140 ul of DMEM/F12. Just before injecting the mice the cell suspension was mixed with 60 ul of hES qualified Matrigel (BD Bioesciences, CA, USA). After 8-10 weeks the tumors were visible. The mice were sacrificed, the tumors were removed and the tissue pieces were fixed in 4 % PFA, embedded in paraffin and serially sectioned into 5 μm sections by the Histology facility of the CIEMAT. Every four slides, one was used to do hematoxylin and eosin staining as previously described (Certo, Gwiazda et al. 2012). The rest of the slides were used for immunostaining as described later.

2.3.4 Promoter demethylation

The study of the presence of demethylated CpGs in *NANOG* and *SOX2* promoters was kindly performed by the Myeloproliferative Syndromes laboratory of the Center for Applied Medical Research (CIMA, Pamplona).

2.3.5 Karyotyping

G-Banding and analysis of metaphasic chromosomes were performed by the Cancer Cytogenetic Core of the Texas Children Cancer and Hematology Center (Houston, Texas). Briefly, cells were incubated overnight with colcemide, then treated with KCl and a mix 3:1 of methanol/Glacial Acetic Acid and finally dropped into slides and Giemsa stained. The 23 sets of chromosomes were counted in 15-20 metaphasic cells in order to see if there was any chromosome aberration.

2.3.6 Immunoglobulin Chain and T cell receptor rearrangement analysis

Genomic DNA from hiPSC was purified and sent to the Histocompatibility and Molecular Biology Laboratory from Madrid Blood Transfusion Centre, where rearrangement analysis was kindly performed following BIOMED-2 strategy (van Dongen, Langerak et al. 2003).

2.3.7 Comparative Genomic Hybridization (CGH)

hiPSC were MEF depleted using mouse feeder removal MicroBeads, an LS Column, and a MidiMACS Separator following manufacturer recommendations (Miltenyi Biotec, Bergisch Gladbach, Germany). Genomic DNA (gDNA) was extracted from purified hiPS using ArchivePure DNA Cell/Tissue Kit (Prime5, Hamburg, Germany), and quantity and quality of the obtained DNA were checked by spectrophotometry (Nanodrop) and agarose gel analysis. Copy Number Variation (CNV) analysis was conducted by Wicell cytogenetic laboratory (Madison, USA). Genomic DNA was labeled and hybridized using the SurePrint G3 ISCA CGH+SNP Microarray Kit, from Agilent, an array of CGH based on competitive differential hybridization of labeled genomic DNA to the 180,000 oligonucleotide whole genome tiling array containing ≈120,000 CGH probes and ≈60,000 SNP probes. It covered the whole genome with some spacing (25.3kb overall, 5kb in ISCA regions). Bioinformatics analysis was performed using Agilent CytoGenomics Edition 2.5.8.1 and infoQuant CGHfusion 6.1.1. Test samples were compared to Agilent's genotyped reference DNA. It was considered as CNV when the aberration was covered with more than 3 probes.

2.3.8 Exome sequencing

Same samples as for CGH were used for exome sequencing, that was performed by Axeq Technologies (MD, USA). Briefly, gDNA was fragmented and a library of sequences corresponding to the human exome was generated by Agilent's SureSelect v4 Exome enrichment kit. Then, the

samples were sequenced to at least 30 fold (30X) average on target depth coverage using Illumina HiSeq 2000 system with 100-bp paired-end sequencing method. Variant calling and gene annotation were performed as previously described (Ju, Kim et al. 2011). In addition, variants had to fulfill these criteria: 1) Read Depth: the number of uniquely mapped reads at the position had to be > 8; 2) The average base quality for the position had to be ≥30; 3) The allele ratio at the position had to be ≥20% for heterozygous variants. The resulting variants that were present also in PKD2 PB-MNC (original sample before reprogramming), even if they were very low represented, were filtered. No variant from corrected PKD2iPS in which the number of reads in PB-MNC was < 8 was considered as we could not discard their presence in the original population. Variants present in the SNP database (SNPdb) were also removed. Resulting variants were validated by PCR amplification (primers 52-65) and Sanger sequencing as explained for patient specific mutations detection.

2.4 Erythroid Differentiation

hiPSC were differentiated according to 2 described protocols, the one of Robert Lanzas group (Lu, Feng et al. 2008) and the one described my Elefantys group (Hatzistavrou, Micallef et al. 2009) with some modifications. hiPSC were forced to form Embryoid Bodies (EBs) in both protocols, and from day 0 to day 1 they were cultured in hES media plus 10 uM of the Rock Inhibitor Y-27632. In the one developed by Dr Lanzas group, the first day, the EBs were cultured with 50 ng/ml of BMP4 and VEGF and then SCF, FLT3L and TPO 40 ng/ml are added for 4 more days. In a second stage of differentiation, EBs were disaggregated and cultured in Blast Growing media (Lu, Feng et al. 2007) in order to form hemato-endothelial blast colonies. In the one described by Hatzistavrou et al; EBs were cultured in APEL differentiation media (Stem Cell technologies) supplemented with 20 ng/ml VEGF, 20 ng/ml BMP4 and 40 ng/ml SCF for the first 10 days, and then they were transferred to gelatin coated plates with APEL plus erythroid differentiation promoting cytokines 10 ng/ml VEGF, 20 ng/ml IL3, 50 ng /ml SCF, 3 U /ml EPO, 20 ng/ml TPO, 20 ng/ml GM-CSF and 20 ng/ml IL6. All cytokines were from Peprotech. At day 21 of differentiation, erythroid markers CD235a and CD71 and the pan leukocyte marker CD45 were analyzed by flow cytometry. In order to disaggregate the 3D structures formed in the gelatin coated plates, we applied the following enzymatic treatment: 20 minutes of collagenase IV 1mg/ml followed by 10 minutes of Trypsin/EDTA 0.25% (Gibco).

3 Homologous Recombination enhanced by Nucleases

3.1 Plasmids

Cellectis therapeutics (Paris, France) designed two different PKLR01 Meganuclease (MG), version 1 and version 2, and one PKLR01 TALEN (TA) to target the intron 2 of *PKLR* gene. The exonuclease TREX2 was also kindly provided by them. These nucleases were shared in the form of DNA plasmids.

The expression of the nucleases was driven by Elongation Factor 1 alpha (EF1α) promoter. Additionally, two plasmids coding for each subunit of the AAVS1 Zinc Finger Nucleases (ZFNs) were kindly provided by Sangamo Biosciences (CA, USA).

3.2 mRNA synthesis

Nucleases were used as mRNA in some experiments. In order to synthesize mRNAs of the two version of PKLR01 MG, two subunits of AAVS1 ZFN and TREX2, their cDNAs were cloned after a T7 promoter. Briefly, cDNAs were amplified by PCR using the following primers that carry the T7 promoter: primers nr 11 and 12 for cloning PKLR01 MG and TREX2, and primers nr 13 and 14 for AAVS1 ZNF subunits. PCR products were purified and cloned into pCR®-Blunt vector by Zero Blunt PCR Cloning Kit (Invitrogen). The sequence and orientation of the cloned products were confirmed by sequencing with M13 primers. PKLR01 TA plasmids were already provided with the T7 promoter. cDNAs were excised from the vectors by the following enzymes: EcoRI or HindIII for MGs and TREX2, and HindIII and XhoI for ZNFs. Two hundred ng. of the digested and purified product were used to synthesize mRNA using the mMESSAGEmMACHINE® T7ULTRAKit (Ambion, Invitrogen) following vendors guidelines and purified by LiCl precipitation. The concentration and quality were analyzed by Agilent 2100 Bioanalyzer. TAs containing plasmids were linearized with HindIII and purified by phenol/chloroform extraction followed by 100% Ethanol and 3M Sodium Acetate precipitation. One μg of the linearized PKLR01 TA plasmid was used for mRNA synthesis as previously explained.

3.3 Nuclease cleavage activity

3.3.1 MG PKLR-specific DSBs generation analysis by deep sequencing

HEK-293H cell line (Invitrogen) was cultured in DMEM medium with Glutamax (Gibco, Life Technologies, California, USA), 10% Hyclone FBS (Thermo Scientific) and 1% penicillin/streptomycin (Gibco). The cells were transfected with or without 3μg of nuclease plasmid with Lipofectamine 2000 (Invitrogen) following vendor guidelines. MG expressing plasmids were co-transfected with a plasmid expressing TREX2 protein, which increases the sensitivity of the assay. After 72 hours, the gDNA was isolated and the surrounding PKLR01 target site was PCR amplified with different forward primers carrying different TAGs for sample identification after deep sequencing (nr 7, 8, 9, TAGs in bold and underlined in Primer Table) and reverse primer nr 10 using Herculase® II Fusion DNA Polymerase (Stratagene, Agilent Technologies, CA, USA). The reverse primer has a common biotin TAG for all the samples. A different combination of primers was used for each sample. The PCR products with the appropriate TAGs were agarose gel purified and sent to Cellectis to perform the deep sequencing analysis.

3.3.2 Surveyor assay

Different amounts of mRNA or DNA of MG, TA, ZFN and TREX2 were transfected by nucleofection according to the optimized protocol for AMAXA Nucleofection technology (Lonza, Basel, Switzerland), with several modifications. K-562 cell line was electroporated in human $CD34^+$ nucleofector buffer using the U08 program. After nucleofection, cells were left in the cuvette for 20 minutes at 37°C in order to be recovered from the electric pulse and then pipetted and transferred into the pre-equilibrated culture medium. For hiPSCs, 1-2 million cells were nucleofected in Nucleofection Solution 1 using A23 program. After nucleofection, cells were seeded into a fresh MEF feeder plate. Seventy-two hours after the nucleofection, gDNA was extracted from the cells and the nuclease target site was amplified by PCR with primers nr 15 and 16. These PCR products were directly used as samples for the Transgenomic SURVEYOR Mutation Detection Kit for Standard Gel Electrophoresis (Transegomic, NE, USA). First, the PCR products were forced to form heteroduplex by the following cycle in the thermocycler: 95°C for 10 min, 95°C to 85°C (2°C/ sec), 85°C to 25°C (0.1°C/ sec), 4°C forever. Then, 1 µl of Surveyor nuclease S and 1 µl of Surveyor enhancer S were added, and they were incubated for 1 hour at 42°C. The digested products were evaluated by separation on a 10% Novex TBE gel with Novex TBE Running Buffer and Novex TBE High density Sample Buffer (Invitrogen). The samples run for 1.5 hours at 100 V and were stained with 1:10000 diluted SyberGold (Invitrogen). Cutting efficiency was evaluated by comparing the non-cleaved Homoduplex against the cut heteroduplex containing mutation sequences by Image J software.

3.4 Generation of the targeting matrix

Figure 9 shows a scheme of all the elements of the donor matrix that served as template for homologous recombination in the intron 2 of the *PKLR* locus. The homology arms (HAs) were designed

to start 25 bp to the right side and 25 bp to the left side of the cutting site of PKLR01 MG. Both HA were designed to be 1 Kilobase (Kb) long. The HAs were chemically synthesized and then included in a pUC57 (pUC-HAs), leaving several restriction sites in the middle in order to clone the additional elements of the matrix. 1) Expression Cassette (EC), composed of the Splicing Acceptor (SA) from intron 1 of *PPP1R12C* gene, the codon optimized cDNA from *RPK* exons 3-11 (coRPKE3-11) with a FLAG tag and SV40 polyA at the end. 2) Selection Cassette (SC), composed of the mouse PGK promoter, which drives the expression of the Puromycin resistance and Thymidine Kinase fusion protein, and the BGH PolyA signal. The whole SC was floxed by LoxP sequences so that it could be removed by ectopic Cre expression.

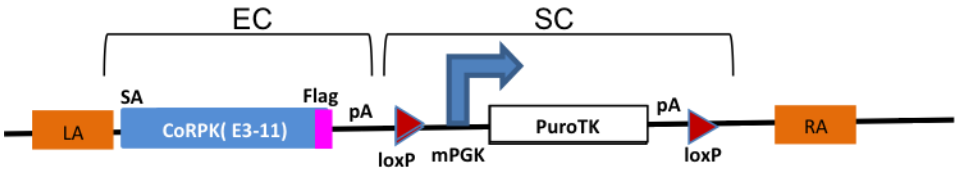


Figure 9: PKLR Donor Matrix elements. Homology Arms are shown in orange, LA (Left Arm) and RA (Right Arm). EC: Expression Cassette: SA, Splicing Acceptor that will anchor splicing donor from endogenous exon 2. CoRPK E3-11, Codon Optimized version of *RPK* cDNA, exons 3 to 11. Before the Stop codon there is a Flag Tag that will allow us to follow the expression of the transgene. pA: SV40 PolyA signal. SC: Selection Cassette, floxed by LoxP signals: Mouse PGK promoter drives the expression of the Puromycin resistance and Thymidine Kinase fusion gene. pA: BGH PolyA signal.

The EC was constructed using Overlapping PCR (OpPCR) procedure, by the amplification of a longer product by using 2 PCR products as templates, which share some homology in their ends. These 2 PCR products were the coRPK cDNA (PCR1) and the SV40polyA (PCR2), amplified by primers 17, 18 and 19, 20 respectively. The Splicing Acceptor was included in the forward primer (primer nr 17) of the PCR1 product and the FLAG tag in the reverse (primer nr 18). All the amplifications were done with Pfx50 and the OpPCR protocol was as follows: PCR1 and PCR2 were agarose gel purified and 50 ng of PCR1 and an equimolecular quantity of PCR2 were included in the OpPCR mixture without primers and the following cycle was applied. 94 °C for 5 min, 14 times: 94 °C for 20 seg., 60 °C for 30 seg., 68 °C for 90 seg. Primers 17 and 20 were added and the previous cycle was repeated 20 more times. Then 68 °C for 5 mins. and 4 °C until stopped. After amplification, the EC was cloned into pCR®-Blunt vector by Zero Blunt PCR Cloning Kit and the quality of the sequence was verified by sequencing with M13 Forward and Reverse primers. Selection Cassette (SC) was kindly provided by Dr. Brian Davis laboratory (Houston, Texas, USA). Both Expression (EC) and Selection Cassette (SC) were included in pUC-HAs by the In-fusion HD Cloning Kit (Clontech Laboratories, CA, USA) following vendors guidelines. In more detail, pUC-HAs was linearized by a NheI restriction enzyme that was in the middle of the two HAs. The EC and the SC were introduced at NheI site in one step by using In-Fusion Technology. According to the In-Fusion provider, each of the pieces to be cloned has to have homologous ends with the other fragments cloned around. The homologous sequences are included in the primers for PCR amplification of the products to include in the plasmid (primers 21-24). The linearized plasmid and the amplified EC and SC were agarose gel purified and the quality and quantity of each product were assessed in another agarose gel. For the In-Fusion reaction 50ng of the linearized plasmid and a molar ratio of 1:2:2 of the other 2 inserts were used in a final volume of 10 µl. After In-Fusion reaction Stellar Competent cells (Clontech) were transformed and 30 colonies were screened by colony PCR as previously described (Machado, Manco et al. 2012) using primers 17 and 20 and Herculase® Fusion II polymerase. The expected PCR product was amplified in one of the 30 colonies. This clone was grown and plasmid DNA was isolated. The presence of the correct

targeting vector was assessed by the restriction enzyme HindIII. Sequencing the whole construct by primers nr 25-35 assessed the integrity of the PKLR targeting matrix.

3.5 Homologous recombination in K-562 cell line

Nucleofection was performed the same way as for Surveyor (or Cel1 assay) but in this case the PKLR01 MG was delivered as plasmid DNA and in combination with PKLR targeting matrix. Two days after transfection, puromycin at 2 μg/ml was added to the culture and it was maintained for 20 days.

3.5.1 PCR assessment of Homologous Recombination

PCR was carried out using two pairs of primers that recognized the *PKLR* locus outside the homology arms and another that recognized inside the integrated targeting matrix (primers nr 36 and 37). This PCR was carried out by Herculase® II Fusion DNA Polymerase (Stratagene).

3.5.2 Erythroid differentiation of K-562

K-562 cells were differentiated by addition of 25 μM of Hemin BioXtra, from Porcine, ≥98.0% (Sigma-Aldrich) as previously described (Baliga, Mankad et al. 1993). Differentiated and undifferentiated cells were analyzed by FACS.

3.6 Homologous recombination in hiPSCs

Nucleofection was performed the same way as for Surveyor assay. Different amounts of nucleases as plasmid DNA or mRNA were nucleofected by Amaxa. After nucleofection, the cells were seeded into a feeder of irradiated puromycin resistant MEFs and 48 hours after transfection, puromycin (0.5 μg/ml) was added to hES media. When performed with MG, 10% of each condition was seeded into one well of a 6 well plate and 7 days post-selection, Phosphatase Alkaline Staining was performed in order to count resistant colonies and compare between the different conditions. The rest of the nucleofection reaction was seeded into a 100 cm tissue culture plate. The selection was carried out for 6-10 days and the resistant colonies were picked individually under the stereoscope and seeded into new Puromycin resistant MEFs to start a clonal population from each single colony. Puromycin was maintained in the culture until the whole homologous recombination analysis was finished.

3.6.1 HR analysis

3.6.1.1 Colony PCR

One puromycin resistant colony per clone was picked into 100 ul of PBS. After spinning down for 5 minutes at 300 g, it was resuspended in 20 μl of lysis buffer: 5mM Tris pH 8.3, 0.45% NP40 and 0.45% Tween20 and 100 μg/μl of Proteinase K and treated for 2 hours at 55°C and 10 minutes at 95°C. 10 μl of the lysis reaction were used per PCR. Two PCR types were performed looking for integration by HR: primers 36 and 37 (Primer pair 1, pp1) and 38 and 39 (primer pair 2, pp2). Both

reactions were carried out by Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific), pp1 with CG rich Buffer and pp2 with high fidelity buffer. Another control PCR to check for DNA presence was carried out to look for an endogenous unmodified gene of the genome (*CFTR* gene) by primers 40 and 41 and GoTaq DNA polymerase (Promega, WI, USA)

3.6.1.2 Southern Blot

5-20 µg of genomic DNA were isolated from each clone and digested with ScaI and SpeI (New England Biolabs, MA, USA) in separate reactions for 8 hours at 37°C. The samples were concentrated by precipitation with ethanol and sodium acetate and resolved on a 0.8% agarose gel. Transfer to a NytranSuPerCharge nylon membrane was carried out overnight by turbo Blotter transfer system (Whatman, GE healthcare) following systems guidelines and then exposed to UV cross-linking (120 mJ/cm^2). Pre-hybridization and hybridization were carried out by quick hybridization solution (Agilent technologies) in rolling cylinders inside the hybridization oven at 65°C. The probe was generated by excising the Targeting Donor by HindIII restriction enzyme and isolating a 1935 bp band from an agarose gel. It was labeled with radioactive P^{32}dCTPs by Prime-It II Random Primer Labeling kit (Agilent) following vendor guidelines. Once labeled, the non-incorporated radioactivity was removed by cleaning the probe with IllustraNICK Columns (GE Healthcare). After 15 minutes of prehibridization, the probe was added to the quick hybridization solution and left overnight. After 12 hours, the membrane was washed twice with 2X SSC, 0.1% SDS for 10 minutes at room temperature, then for 15 minutes with pre warmed 1X SSC, 0.1% SDS at 65°C and finally for 50 minutes with 0.1X SSC, 0.1% SDS at 65°C with two media changes. Then, the membrane was covered with plastic wrap and introduced in a Hyperscreen autoradiography cassette together with an Amersham Hyperfilm ECL (GE Healthcare) that was exposed overnight at -80°C covered. After a minimum of 10 hours, the film was developed in an automated processor.

3.6.2 Western Blot

The presence of untargeted RPK and flagged RPK was analyzed in hiPSC erythroid differentiation protein lysates. For RPK, a polyclonal antibody (Ab. nr 21) developed by Dr Meza (Táchira Medicine School, University of Los Andes, San Cristóbal, Venezuela) by rabbit immunization with the human recombinant enzyme (Diez, Gilsanz et al. 2005) was used. Flagged RPK was detected by Ab. nr 20 and β-Actin, used as loading control, with Ab. nr 22. Cell lysates were generated treating the cells with a buffer consisted of 50mM Tris pH 7.5, 150 mM NaCl, 1% NP40, 2mM EDTA supplemented with Protease Inhibitors (Roche Diagnostics, Switzerland). Total protein concentration was determined using Quick Star Bradford Protein Assay kit (Bio-Rad, CA, USA). Samples were mixed with NuPAGE LDS Sample Buffer (Invitrogen) and 2-mercaptoetanol and loaded into SDS polyacrylamide gel (4-

12% Bis-Tris Nu-PAGE, XCell II Blot, Invitrogen) and run at 125 V for 1h. A pre-stained protein ladder was used (PageRuler, Thermo Scientific). Proteins were then transferred using a humid transfer system into nitrocellulose membranes (mb) (Hybond-ECL, Amersham Bioscience) at 150V for 1.5 hours at 4°C. Protein containing mb was blocked with 5% milk in TBS for 1 hour at RT and incubated with primary Ab diluted in blocking buffer overnight at 4°C. The mb was washed three times with TBS-T (Tris-PBS, 0.05% Tween20) and incubated with secondary Ab conjugated to horseradish peroxidase (HRP) for 1 hour at RT. Detection was carried out using Amersham ECL Select Western Blotting detection Kit (GE Healthcare) and Kodak Medical X-Ray Films.

4 General techniques

4.1 Immunohistochemistry

- For pluripotency analysis: hiPSCs were seeded on feeder pre-coated 2 well Lab Tek chamber slides (Nunc, Thermo Fischer, Thermo Fischer Scientific, MA, USA). When the colonies had the appropriate size (≈6 days), they were fixed with 4% PFA for 10 minutes at RT. Fixed colonies were washed twice and blocked and permeabilized with the following solution for 1 hour at room temperature: 1% BSA, 10% FBS, 0.3M Glycine, 0.1% Tween20 in PBS. Two different double detections for presence of embryonic proteins were performed by immunostaining with: 1) rabbit anti-NANOG (Ab nr 6) + mouse anti-Tra160 (Ab nr 7) and 2) rabbit anti-OCT4 (Ab nr 8) + mouse anti-SSEA4 (Ab nr 9) (see antibodies table). The samples were incubated overnight in PBS, washed 3 times with PBS and incubated for 1 hour with anti Rabbit Ab conjugated to Alexa488 (Molecular Probes), anti Mouse Ab conjugated to Texas Red (Jackson Immunoresearch Laboratories, UK) and DAPI or Topro3 for DNA staining. After 3 washes with PBS, Vectashield mounting solution (Vector Laboratories) was added and the slides were covered and let dry overnight. Immunofluorescence images were acquired with Zeiss Axioplan2 epifluorescence microscope (Carl Zeiss) equipped with an AxioCam MRm camera (Carl Zeiss) and the pictures obtained were processed with AxioVision version 4.6.3 (Carl Zeiss) and Corel Photo-Paint 11 (Corel).

- For teratoma analysis: First, slides were deparaffinizated at 54°C for 30 minutes followed by 20 immersions in Histoclear followed and then rehydration in decreasing ethanol solutions of 100%, 95%, 70%, 50% (v/v) and finally in water. Antigen retrieval was carried out in 10mM Sodium Citrate, 0.05% Tween 20, pH 6.0 at 95°C for 20-30 minutes followed by cooling for another 20 minutes and 2 washes with PBS. Slides were blocked for 30 minutes at RT with PBS/ 4% FBS/ 0.2% TritonX10. For immunostaining, slides were incubated overnight at 4°C with rabbit anti-neuron specific beta III tubulin (Ab. Nr 15). The rest of the protocol continued as previously explained for pluripotency analysis.

4.2 Flow cytometry

- For population profile analysis: PB-MNC were stained for 30 minutes at 4°C in the dark using the following cocktail of conjugated antibodies: anti CD3-PECy5 (Ab nr 1), anti CD19-PE (Ab nr 2), anti CD14-PE (Ab nr 3), anti CD15-PECy5 (Ab nr 4) and anti CD34-PE-Cy7 (Ab nr 5) (See antibodies table for details). Then, they were washed and resuspended in DAPI containing PBA (1.0% BSA/0.1% sodium azide in PBS) buffer. Azami green expression was also analyzed in each positive population by flow cytometry in BD LSR Fortessa (Becton Dickinson, CA, USA). FSC files were analyzed with Flow Jo software v10 (TreeStar, OR, USA).

- For pluripotency analysis of hiPSC: cultures were washed twice with PBS and then treated with Accutase for 5 minutes at 37°C. Then they were collected, washed with PBS, 2% FBS (FACS Buffer) and stained first with the following mix of Antibodies for 30 minutes at 4°C in a final volume of 100 µl: anti-CD29-Alexa 488 (Ab nr 10) and anti-SSEA4-APC (Ab nr 11). Then they were washed with FACS buffer, fixed with 2% PFA for 30 minutes at RT, washed again and permebilized with permeabilization buffer (0.1% Saponine and 0.1% BSA in PBS) for 30 minutes at RT in the dark. After washing, staining with anti-OCT4-PE (Ab nr 12) was done. Cells were finally analyzed by flow cytometry

- For erythroid analysis: cells at days 15 and 20 of differentiation were collected and disaggregated, washed twice, resuspended in PBA buffer and stained with the following antibodies for 30 minutes in the dark at 4°C: anti CD235a-PE (Ab nr 17), anti CD71-FITC (Ab nr 18), and anti CD45-APC (Ab nr 19).

- FLAG expression was also assessed by flow cytometry. Cells were fixed, permebilized and washed by BD Cytofix/Cytoperm Fixation and Permeabilization Kit (BD Biosciences, NJ, USA) following vendor guidelines and stained with a monoclonal antibody against DYKDDDDK FLAG epitope conjugated to Alexa Fluor 647 (Cell signaling Technology, MA, USA) (Ab nr 20). One million cells were stained with 2 µl of Ab in a final volume of 50 µl for 30 minutes at room temperature protected from the light. Cells were washed and analyzed as previously described.

4.3 Gene expression analysis

- For the study of Erythroid differentiation, RNA was extracted at day 21, and RPK and MPK expression were analyzed by quantitative reverse transcriptase PCR (qRT-PCR). RNA was isolated by TRIzol® (Life Technologies) extraction and isopropanol precipitation. For cDNA generation AmbionRETROscript First Strand Synthesis Kit or SuperScript VILO cDNA Synthesis Kit (Invitrogen, Life technologies) was used as recommended by the manufacturer. Thirty ng or more of equivalent cDNA were used in each reaction. The following primers were used for each transcript detection;

RPK: primers nr 46 and 47, *MPK*: nr 48 and 49 for and β-Tubulin: primers 50 and 51 (used as housekeeping gene). The specificity and efficiency of these primers was validated prior using them. qRT-PCR was carried out in an Applied Biosystems 7500 Real-Time PCR System (Life Technologies) using Fast SYBR Green Master Mix (Applied Biosystems, Life Technologies) as described by manufacturer.

- Sendai viral transcripts presence was analyzed by reverse transcription PCR using a pair of primers that bind a sequence present in all the vectors used in our experiments. The detection of SeV cDNAs was performed by Pfx50 polymerase (Invitrogen, Life Technologies) using 2 μl of cDNA and primers nr 42 and 43. To assess the quality of the cDNA sample, GAPDH presence (primers 44 and 45) was also analyzed. cDNA from PB-MNCs transduced with azami green SeV was used as positive control.

5 Primers Table

Primer Nr	Primer Name	Sequence (5'--> 3')
1	Ex4_Fw 2	TTGGGTTTGGTTGCCTCTCA
2	Ex4_Rv2	TGAGTGGGGAACCTGCAAAG
3	Ex9_Fw1	GTCCTACAACTTTGACATCC
4	Ex9_Rv 1	TAGCTCCTCAAACAGCTGC
5	PKD3mut 4F	AGAGACAAGCGATGTCGCCAATG
6	PKD3mut 4R	TGTTCATGAGGAAAGACAGCAGGCT
7	PKLR01v0 DeepS 2F	**CCATCTCATCCCTGCGTGTCTCCGACTCA**Gagtaggtaaatggcaaaacccatc
8	PKLR01v1 DeepS 2F	**CCATCTCATCCCTGCGTGTCTCCGACTCAG**ctacggtaaatggcaaaacccatc
9	PKLR01v2 DeepS 2F	**CCATCTCATCCCTGCGTGTCTCCGACTCAG**cgagggtaaatggcaaaacccatc
10	PKLR01 DeepS 3R	**5'BioTEG/CCTATCCCCTGTGTGCCTTGGCAGTCTCAG**gcctgcttgctttctttcc
11	Fw_RNA_MG_pCl	CGAAGGATCCTAATACGACTCACTATAGGGAGATACACAGCGGCCTTGCCAC.
12	Rv_RNA_MG_pCL	ATCGGATCCAAAGCTGGGTGCTAGCGCTCGAG
13	Fw_RNA_AAVS1s	CGAAGGATCCTAATACGACTCACTATAGGGAGACTTAAGCTGATCCACTAGTCCAG
14	Rv_RNA_AAVS1s	ATCGGATCCTCCAGAGGTTGATTGTCG
15	PKLR01_DS_3F	GGTAAATGGCAAAACCCATC
16	PKLR01_DS_3R	GCCTGCTTGCTTTCTTTCC
17	OPPr1_RPKSV40pA.2	TCGCGGCCGCCTCTTCCTCCCACAGGCC
18	OPPr2_RPKSV40pA.2	CTCACGCGTTCATTTATCGTCATCGTCTTTGTAGTCGCTGATGGACAGCACCCTCATG
19	OPPr3_RPKSV40pA.2	GACTACAAAGACGATGACGATAAATGAACGCGTGAGTTACAAATAAAGCAATAGCATC
20	OPPr4_RPKSV40pA.2	CACAACTAGAATGCAGTGAAAAAAATGC
21	Inf_1.3_Fw	CCAGCCATGGGCTAGCTCGCGGCCGCCTCTTCCTC
22	Inf_1.3_Rv	GTATGCTATACGAAGTTATCCTCTCGAGCACAACTAGAATGCAGTG
23	Inf_2.3_Fw	CACTGCATTCTAGTTGTGCTCGAGAGGATAACTTCGTATAGCATAC
24	Inf_2.3_Rv	CAATTTAAATGCTAGCAGGATAACTTCGTATAATGTATGC
25	Seq 3 Fw	CCAGCCTGGGCAACAGGAG
26	Seq 18 Fw	CAGCCAAGGGCAACTTCCCC
27	Seq 5 Fw	CGTGGAGCACGGCGTGGAC
28	Seq 7 Fw	GACAGGCTGGCGGCCTGGCAG
29	Seq 12 Fw	CAAGAACTCTTCCTCACGCG
30	Seq 10 Fw	CGCGCGGTACCTTATGGGCAG
31	Seq 8 Rv	CTTGGCCAAACGCCTCCGTTC
32	Seq 15 Fw	GAAAGAAAGAAAGGAGTGAAAG
33	Seq 21 Rv	CATCGAGCGGGTCACCGAGCTG
34	Seq 20 Rv	GCCCTCCTGTGCTACCCGGC
35	Seq 9 Rv	CTTGGCCAAACGCCTCCGTTC
36	HR PKLR Fw6	AGGGTATGCTGAGAGACGAAG
37	HR PKLR Rv6	CCGGGCGATATTCATGCCG
38	HR Hou2 Fw1	TGGGATGGGTCAAGCTACAG
39	HR Hou2 Rv1	CAATGCTCTCTGCGTGGTACT
40	CF Fw1	CTATCTGAATCATGTGCCC
41	CF Rv2	CTCATTAGTGAGACAAACGTCC
42	sendai detection Fw	GGATCACTAGGTGATATCGAGC
43	Sendai detection Rv	ACCAGACAAGAGTTTAAGAGATATGTATC
44	qGAPDH Fw	GCTCTCTGCTCCTCCTGTTC
45	qGAPDH Rv	ACGACCAAATCCGTTGACTC
46	qRPK4 Fw	ATATCATCCCTGCAGCTTCG
47	qRPK4 Rv	CAGCTCCTGGGTCAGTTGG
48	qPKM Fw	ATATCATCCCTGCAGCTTCG
49	qPKM Rv	CAGCTCCTGGGTCAGTTGG
50	qBTUB Fw	ATATCATCCCTGCAGCTTCG
51	qBTUB Rv	CAGCTCCTGGGTCAGTTGG
52	PHF2Fw	TGCCATCCTACAGGCTCAAC
53	PHF2Rv	TTGTTTGGAGACGATGCGGA
54	ZNF747Fw	GCTAAGCAGGTAGCTGGAGG
55	ZNF747Rv	CGGGATCCGAGTGGAGAAAG
56	SNX3Fw	ACGTTTTTGCCCCTTCTTGC
57	SNX3Rv	AGTGCATAGCTTTATGGACAGT
58	TNRC18Fw	TTGAGCAGCTCCTTGGCAG
59	TNRC18Rv	CGCTTTTCGGCAAGAAGGAC
60	TUBGCP6Fw	GGGACCCTGTGTGCCAATTA
61	TUBGCP6Rv	TCTGGCTTCTTGTGGATGGG
62	APOA5Fw	TTGCTCAGAACCTTGCCACT
63	APOA5Rv	CTTCCACCCATACGCCGAG
64	RUSC2Fw	TGATGCCAACTGCAACTCCT
65	RUSC2Rv	CAACATACGTTCCAGCGAGC

6 Antibodies Table

Antibody Nr	Antibody Name	Flourochrome	Brand, Cat #	Dilution
1	Anti-CD3	PECy5	Immunotech, A07749	1:10
2	Anti-CD19	PE	Immunotech, A07769	1:10
3	Anti-CD14	PE	BD, 345785	1:10
4	Anti-CD15	PRCy5	Beckman Coulter, IM2641	1:10
5	Anti-CD34	PECy7	BD, 348791	1:10
6	Anti-Nanog	2ary Ab; Anti Rb Alexa 488	ab21624	1:50
7	Anti-Tra160	2ary Ab; Anti Ms texas Red	ab16288	1:50
8	Anti-Oct4	2ary Ab; Anti Rb Alexa 488	ab19857	1:500
9	Anti-SSEA4	2ary Ab; Anti Ms texas Red	ab16287	1:50
10	Anti-CD29	ALEXA488	Serotec, MCA2298A488	1:10
11	Anti-SSEA4	APC	R&D, FAB1435A	1:10
12	Anti-Oct4	PE	BD, 560186	1:10
15	Anti-β III Tubulin	2ary Ab; Anti Rb Alexa 488	Abcam, ab18207	1:1000
17	Anti-CD253a	PE	Biolegend, 349106	1:25
18	Anti-CD71	FITC	Beckman Coulter, IM0483	1:10
19	Anti-CD45	APC	Ebioscience, 17-9459	1:10
20	Anti-FLAG	Alexa 647	CellSignalling, 3916S	1:50 (FC), 1:500 (WB)
21	Anti-RPK	2ary Ab; Anti Rb HRP	Home-made	1:10000
22	Anti-B-Actin	2ary Ab; Anti Ms HRP	Abcam, Ab6276	1:4000

VII Results

1. Cell reprogramming

In order to have an unlimited source of stem cells from PKD patients, our first goal was the generation of hiPSC from PKD patients that could be used for disease modeling and gene correction of the disease.

1.1. Generation of hiPSCs from Peripheral Blood Monuclear Cells (PB-MNC)

We investigated several procedures for the generation of hiPSC from PKD patients. In order to obtain a clinically oriented reprogramming platform, the efficiency of two available non-integrative/excisable technologies were tested: Excisable reprogramming lentiviral platform (STEMCCA (Sommer, Stadtfeld et al. 2009)), which expresses the four Yamanaka´s Factors in a polycistronic construct and can be excised using Cre Recombinase, and reprogramming Sendai Viruses (SeV), non-integrative RNA vectors expressing the same factors in four independent vectors. In order to evaluate their efficiency, a comparison of both methodologies was carried out by reprogramming healthy donor PB-MNCs. The protocol used for reprogramming is represented in figure 10A. Non-lymphoid cells were pre-stimulated by a specific cytokines combination. Five days post-transduction, 15,000 cells/well were plated in each well of a 6 well plate pre-coated with human irradiated fibroblast feeders (HFF). After 4-6 days the first colonies appeared, that turned out to be unstable. During the second week, colonies with a clear morphology of pluripotent colonies came out: flat, dense and with well-defined colony edges. After 12 days, the cells were fixed and Alkaline Phosphatase (AP) staining was performed in order to estimate the quantity of reprogrammed colonies (Figure 10B).

Efficiency was calculated as the number of colonies over the number of seeded cells, assuming that each colony was originated from a single reprogrammed cell, so ($\frac{colonies}{15{,}000} x\ 100$). As the number of cells seeded is 15,000 the lowest efficiency that could be detected was 0.007% (= $\frac{1}{15.000} x\ 100$). The efficiency of Sendai reprogramming vectors was 0.4-0.6%, whereas no pluripotent colony was formed after STEMCCA transduction. Thus, SeV reprogramming technology was selected for the generation of hiPSC from patient PB-MNC samples.

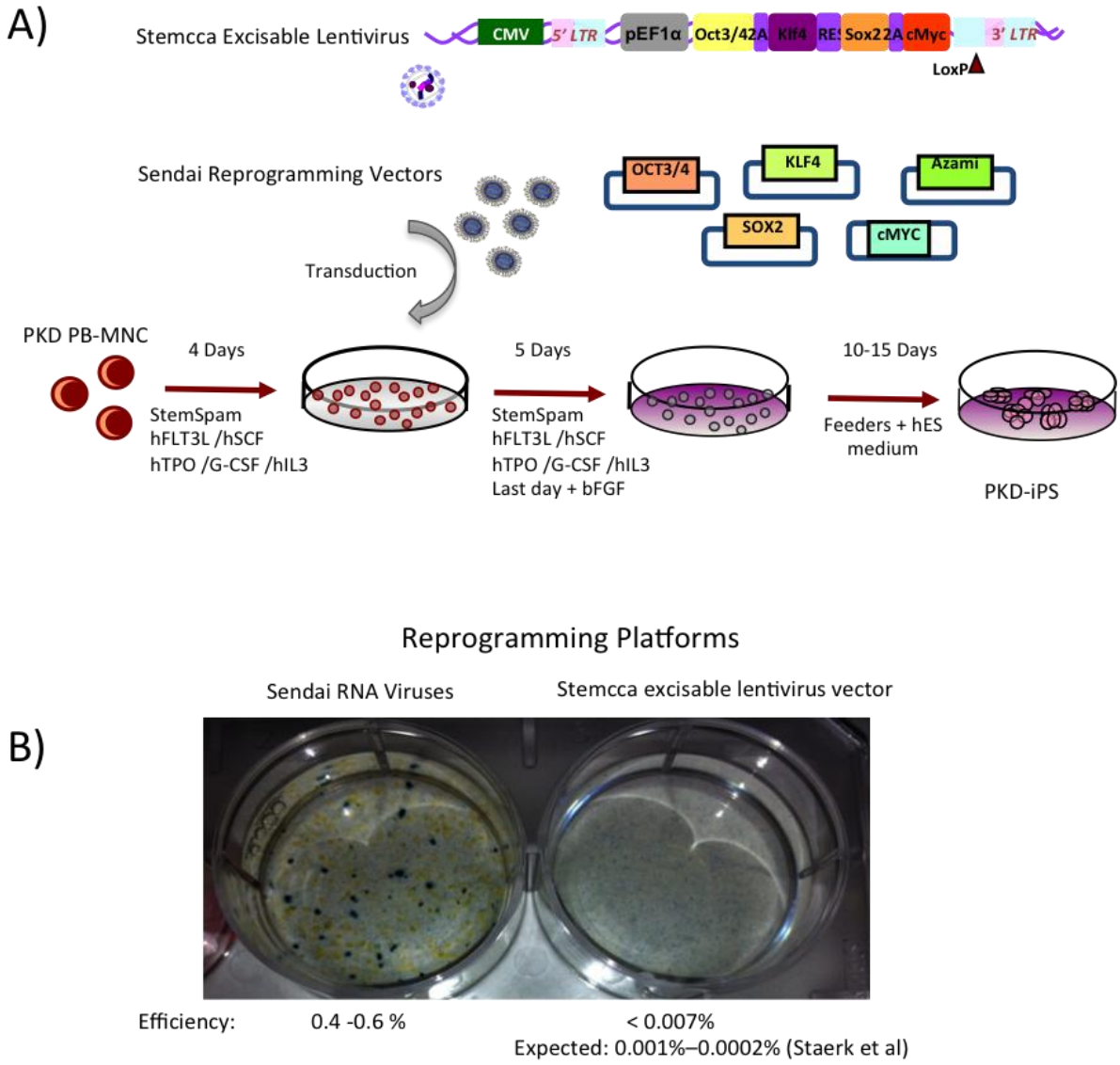


Figure 10: Peripheral Blood Mononuclear Cells (PB-MNC) reprogramming to pluripotency. A) Reprogramming methodology: Cells are prestimulated with the shown cocktail of cytokines and then transduced with either Stemcca excisable lentiviral polycistronic vector or Sendai reprogramming vectors. B) Reprogramming efficiencies comparison between the two reprogramming platforms is shown. Twelve days after seeded in HFF feeders, AP was performed to detect pluripotent colonies.

1.1.1 Targeting non-lymphoid cells for reprogramming

As PB-MNCs is a heterogeneous population composed of many different hematopoietic lineages and we would like to avoid any selection step to minimize the cells manipulation, reprogramming of lymphoid cells should be taken in consideration. If a T or B cell is reprogrammed, the genetic information corresponding to the T cell receptor (TCR) or to the immunoglobulin (Ig) chains is going to be rearranged and therefore, the resulting hiPSCs are not going to have the ability to generate a complete repertoire of T or B lymphocytes with different specificities, limiting its differentiation ability.

Trying to avoid lymphoid reprogramming, we added a combination of cytokines to prompt transduction, proliferation and survival of the non-lymphoid cell subsets present in human PB-MNC. To analyze SeV transduction efficiency under the stimulation conditions proposed, we looked at the proportion of transduced cells within the different hematopoietic subsets by a control SeV encoding for the Azami green fluorescent marker. The presence of marked cells in the different hematopoietic cell types was assessed five days after transduction (Fig. 11). We stained the cells with the following cocktail of antibodies and looked for Azami green expression in each positive population (T Lymphocytes: anti-CD3, B Lymphocytes: anti-CD19, Myeloid cells: anti-CD14 and anti-CD15 and Hematopoietic Progenitors: anti-CD34). A high proportion of the total transduced cells were lymphoid cells (50%) due to their high presence of in the original sample. However, in the myeloid and CD34 positive subset, the transduction percentage was much higher than in lymphoid cells: 54% of transduction in macrophages and granulocytes and 76% of hematopoietic progenitors whereas 13% in lymphocytes, which indicates a preferential transduction of myeloid cells and hematopoietic progenitors under the selected culture conditions. These conditions were selected for the following experiments.

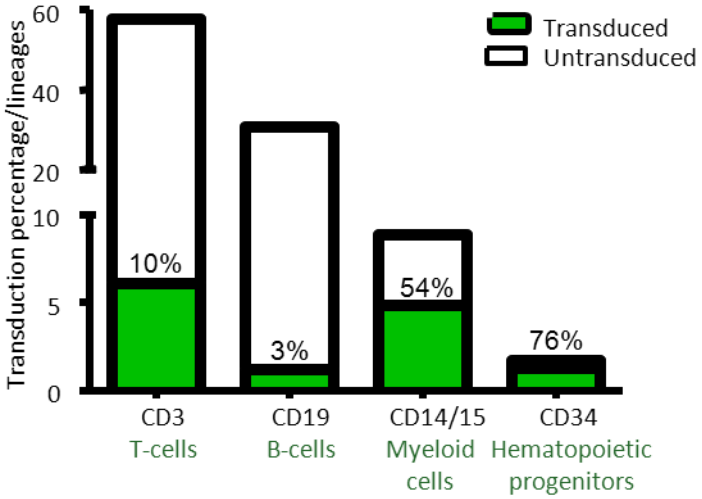


Figure 11: Analysis of transduction efficiency within different PB-MNC populations after stimulation with cytokines and transduction with an Azami green expressing SeV. Green plus white bars show percentage of each cell type within the pool. Green bars show transduced cells within the specific population and the values in percentages are presented on the top of each green bar.

1.1.2 Pyruvate Kinase Deficient (PKD) patient and healthy donor reprogramming by SeV

For the generation of patient specific hiPSCs, PB-MNCs were isolated from 3 PKD patients (PKD1, PKD2 and PKD3) and 2 healthy donors (PB1 and PB2) (Table 5 and Figure 12) and subjected to the selected stimulation and transduction protocol. Several clones (20-100) were successfully picked from 2 patients (PKD2 and PKD3) and 1 healthy donor (PB2). There was no apparent difference between the healthy and diseased hiPSCs, the kinetic of colonies appearance was as previously

described, starting to appear after one week but gaining pluripotent characteristics with time. Several clones from PKD1 (PKD1iPSC) could be selected and expanded for 2-3 passages but they disappeared due to technical problems. We were not able to reprogram PB1. In a second experiment using PKD1 and PB1, reprogramming was neither achieved. Bright field images of hiPSC cultures growing over a feeder layer are shown in figure 12 from the following lines: PKD2iPC c78, PKD3iPC c1 and PB2iPSC c33.

		Experiment nr 1	Experiment nr 2
Patient sample	PKD1	Isolated clones but not survived	Not reprogrammed
	PKD2	100 clones	
	PKD3	50 clones	
Healthy Donor	PB1	Not reprogrammed	Not reprogrammed
	PB2	20 clones	

Table 5: Reprogramming attempts in Pyruvate Kinase Deficiency patients and healthy donors

To verify the origin of the generated hiPSCs, disease specific genotypes were amplified by PCR and analyzed by Sanger sequencing (Fig. 12). Genetic alteration of PKD2 patient is a compound heterozygous mutant showing mutations in exon 4 (359C>T (Ser120Phe)) in one allele and in exon 9 (1168G>A (Asp390Asn)) in the other allele of the *PKLR* gene, leading to a protein with no biological functionality (Rouger, Valentin et al. 1996, Zanella, Bianchi et al. 1997). PKD3 patient genotype is a IVS9(+1)G>C splice mutation in the invariant sequence GT of the splicing donor of intron/exon 9/10. These mutations usually generate unstable and rapidly degraded transcripts or truncated proteins (Manco, Ribeiro et al. 1999). The analyzed clones carried the mutations defined in the patients demonstrating their origin. On the other hand, the sequencing of the same location in the healthy donor (PB2iPSC) showed the wild type sequences, as expected.

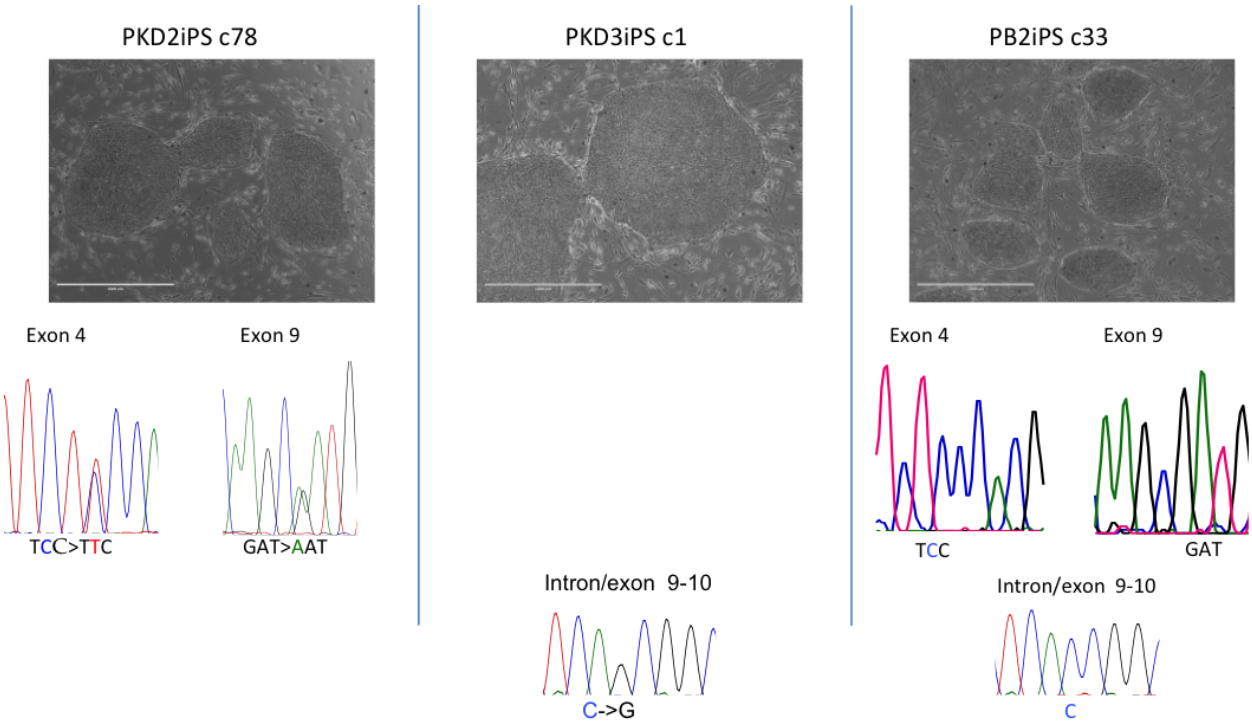


Figure 12: Bright field images and PKD genotype of the hiPSC lines from PKD patients (PKD2 and PKD3) and healthy donor (PB2). Colonies growing over a MEF feeder showing typical ES morphology: flat, defined colony edges and dense. Mutant and wild type genotype of exons 4 and 9 can be seen in patient and healthy samples in the upper row. Lower row, mutant and wild type genotype of the splicing donor of intron9/exon10.

1.1.2.1 hiPSC characterization

A) *Pluripotency analysis*

Pluripotency of hiPSC lines from PKD2, PKD3 and PB2 was analyzed by immunofluorescence, flow cytometry, RT-PCR array and teratoma formation. The expression of a broad set of differentiation and undifferentiation-associated genes was analyzed by quantitative RT-PCR (qRT-PCR) using an hESC RT2 Profiler PCR Array, which contains primers for the analysis of 84 key genes involved in the maintenance of pluripotency and the self-renewal status of embryonic stem cells, such as transcription factors involved in stemness, signaling molecules, cytokines, growth factors and differentiation markers. The level of expression of the analyzed genes in the sample (Y axis) was compared to the one of the reference cell line, the human Embryonic Stem Cell (hESC) line H9 (X axis) in the histograms of figure 13. The level of expression of the analyzed genes was equivalent to the one seen in the control in the majority of the genes, including the ones related to pluripotency and self-renewal (Fig. 13).

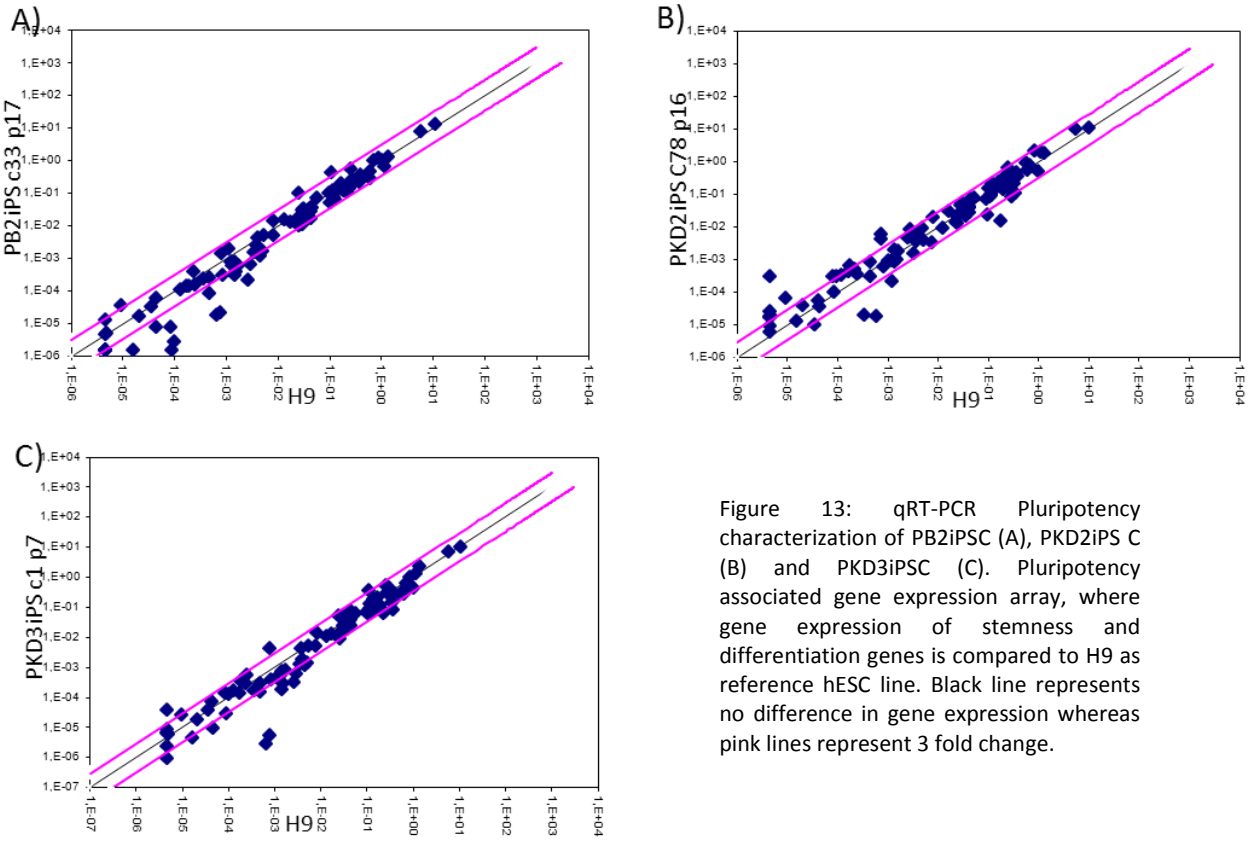


Figure 13: qRT-PCR Pluripotency characterization of PB2iPSC (A), PKD2iPS C (B) and PKD3iPSC (C). Pluripotency associated gene expression array, where gene expression of stemness and differentiation genes is compared to H9 as reference hESC line. Black line represents no difference in gene expression whereas pink lines represent 3 fold change.

Four pluripotency-associated markers were also analyzed at the protein level by immunofluorescence. Figure 14 shows the correct colony morphology in which the cells were stained positive for OCT4 and NANOG or for TRA 1-60 and SSEA4. Surrounding fibroblasts were negatively stained for the assayed markers. OCT4 and SSEA4 were also analyzed by flow cytometry to test the status of the whole culture, showing that in the three cell lines more than 90% of the culture is maintained under an undifferentiation status.

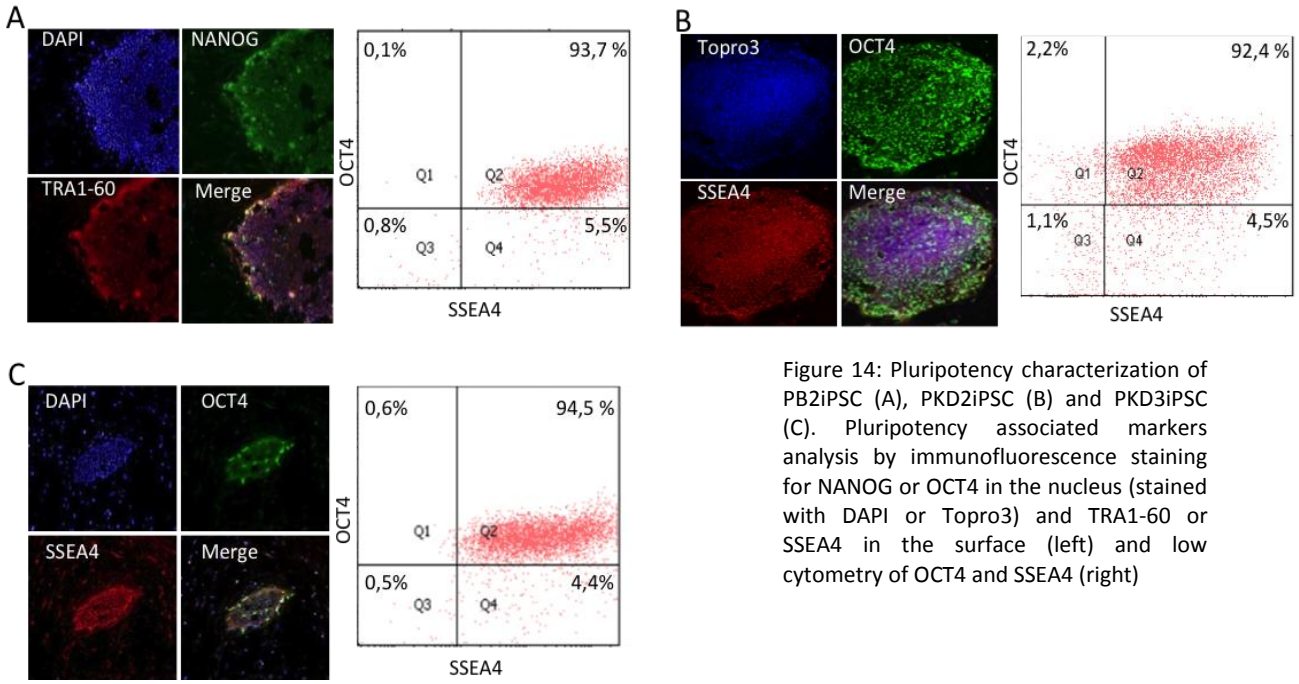


Figure 14: Pluripotency characterization of PB2iPSC (A), PKD2iPSC (B) and PKD3iPSC (C). Pluripotency associated markers analysis by immunofluorescence staining for NANOG or OCT4 in the nucleus (stained with DAPI or Topro3) and TRA1-60 or SSEA4 in the surface (left) and low cytometry of OCT4 and SSEA4 (right)

In order to test the *in vivo* pluripotency of the reprogrammed lines, 10^6-10^7 cells from each hiPSC clone were injected subcutaneously into immunocompromised NSG mice. The three cell lines generated fully differentiated teratomas within 8 to 12 weeks and all of them showed tissues corresponding to the three germ layers (mesoderm, endoderm and ectoderm) (Fig. 15). In more detail, by staining the teratomas with hematoxilin/eosin it was possible to see structures resembling Retinal Pigmented Epithelium (RPE) (stained very dark due to pigmentation), muscle (stained in pink), cartilage, secreting and glandular epithelium and neural rosettes. The teratoma generated by PKD2iPSC line showed no clear morphological evidence for ectodermal tissue, but positive areas for neural marker β-III Tubulin of ectodermal origin were identified by immunofluorescence analysis (Fig. 15).

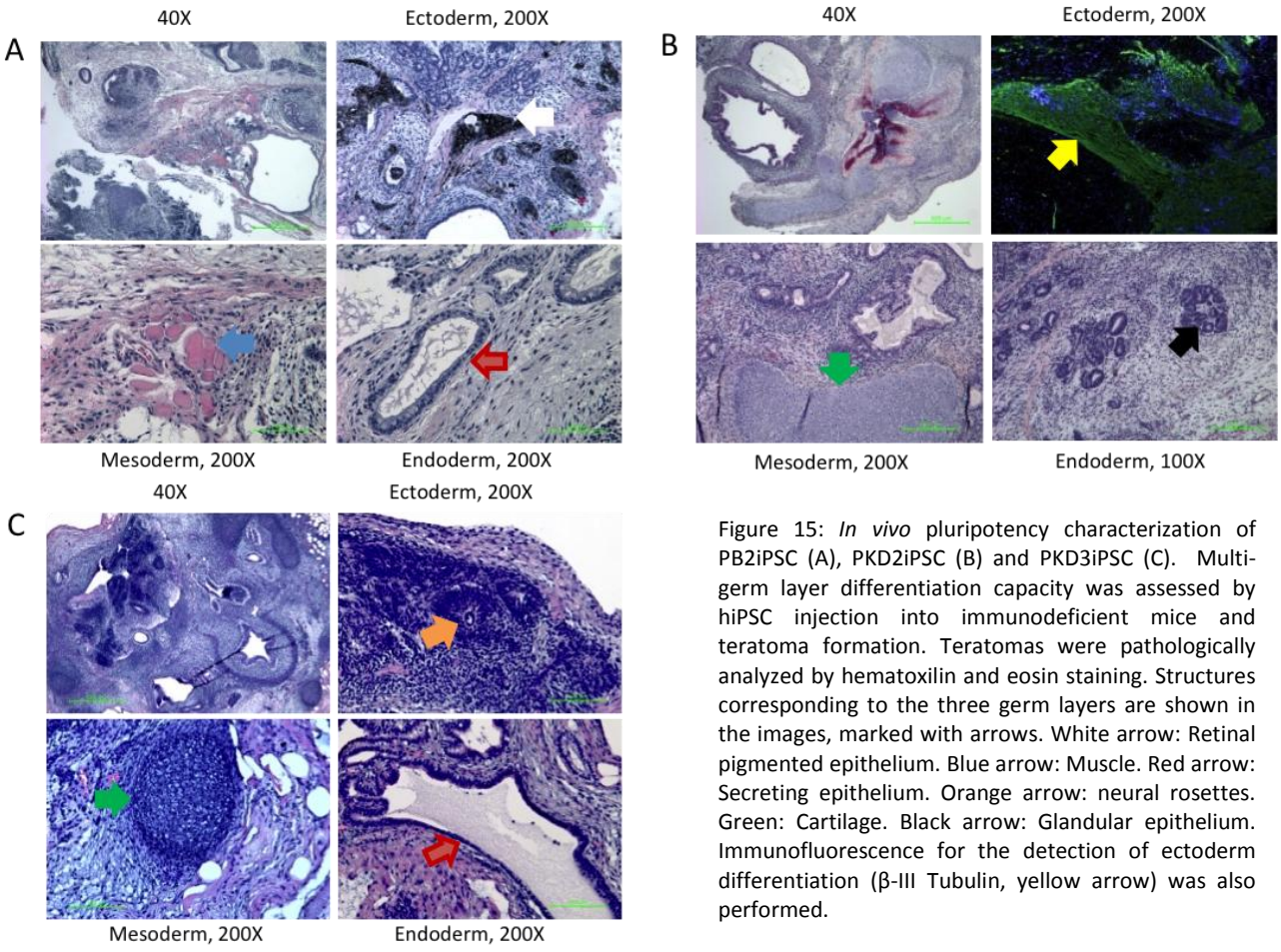
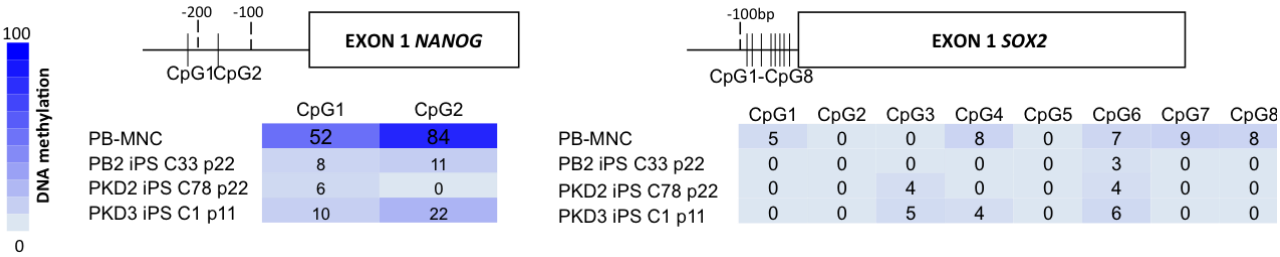


Figure 15: *In vivo* pluripotency characterization of PB2iPSC (A), PKD2iPSC (B) and PKD3iPSC (C). Multi-germ layer differentiation capacity was assessed by hiPSC injection into immunodeficient mice and teratoma formation. Teratomas were pathologically analyzed by hematoxilin and eosin staining. Structures corresponding to the three germ layers are shown in the images, marked with arrows. White arrow: Retinal pigmented epithelium. Blue arrow: Muscle. Red arrow: Secreting epithelium. Orange arrow: neural rosettes. Green: Cartilage. Black arrow: Glandular epithelium. Immunofluorescence for the detection of ectoderm differentiation (β-III Tubulin, yellow arrow) was also performed.

Total pluripotency induction is also associated to endogenous gene expression activation and therefore promoter CpGs demethylation of transcription factors belonging to the pluripotency core. By pyrosequencing technique we analyzed different CpG dinucleotides, 2 belonging to *NANOG* and 8 to *SOX2* promoters in PB-MNC, which is the sample before reprogramming, and in the following reprogrammed samples: PB2iPSC, PKD2iPSC, and PKD3iPSC. Based on the number of methylated

events detected in each sample a number and a color were assigned (darker color and higher number means more methylated events) to each of the CpGs of the two analyzed promoters (Fig. 16). The 2 assayed CpGs from *NANOG* promoter were demethylated in comparison to PB-MNC verifying the induction of activation of endogenous *NANOG* in the different analyzed hiPSCs. Same was observed in the 8 assayed CpGs from *SOX2* promoter, which unexpectedly was not strongly methylated in PB-MNC.



	CpG1	CpG2
PB-MNC	52	84
PB2 iPS C33 p22	8	11
PKD2 iPS C78 p22	6	0
PKD3 iPS C1 p11	10	22

	CpG1	CpG2	CpG3	CpG4	CpG5	CpG6	CpG7	CpG8
PB-MNC	5	0	0	8	0	7	9	8
PB2 iPS C33 p22	0	0	0	0	0	3	0	0
PKD2 iPS C78 p22	0	0	4	0	0	4	0	0
PKD3 iPS C1 p11	0	0	5	4	0	6	0	0

Figure 16: Pluripotency characterization of PB2iPS, PKD2iPSC and PKD3iPSC. *NANOG* and *SOX2* promoters demethylation was analyzed by pyrosequencing and compared to PB-MNC. Based on the number of methylated events a numeric number associated to a color scale was given to each sample. Darker color and higher number mean more level of methylation in the promoter CpG. The demethylation status of *NANOG* promoter is higher in the assayed samples than in PB-MNC. *SOX2* promoter is demethylated not only in hiPSC but also in PB-MNC.

B) Karyotype analysis

As previously explained, reprogramming process and pluripotent cultures maintenance is associated to genomic instability leading to chromosomal aberrations. In order to test the genetic stability of our hiPSC lines, karyotype analysis by G banding of metaphasic chromosomes was performed in PB2iPSC c33 p17, PKD2iPSC c78 passage 15 and PKD3iPSC c1 p10, showing no aneuploidies or large duplications/deletions (> 5 Megabases) in any of them (Fig. 17).

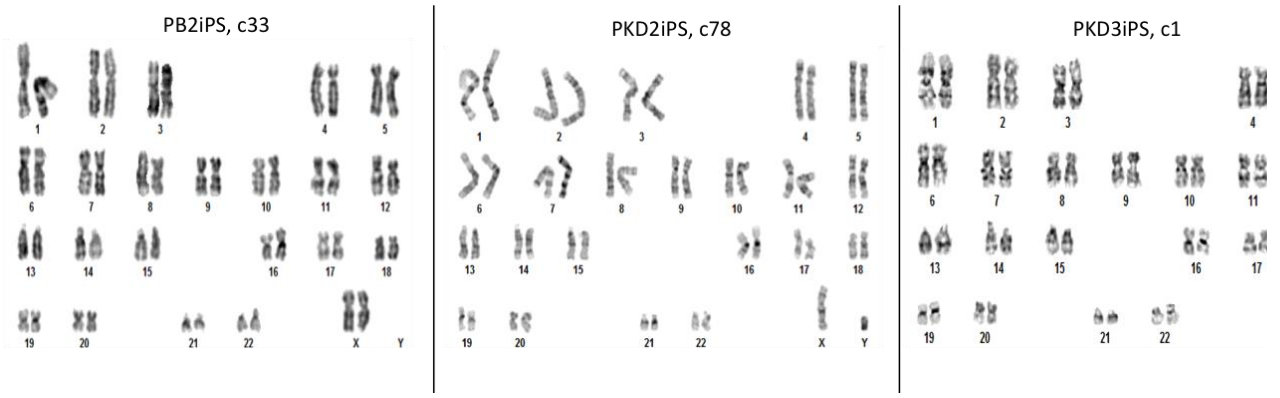


Figure 17: Chromosomal stability of PB2iPS, PKD2iPSC, and PKD3iPSC. Karyotype G-Banding of the 23 pairs of chromosomes showing that the three samples have the expected number of chromosomes and no detectable aneuploidy.

C) T cell receptor and Immunoglobulin genes somatic rearrangement analysis

The analysis of somatic rearrangements of the immunoglobulin chains and T cell receptor (TCR) confirmed that the generated hiPSCs were reprogrammed from non-lymphoid cells. gDNA from five different hiPSC lines, PKD3 c14, PKD3 c10, PKD3 c35, PKD2 c78 and PB2 c33 was isolated and rearrangement analysis was performed following BIOMED-2 strategy. For Immunoglobulin H (IgH), the 3 hyper variable regions and for T cell Receptor, gamma rearrangements VJA and VJB were analyzed. None of the assayed cell lines showed any rearrangements in any of the analyzed regions, discarding that they came from a mature T or B lymphocyte. Figure 18 shows a representative image of gamma VJA rearrangement analysis of two PKD3iPSC clones, clone 10 and clone 14 and 2 controls: a monoclonal population and a polyclonal population with more than one TCR rearrangement. Whereas the monoclonal control showed one peak and the polyclonal several, the hiPSC lines did not show any amplification, confirming that none of them had any somatic rearrangement. The peaks at the left of the image of hiPSC lines are not real products, as the intensity (height of the peak) is not comparable to the one seen in the controls.

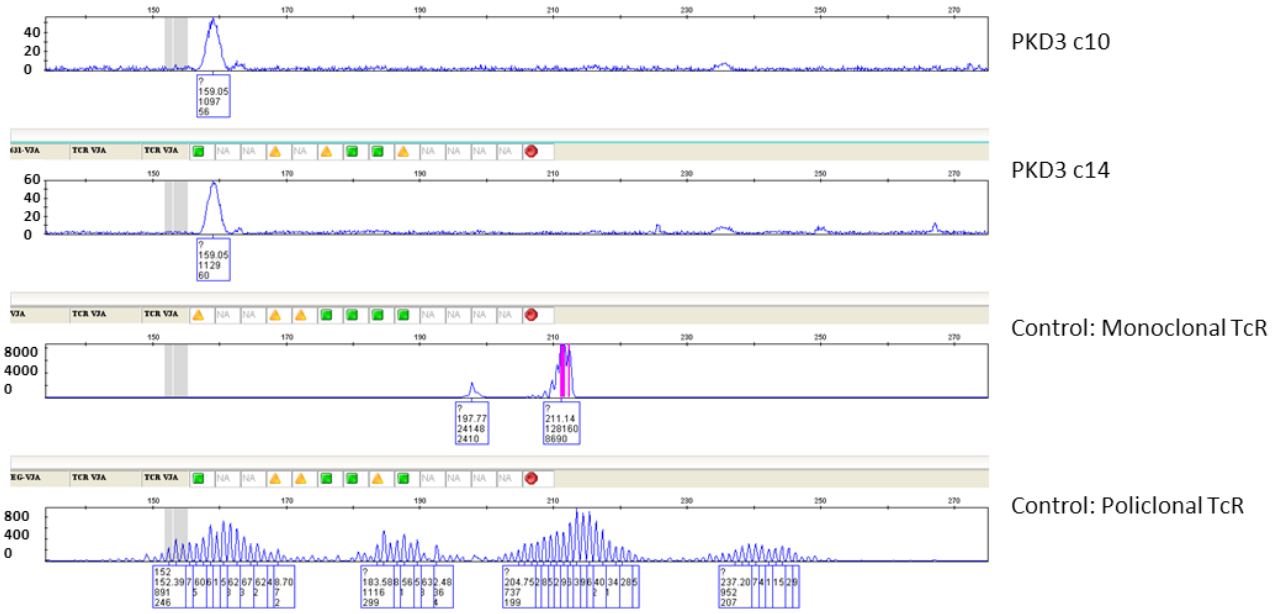


Figure 18: T cell receptor VJA gamma rearrangement analysis in PKD3iPSC clone 10 and clone 14 and in 2 controls: one monoclonal and a polyclonal control. The monoclonal sample shows one peak corresponding to one PCR product whereas the polyclonal shows multiple peaks. In the analyzed samples there are no TCR rearrangements as no amplification is observed. The peaks shown in the analyzed hiPSC samples are not real PCR products (see height peak scale).

D) Reprogramming vectors disappearance

The main advantage of using SeV for reprogramming is that although they are auto-replicative and can be maintained in the cell for long periods of time, once hiPSCs are established and start to divide, SeV vectors tend to disappear because hiPSC divide faster than SeV replicate (Fusaki, Ban et al. 2009). In order to follow SeV disappearance during reprogramming, a vector expressing fluorescent Azami green was co-transduced with the reprogramming vectors and green florescence disappearance was followed once the hiPSC were established. The image in figure 19A shows a colony of one of the clones of PKD2iPS at passage 2, where the fibroblasts differentiated from the colony at the beginning of reprogramming retained Azami green expression, as they do not divide as much as the pluripotent cells inside the colony, which have lost the green fluorescence. In order to verify the complete disappearance of SeV transcripts, RT-PCR was performed to detect the presence of SeV mRNA in two hiPSC lines, PKD2iPS at passage 15 and PB2iPS at passage 30. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) RT-PCR was used as a positive control of each sample. As seen in figure R.19B, no SeV specific RT-PCR product was amplified in any of the analyzed lines whereas GAPDH was correctly amplified.

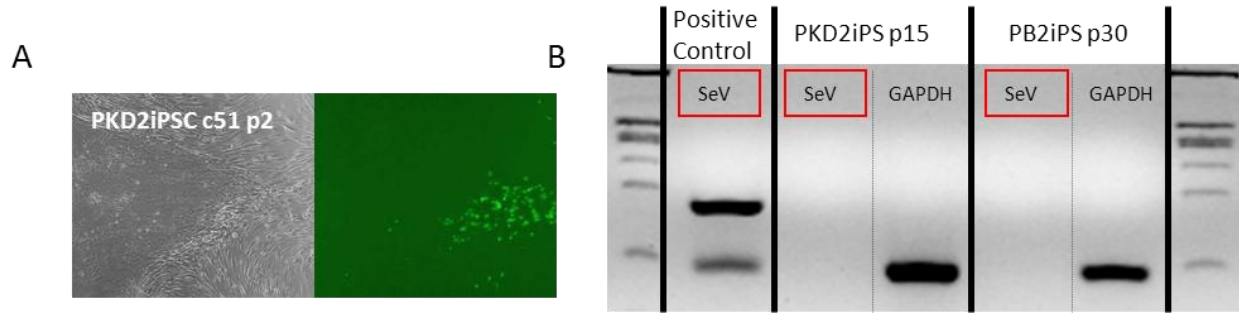


Figure 19: SeV disappearance. A) Image showing a colony from PKD2 at passage 2. While colony derived fibroblasts still harbor Azami green expressing viral vectors, rapidly dividing hiPSC within the colony have lost them. B) RT-PCR analysis showing that in hiPSC at passage 15 or 30 no SeV transcript can be detected.

1.2 Erythroid differentiation of PKDiPSC

We then established an *in vitro* differentiation protocol for the generation of erythrocytes, the affected cells in this disease.

1.2.1 RPK expression during in vitro erythroid differentiation

Once we developed and characterized the patient-specific cell lines, we wanted to know if these cells showed phenotypic characteristics of PKD. There were many reported publications that show the generation of Glycophorin A (also called CD235a) positive cells, a differentiation marker expressed during erythrocyte differentiation and in mature erythrocytes (Lu, Feng et al. 2008,

Hatzistavrou, Micallef et al. 2009, Lapillonne, Kobari et al. 2010, Dias, Gumenyuk et al. 2011) from hESCs and hIPSCs. But the different protocols are very heterogeneous in other parameters such as purity of erythroid cells (%), enucleation, expression of fetal, embryonic or adult globins and what is crucial to us, the level of expression of RPK versus PKM proteins. We assayed the protocols described by Robert Lanza's group (Lu, Feng et al. 2008) (Fig. 20A, method 1) and the one described by Elefanty's group (Hatzistavrou, Micallef et al. 2009) (Fig. 20A, method 2), both of them based on Embryoid Body (EB) formation and explained in figure 20. *RPK* expression was higher in method 1 than in method 2 and so was the relation *RPK/MPK* (1.35 for method 1 and 1.1 for method 2) (Fig. 20B). However, method 1 showed a big disadvantage, the cell number amplification. Starting with same amount of cells, we ended up having 10-fold more cells with method 2 than with method 1. As *RPK* was also expressed in method 2 and cell numbers were much higher, we selected method 2 for PKDiPSCs differentiation and therefore we performed a time course experiment of *RPK* expression, being the highest expression at day 20 of culture (Fig 20C).

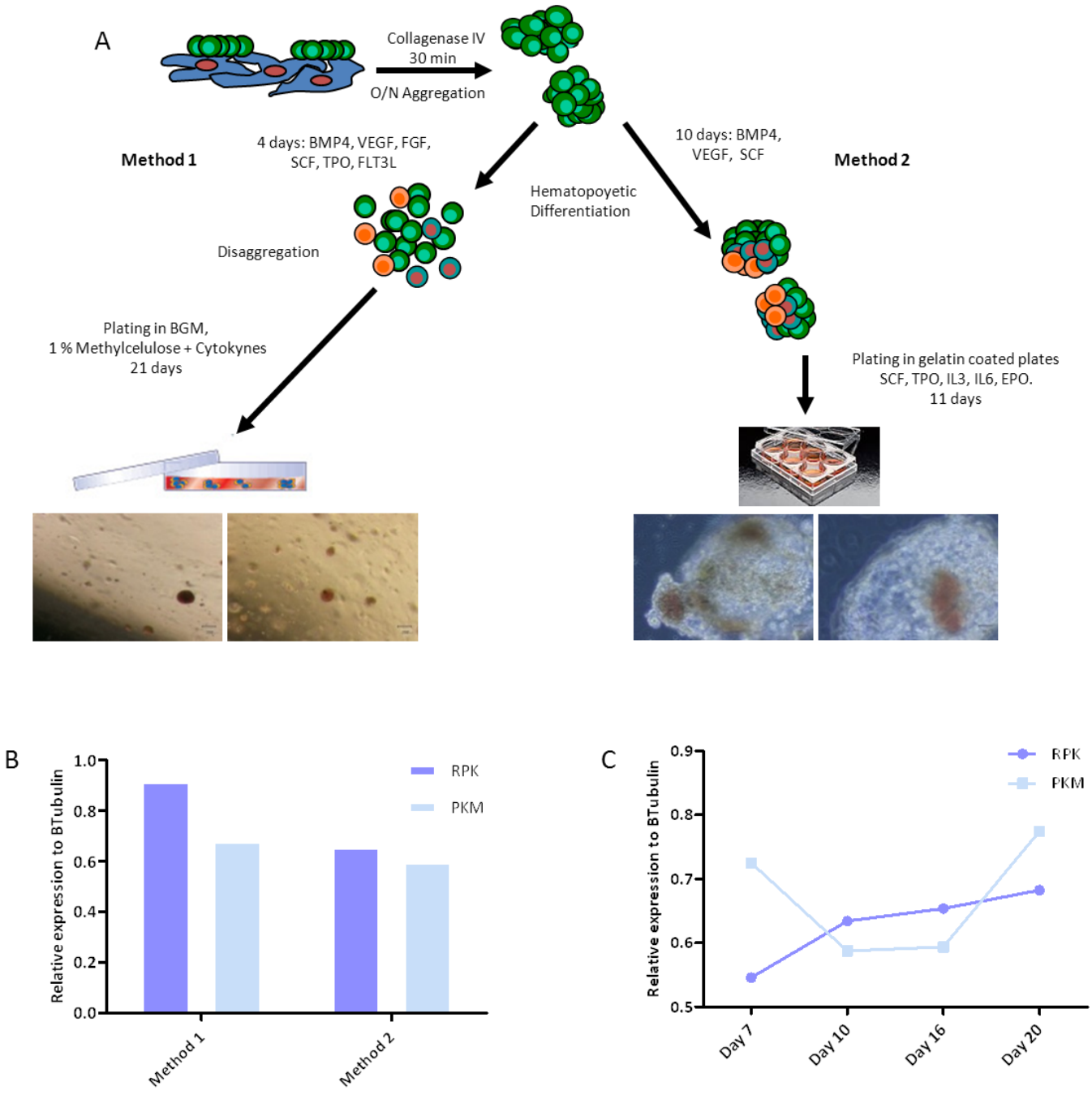


Figure 20: Erythroid differentiation of hiPSC: A) Scheme of two erythroid differentiation procedures. Images of the PB2iPSC differentiation process at day 20 are shown. Method 1: Erythroid colonies growing in methylcellulose culture. Method2: EBs are attached to the plate and red cells can be seen inside the three dimensional structure. B) Relative *RPK* and *PKM* expression comparison in both differentiation protocols at day 20. C) Relative *RPK* and *PKM* expression at different time points in method 2.

1.2.2 Pyruvate Kinase deficiency phenotype in PKDiPSC derived erythrocytes

When peripheral blood of PKD patients is analyzed for erythroid markers (CD235a and CD71), there is a clear phenotype associated with the disease, which is the accumulation of immature red blood cells, double positive for both markers, something not observed in healthy individuals, where almost

all the CD235a positive cells in peripheral blood are negative for CD71 (Fig. 21A). We asked if this phenotype could be reproduced *in vitro* after PKDiPSC differentiation. Therefore, we performed erythroid differentiation of 2 PKD-iPSC (PKD2 and PKD3iPSC) and the control hiPSC (PB2iPSC) and saw in an experiment that the phenotype was reproduced *in vitro,* being the proportion of CD71 negative cells within the CD235a positive population much higher in PB2iPSC (69%) than in PKD2iPS (17%) and PKD3iPS (15%) (Fig 21B).

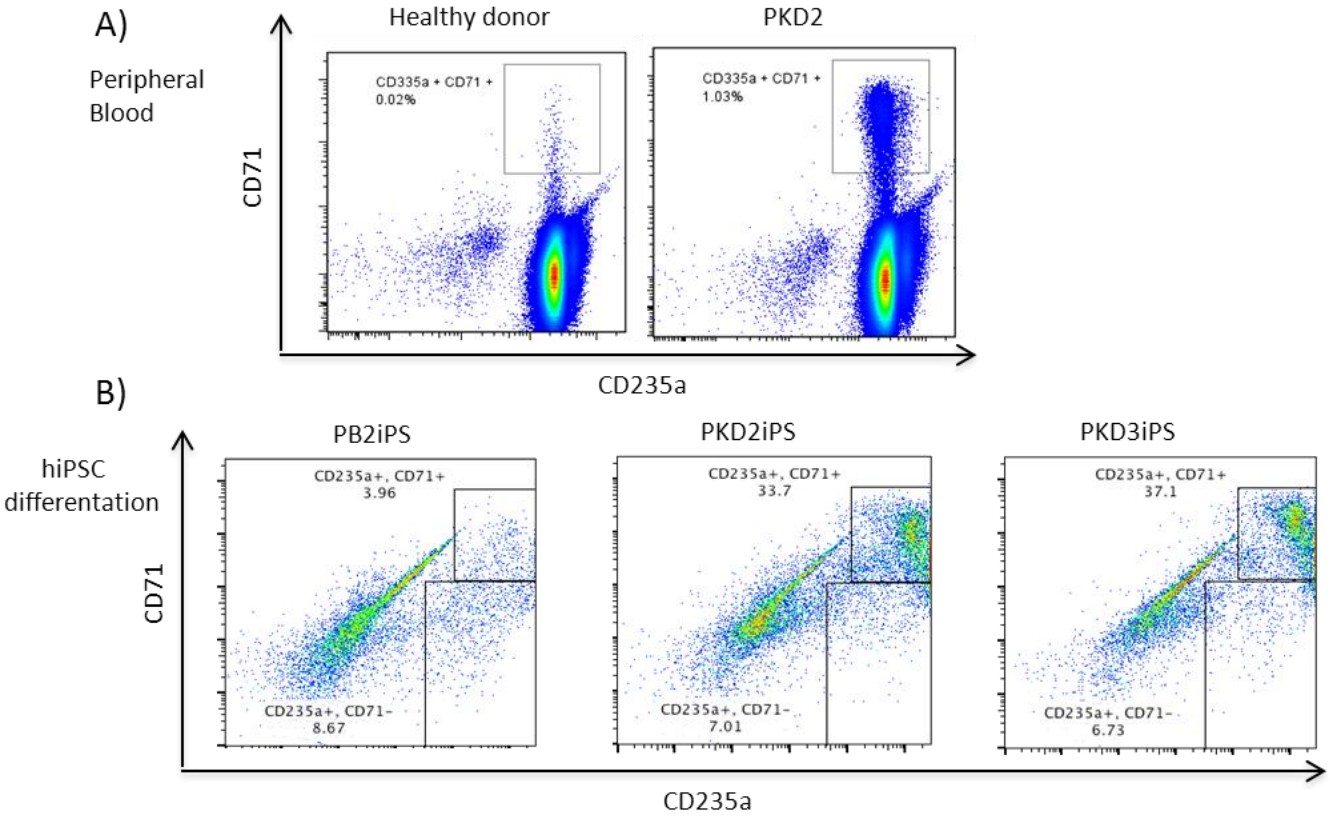


Figure 21: Erythroid phenotype in PKD samples. A) Peripheral blood analysis of PKD2 and a healthy individual showed the presence of a double positive ($CD235a^{+}CD71^{+}$) immature population in PKD, which was much less represented in the healthy sample. B) PKDiPSC differentiation (PKD2 and PKD3) showed a similar phenotype than the observed in patient peripheral blood. Whereas 70% of erythroid cells are CD71 negative in PB2iPSC, in PKDiPSC there is an accumulation of immature $CD235a^{+}/CD71^{+}$ cells.

2 Nuclease mediated *Knock In* gene correction in *PKLR* locus

2.1 Homologous recombination (HR) strategy design

In order to restore the mutated *PKLR* gene in the PKD derived hiPSCs and take advantage of the endogenous gene regulation, a Knock-In gene correction strategy was designed (Fig. 22). Thanks to an established collaboration, the company Cellectis, which is specialized in the design and development of site-specific nucleases, developed DNA nucleases (MG and TA) that specifically generated DSB in intron 2 of the *PKLR* gene. The sequences targeted by each DNA nuclease are described in table 6. Once DSB have been generated, the HR cell machinery uses the Homology Arms (HAs) of the donor matrix (DM) as a template for restoration of the break and introduces the transgenic elements (Fig. 22A and B) cloned between the HAs of the DM. This DM included:

1) Homology Arms, 2) the expression Cassette (EC): composed by the splicing acceptor (SA), the codon optimized cDNA from *RPK* exons 3-11, followed by a FLAG tag and the SV40 polyA at the end, and 3) the Selection Cassette (SC), composed by the mouse PGK promoter driving the expression of a puromycin resistance (PuroR) and Thymidine Kinase (TK) fusion protein flanked by LoxP sequences.

Once the cassette is correctly integrated in intron 2 of the *PKLR* locus (Fig. 22C), the SA will tie to the endogenous splicing donor of the endogenous exon 2 and a chimeric mRNA is generated, including endogenous exons 1 and 2 and exogenous codon optimized exons 3-11 and the FLAG tag at the end. The presence of the FLAG tag will help us to easily follow the expression of the chimeric protein. Additionally, the presence of a polyA sequence at the end of the expression cassette should avoid the transcription of endogenous exons 3 to 11. Moreover, once targeted, the PuroR allows the selection of the cells in which the selection cassette is integrated. Once the correct targeting is confirmed, the selection cassette can be removed by ectopic Cre recombinase expression that will excise the sequences between the LoxP sites, leaving only one copy of the LoxP sequences.

PKLR01 Meganuclase	TCGAGCCACTGTACTCCAGCCTAG
PKLR1 Talen	TGATCGAGCCACTGTACTCCAGCCTAGGTGACAGACGAGACCCTAGAGA

Table 6. PKLR specific DNA nucleases targeting sequences.

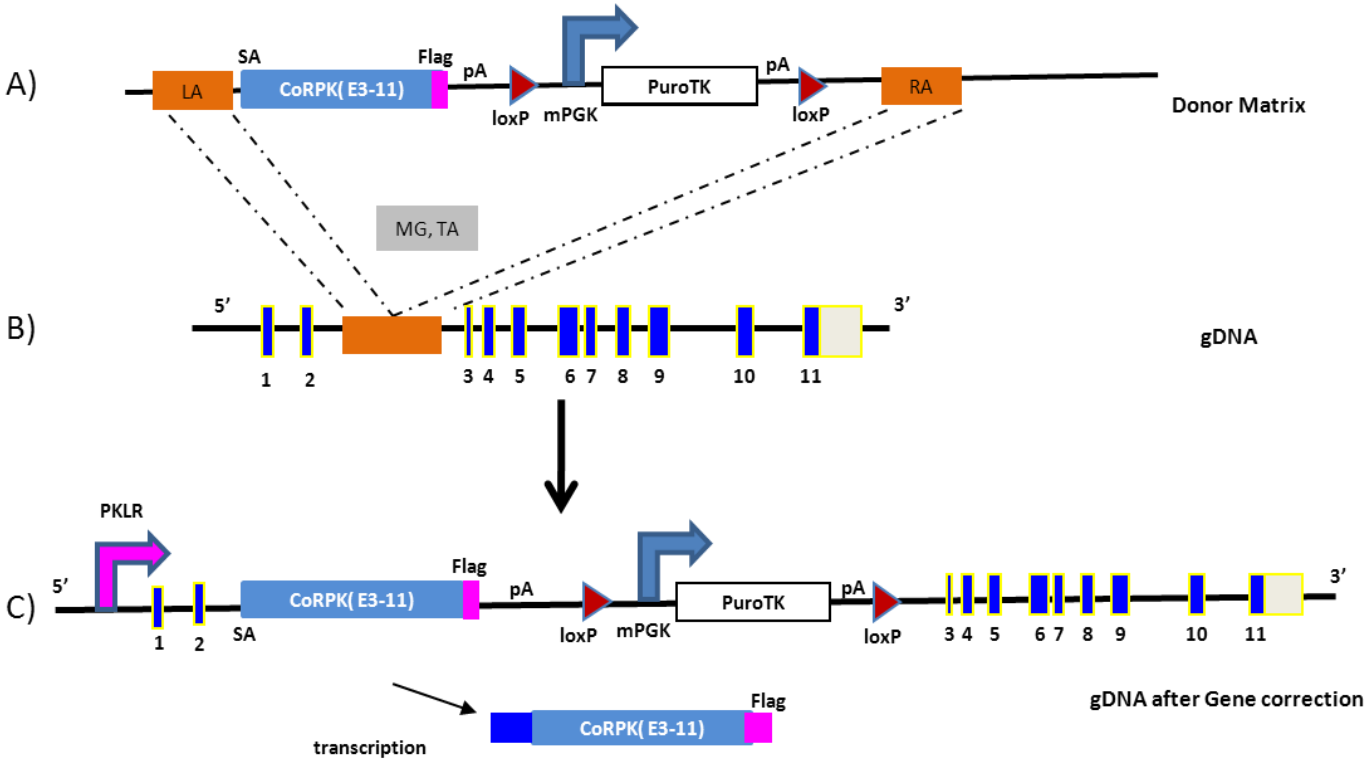


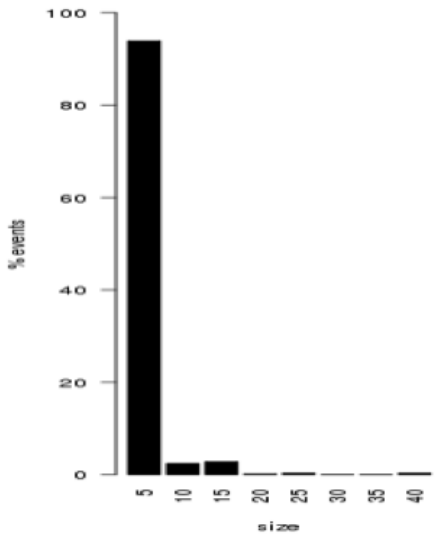
Figure 22: Knock In gene correction in *PKLR* locus. A) Scheme of the donor matrix used for the homologous recombination. A partial wild type cDNA (exons 3-11) of the *RPK* cDNA is going to be introduced in intron 2 of the *PKLR* locus by HR; the cDNA is flagged so that when the cassette is inserted a flagged transcript will be generated; floxed mPGK driven PuroR-Tk cDNA is also inserted for selection. B) Representation of the PKLR genomic locus indicating the place where the homologous recombination is going to take place, C) Diagram showing the PKLR genomic locus once the exogenous DNA has been integrated. At the bottom, scheme of the mRNA generated including endogenous exons 1 and 2 exogenous (blue), codon optimized exons 3-11 (light blue) and flag tag sequence (pink) at the end of the chimeric protein. Abbreviations: LA: Left Arm, RA: Right Arm, MG: Meganuclease, TA: TALE nuclease. Blue Boxes: Endogenous PKLR exons.

2.2 PKLR Meganuclease Efficiency

First, we started to work with the homing MG that specifically generated DSBs in intron 2 of *PKLR* gene, the PKLR01 MG.

2.2.1 PKL01 MG versions comparison in 293T cell line by deep sequencing

When a DSB is induced in the genome, the most common way of the cellular machinery to resolve the damage is by Non Homologous End Joining (NHEJ), which is highly error prone and usually generates insertions and/or deletions. We took advantage of this cellular function to test the efficacy of different versions of the developed MG. We compared two available versions of this MG that targeted intron 2 by transfecting them in HEK-293H cells and analyzing the percentage of insertions or deletions (Indels), consequence of the DSB generation, by deep sequencing of a 300 bp around the cutting site of the MG (Fig 23). Both versions generated NHEJ resolved DSB (more than 90% of the events were of 5 bp long but there were also longer events, like 40 pb long). Version 2 (V2) was slightly more efficient, showing 6.7% of indels whereas version 1 generated 6.0% and therefore it was the chosen version for the next experiments.



	Percentage of Indels
MG PKLR01 V1	5.966
MG PKLR01 V2	6.693

Figure 23: Comparison of two versions of PKLR01 MG for DSB generation in HEK-293H cell line by deep sequencing. After treatment with MG, the 300 bp surrounding PKLR MG recognition site were PCR amplified and analyzed by deep sequencing. Left graph shows the sizes of deletions generated as a consequence of the Double Strand Breaks (DSBs) by version 1 MG and the number of events of each size. The upper table shows the overall percentage of insertions and deletions generated by version 1 and 2 within the pool.

2.2.2 PKLR Meganuclease dose optimization

We optimized the dose of PKLR01 MG version 2 by nucleofecting different doses of MG in the form of mRNA into K-562 cell line, because the nucleofection of mRNA into more sensitive cells was less cytotoxic than the DNA nucleofection (Wiehe, Ponsaerts et al. 2007). DSB generation efficiency was measured by Surveyor assay (explained in figure 24A). In this case the indels and genomic modifications consequence of NHEJ are not detected by deep sequencing but by treatment by the Cel1 enzyme, which recognizes and cuts non-annealed sequences. In this concrete case the PCR product around the MG target site was 300 bp with the cutting site in the middle. As a consequence of Cel1 cutting, 2 products of around 150 bp were generated (Fig. 24B). As a control of the technique, we used the AAVS1 ZFN, which we already know that cuts efficiently in *AAVS1* locus (Hockemeyer, Soldner et al. 2009). In order to increase the assay sensitivity, the MG was co-transfected with mRNAs expressing TREX-2 protein, a non-processive 3′ exonuclease shown to degrade the 3′ DNA overhangs and therefore increasing NHEJ (Certo, Gwiazda et al. 2012). The following quantities of each mRNA were nucleofected: a) 3 µg of each AAVS1 ZNF subunit. b) 0.5 µg of PKLR01 MG V2 + 3 µg of TREX c) 3 µg PKLR01 MG V2 + 3 µg of TREX d) 6 µg PKLR01 MG V2 + 3 µg of TREX. A plasmid expressing GFP under the control of EF1α promoter was used as transduction efficiency control, showing 65% of the cells expressing GFP. DNA electrophoresis in figure 24B shows surveyor assay results. AAVS1 ZFN cutting efficiency proofed that the system was working as expected (Fig. 24B, lane 2). After comparing the different PKLR MG doses, the highest DSB

generation efficiency was obtained with 0.5 and 3 µg of MG (35% for both doses). Both doses were used in the next experiments.

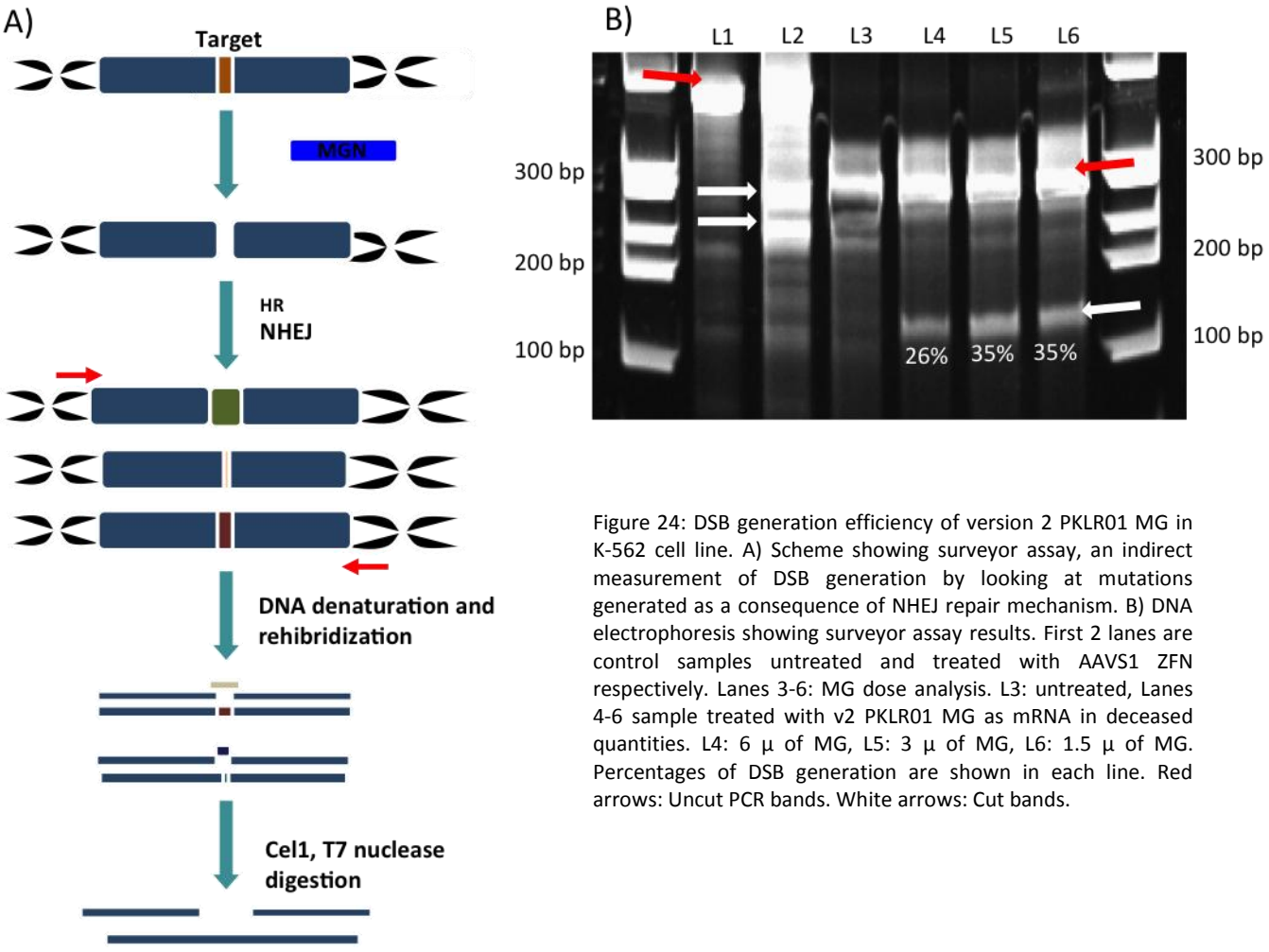


Figure 24: DSB generation efficiency of version 2 PKLR01 MG in K-562 cell line. A) Scheme showing surveyor assay, an indirect measurement of DSB generation by looking at mutations generated as a consequence of NHEJ repair mechanism. B) DNA electrophoresis showing surveyor assay results. First 2 lanes are control samples untreated and treated with AAVS1 ZFN respectively. Lanes 3-6: MG dose analysis. L3: untreated, Lanes 4-6 sample treated with v2 PKLR01 MG as mRNA in deceased quantities. L4: 6 µ of MG, L5: 3 µ of MG, L6: 1.5 µ of MG. Percentages of DSB generation are shown in each line. Red arrows: Uncut PCR bands. White arrows: Cut bands.

2.3 HR analysis in *PKLR* locus using K-562 cell line

In order to test the correct splicing in the targeted cells it is important to have an active locus, so that the transgenic FLAG tagged RPK protein can be expressed and detected. In this respect, K-562, an erythroleukemic cell line expressing RPK, was an appropriate cell line to test the developed HR tools. To induce HR in *PKLR* locus, K-562 was nucleofected with 0.5 and 3 µg of mRNA of PKLR01 v2 MG in combination with 4 µg of the donor matrix (DM). Puromycin selection at 2 µg/ml was carried out for 20 days generating a pool of resistant K-562 cells to set up analysis techniques such as specific PCR to check for HR and flow cytometry of the FLAG tag for studying the expression of the chimeric protein using the exogenous splice acceptor introduced by HR.

2.3.1 PCR assessment of Homologous Recombination

A pair of primers that recognized the *PKLR* locus outside the HA and inside the integrated targeting matrix was used (Fig. 25A). The expected 1176 bp band appeared just in the conditions in which MG was co-nucleofected with the DM, demonstrating the correct integration of the exogenous sequences in the desired place, the intron 2 of the *PKLR* gene (Fig. 25B).

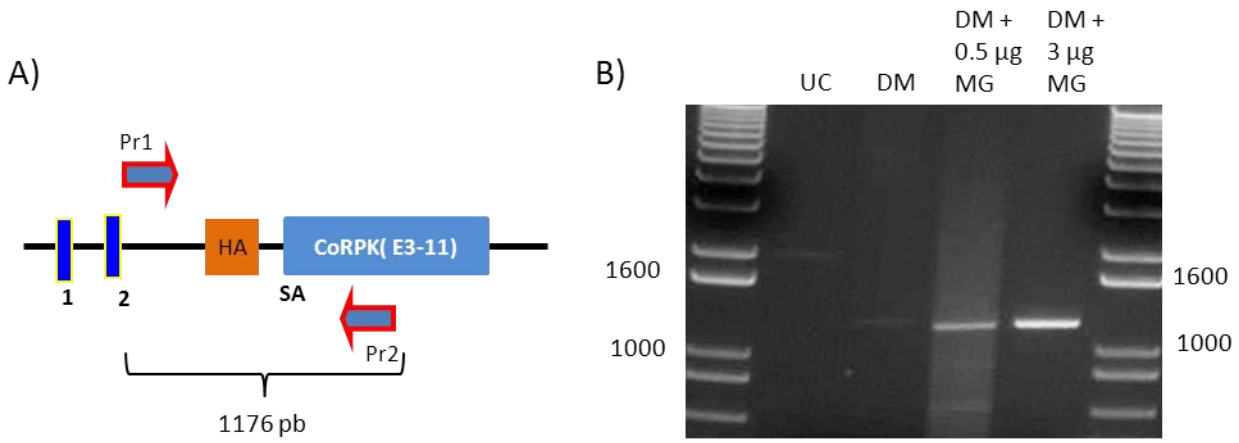


Figure 25: HR analysis in K-562. A) Diagram showing HR-specific PCR strategy. Pr1: Forward primer, Pr2: Reverse primer, HA: Homology Arm, SA: splicing acceptor. B) DNA electrophoresis showing HR-PCR amplification of an 1176 band. Lane 1: Uncorrected (UC) K-562, Lane 2: Donor Matrix (DM), Lane 3: DM + 0.5 µg of mRNA of v2 PKLR01 MG and Lane 4: DM + 3 µg of mRNA of v2 PKLR01 MG.

2.3.2 Correct splicing assessment

Once we confirmed the targeting in the *PKLR* locus, we evaluated if the splicing between the endogenous donor and the exogenous acceptor was taking place properly by analyzing the expression of the FLAG tag located at the end of the chimeric RPK protein in the targeted cell pools. By looking at the basal expression difference between the cell pools, the only condition in which we had FLAG tag expressing cells was DM + 3 µg of PKLR01 v2 MG (5,75%, Fig. 26, upper histograms). K-562, being an erythromyeloblastoid leukemia cell line, is a heterogeneous population, where a reduced subset of the cells express *PKLR* gene. In order to increase the expression of *RPK*, we induced its differentiation by Hemin treatment, which increased the percentage of cells expressing the FLAG tag, being even higher when the dose of 0.5 µg of MG was used, supporting the targeting at *PKLR* locus and the correct splicing after HR (Fig. 26, lower histograms).

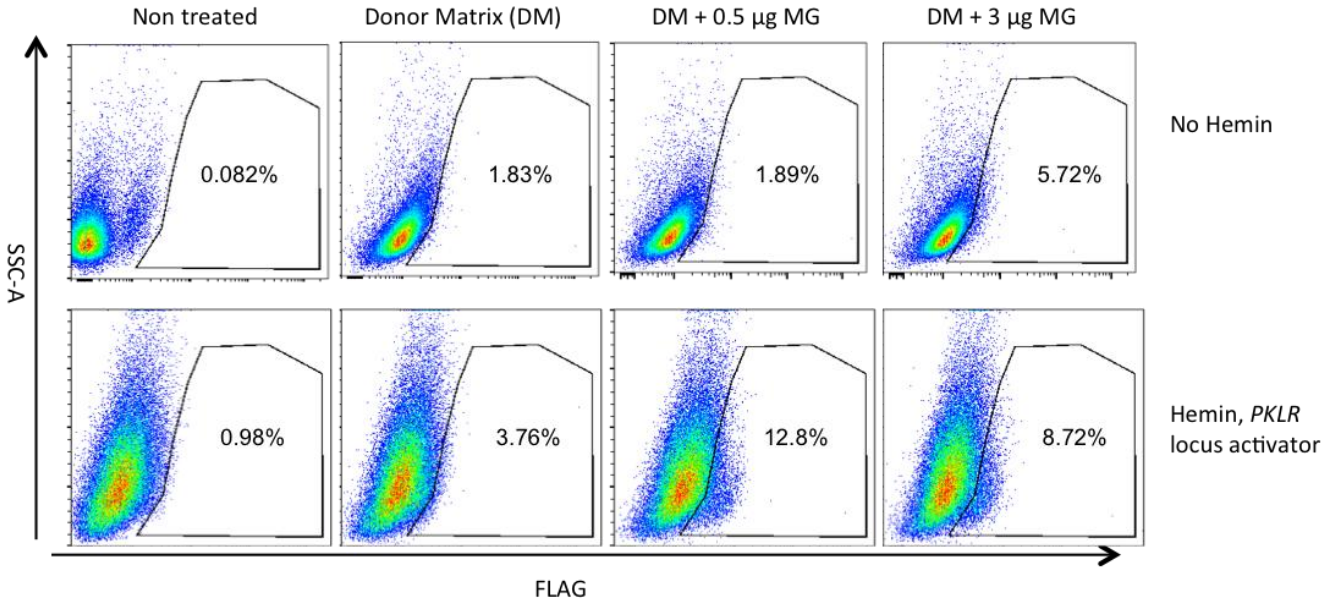


Figure 26: Flow cytometry analysis of FLAG expression in K-562 cell line after targeted integration and puromycin selection. Cells were analyzed after culture in the absence (upper row) or presence (lower row) of hemin in order to increase RPK expressing cells.

2.4 HR in PKD2iPSC facilitated by PKLR MG

So as to genetically correct the defect in the patient specific cell line PKD2iPSC, cells were nucleofected with 4 µg of the DM alone, or in combination with 0.5 µg of PKLR01 v2 MG mRNA. We choose 0.5 because cutting efficiency was equally high as 3 µg, the HR-PCR was positive and the percentage of FLAG expressing cells was higher after differentiation with hemin (Fig. 26). As control, a targeting matrix for AAVS1 locus that constitutively expresses GFP was used in combination with mRNA of the AAVS1 ZFNs. In order to select those cells in which the matrix was integrated, puromycin selection was performed. The AAVS1 control gave us information about nucleofection efficiency, which was 29% (mean value, n=3) after 48 hours. A representative part of each reaction was seeded for alkaline phosphatase staining after one week of puromycin selection (Fig 27). In the condition in which the DM and MG was used, 21 colonies were scored whereas in the absence of the MG there were only 3 surviving colonies. Targeting efficiency could not be defined, as the number of colonies in the absence of selection was uncountable.

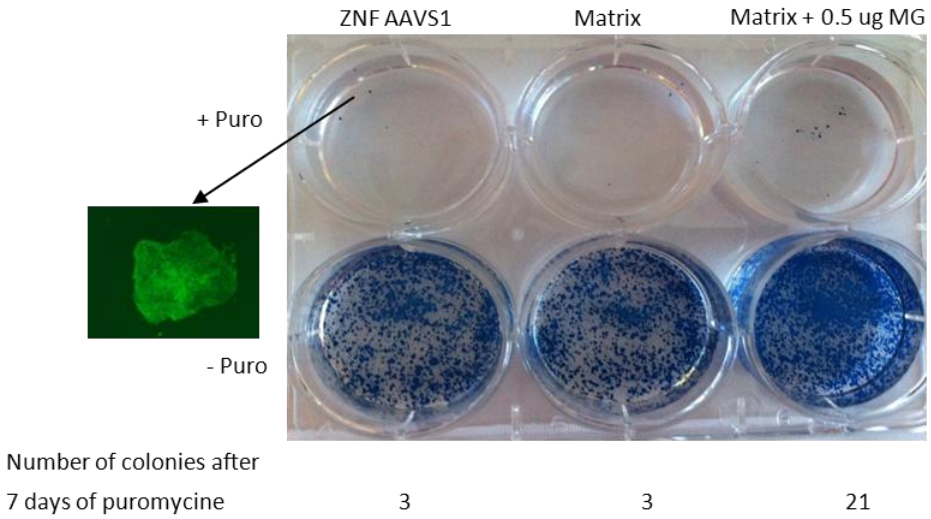


Figure 27: HR in PKD2iPSC mediated by MG. Alkaline phosphatase staining after 7 days of puromycin selection. Puromycin was added to the upper wells and not to the lower ones. A representative image of green florescence in one of the resistant colonies is shown.

Twenty clones were manually picked and expanded for further analysis. As explained for K-562, a PCR in which the 1176 bp band would appear just in the case of HR was performed (Fig 28A). In the case of PKD2iPSC targeting, no unique bands appeared in any of the selected clones and the results were not reproducible between different PCR analyses and by using other alternative primer pairs. To further analyze if HR took place in the selected place, we selected 3 clones in which the PCRs were most reproducible and performed Southern Blot analysis (Fig 28 B, C). With this technique, single integrations can be distinguished from multiple integrations, as the radioactive probe hybridizes a sequence in the DM (Fig. 28B). By cutting the gDNA with the restriction enzymes SpeI and ScaI in separate reactions and assuming that the DM had been integrated in the right place, a band of 4410 bp (SpeI) and of 14273 (ScaI) should appear. Two of the clones (c9 and c12) showed multiple bands different than the expected ones, meaning that the donor matrix was integrated in more than one place of the genome. In the remaining clone, clone number 5, there might be integration in *PKLR* but also in other loci. (Fig. 28C). The explanation for the multiple targeting is that PKLR01 v2 MG was generating DSBs in other sequences of the genome in addition to intron 2 of *PKLR* gene, facilitating the HR in these non-desired places as well. When a blast search using the MG targeting site against human genome was performed, 26 additional targeted sequences were found.

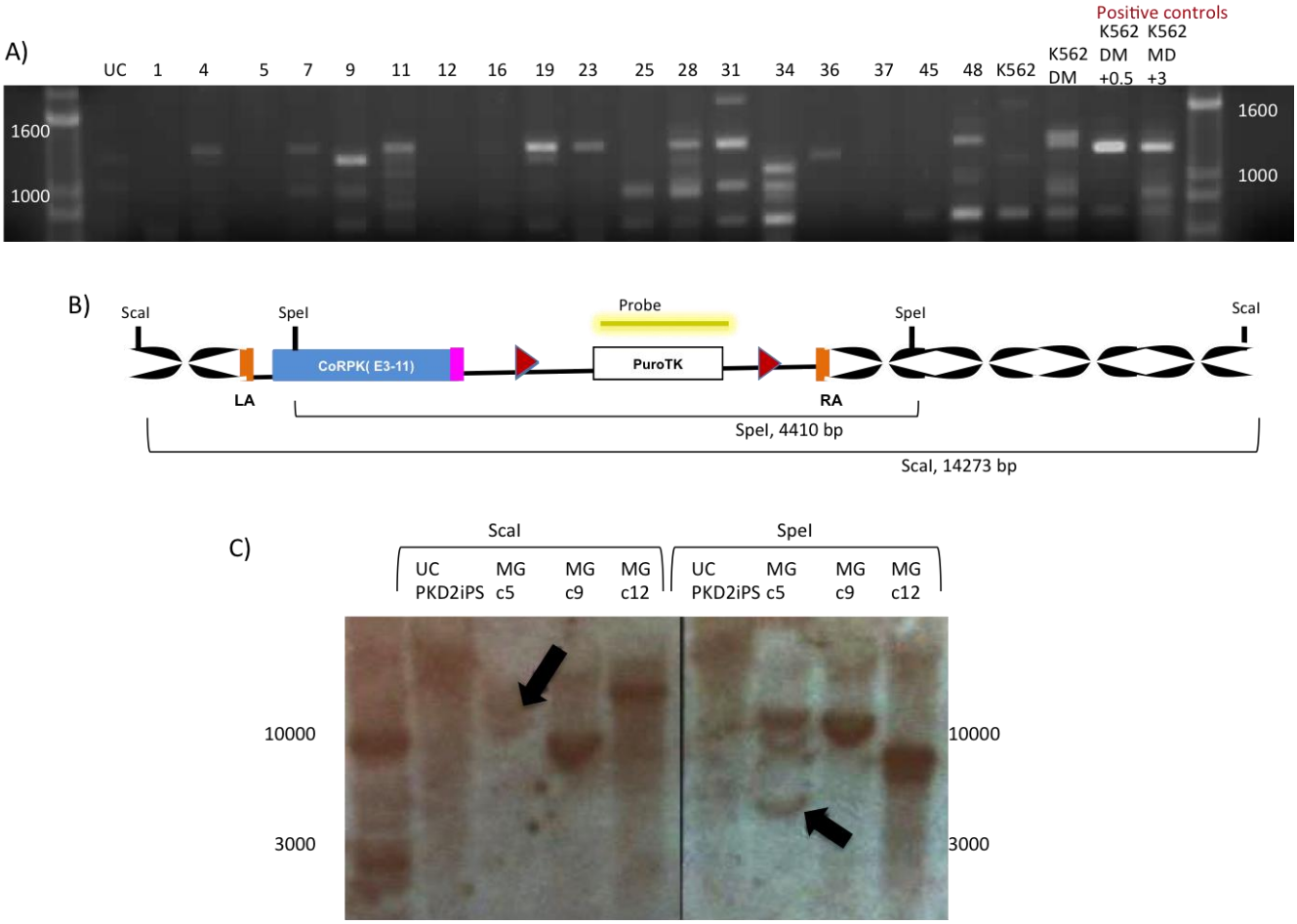


Figure 28: MG assisted HR analysis in PKD2iPSC. A) DNA electrophoresis showing HR specific PCR in 18 puromycin resistant PKD2iPS clones and in puro resistant K-562 gDNAs UC: Uncorrected cells. B) Southern blot strategy: A radioactive labeled probe that binds the PuroR-TK was used and the following unique band sizes should be seen if the targeting had happened uniquely in *PKLR* locus: SpeI: 4410 bp, ScaI: 14273 bp. C) Southern blot film corresponding to digested gDNA of 3 puro resistant PKD2iPS clones. As seen in analyzed clones, multiple bands indicate integration in additional loci. Arrows mark expected size bands in clone 5.

2.5 HR by Transcription activator-like effector nucleases (TALEN™)

In order to increase the specificity of the PKLR nuclease and reduce the frequency of off targets, our collaborators designed a TALEN™ (TA) that specifically recognized a sequence that included the one that targeted the MG and that we could combine with the DM designed for the previous explained approaches. By using PKLR1 TA, same DSB generation efficiency analysis and HR was performed in PKDiPSCs.

2.5.1 DSB generation efficiency

First of all we tested the DSB generation efficiency in PKD2iPSC and PKD3iPSC by nucleofecting 2 doses of DNA of PKLR1 specific TA (3 or 6 ug of each subunit), and performed the previously explained surveyor assay. We choose DNA for electroporation and the mentioned amounts because were the conditions suggested by our collaborators and reported in literature (Hockemeyer, Wang et al. 2011). As a consequence of the DSB correction by NHEJ, the two 150 bp bands would appear

(Fig 24). It was possible to detect the bands (indicated with a white arrow) in both PKDiPSC lines but it was below the quantification range as we had to overexpose the gel with Ultraviolet light and the uncut band became saturated. Then, we could not quantify the different bands properly to get an efficiency value.

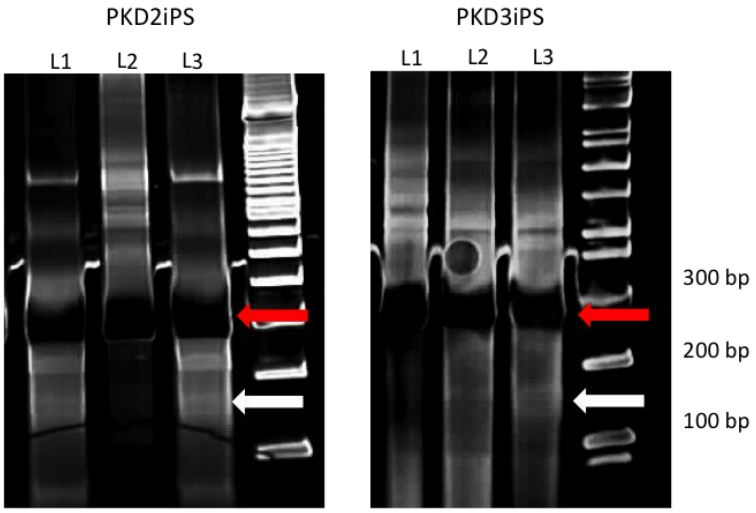


Figure 29: PKLR01 TALEN™ (TA) DSB generation efficiency in PKD2iPSC and PKD3iPSC. PKD2iPSC: L1: treated with 3 ug of each TA subunit L2: untreated, L3: treated with 6 ug of each TA subunit. PKD3iPSC: L1: untreated L2: treated with 3 ug of each TA subunit, L3: treated with 6 ug of each TA subunit. Normal PCR product 300 pb (red arrow). After cutting, 2 bands of approx. 150 bp,(white arrow). The 230 and 71 bp bands that appeared during PKD2iPSC surveyor assay will be explained later in 2.5.2.3.

2.5.2 HR mediated gene correction in PKDiPSC from two different patients

As performed when the PKLR MG was used, TA induced HR was performed in PKD2iPSC and PKD3iPSC by nucleofecting the two TA subunits in combination with 4 ug of the PKLR DM. When DSB efficiency was measured, no big difference was detected between nucleofecting 3 or 6 ug of each subunit and therefore both doses were applied for HR. In PKD2iPSC targeting, after one week of puromycin selection less resistant colonies appeared than when the MG was used. Thirteen colonies from each reaction with different doses of TA were picked (26 in total). From them, 14 were expanded and colony PCR was performed. Ten out of fourteen gave the expected PCR product corresponding to *PKLR* Knock-In HR (Fig. 30A). In this particular case the PCR was performed by lysing a single colony and to be sure that negative clones were not due to gDNA absence, *CFTR* gene was amplified as control (Fig. 30A, lower bands). In all of them, excepting clone 17 the expected band was amplified, assuring the gDNA quality. As we had 10 positive clones from 13 clones in which the DNA quality was enough, we estimated an HR efficiency of 77% over the puromycin resistant clones. A similar protocol was performed in PKD3iPSC, resulting in similar targeting efficiency, 76% of puromycin resistant clones were correctly targeted, also assessed by HR-PCR. A description of the targeted clones generated during PKD2iPSC and PKD3iPSC Knock-In gene correction is showed in

table 7. In order to analyze if there were additional copies of the DM integrated in other places of the genome, five of the corrected PKD2iPSC clones were expanded and Southern Blot analysis was performed as previously explained (Fig. 30B). After gDNA isolation, digestion, and probe hybridization in these 5 clones, all of them showed the single and expected band. The picture of the whole blot is shown in order to verify that no other bands could be detected.

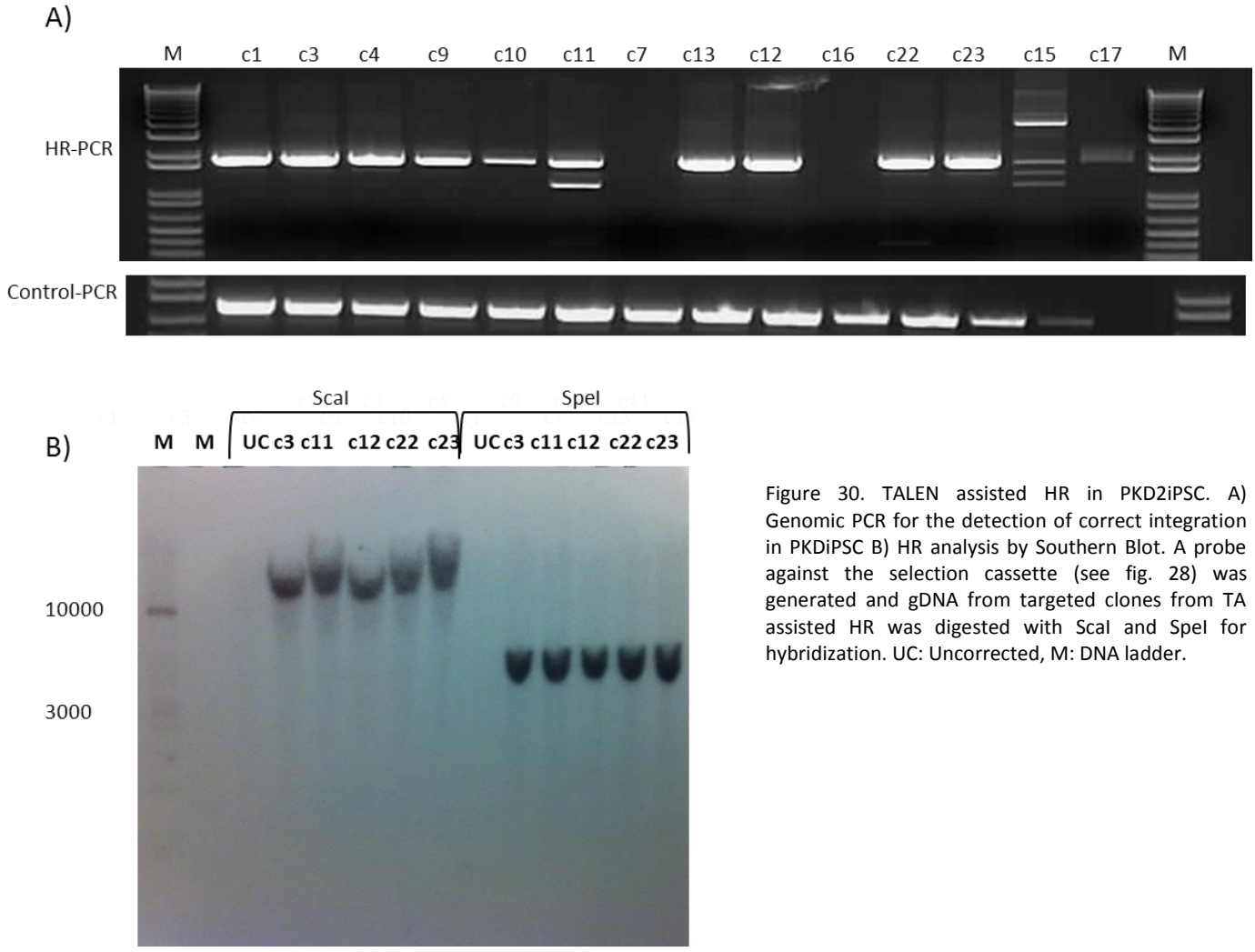


Figure 30. TALEN assisted HR in PKD2iPSC. A) Genomic PCR for the detection of correct integration in PKDiPSC B) HR analysis by Southern Blot. A probe against the selection cassette (see fig. 28) was generated and gDNA from targeted clones from TA assisted HR was digested with ScaI and SpeI for hybridization. UC: Uncorrected, M: DNA ladder.

2.5.2.1 Non Homologous End Joining (NHEJ) in untargeted *PKLR* locus

A PCR surrounding the PKLR1 TA cutting site was performed in order to analyze heterozygous or homozygous targeting and integrity of the untargeted *PKLR* locus. For this purpose the same primers as for the surveyor assay were used. First, for heterozygous vs. homozygous targeting, this PCR would not generate any product if both alleles were targeted, as it is a short PCR (301 pb) between the end of the left and the beginning of the right Homology Arm (HA). The PCR product was generated in all the targeted clones, so no homozygous targeting clones were present in PKD2iPSC HR (Fig. 31A). In PKD3iPSC, in contrast, three clones seemed to be homozygous (11.5%) (Table 7).

The homozygosity of targeted PKD3iPSC clones needs to be verified by Southern blot. Second, we performed Sanger sequencing of the amplified product in order to check the genomic integrity of the untargeted *PKLR* locus. From the picture of the DNA electrophoresis, it was possible to notice that the sizes of the PCR products varied among the clones, suggesting the presence of indels as a consequence of NHEJ resolved DSBs. This was confirmed after sequencing, which revealed that in 40% of the assayed clones from PKD2iPSC and 31% from PKD3iPSC, NHEJ related modifications were found in the untargeted alleles. In targeted PKD2iPSC, clone 12 showed a 14 nucleotides (nts) insertion and a 2 single base deletions, clone 22 had a 12 nts deletion, clone 23 showed 32 nts insertion and clone 10 had two insertions, one of a single nt and another of 20 nts, and 2 deletions: one of a single nt and other of 3 nts (Fig. 31B).

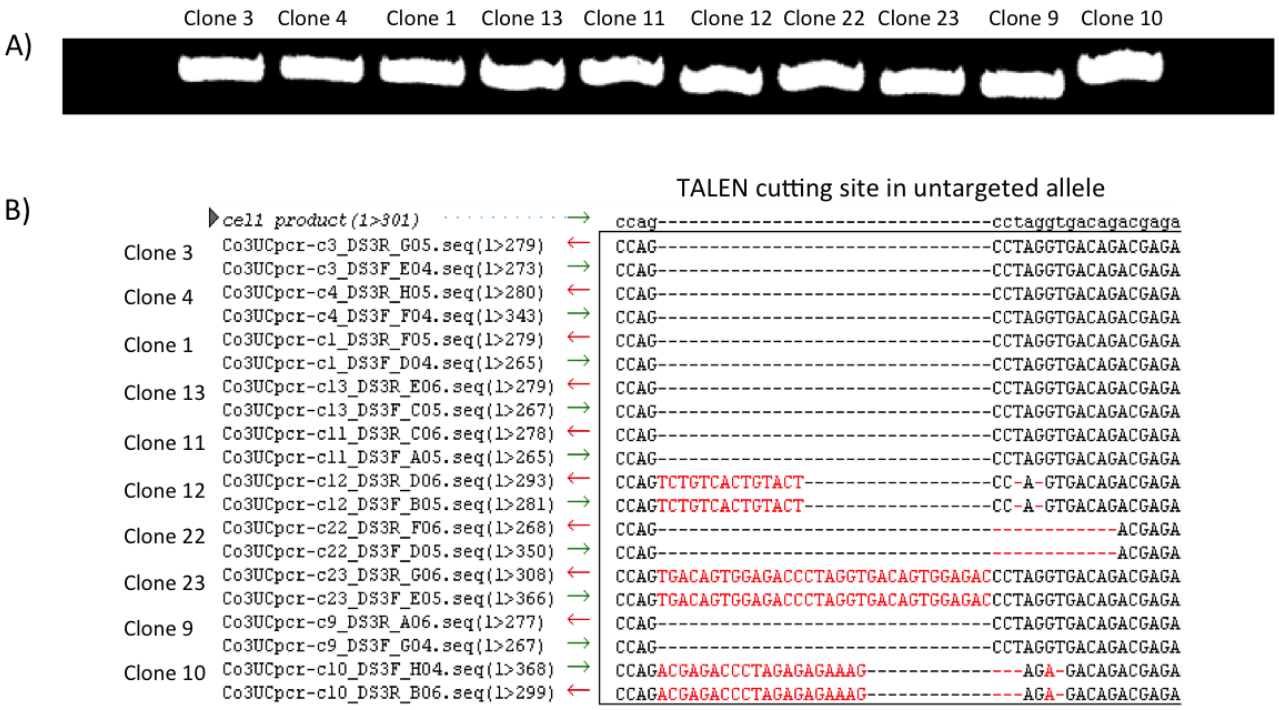


Figure 31. A) PCR surrounding the PKLR1 TA cutting site confirmed no homozygous targeting during PKD2iPSC HR. B) PCR product sequencing for analysis of NHEJ in *PKLR* untargeted allele. Overall, 40% of the clones showed genomic modifications (insertions and/or deletions). Sequences in red are insertions and red dashes represent deletions. Green and red flags represent same sequencing with forward and reverse primers respectively.

	Nr. Analyzed PuroR clones	Correctly targeted	Homozygously targeted	NHEJ in untargeted allele
PKD2iPSC	13	77%	0%	40%
PKD3iPSC	40	76%	11%	31%

Table 7: Targeted clones from PKD2 and PK3iPSC HR.

2.5.2.2 Off target analysis

One of the main concerns when using DNA nucleases to increase HR efficiency in a particular site of the genome is the off target recognition of other sequences of the genome and consequent cutting. This has been described in genome regions in which the nuclease finds a similar sequence with some mismatches and generally generates NHEJ mediated genomic modifications. In order to check this possibility in the genome of several *PKLR* targeted clones from PKD2iPSC and PKD3iPSC, sequence integrity was verified by amplifying the region surrounding PKLR1 TA targeting site with 6 mismatches or less and Sanger sequencing.

The sequence that each subunit of PKLR1 TA recognizes in the human genome is described in the first row of table 8. As this particular TA is not an obligated heterodimer, a homodimer with a recognition sequence composed of a duplicate of one of the sites can be generated (left-spacer-left or right-spacer-right). A blast against the human genome using these sequences and its complementary reverse as queries was performed and no possible off targets with less than 5 mismatches were found. Three sites with 6 mismatches and 2 with 5 were the final list of sequences that were found (table 8) and interrogated in the following targeted clones: corrected PKD2iPSC clone 11 and 3 and corrected PKD3iPSC clones 80 and 31. Figure 32 shows a representative alignment of one of the possible off target sequences, off target 1, which genomic region was amplified by PCR and the sequence integrity was analyzed by Sanger sequencing. None of the analyzed clones showed unexpected genomic modifications in this site and neither in off targets 2, 4 and 5. The sequence surrounding off target 3 could not be amplified by PCR due to the complexity of the sequence. Nevertheless, as the first base in 5` is an A and not a T as it is recommended for a TALE recognition site, it should not be recognized (Boch, Scholze et al. 2009).

	Location	Left_Sequence	Spacer	Right_Sequence	Type	Total Missmatches
PKLR01 Target	Chr1: 156664442-156664490	TGATCGAGCCACTGTAC	15	AGACGAGACCCTAGAGA	LR	0
Talen PKLR1 off target 1	Chr10: 54947560-54947622	TGATTGTGCCACTGCA	29	TGCAGTGGCACAATCA	LL	6
Talen PKLR1 off target 2	Chr15: 24745773-24745821	TGATCGTGCCACTGCA	15	TGCAGTGGCACGATCT	LL	5
Talen PKLR1 off target 3	Chr19: 10933216-10933277	AGATCGTGCCACTGCA	28	TGCAGTGGCACGATCT	LL	6
Talen PKLR1 off target 4	Chr 4:91614547-91614608	TGATGGAGCAATTGTA	28	GAGGGGACCCTAGGGA	LR	6
Talen PKLR1 off target 5	Chr 6: 26393830-26393875	GGATTGTGCCACTGTA	12	TACAGTGGCACGATCT	LL	5

Table 8: TA PKLR01 target and probable off target sequences found in human genome.

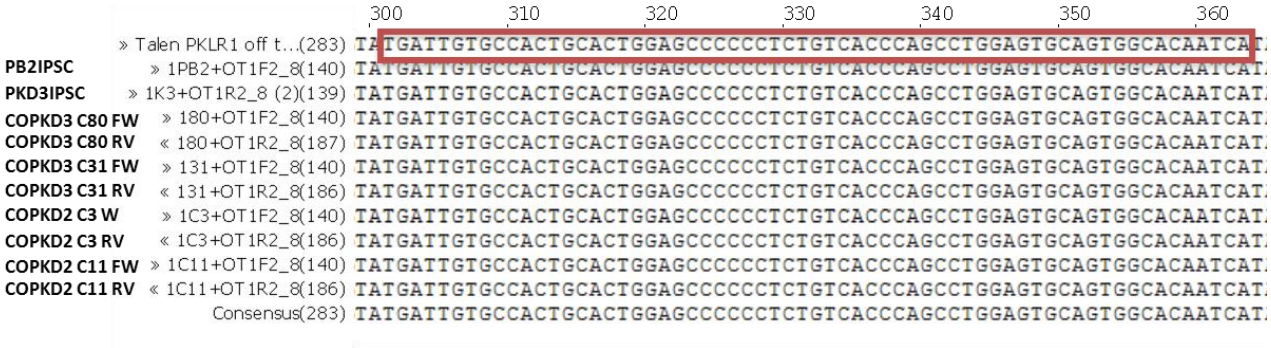


Figure 32: Representative alignment of sequencing results of 4 PKLR targeted clones from the PKLR TA off target 1 surrounding sequence. None of the targeted clones showed NHEJ derived genomic modifications. Red square indicates the genomic sequence corresponding to off target 1.

2.5.2.3 Allele specific targeting

Two unexpected bands from approximately 230 and 71 bp were identified in the surveyor assay of original PKD2iPSC (Fig 29). We speculated that these additional bands might have appeared due to the presence of a Single Nucleotide Polymorphism (SNP) in one of the alleles of PKD2 sample. The consequence might have been that after the annealing/reannealing process, the PCR form one of the alleles would hybridize with the PCR from the other, giving rise to a 1 base mismatch susceptible to be cut by Cel1 enzyme. By Sanger sequencing we analyzed the surveyor assay PCR product obtained from PKD2iPSC gDNA and found that 50% of the PCR products showed a G (expected) whereas 50% showed an A (unexpected) in the position 190 of the PCR product (Fig R. 33A).

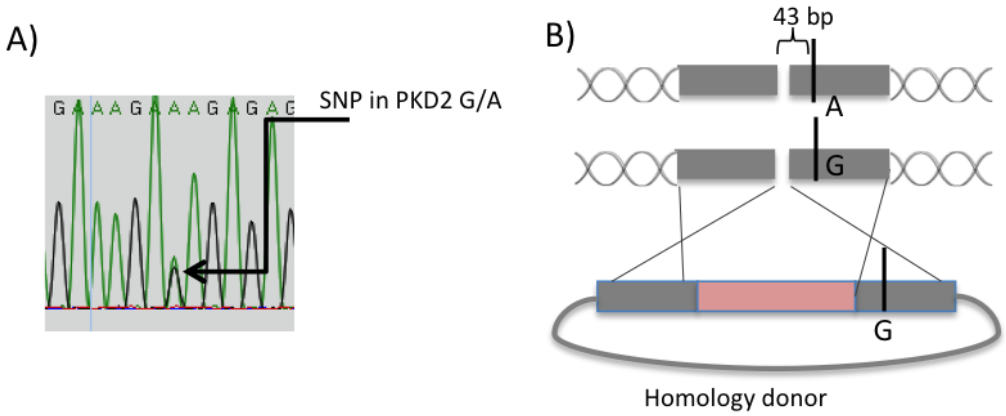


Figure 33. A) Chromatogram of sequencing of Single Nucleotide Polymorphism in PKD2. One allele shows a G whereas the other is an A. B) Possible HA harboring SNP effect in HR process.

This SNP is 43 base pairs away from the expected PKLR1 TA cutting site and therefore is part of the right HA of the DM (Fig. 33B). The sequence of our DM contained a G (Fig 33B) and we studied whether the difference in a single nucleotide in an allele nearby the DSB can promote allele specificity. As all the targeted clones from corrected PKD2iPSC (coPKD2iPSC) were heterozygous, we amplified the untargeted allele around the SNP by PCR and sequenced it to know which allele had undergone HR. Ten out of ten untargeted amplified products carried an A in the mentioned position,

demonstrating that all the clones were targeted in the allele that had the G as it is shown in figure 34 and confirming that the presence of a SNP can lead to allele specific HR. On the contrary, this SNP was no present in PKD3 and therefore no allele specific targeting in any coPKD3iPSC clones was observed.

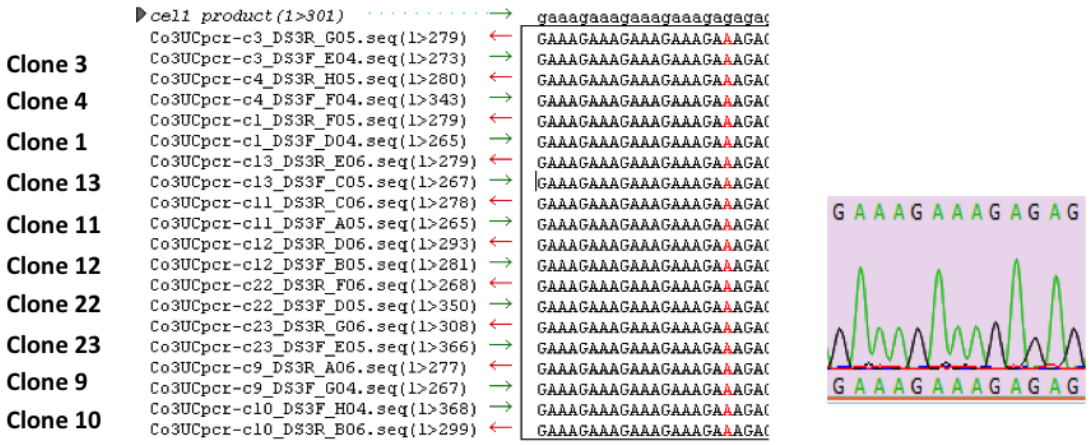


Figure 34. Sequencing around SNP in untargeted alleles of corrected PKD2iPSC, showed that all the clones had an A. So all the clones were targeted in the allele that had the G, showing allele specific targeting

2.5.2.4 Pluripotency maintenance after HR

In order to verify the maintenance of pluripotency in coPKD2iPSC, gene expression of 84 key genes was measured by hESC RT2 Profiler PCR Array, as previously explained. As seen in figure 35A, the gene expression profile after gene correction correlated to the one shown by the reference hESC line H9. In addition, demethylation of *SOX2* and *NANOG* promoters in clone 11 of coPKD2iPSC was also verified (figure 35B). Additionally, clone 31 of coPKD3iPSC was able to generate a teratome in NSG mice (data not shown). Altogether, these results demonstrate that gene correction process did not affect pluripotency status.

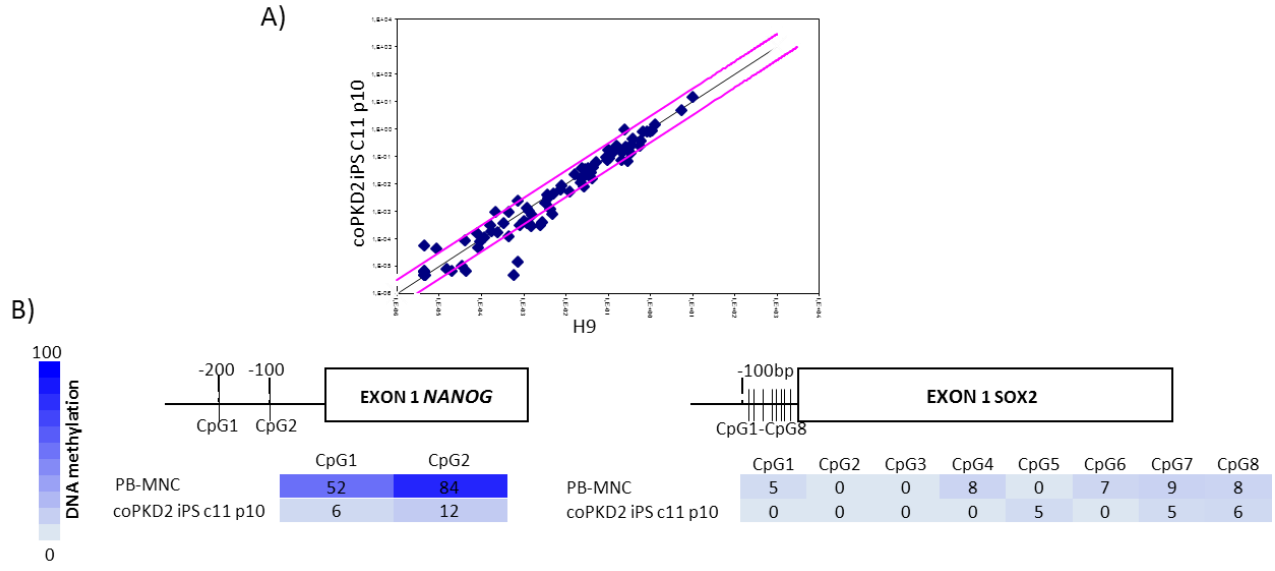


Figure 35: Pluripotency maintenance after HR. A) Pluripotency associated gene expression profile of corrected PKD2iPSC clone 11 correlates with the one observed in the hESC line, H9. B) *NANOG* and *SOX2* promoters CpGs dinucleotide demethylation.

2.5.2.6 Flagged RPK expression in differentiated coPKD2iPSC

The presence of the flagged RPK as a result of a correct targeting was analyzed in coPKD2iPSC by western blot. Wild type PB2iPSC, PKD2iPS and coPKD2iPSC clone 11 were differentiated into the erythroid lineage (Fig 37A) in order to activate *PKLR* promoter and at day 20 protein lysates were generated. Two immunoblots were performed; one with anti-RPK polyclonal antibody (Fig 37B, upper blot) and the other one with Anti-FLAG TAG monoclonal Antibody (Fig 37B, lower blot). β-tubulin was used as loading control. The percentage of erythropoietic cells (CD235a positive) was variable among the different assayed cell lines (Fig 37A) and therefore we loaded protein quantities equivalent to the erythroid differentiation efficiency to have similar RPK levels in the three cell lines. As seen in the upper image of figure 37B, RPK is expressed in the three cell lines and in the two reported protein isoforms: the L isoform (upper band) and the proteolyzed Lc isoform (lower band) (see introduction 1.1 for details). The anti-FLAG antibody bound the two RPK isoforms in the coPKD2iPSC lysate, giving exactly the same protein sizes pattern than when using the anti-RPK antibody (blue arrows in Fig 37B), confirming correct targeting by Knock-In and the expected splicing and protein processing of chimeric RPK.

A)

	CD235a$^+$ cells
PB2iPSC	13%
PKD2iPSC	41%
coPKD2iPSCc11	5%

B)

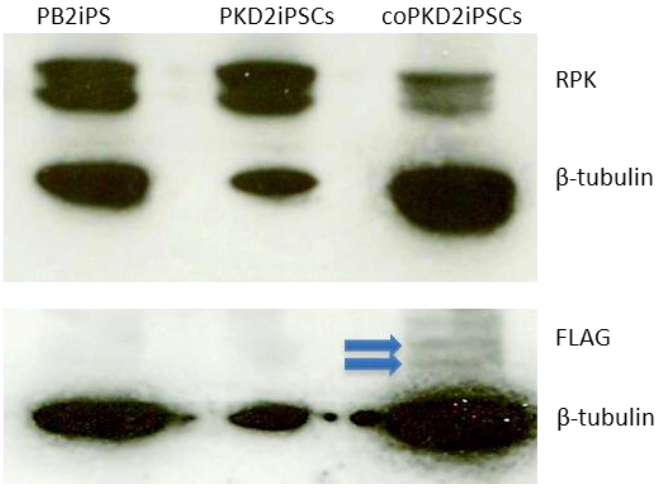


Figure 37: Healthy donor, PKD2 and corrected PKD2 iPSC erythroid differentiation and flagged RPK analysis by western blot. A) Percentage of CD235a$^+$ cells present in the *in vitro* cultures determined by flow cytometry. B) Western blot analysis of protein extracts from the in vitro differentiated cells. See text for details about protein quantities loading. Blue arrows mark the bands corresponding to flagged RPK detection in coPKD2iPSC.

2.5.2.5 Erythroid differentiation in corrected PKD2iPSC

In order to verify the phenotype consequence of gene correction, coPKD2iPSC clone 11 was differentiated into the erythroid lineage and at day 21 CD235a and CD71 markers were analyzed to confirm the erythrocyte maturation status (Fig. 36). As seen in hiPSC from a healthy donor (PB2iPSC)

in figure 21, in which 69% of erythroid cells positive for CD235a had lost CD71 at day 21, differentiation of coPKD2iPSC ended up generating a 82% of cells positive for CD235a and negative for CD71 (Fig. 36A and B). The presence of these mature erythroid cells confirms that impaired erythropoiesis of PKDiPSC is restored after KI correction mediated by PKLR1 TA.

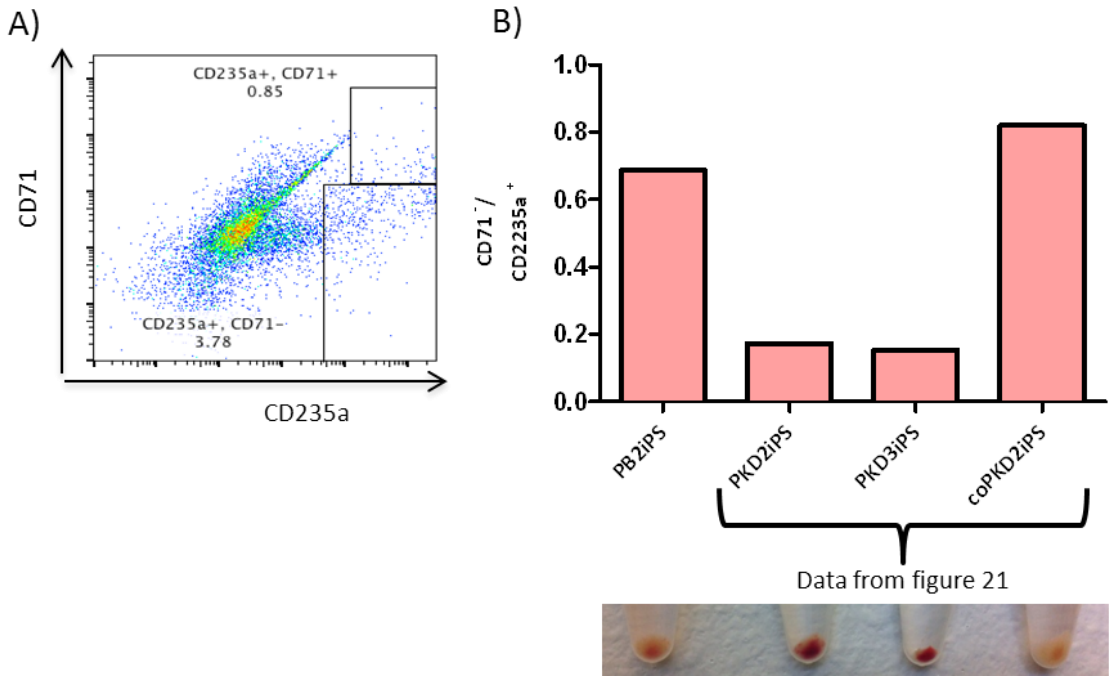


Figure 36. Erythroid maturation profile of coPKD2iPSC. A) Flow cytometry of erythroid markers CD235a and CD71 showing a high percentage (82%, (3.78 / (3.78 + 0.85)) of the total erythroid cells ($CD235a^{+}$) being negative for CD71 as a consequence of maturation. B) CD71 loss at day 21 of differentiation is represented as CD71 negative cells/ total CD235a positive cells. Images show red pellets at day 21 from all the cell lines.

2.6 Genome Stability analysis

As a consequence of reprogramming, prolonged culture, correction process and clonal selection of pluripotent cells, many genomic aberrations could be generated. PKD2iPSC line was already proven to be karyotipically normal (Fig. 17) but, as reported by many authors, numerous genomic abnormalities could have been generated that are not detectable by G banding of metaphasic chromosomes, which has a resolution of 3-10 Mb. Therefore, we analyzed Copy Number Variations (CNV) by array based Comparative Genomic Hybridization (aCGH) and somatic mutations by exome sequencing of the following samples: Peripheral blood mononuclear cells of PKD2 patient (PKD2MNC), the hiPSC line generated from this sample (PKD2iPSC) and one of the corrected clones generated from PKD2iPSC (coPKD2iPSC).

Copy Number Variation (CNV) analysis

For CNV analysis an array of probes that covered the whole human genome with some spacing was used. CNV were detected as DNA segments that have gained (CN ≥3) or loss (CN ≤ 1) a copy number

state in comparison with the reference genomic DNA. Losses of Heterocygosity (LOH) events are also detected due to collection of probes for Single Nucleotide Polymorphisms (SNPs) genotyping. Thirty two CNVs were found in coPKD2iPSC clone 11 (Table 9). The majority of them were already present in the uncorrected PKD2iPSC (91%) and many of them also in the original population of mononuclear cells (PKD2MNC) (75%). An interesting and surprising observation was the one seen in CNV nr. 7, which is a LOH of 6.3 Megabases (Mba) found in PKD2MNC that was not detected in PKD2iPSC and detected back in coPKD2iPSC. Only CNV nr. 1 and 29 were generated during gene correction or during its consequent clonal selection and culture. The first one is a deletion of 60.6 kb that includes the following described genes: *OR2T10, OR2T11, OR2T35*, which are a family of olfactory receptor genes. The second is an amplification of 6 kb that includes the *FGD1* gene. None of them have been associated to survival advantage related to pluripotent cultures. They could also be a consequence of TA off target cutting and therefore we analyzed the sequences surrounding the CNV for homology with the PKLR TALEN targeting sites. No sequences with less than 8 mismatches were found in the vicinity of these CNVs. CNV marked in light red were detected in PKD2iPSC and not in PKD2MNC, probable being a consequence of reprogramming induction and prolonged culture. None of the genes present in these amplified/deleted regions have ever been associated neither to a proliferative or survival advantage nor to be involved in hematopoietic differentiation.

Nr	Chr	Cytoband	Size(bp)	Type	Present in PKD2iPSC	Present in PKD2-MNC	Involved genes
1	**1**	**q44**	**60.641**	**DEL**	**No**	**No**	***OR2T10, OR2T11, OR2T35***
2	2	p21	225.073	AMP	Yes	Yes	
3	2	p11.2	159.607	AMP	Yes	Yes	
4	3	p12.2 - p12.1	3.931.633	LOH	Yes	No	GBE1
5	4	p16.3	81.411	AMP	Yes	Yes	
6	5	p15.33	109.982	AMP	Yes	Yes	
7	5	q23.2 - q31.1	6.301.808	LOH	No	Yes	
8	7	p21.1	433.337	AMP	Yes	Yes	
9	8	p22	58.286	AMP	Yes	Yes	
10	8	p11.22	122.621	AMP	Yes	Yes	
11	8	q11.23	169.460	AMP	Yes	No	TCEA1, LYPLA1
12	9	p23	345.326	DEL	Yes	Yes	
13	10	q11.21	152.536	DEL	Yes	Yes	
14	11	p15.5	548	AMP	Yes	Yes	
15	11	q14.1	113.264	DEL	Yes	No	DLG2
16	12	p13.31	56.626	AMP	Yes	Yes	
17	12	p12.3	1.182.747	AMP	Yes	No	PIK3C2G, PLCZ1, CAPZA3, PLEKHA5, AEBP2
18	14	q11.2	627.101	AMP	Yes	Yes	
19	14	q32.33	132.778	AMP	Yes	Yes	
20	14	q32.33	355.684	AMP	Yes	Yes	
21	14	q32.33	136.254	DEL	Yes	Yes	
22	15	q11.1 - q11.2	2.077.055	AMP	Yes	Yes	
23	15	q14	49.134	DEL	Yes	Yes	
24	17	q21.31	199.747	AMP	Yes	No	KIAA1267, LRRC37A, ARL17A, ARL17B, NSFP1
25	22	q11.22	171.922	AMP	Yes	Yes	
26	X	p22.33	2.630.127	AMP	Yes	Yes	
27	X	p11.23	456	AMP	Yes	Yes	
28	X	p11.23	5.761	AMP	Yes	Yes	
29	**X**	**p11.22**	**6.030**	**AMP**	**No**	**No**	**FGD1**
30	X	q28	184.649	AMP	Yes	Yes	
31	X	q28	206.750	AMP	Yes	Yes	
32	X	q28	288.442	AMP	Yes	Yes	

Table 9: Copy Number Variations (CNVs) in coPKD2iPSC, clone 11.

Somatic mutation analysis

The whole exome of the three samples (PKD2MNC, PKD2iPSC and coPKD2iPSC) was interrogated by Illumina HiSeq 2000 system. After bioinformatics analysis by comparing the sequencing data with the human genome reference, variant calling was performed and a list of 67729 variants was generated. After removing the ones that were already present in PKD2MNC, the list was reduced to 5797 variants. After selecting those included only in exonic regions the list decreased to 420 variants and after removing the ones that were included in the SNP database to 202. Then the ones in which the number of reads was lower than 8 and the variants that were present in less than 20% of the reads were removed, generating a list of 76 variants. By looking at the sequencing raw data of these 76 genomic regions in the IGH (Integrative Genomic Viewer, Broad institute) genome viewer, some of them were removed for having less than 8 reads in the original PKD2MNC and therefore being impossible to discard if they were already present in the original population before reprogramming. The final list of variants (Table 10) included 10 variants and 4 of them were also detected in PKD2iPSC. From them, just 3 of them were present in around 40% of the reads. In order to verify the presence of these mutations by Sanger sequencing, these regions were PCR amplified and sequenced. The mutations in *RUSC2, TACR2* and in *APOA5* were confirmed (Fig 38). The rest of the mutations were also analyzed by sequencing, being none of them verified.

Position	Ref_base	Alt_base	Tot_depth	Alt_depth	% reads	Gene	Change type	Present in PKD2iPSC
chr09_96439006_96439006	-	TGCCTCCACCACACC	9	2	22	PHF2	nonframeshift_insertion	No
chr16_30545672_30545672	G	T	10	2	20	ZNF747	nonsynonymous_SNV	No
chr06_108533436_108533436	G	C	9	2	22	SNX3	nonsynonymous_SNV	No
chr22_50666429_50666429	A	T	9	2	22	TUBGCP6	nonsynonymous_SNV	No
chr10_71168779_71168779	A	G	26	12	46	TACR2	nonsynonymous_SNV	No
chr07_5428308_5428308	C	A	10	2	20	TNRC18	stopgain_SNV	No
chr18_51750555_51750555	C	A	18	4	22	MBD2	nonsynonymous_SNV	Yes
chr18_51750571_51750571	C	A	12	3	25	MBD2	nonsynonymous_SNV	Yes
chr09_35547841_35547841	G	T	65	28	43	RUSC2	nonsynonymous_SNV	Yes
chr11_116661080_116661080	G	A	43	20	47	APOA5	nonsynonymous_SNV	Yes

Table 10: Genetic variants in coPKD2iPSC clone 11

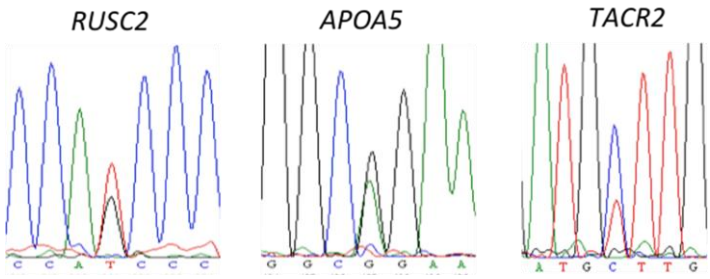


Figure 38 Single nucleotide variants (SNV) verified by Sanger sequencing in coPKD2iPSC clone 11. *RUSC:* G>T, *APOA5:* G>A and *TACR2,* A>G (complementary reverse sequence is displayed in chromatogram, T>C).

VIII Discussion

During the present thesis we have generated hiPSCs from PKD patients, with high efficacy and on a safe manner, which recapitulated the PKD phenotype once differentiated into the erythroid lineage. Moreover, we have accomplished their gene correction through a Knock-In approach in *PKLR* locus that restored the PKD phenotype and analyzed the genome alterations using state of the art techniques for genome sequence analysis.

1 Cell Reprogramming

The idea of generating an unlimited source of cells with the ability to differentiate into any cell from an adult somatic tissue is presented as a very promising approach for regenerative medicine (Robinton and Daley 2012, Cherry and Daley 2013, Svendsen 2013). The establishment of the conditions for the generation of hiPSC, by means of reprogramming, has opened a big window in this new field. Consequently, this discovery was awarded with the 2012 Nobel Prize for Physiology or Medicine to two researchers that were crucial for the development of this technology; Shynya Yamanaka (Takahashi and Yamanaka 2006) and John Gurdon (Gurdon 2006). The two main applications of reprogrammed cells are the already under development use for disease modeling and drug discovery (Egawa, Kitaoka et al. 2012) and as autologous cell source for regenerative medicine, which will probably be demonstrated in a near future.

1.1 Generation of hiPSCs from PKD patients for gene correction

First steps in gene therapy field are based in additional gene therapy (GT), which is the incorporation of a therapeutic copy of the affected gene, which will commonly integrate in a genomic locus different from the affected one. In this regard, previous work from our group proposed an innovative therapy based on additional GT for PKD by using corrected mouse HSC, which reverted many PKD symptoms in PKD mice (Meza, Alonso-Ferrero et al. 2009). Nevertheless, additional GT has important drawbacks. In addition to the already mentioned random integration that could potentially end up in insertional mutagenesis events, for diseases in which the correction does not impair a selective advantage such as PKD, a high chimerism might be established that could assure GT success. In this respect, genes that confer advantage can be included in the GT vector, as it was shown in the latest reported GT attempt for PKD (Trobridge, Beard et al. 2012). However, by using patient specific hiPSC for gene correction instead of patient HSCs, as it is a clonal population in which corrected clones can be selected and expanded, these drawbacks could be solved. Being a clonal and unlimited source of cells, safety analysis could be performed such as integration sites interrogation,

or therapeutic transgene and surrounding genes expression analysis. This was performed by Tubsuwan et al, by conducting a deep biosafety analysis of hiPSC generated from a patient with β-thalassemia, that was cured 5 years ago by lentiviral gene transfer in HSC while showing a myeloid-biased cell clone (Tubsuwan, Abed et al. 2013). The idea of using corrected hiPSC derived HSC for transplantation would also solve one of the main bottlenecks for HSC GT success, which is the available number of cells for transplantation, which in some cases is limited.

We are proposing the use of gene corrected patient specific hiPSC as therapeutic option and thus we believe that there are many aspects of the reprogramming process that are crucial to be taken into account, such as the safety and efficiency of the reprogramming process, the cell source for reprogramming and the differentiation into the desired tissue.

1.1.1 Reprogramming on a safe manner

hiPSC suitable for clinical pursues should not include integrated forms of reprogramming genes mainly because of safety issues, but also because they are more similar to hESCs in terms of gene expression (Cheng, Hansen et al. 2012), genome stability (Liu, Cheng et al. 2012) and differentiation ability (Ramos-Mejia, Montes et al. 2012, Sommer, Christodoulou et al. 2012). Up until now, there have been many described non-integrative reprogramming methods. Among them, two were preferred in terms of combining efficiency and biosafety: 1) Excisable reprogramming polycistronic lentivirus (STEMCCA), that showed a high efficiency when reprogramming different cell types, including CD34 progenitors (Ramos-Mejia, Montes et al. 2012) and can be removed by Cre recombinase excision; and 2) Sendai viral vectors (SeV), that are non-integrative RNA vectors and highly efficient in transduction (Fusaki, Ban et al. 2009). There was a report on reprogramming PB-MNC by lentiviral vectors showing an efficiency of 0.001-0.0002 % (Staerk, Dawlaty et al. 2010), but at that point no one ever described PB-MNC reprogramming by SeV. When we assayed STEMCCA lentiviral system in PB-MNCs, due to technical problems, our detection range was below the described efficiency. So, we could not reprogram any clone. Nevertheless, by plating same amount of cells, reprogramming of PB-MNC by SeV reached up to 0.4-0.6% of reprogramming efficiency, being much more efficient than with lentiviral vectors. Nevertheless, although the STEMCCA can be excised, one of the LoxP site will remain, and this implies a genome modification that could be avoided by using SeV. Also, in the scenario in which more than one copy of STEMCCA would have been integrated, they could recombine in the presence of Cre recombinase.

There is a very recently described methodology that could be equivalent to reprogramming by SeV, which are the so called self-replicating RNAs (SR-RNAs), based on the replicative form of Venezuelan Equine Virus (VEE-RF) (Yoshioka, Gros et al. 2013). As it is in the case of SeV reprogramming, there is no chance of viral integration and the transcription factors expression is so

high that makes the process very efficient (equivalent to retroviruses). Authors claim that in comparison with SeV, there is an important biosafety advantage related to reprogramming sequences disappearance. As the SR-RNAs are maintained in the presence of puromycin, once the drug is removed, they disappear due to cellular degradation. Based on our results on SeV transcripts amplification by PCR, the vectors dilution consequence of rapid cell division works in this system and there are no residual reprogramming genes at passage 15 of culture. We also believe that SeVs show and advantage in comparison to SR-RNAs which is the high viral transduction efficiency that we have also proven, especially in hematopoietic progenitors ($CD34^+$) within PB-MNCs, in which 76% of the cells within this lineage were transduced. When using SR-RNAs, the delivery method is transfection and there are many cell types in which this method is not as efficient as it is in fibroblasts.

1.1.2 Use of PB-MNC as cell source for reprogramming

We chose to use PB-MNC for three main reasons: 1) They can be obtained by a minimally invasive procedure, 2) Can be directly obtained in enough numbers without the need for expansion, which could cause genomic instability and increases the time for hiPSC generation, and 3) Although unproven, they could have hematopoietic differentiation ability advantage.

The most important disadvantage of the PB-MNC would be its low transduction efficiency and this might be the reason why the reprogramming efficiency when using lentiviral vectors decreases from 1% described for fibroblast to 0.001-0.0002% (Staerk, Dawlaty et al. 2010). Nevertheless, by using SeV vectors in combination with the cytokine stimulation conditions described, we obtained high transduction efficiency, increasing reprogramming efficiency to 0.4-0.6%. More interestingly, the combination of our system with the recently described Knock-Down of MBD3 (Rais, Zviran et al. 2013), should be tested to improve the efficiency of PB-MNC reprogramming by SeV. Rais et al showed that the Knock-Down of this gene, a core member of the Mbd3/NuRD (nucleosome remodeling and deacetylation) repressor complex, lead to a deterministic and synchronized reprogramming in the human and the mouse system.

Another point that should be taken into account when using PB-MNCs is that, as it is a hematopoietic heterogeneous population, there is a high percentage of cells belonging to the lymphoid lineage that have undergone somatic rearrangement. This could be used as a genomic signature for hiPSC tracking or for specific immunotherapy (Themeli, Kloss et al. 2013) but there are situations in which the ability to generate a complete repertoire of T or B cells would be desirable. In this regard, as also shown by Staerk et al (Staerk, Dawlaty et al. 2010), we have proved that by stimulating proliferation and promoting the transduction of a specific subset within PB-MNCs with specific cytokines, none of the resulting hiPSC lines showed neither T cell receptor (TCR) nor immunoglobulin (Ig) chain genomic rearrangements. By analyzing SeV transduction in the different

cell lineages within PB-MNCs after cytokines stimulation, we showed that half of the transduced cells were T or B cells, but this was not enough to prompt reprogramming, as when somatic genomic rearrangements were analyzed none of the lines showed neither TCR nor IgH rearrangements. This data suggest that these cytokines prompting survival and proliferation might be affecting reprogramming efficiency, as it has been already shown in the case of the IL-6 (Brady, Li et al. 2013). The effect of cytokines stimulation on reprogramming should be deeply studied as it could provide insight into pluripotency induction. Others have proposed the purification of $CD34^+$ from normal (Merling, Sweeney et al. 2013) or mobilized PB-MNC (Ye, Muench et al. 2013) in combination with SeV vectors for reprogramming . We believe that by using our proposed cocktail of cytokines, we are targeting the reprogramming of non-lymphoid cells without any need for previous cell purification.

1.1.3 PKD and healthy samples reprogramming

Our initial goal was to generate iPSCs from five independent human samples, three from PKD patients and two from healthy donors. We processed PKD1 and PB1 and kept them frozen for later reprogramming. Once thawed and transduced, we could not isolate any clone from PB1 and we isolated some from PKD1 but they did not survive due to technical problems unrelated to the sample. We tried a second reprogramming experiment with both samples but no clones could be isolated. In order to analyze if it could be related to the freshness of the sample, PB2 and PKD2 were reprogrammed right after isolation and many lines were established from both samples. In order to explore the hypothesis that reprogramming is more efficient in fresh samples, we tried to reprogram PKD3 from frozen sample and successfully isolated many lines. Another important fact that could have influenced PB1 reprogramming was the high percentage of lymphoid cells within this PB-MNC sample (98%). Little is known about cellular barriers for reprogramming (Qin, Blaschke et al. 2012), but it has been reported that there are some samples in which the establishment of pluripotency is more difficult than others. So far, difficulties associated to DNA damage (Marion, Strati et al. 2009), high doses of tumor suppressor genes (Li, Collado et al. 2009) or to defects in homologous recombination (Gonzalez, Georgieva et al. 2013) and non-homologous end joining (Molina-Estevez, Lozano et al. 2013, Tilgner, Neganova et al. 2013) have been reported. We believe that there might be sample specific characteristics that could have made more difficult PKD1 or PB1 reprogramming but that the sample was frozen could also have influenced the outcome. In order to verify this hypothesis, new blood samples should be extracted and a new reprogramming with fresh PKD1/PB1 PB-MNC should be attempted.

None of the obtained patient hiPSCs had any defect in proliferation and all the lines grew as expected showing the typical hESC morphology. One clone from each sample was randomly selected for expansion and characterization. The first thing we made was to analyze if the patient iPSC had

the specific patient mutations and, once verified, we performed a deep characterization in which the pluripotency of the 3 lines was assessed. We analyzed the presence of many pluripotency related molecules at gene expression level and at protein level, including the ones belonging to the pluripotency core (Heng, Orlov et al. 2010). Promoter demethylation of two of them, *NANOG* and *SOX2* was also assessed, confirming the total and stable induction of pluripotency in the three cell lines. Surprisingly, *SOX2* promoter was also demethylated in PB-MNCs, which might have facilitated its reprogramming. By looking at the undifferentiation status of the whole culture by flow cytometry, we could verify that there was no tendency to differentiation and that there was a high confluency (MEF percentage was always lower than 10%). Having a pure and confluent culture is a crucial point for hiPSC maintenance and for performing reproducible differentiation experiments. The final proof that pluripotent cell lines have the ability to generate any cell type from the body is to generate a chimera as it is made in mouse iPSCs (Carstea, Pirity et al. 2009). For obvious reasons, this is not possible in the human system and therefore an *in vivo* differentiation system based on teratoma formation was performed. The presence of tissues belonging to the three germ layers (mesoderm, endoderm and ectoderm) confirmed the capacity of the generated cell lines to produce any tissue and consequently its pluripotency. Some authors have argued about the need to establish a standardized minimal set of assays for hiPSC characterization (Loring and Rao 2006, Luong, Auerbach et al. 2011) but the tendency is that each cell bank determines the assays needed in order to keep the cell line. Even a unique bioinformatics assay based on gene expression that would substitute the other assays has been proposed (Muller, Schuldt et al. 2011). We have performed an extensive characterization, with which we are confident on the pluripotency of the PKD patient and healthy hiPSCs.

1.1.4 PKD phenotype recapitulation during PKDiPSC erythroid differentiation

Once we generated the PKDiPSCs, we induce its differentiation into the erythroid lineage and evaluated the disease-associated phenotype. Therefore, we assayed two of the many described erythroid differentiation protocols using the healthy donor PB2iPSC in order to verify that the level of maturation was enough for *RPK* to be expressed. One of the main limitations of all the described approaches is the absence of efficient and full maturation and enucleation of erythrocytes (Sommer, Stadtfeld et al. 2009). In method number 1 the level of *RPK* expression was higher but it showed a big disadvantage in our hands which was the cell expansion. In this regard the method nr. 2, although *RPK* level was lower, promoted cell expansion, allowing the availability of enough material for further biochemical analysis. When *RPK* and *PKM* expression levels were analyzed in method 2 at different time points, both genes were expressed, being *PKM* even higher than *RPK* at day 20. One

possible explanation for this is that there are many highly proliferative cell types different than erythrocytes generated during differentiation that might express high levels of *PKM* and that might increase in numbers during the differentiation process. In order to avoid this interference, erythrocyte-like cells could be purified either by fluorescent or magnetic cell sorting. Overall, it was important for us to have enough material to analyze the different parameters related to disease correction and thus method nr. 2 was the chosen one for PKDiPSC differentiation.

Previous work from our group analyzed the different erythroid developmental stages in PKD mice by looking at the mouse erythroid markers Ter119 (equivalent to human CD235a) and CD71 (Meza, Quintana-Bustamante et al. 2007, Meza, Alonso-Ferrero et al. 2009). They described 4 populations: early proerythroblasts ($Ter119^{med}CD71^{high}$); II, basophilic erythroblasts ($Ter119^{high}CD71^{high}$); III, late basophilic and polychromatophilic erythroblasts ($Ter119^{high}CD71^{med}$); and IV, orthochromatophilic erythroblasts and mature erythroid cells ($Ter119^{high}CD71^{low}$) (see intro. 1.1) When bone marrow and spleen of PKD mice were analyzed (Meza, Alonso-Ferrero et al. 2009), a predominance of immature erythroid precursor cells was observed (proerythroblasts, basophilic erythroblasts, and polychromatophilic erythroblasts) and late erythroid populations were significantly lower than in wild type mice. This anemic phenotype has been also described in thalassemic mice (Somers, Jean et al. 2010). Whereas in bone marrow of wild type mice there is a small percentage of immature erythroid cells (populations II and III), in peripheral blood of healthy humans the majority of erythroid cells belong to population IV. When we analyzed these cell subsets in peripheral blood of one of the PKD patients (PKD2), cells belonging to population II and III were much more prominent than in healthy donors. Importantly, when the differentiation was carried out in PKDiPSC, we observed a high accumulation of cells belonging to groups II and III after 20 days of PKD2iPSC and PKD3iPSC differentiation. The percentage of cells within these two subsets was ten-fold higher than the ones present in PB2iPSC differentiation, confirming the feasibility of hiPSC obtained from PKD patients for modeling the disease.

2 PKDiPSC Knock-In gene correction

There are three main strategies for gene editing mediated gene correction: 1) targeted gene correction, 2) Knock-In gene correction and 3) safe harbor integration. The selection of the best strategy depends on the disease and type of mutation to be corrected. Targeted correction is the cleanest choice and the one that has been more frequently reported in patient hiPSCs (Garate, Davis et al. 2013). However, this approach is only appropriate for one specific patient or group of patients carrying the same mutation, which limits its use. Alternatively, safe-harbor integration is applicable to treat all the genetic diseases already addressed by gene addition therapies, as it is the same strategy directed into a specific genomic site. However, since the therapeutic gene loses its

endogenous regulation, a specific promoter may be required to obtain a tight physiological regulation; additionally, the definition of a safe-harbor locus might not be accurate. The Knock-In strategy is an intermediate possibility in which a large number of patients of a specific disease might be treated. Additionally, the endogenous elements of the locus will regulate the expression of the therapeutic gene. However, there are two concerns regarding the Knock-In strategy: the first one, shared with the safe harbor approach, is that the use of a specific cDNA transgene may exclude the co-expression of various splicing variants; and the second one is the unfeasibility to correct promoter mutations. In our case, we were dealing with a disease in which many different mutations have been described (more than 190), being the majority of them in exons three to eleven. Particularly, in the case of PKD, from the reported mutations described so far, just two were in promoter regions and one in exon two. In addition, a tight physiological gene regulation is very important, as RPK expression is restricted to very specific cell types in which expression is needed. That is why we decided that Knock-In was the best strategy to genetically correct PKD.

We included a codon optimized version of *RPK* cDNA as it has been shown to increase gene expression level (Wiehe, Ponsaerts et al. 2007), followed by a FLAG tag that allowed us to follow the expression of the correctly spliced therapeutic cDNA.

An element of the donor matrix that is subject of controversy is the Selection Cassette (SC). Although its use is common, when Soldner et al generated isogenic lines for Parkinson disease modeling (Soldner, Laganiere et al. 2011), in one of the approaches they did not include a SC for two reasons. i) they wanted to modify the locus minimally and ii) they wanted to avoid the cloning steps associated to a selection process as it will increase the selective pressure into the culture. But we believed that even by using DNA nucleases, the HR efficiency remains too low, as it is shown in figure 27. The ratio of cells that grew in the presence of puromycin was very low compared to the number of colonies in the absence of it, meaning that the number of cells that incorporated the matrix was low. That is why we believed that the inclusion of a SC in the targeting matrix is very important for later selection of correctly targeted clones. The number of clones that should be picked and expanded in the absence of selection in order to have a correctly targeted clone is very high (1 out of 240 in the case of Soldner et al.), and that is why be decided to include this element. Nevertheless, our SC is floxed by LoxP sequences in order to be removed by transient expression of Cre recombinase. Even though, one of the LoxP sequences will remain after the excision, but from our point of view and taking into consideration the fact that it is included in an intron, the genetic modification will most probably going to be harmless.

2.1 Meganuclease mediated homologous recombination

The first assayed nuclease was the PKLR01 MG. As engineered MGs are developed from naturally occurring rare-cutting endonucleases, we thought that the possibility that they could show a reduced toxicity and a higher specificity should be explored. Additionally, as they cleave as monomers, they could avoid the problems associated with the delivery, coordinated expression and dimerization of two molecules as it is the case of other engineered DNA nucleases (Grizot, Epinat et al. 2010). We started analyzing the cleavage efficiency of two of the versions that Cellectis kindly provided us in HEK-293H cell line. Both showed an acceptable mutagenesis being version 2 slightly more efficient, so we decided to use it to correct PKDiPSC by Knock-In approach, since MG-induced targeted mutagenesis and HR efficiencies are strongly correlated (Daboussi, Zaslavskiy et al. 2012).

For the next experiments, we decided to move to the eryhtroleukemic K-562 cells, as it is a cell line in which *PKLR* locus is active and *RPK* is expressed, so we could analyze the splicing after HR. We assayed different doses of PKLR01 MG in the form of mRNA, as it has been shown that nucleofection of mRNA is more efficient, less toxic and yields a higher protein expression in human primary cells (Wiehe, Ponsaerts et al. 2007). The cutting efficiency decreased when using high doses of MG (6 µg, 26%), whereas it was constant (35%) with 0.5 and 3 µg. As 6 µg is not such a high quantity of mRNA to be toxic, it might be due to a high toxicity of the MG. Once we confirmed mutagenesis in *PKLR* locus, we performed HR by combining the MG with the DM and selecting for 20 days with puromycin. In the condition in which just the matrix was nucleofected, a pool of puro^{R} K562 cells was also generated, most probably due to random integration of the matrix. HR-specific PCR confirmed targeting in *PKLR* locus, and our next question was if the therapeutic transgene was being generated properly. Thus, we analyzed the presence of the FLAG tag by flow cytometry. K-562 is a heterogeneous cell line in which just a subset of cells has an active *PKLR* promoter and the level of expression might not be high. That is why, the detection of the FLAG tag was observed only in a small subset of the cells, and at very low levels. We stimulated the erythroid differentiation by hemin (Baliga, Mankad et al. 1993), which increased the number of cells expressing *PKLR* gene and therefore the expression FLAG tag, and confirming gene targeting in at least a subset of cells from the selected pool. This preliminary set-up of the HR conditions was very useful for its later application in patient iPSCs.

Next, we performed HR experiments in PKD2iPSC and isolated and expanded puromycin resistant clones. As the HR-specific PCR was not as clean and resolving as expected, Southern Blot (SB) analysis was performed, which gave us information about integration sites. SB revealed that the DM was not integrated where expected, at least not uniquely, suggesting that the MG could have targeted other places of the genome. This was confirmed by reanalyzing the target sequence of

PKLR01 MG in the human genome, revealing that it was not uniquely found in intron 2 of *PKLR* gene, but also in other 26 sites of the genome. This would also explain the low percentage of cells expressing the flagged RPK in K-562, as many of them would be targeted in other *loci* different than *PKLR*. The SB expected pattern was only found in one of the three analyzed clones and it was not unique. This unexpected setback points out the importance of a preliminary *in silico* engineering of nucleases in order to reduce off targets.

The predominance of integration in other loci different than *PKLR* is surprising even if the MG was cutting at other places, as the homology arms from the DM should prompt the integration in *PKLR* locus. Maybe the generation of DSBs was prompted in other loci more than in *PKLR*. This could be related to the presence of methylated CpGs in the targeting sequence of the MG, as it has been previously reported to affect MGs (Daboussi, Zaslavskiy et al. 2012), and TAs (Shijia Chen 2013) efficiency. It was established that in order to have an effect on DSB generation efficiency by MGs, the methylated CpGs had to be in the central tetrabase of the recognition site (Valton, Daboussi et al. 2012), but this effect of methylation should be deeply analyzed as it could vary depending on the used nuclease and the targeted locus. We analyzed the methylation status of the CpG present in the left extreme of the targeting sequence in hiPSCs and it was indeed methylated, but we cannot conclude the effect of methylation because we had no sample in which it was demethylated.

The reason for HR-specific PCR working better in K-562 than in PKDiPSC might be the fact that in the first one, as we were working with a pool of different targeted cells, among them, there might be some targeted cells in *PKLR* locus that could have generated the PCR product.

The unspecificity of PKLR01 MG should not be generalized to other MGs, as it has been shown that it was due to a design mistake. In addition, there are many studies in which successful MG mediated targeting in human cells have been reported (Grizot, Smith et al. 2009, Cabaniols, Ouvry et al. 2010, Popplewell, Koo et al. 2013). One disadvantage of MGs is that they cannot be designed against any desired genome sequence, and there was no other MG available that could be combined with our generated targeting matrix. That is why we, in collaboration with Cellectis, decided to use another type of DNA nuclease, the PKLR1 TALEN™ (TA).

2.2 TALEN™ mediated homologous recombination

The simplicity of TAs design allows a fast engineering, being even possible to be designed by the user without the need to purchase them. This is the reason why, in just two years, TALENs have been used in many of the applications in which ZFN had already being used: hPSCs gene editing and rat, mouse, zebrafish, and worm transgenic models generation. In an attempt to compare TAs and ZFNs in hESCs and hiPSCs (Hockemeyer, Wang et al. 2011), several loci have been targeted and HR efficiencies with both types of nucleases have been compared. It was concluded that both ZFNs and

TAs show similar efficiencies and accuracy. Nevertheless, an *in vitro* gene disruption comparison between ZFN and TAs showed that TAs were more efficient and less cytotoxic (Mussolino, Morbitzer et al. 2011).

The TA designed for targeting intron 2 of *PKLR* included the targeting sequence of PKLR01 MG, being 25 bp longer. By bioinformatics analysis it was not possible to find the targeted sequence in other places of the genome in addition to the *PKLR* locus. We assayed its cutting efficiency by the surveyor assay directly in PKDiPSC. We performed an attempt by nucleofecting both TA subunits as mRNA (data not shown) but we detected no cutting whereas as plasmid DNA we did. The reason for this might be that an optimization of the *in vitro* mRNA synthesis assay should have been performed as the length of each TA subunit was 3 Kb whereas the MG was 1 kb. Even when using plasmid DNA, cutting efficiency was lower than the one observed in K-562, actually being below the quantification range. One factor that could be related to this decrease is the DNA nucleofection efficiency. When a GFP expressing plasmid was nucleofected in PKD2iPSCs and GFP was analyzed, 29% of the $SSEA4^{+}$ population was positive for GFP, whereas in K-562 it was 65%. Being the TA expressing plasmids larger, and taking into account that the two plasmids expressing both subunits should enter the cell, it is understandable that the cutting efficiency decreases. Regarding the effect of methylation on DSB generation efficiency, it should also be taken into account than PKLR1 TA recognition site has two CpGs that resulted to be methylated in hiPSCs.

Using the PKLR1 TA, we performed HR experiments in PKD2iPSC and PKD3iPSC, resulting in a targeting efficiency of 77% and 76% within the total of puromycin resistant clones, respectively. Unique targeting was confirmed in 5 of the PKD2iPSC targeted clones, verifying the specificity of PKLR1 TA. Although a proper evaluation of PKLR1 TA efficiency would include data in the absence of puromycin, we believe that the presented data clearly demonstrates the accuracy and applicability of the PKLR1 TA.

PKD is a recessive disease and thus, correcting just one allele would be enough to restore the phenotype. Nevertheless, we believe that the capacity of the designed HR approach to generate homozygous corrected clones is important, as there are many other scenarios in which both alleles should be targeted. When we analyzed if the targeting happened in homozygosis or heterozygosis, in PKD2iPSC there was not any homozygous clone, whereas in PKD3iPSC there were 3 potential clones (11%). Still, the integration in both alleles in these three clones needs to be confirmed by southern blot.

The presence of NHEJ mediated genome modifications in the untargeted allele has been analyzed in many gene correction studies without finding any evidence of it (Sebastiano, Maeder et al. 2011, Soldner, Laganiere et al. 2011, Zou, Mali et al. 2011). This is especially important in targeted gene

correction approaches as the matrix is inserted in the gene and a genomic modification of the untargeted locus might be deleterious. In our case we were targeting an intron and therefore, a genome modification would probably have no effect, but we believed it was interesting to analyze its integrity for HR strategy design purposes. This analysis revealed genome modifications in the untargeted allele of several clones (40% in PKD2iPSC and 31% in PKD3iPSC) pointing out another characteristic that should be considered when using DNA nucleases. The high percentage of homozygous targeted clones in PKD3iPSC together with the high NHEJ in the untargeted alleles in both PKDiPSCs imply a high DSB generation by the PKLR1 TA. It also suggests that the outcome of the PKLR1 TA surveyor assay in PKDiPSC is more related to difficulties for proper nucleofection of both TA subunits rather than to a reduced DSB generation efficiency.

Another very important possible consequence of using DNA nucleases is the possibility of cutting other sequences of the genome that are similar to the intended one, the so called off targets. PKLR1 TA targeting sequence was very unique being impossible to find by bioinformatics approaches other sequences in the genome with less than 5 mismatches. All the found sequences with 5 and 6 mismatches were interrogated by Sanger sequencing in several clones from coPKD2iPSC and coPKD3iPSC, being all of them intact. In other studies, this type of analysis was performed by next generation sequencing, so that very low represented events could be detected. In our case, as the corrected clones come from a unique cell due to the clonal nature of hiPSC lines, if in this cell a NHEJ-modification would have been produced in any off target, it should be present in all the clonal population. Some authors have suggested that bioinformatics approaches based on sequence homology are not accurate enough for off targets prediction and have proposed alternative *in vitro, in vivo* and biochemical assays (Gabriel, Lombardo et al. 2011, Pattanayak, Ramirez et al. 2011). We have not considered them here, but this additional biosafety analysis should be performed in the case the cells would be used for clinical settings.

2.2.1 Allele specific targeting

The reason for having no homozygous targeted clones in PKD2iPSC might have been in the sequence of the Homology Arm (HA) of the Door Matrix (DM). When the surveyor assay was performed in PKD2iPSC, two unexpected and high intensity bands appeared. They were consequence of an SNP close to the TA targeting site in one of the alleles of PKD2iPSC that was confirmed by Sanger sequencing. We thought that the presence of a mismatch between the HA and the genomic DNA to be targeted, could influence HR efficiency. Two studies reported the effect of variable sequences within different strains in mouse ESC on gene targeting (te Riele, Maandag et al. 1992, Zhou, Rowley et al. 2001) but, as far as we know, there are no studies analyzing the effect of SNPs in the HAs of the donor matrix. We concluded that the targeted allele in all the clones was the one that had 100%

homology with the HA. In order to proof the effect of this SNP on HR, we should generate a DM to target specifically the allele where the SNP is. As in PKD3iPSC there is no SNP close to TA target site, HR took place in both alleles, as we have observed. The use of this potential new DM should theoretically strongly reduce HR in PKD3iPSCs. There are many questions about this process such us how the SNP would influence if it would have been in the other arm, or how far can it be from the beginning of the HA in order to have an impact on HR efficiency. Another fact that could also have influenced HR in this locus is the repetitive nature of the right HA sequence.

For PKD gene correction, this is not crucial as by correcting any of both alleles it would be enough to cure this disease. This would not be the case of a dominant disease such as sickle cell anemia, in which the correction should be done preferentially in the allele that harbors the disease causing mutation. Another reported disease correction approach that would have taken advantage of an SNP in a specific chromosome, would be the one in which down syndrome hiPSC were corrected by a Xist mediated inactivation of one of the extra chromosomes (Jiang, Jing et al. 2013). They selected the clone in which just one chromosome was inactivated, but they could have targeted the expression of Xist lncRNA into one of the three chromosomes in the case it would have been and SNP in the vicinity of the ZFN targeting sequence. From our point of view, a deep study of HAs SNPs influence in HR process would be very useful for future gene targeting strategies design.

2.2.2 Transgene expression and phenotype restoration in corrected PKDiPSC

HR process had no deleterious effect on pluripotency maintenance in corrected PKD2iPSC (coPKD2iPSC) as it was shown by pluripotency gene expression array and *SOX2* and *NANOG* promoter demethylation. When coPKD2iPSC was differentiated into the erythroid lineage, the expression of the flagged RPK was assessed by western blot and the two bands corresponding to both RPK isoforms could be detected using an anti-FLAG antibody, assessing the correct targeting in PKD2iPSC. Once we had the proof that the wild type transcript was expressed once the cells were differentiated, what was crucial to be analyzed was the effect of gene correction in PKD phenotype. The maturation profile of coPKD2iPSC, expressed as CD71 loss within the CD235a positive cells, was similar to the one described for PB2iPSC, being a high percentage of erythroid cells negative for CD71.

2.3 Genome stability analysis

Since the derivation of hESC in 1998 and the generation of hiPSCs in 2007, many studies have reported the appearance of genomic variations in pluripotent cultures such as aneuploidies, subchromosomal copy number variations (CNV) and single nucleotide variations (SNV) (Liang and Zhang 2013). These might be generated due to the selective pressure established during the

reprogramming process or prolonged culture. In addition, we should take special attention to the ones that could have been generated by the process of HR, especially by the use of DNA nucleases and the clonal selection procedures.

A comprehensive karyotyping study of many hESC and hiPSC lines conducted by the WiCell Research Institute concluded that 13% of the analyzed lines showed aberrant karyotypes (Taapken, Nisler et al. 2011). Gaining an extra copy of a chromosome might increase the dose of a gene beneficial for self-renewal, clonogenicity or proliferation rate and therefore be selected and fixed in the culture. An example of this hypothesis is the commonly found trisomy of chromosome 12 (Draper, Smith et al. 2004), in which *NANOG* and *GDF3* genes are located. In our case, none of the analyzed cell lines showed any aneuploidy. Nevertheless, all of them were below passage 17 and although culturing for longer periods would increase the chances of genomic aberrations, it does not have to, as high passage hiPSC have been reported in which no aneuploidies were detected (Taapken, Nisler et al. 2011). Corrected clone 11 from PKD2iPSC was analyzed 16 passages later for chromosome gains or losses but in this case by CGH, which has a higher resolution (25.3kb overall, 5kb in ISCA (International Standards of cytogenetic arrays) regions, which are regions shown to be involved in cancer). Six CNVs and one Loss of Heterozygosity (LOH) events were detected that were not present in the original tissue, PKD2PB-MNC. There were two deletions and four amplifications and their mean size was 943 Kb. None of them were the ones commonly described e.g. around *NANOG* gene in chromosome 12 or around *DNMT3B* in chromosome 20. From these six, two CNV were detected after correction, CNV 1 and 29. There were no genes associated to a survival or proliferation advantage either in the amplified or deleted regions. Although the probability for generating such big modifications by the TA is very low, we also looked for sequence homology surrounding the CNV without finding any evidence of it.

One of the main limitations of this and other similar genomic techniques is that the outcome of the analysis will be the one harboring the majority of cells and low frequency variations in the sample before reprogramming might be beyond the detection limit. A study in which deep sequencing techniques were used for genome analysis concluded that the majority of observed CNV in hiPSC were already present in the original source but at low frequency, consequence of the somatic mosaicism present in many human tissues (Abyzov, Mariani et al. 2012). It might be a possibility that the four CNV that appeared in the PKD2iPSC were not truly *de novo.* The two that appeared after correction were actually *de novo* as they were not detected in the PKD2iPSC sample.

In this regard, the whole exome sequencing is a more quantitative technique; generally the number of reads from a concrete region is higher than 10 and consequently, a low represented Single Nucleotide Variant (SNV) could be detected but in a lower number of reads. The higher the read

depth, the better low represented SNVs can be detected. We used a 30X platform in which the average number of reads was 30 but there are other sequencing platforms in which it could be much higher, for example the Ion Torrent semiconductor sequencer, which could give a 50X read depth (Rothberg, Hinz et al. 2011). In our whole exome sequencing study, ten SNVs were detected that were not present in PKD2PB-MNC, being four of them also present in the uncorrected PKD2iPSCs. From them, just 3 could be verified by Sanger sequencing. The rest, being all of them present in around 20% of the reads, could not be verified. This might be because their dose is below the detection range by Sanger sequencing or because they were wrongly detected in the exome sequencing. Nevertheless, as they have not been fixed in the culture, their presence might not be as important as a mutation present in all the cells.

The observed SNVs in our samples could corroborate the hypothesis that hiPSC related mutagenesis might be stochastic, as the in majority of studies in which they have been analyzed, there were not shared (Liang and Zhang 2013).

As it was in the regions affected by CNVs, none of the verified SNVs affected genes already described to be involved in essential biological processes and neither in hematopoiesis. Nevertheless, special attention should be taken to the genes present in these regions as their function for the intended application might be crucial. In our case, they are not genes that have been described to be involved in essential processes or related to the hematopoietic system, but there are many genes that have not been described yet or gene functions that might not be accurate and therefore this is a fact that should be deeply analyzed before hiPSC can be used in the clinics.

Another fact that should be studied is the acquisition of genome modifications during the differentiation process. In a study by Laurent el al., the most rapidly arising genomic aberrations were the ones generated during a directed differentiation experiment (Laurent, Ulitsky et al. 2011), pointing out that it is a highly selective process that should be deeply studied before the differentiated cells could be used for clinical pursues. On the other hand, and taking into consideration the reported difficulties to generate transplantable HSCs, the treatment of erythroid diseases could be performed by using autologous mature enucleated erythrocytes. In this regard, the absence of a nucleus would reduce the risks associated to the acquired genomic modifications.

3 Future perspectives

Our results corroborate the feasibility of hiPSC and HR technology for disease modeling and as a platform for the development of safe innovative therapies that could be used for many genetic diseases in the future. The synergy between reprogramming and gene editing is prompting the progress of this new field and our work adds valuable information to the different described strategies that are needed before this type of therapeutic approaches reach the clinics. As there are

many technical limitations that need to be solved, mainly genomic modifications and differentiation limitations, we believe that a big effort needs to be done to understand all the steps involved in the development of gene corrected differentiated cells so that such a powerful tool for regenerative medicine could be fully explored.

IX Conclusions

The aim of this thesis is the feasibility analysis of the combination of hiPSC and gene editing technologies as a therapeutic tool for PKD treatment, and from the results obtained during its development several conclusions can be outlined:

1. Induced Pluripotent Stem Cells (iPSCs) from Pyruvate Kinase Deficient (PKD) patients (PKDiPSC) and healthy donors (PBiPSC) have been generated on a safe manner, by transduction of peripheral blood mononuclear cells (PB-MNC) using Sendai viral vectors (SeV) expressing reprograming factors, showing a higher efficiency than the ones reported with similar methodologies.

2. Erythroid differentiation of PKDiPSCs resembles the immature phenotype observed in peripheral blood of PKD patients, assessing the use of hiPSC for *in vitro* modeling of hematopoietic diseases.

3. By the use of PKLR1 TALEN^{TM}, PKDiPSC derived from two different patients have been targeted by homologous recombination uniquely in *PKLR* locus, showing no off target genomic modifications, and leading to the expression of the therapeutic recombinant RPK and the restoration of the erythroid phenotype.

4. The presence of a single nucleotide polymorphism in PKD2iPSC lead to an allele specific targeting, inserting the therapeutic matrix exclusively in the allele that perfectly matched the donor matrix.

5. Few genomic variants have been acquired in corrected PKD2iPSC, being none of them the commonly described ones in hiPSC and not affecting genes reported to be involved in survival or hematopoiesis.

6. The correction of hiPSCs by HR is revealed as a powerful tool for gene replacement therapy for PKD.

IX Conclusiones

El objetivo de la presente tesis es el estudio de combinación de *iPSC* de pacientes con su corrección genética mediante recombinación homóloga, como una alternativa terapéutica para el tratamiento de la Deficiencia en Piruvato Quinasa (DPQ). De los resultados obtenidos durante su desarrollo destacan las siguientes conclusiones:

1. Se han obtenido células madre pluripotentes inducidas (*iPSC*) a partir de muestras de pacientes con DPQ (DPQiPSC) y de donantes sanos de forma segura, mediante la transducción de células mononucleares de sangre periférica con vectores virales Sendai (SeV) que expresan factores de reprogramación, mostrando una eficiencia de reprogramación superior a metodologías similares.

2. La diferenciación eritroide de DPQiPSC mostró un fenotipo inmaduro, similar al observado en los pacientes con DPQ, confirmando el uso de *hiPSCs* para modelado *in vitro* de enfermedades del sistema hematopoyético.

3. Mediante el uso de PKLR1 TALEN™, se corrigieron *iPSC* de dos pacientes de DPQ mediante recombinación homóloga en el locus *PKLR*, sin detectarse ninguna modificación genética indeseable debida a la nucleasa. Se confirmó la expresión de la proteína terapéutica RPK recombinante, lo cual restauró el fenotipo eritroide.

4. La presencia de un polimorfismo de un solo nucleótido en una de las muestras dio lugar a una integración específica de alelo, introduciéndose la matriz terapéutica exclusivamente en el alelo que era perfectamente homólogo.

5. Se han adquirido algunas modificaciones genómicas durante el proceso de reprogramación, cultivo y corrección. Ninguna variante adquirida ha sido previamente descrita en *hiPSC* ni se encuentra en genes involucrados en supervivencia o en hematopoyesis.

6. La corrección genética de *hiPSCs* mediante recombinación homóloga se presenta como una poderosa herramienta de terapia génica de reemplazo para el tratamiento de DPQ.

X References

Aasen, T. and J. C. Izpisua Belmonte (2010). "Isolation and cultivation of human keratinocytes from skin or plucked hair for the generation of induced pluripotent stem cells." Nat Protoc **5**(2): 371-382.

Abad, M., L. Mosteiro, C. Pantoja, M. Canamero, T. Rayon, I. Ors, O. Grana, D. Megias, O. Dominguez, D. Martinez, M. Manzanares, S. Ortega and M. Serrano (2013). "Reprogramming in vivo produces teratomas and iPS cells with totipotency features." Nature.

Abyzov, A., J. Mariani, D. Palejev, Y. Zhang, M. S. Haney, L. Tomasini, A. F. Ferrandino, L. A. Rosenberg Belmaker, A. Szekely, M. Wilson, A. Kocabas, N. E. Calixto, E. L. Grigorenko, A. Huttner, K. Chawarska, S. Weissman, A. E. Urban, M. Gerstein and F. M. Vaccarino (2012). "Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells." Nature **492**(7429): 438-442.

Aizawa, E., Y. Hirabayashi, Y. Iwanaga, K. Suzuki, K. Sakurai, M. Shimoji, K. Aiba, T. Wada, N. Tooi, E. Kawase, H. Suemori, N. Nakatsuji and K. Mitani (2012). "Efficient and accurate homologous recombination in hESCs and hiPSCs using helper-dependent adenoviral vectors." Mol Ther **20**(2): 424-431.

Alberghini, M., G. Pasquinelli, L. Zanella, G. Pignatti, S. Benini, P. Bacchini and F. Bertoni (2007). "Primary malignant myoepithelioma of the distal femur." APMIS **115**(4): 376-380.

Amit, M., M. K. Carpenter, M. S. Inokuma, C. P. Chiu, C. P. Harris, M. A. Waknitz, J. Itskovitz-Eldor and J. A. Thomson (2000). "Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture." Dev Biol **227**(2): 271-278.

An, M. C., N. Zhang, G. Scott, D. Montoro, T. Wittkop, S. Mooney, S. Melov and L. M. Ellerby (2012). "Genetic correction of Huntington's disease phenotypes in induced pluripotent stem cells." Cell Stem Cell **11**(2): 253-263.

Anokye-Danso, F., C. M. Trivedi, D. Juhr, M. Gupta, Z. Cui, Y. Tian, Y. Zhang, W. Yang, P. J. Gruber, J. A. Epstein and E. E. Morrisey (2011). "Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency." Cell Stem Cell **8**(4): 376-388.

Arnould, S., P. Chames, C. Perez, E. Lacroix, A. Duclert, J. C. Epinat, F. Stricher, A. S. Petit, A. Patin, S. Guillier, S. Rolland, J. Prieto, F. J. Blanco, J. Bravo, G. Montoya, L. Serrano, P. Duchateau and F. Paques (2006). "Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets." J Mol Biol **355**(3): 443-458.

Baliga, B. S., M. Mankad, A. K. Shah and V. N. Mankad (1993). "Mechanism of differentiation of human erythroleukaemic cell line K562 by hemin." Cell Prolif **26**(6): 519-529.

Ban, H., N. Nishishita, N. Fusaki, T. Tabata, K. Saeki, M. Shikamura, N. Takada, M. Inoue, M. Hasegawa, S. Kawamata and S. Nishikawa (2011). "Efficient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus vectors." Proc Natl Acad Sci U S A **108**(34): 14234-14239.

Ban, H., N. Nishishita, N. Fusaki, T. Tabata, K. Saeki, M. Shikamura, N. Takada, M. Inoue, M. Hasegawa, S. Kawamata and S. Nishikawa (2011). Efficient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus vectors. Proc Natl Acad Sci U S A. **108:** 14234-14239.

Bar-Nur, O., H. A. Russ, S. Efrat and N. Benvenisty (2011). "Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells." Cell Stem Cell **9**(1): 17-23.

Baron, M. H. (2013). "Early Embryonic Erythropoiesis ... Not so Primitive After All." Stem Cells.

Beard, B. C. and H. P. Kiem (2009). "Canine models of gene-modified hematopoiesis." Methods Mol Biol **506**: 341-361.

Bedel, A., M. Taillepierre, V. Guyonnet-Duperat, E. Lippert, P. Dubus, S. Dabernat, T. Mautuit, B. Cardinaud, C. Pain, B. Rousseau, M. Lalanne, C. Ged, Y. Duchartre, E. Richard, H. de Verneuil and F. Moreau-Gaudry (2012). "Metabolic correction of congenital erythropoietic porphyria with iPSCs free of reprogramming factors." Am J Hum Genet **91**(1): 109-121.

Beurdeley, M., F. Bietz, J. Li, S. Thomas, T. Stoddard, A. Juillerat, F. Zhang, D. F. Voytas, P. Duchateau and G. H. Silva (2013). "Compact designer TALENs for efficient genome engineering." Nat Commun **4**: 1762.

Beutler, E. and T. Gelbart (2000). "Estimating the prevalence of pyruvate kinase deficiency from the gene frequency in the general white population." Blood **95**(11): 3585-3588.

Bibikova, M., K. Beumer, J. K. Trautman and D. Carroll (2003). "Enhancing gene targeting with designed zinc finger nucleases." Science **300**(5620): 764.

Bibikova, M., M. Golic, K. G. Golic and D. Carroll (2002). "Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases." Genetics **161**(3): 1169-1175.

Blackburn, P. R., J. M. Campbell, K. J. Clark and S. C. Ekker (2013). "The CRISPR system--keeping zebrafish gene targeting fresh." Zebrafish **10**(1): 116-118.

Blasco, M. A., M. Serrano and O. Fernandez-Capetillo (2011). "Genomic instability in iPS: time for a break." EMBO J **30**(6): 991-993.

Boch, J., H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt and U. Bonas (2009). "Breaking the code of DNA binding specificity of TAL-type III effectors." Science **326**(5959): 1509-1512.

Bogdanove, A. J. and D. F. Voytas (2011). "TAL effectors: customizable proteins for DNA targeting." Science **333**(6051): 1843-1846.

Braam, S. R., C. Denning, S. van den Brink, P. Kats, R. Hochstenbach, R. Passier and C. L. Mummery (2008). "Improved genetic manipulation of human embryonic stem cells." Nat Methods **5**(5): 389-392.

Braam, S. R., L. Tertoolen, S. Casini, E. Matsa, H. R. Lu, A. Teisman, R. Passier, C. Denning, D. J. Gallacher, R. Towart and C. L. Mummery (2013). "Repolarization reserve determines drug responses in human pluripotent stem cell derived cardiomyocytes." Stem Cell Res **10**(1): 48-56.

Brady, J. J., M. Li, S. Suthram, H. Jiang, W. H. Wong and H. M. Blau (2013). "Early role for IL-6 signalling during generation of induced pluripotent stem cells revealed by heterokaryon RNA-Seq." Nat Cell Biol **15**(10): 1244-1252.

Cabaniols, J. P., C. Ouvry, V. Lamamy, I. Fery, M. L. Craplet, N. Moulharat, S. P. Guenin, S. Bedut, O. Nosjean, G. Ferry, S. Devavry, C. Jacqmarcq, C. Lebuhotel, L. Mathis, C. Delenda, J. A. Boutin, P. Duchateau, F. Coge and F. Paques (2010). "Meganuclease-driven targeted integration in CHO-K1 cells for the fast generation of HTS-compatible cell-based assays." J Biomol Screen **15**(8): 956-967.

Carey, B. W., S. Markoulaki, J. Hanna, K. Saha, Q. Gao, M. Mitalipova and R. Jaenisch (2009). "Reprogramming of murine and human somatic cells using a single polycistronic vector." Proc Natl Acad Sci U S A **106**(1): 157-162.

Carroll, D. (2011). "Zinc-finger nucleases: a panoramic view." Curr Gene Ther **11**(1): 2-10.

Carstea, A. C., M. K. Pirity and A. Dinnyes (2009). "Germline competence of mouse ES and iPS cell lines: Chimera technologies and genetic background." World J Stem Cells **1**(1): 22-29.

Carvajal-Vergara, X., A. Sevilla, S. L. D'Souza, Y. S. Ang, C. Schaniel, D. F. Lee, L. Yang, A. D. Kaplan, E. D. Adler, R. Rozov, Y. Ge, N. Cohen, L. J. Edelmann, B. Chang, A. Waghray, J. Su, S. Pardo, K. D. Lichtenbelt, M. Tartaglia, B. D. Gelb and I. R. Lemischka (2010). "Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome." Nature **465**(7299): 808-812.

Cathomen, T. and J. K. Joung (2008). "Zinc-finger nucleases: the next generation emerges." Mol Ther **16**(7): 1200-1207.

Cermak, T., E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove and D. F. Voytas (2011). "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting." Nucleic Acids Res **39**(12): e82.

Certo, M. T., K. S. Gwiazda, R. Kuhar, B. Sather, G. Curinga, T. Mandt, M. Brault, A. R. Lambert, S. K. Baxter, K. Jacoby, B. Y. Ryu, H. P. Kiem, A. Gouble, F. Paques, D. J. Rawlings and A. M. Scharenberg (2012). "Coupling endonucleases with DNA end-processing enzymes to drive gene disruption." Nat Methods **9**(10): 973-975.

Colleaux, L., L. D'Auriol, F. Galibert and B. Dujon (1988). "Recognition and cleavage site of the intron-encoded omega transposase." Proc Natl Acad Sci U S A **85**(16): 6022-6026.

Cradick, T. J., E. J. Fine, C. J. Antico and G. Bao (2013). "CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity." Nucleic Acids Res.

Chen, Z. and H. Zhao (2005). "A highly sensitive selection method for directed evolution of homing endonucleases." Nucleic Acids Res **33**(18): e154.

Cheng, L., N. F. Hansen, L. Zhao, Y. Du, C. Zou, F. X. Donovan, B. K. Chou, G. Zhou, S. Li, S. N. Dowey, Z. Ye, N. C. S. Program, S. C. Chandrasekharappa, H. Yang, J. C. Mullikin and P. P. Liu (2012). "Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression." Cell Stem Cell **10**(3): 337-344.

Cherry, A. B. and G. Q. Daley (2013). "Reprogrammed cells for disease modeling and regenerative medicine." Annu Rev Med **64**: 277-290.

Daboussi, F., M. Zaslavskiy, L. Poirot, M. Loperfido, A. Gouble, V. Guyot, S. Leduc, R. Galetto, S. Grizot, D. Oficjalska, C. Perez, F. Delacote, A. Dupuy, I. Chion-Sotinel, D. Le Clerre, C. Lebuhotel, O. Danos, F. Lemaire, K. Oussedik, F. Cedrone, J. C. Epinat, J. Smith, R. J. Yanez-Munoz, G. Dickson, L. Popplewell, T. Koo, T. VandenDriessche, M. K. Chuah, A. Duclert, P. Duchateau and F. Paques (2012). "Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases." Nucleic Acids Res **40**(13): 6367-6379.

Delacote, F. and B. S. Lopez (2008). "Importance of the cell cycle phase for the choice of the appropriate DSB repair pathway, for genome stability maintenance: the trans-S double-strand break repair model." Cell Cycle **7**(1): 33-38.

Delacote, F., C. Perez, V. Guyot, M. Duhamel, C. Rochon, N. Ollivier, R. Macmaster, G. H. Silva, F. Paques, F. Daboussi and P. Duchateau (2013). "High frequency targeted mutagenesis using engineered endonucleases and DNA-end processing enzymes." PLoS One **8**(1): e53217.

Dias, J., M. Gumenyuk, H. Kang, M. Vodyanik, J. Yu, J. A. Thomson and Slukvin, II (2011). "Generation of red blood cells from human induced pluripotent stem cells." Stem Cells Dev **20**(9): 1639-1647.

Diez, A., F. Gilsanz, J. Martinez, S. Perez-Benavente, N. W. Meza and J. M. Bautista (2005). "Life-threatening nonspherocytic hemolytic anemia in a patient with a null mutation in the PKLR gene and no compensatory PKM gene expression." Blood **106**(5): 1851-1856.

Doyon, J. B., V. Pattanayak, C. B. Meyer and D. R. Liu (2006). "Directed evolution and substrate specificity profile of homing endonuclease I-SceI." J Am Chem Soc **128**(7): 2477-2484.

Draper, J. S., K. Smith, P. Gokhale, H. D. Moore, E. Maltby, J. Johnson, L. Meisner, T. P. Zwaka, J. A. Thomson and P. W. Andrews (2004). "Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells." Nat Biotechnol **22**(1): 53-54.

Egawa, N., S. Kitaoka, K. Tsukita, M. Naitoh, K. Takahashi, T. Yamamoto, F. Adachi, T. Kondo, K. Okita, I. Asaka, T. Aoi, A. Watanabe, Y. Yamada, A. Morizane, J. Takahashi, T. Ayaki, H. Ito, K. Yoshikawa, S. Yamawaki, S. Suzuki, D. Watanabe, H. Hioki, T. Kaneko, K. Makioka, K. Okamoto, H. Takuma, A. Tamaoka, K. Hasegawa, T. Nonaka, M. Hasegawa, A. Kawata, M. Yoshida, T. Nakahata, R. Takahashi, M. C. Marchetto, F. H. Gage, S. Yamanaka and H. Inoue (2012). "Drug screening for ALS using patient-specific induced pluripotent stem cells." Sci Transl Med **4**(145): 145ra104.

Fischer, A., S. Hacein-Bey-Abina and M. Cavazzana-Calvo (2010). "20 years of gene therapy for SCID." Nat Immunol **11**(6): 457-460.

Fong, H., C. Wang, J. Knoferle, D. Walker, Maureen E. Balestra, Leslie M. Tong, L. Leung, Karen L. Ring, William W. Seeley, A. Karydas, Mihir A. Kshirsagar, Adam L. Boxer, Kenneth S. Kosik, Bruce L. Miller and Y. Huang (2013). "Genetic Correction of Tauopathy Phenotypes in Neurons Derived from Human Induced Pluripotent Stem Cells." Stem Cell Reports **1**(3): 226-234.

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

Fusaki, N., H. Ban, A. Nishiyama, K. Saeki and M. Hasegawa (2009). "Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome." Proc Jpn Acad Ser B Phys Biol Sci **85**(8): 348-362.

Gabriel, R., A. Lombardo, A. Arens, J. C. Miller, P. Genovese, C. Kaeppel, A. Nowrouzi, C. C. Bartholomae, J. Wang, G. Friedman, M. C. Holmes, P. D. Gregory, H. Glimm, M. Schmidt, L. Naldini and C. von Kalle (2011). "An unbiased genome-wide analysis of zinc-finger nuclease specificity." Nat Biotechnol **29**(9): 816-823.

Garate, Z., B. R. Davis, O. Quintana-Bustamante and J. C. Segovia (2013). "New frontier in regenerative medicine: site-specific gene correction in patient-specific induced pluripotent stem cells." Hum Gene Ther.

Garcia-Gomez, M., Quintana-Bustamante, O., Garcia-Bravo, M., Navarro, S., Garate, Z. & Segovia, J.C. (2013). "Gene Therapy for Erythroid Metabolic Inherited Diseases. In Gene Therapy for Erythroid Metabolic Inherited Diseases ": 511-538.

Garcia, S. C., A. C. Moragon and M. E. Lopez-Fernandez (1979). "Frequency of glutathione reductase, pyruvate kinase and glucose-6-phosphate dehydrogenase deficiency in a Spanish population." Hum Hered **29**(5): 310-313.

Garcon, L., J. Ge, S. H. Manjunath, J. A. Mills, M. Apicella, S. Parikh, L. M. Sullivan, G. M. Podsakoff, P. Gadue, D. L. French, P. J. Mason, M. Bessler and M. J. Weiss (2013). "Ribosomal and hematopoietic defects in induced pluripotent stem cells derived from Diamond Blackfan anemia patients." Blood **122**(6): 912-921.

Giarratana, M. C., H. Rouard, A. Dumont, L. Kiger, I. Safeukui, P. Y. Le Pennec, S. Francois, G. Trugnan, T. Peyrard, T. Marie, S. Jolly, N. Hebert, C. Mazurier, N. Mario, L. Harmand, H. Lapillonne, J. Y. Devaux and L. Douay (2011). "Proof of principle for transfusion of in vitro-generated red blood cells." Blood **118**(19): 5071-5079.

Gilsanz, F., M. A. Vega, E. Gomez-Castillo, J. A. Ruiz-Balda and F. Omenaca (1993). "Fetal anaemia due to pyruvate kinase deficiency." Arch Dis Child **69**(5 Spec No): 523-524.

Gonzalez, F., D. Georgieva, F. Vanoli, Z. D. Shi, M. Stadtfeld, T. Ludwig, M. Jasin and D. Huangfu (2013). "Homologous recombination DNA repair genes play a critical role in reprogramming to a pluripotent state." Cell Rep **3**(3): 651-660.

Gore, A., Z. Li, H. L. Fung, J. E. Young, S. Agarwal, J. Antosiewicz-Bourget, I. Canto, A. Giorgetti, M. A. Israel, E. Kiskinis, J. H. Lee, Y. H. Loh, P. D. Manos, N. Montserrat, A. D. Panopoulos, S. Ruiz, M. L. Wilbert, J. Yu, E. F. Kirkness, J. C. Izpisua Belmonte, D. J. Rossi, J. A. Thomson, K. Eggan, G. Q. Daley, L. S. Goldstein and K. Zhang (2011). "Somatic coding mutations in human induced pluripotent stem cells." Nature **471**(7336): 63-67.

Grahn, R. A., J. C. Grahn, M. C. Penedo, C. R. Helps and L. A. Lyons (2012). "Erythrocyte pyruvate kinase deficiency mutation identified in multiple breeds of domestic cats." BMC Vet Res **8**: 207.

Greenhough, S., C. N. Medine and D. C. Hay (2010). "Pluripotent stem cell derived hepatocyte like cells and their potential in toxicity screening." Toxicology **278**(3): 250-255.

Grizot, S., J. C. Epinat, S. Thomas, A. Duclert, S. Rolland, F. Paques and P. Duchateau (2010). "Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds." Nucleic Acids Res **38**(6): 2006-2018.

Grizot, S., J. Smith, F. Daboussi, J. Prieto, P. Redondo, N. Merino, M. Villate, S. Thomas, L. Lemaire, G. Montoya, F. J. Blanco, F. Paques and P. Duchateau (2009). "Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease." Nucleic Acids Res **37**(16): 5405-5419.

Gurdon, J. B. (2006). "From nuclear transfer to nuclear reprogramming: the reversal of cell differentiation." Annu Rev Cell Dev Biol **22**: 1-22.

Hacein-Bey-Abina, S., A. Garrigue, G. P. Wang, J. Soulier, A. Lim, E. Morillon, E. Clappier, L. Caccavelli, E. Delabesse, K. Beldjord, V. Asnafi, E. MacIntyre, L. Dal Cortivo, I. Radford, N. Brousse, F. Sigaux, D. Moshous, J. Hauer, A. Borkhardt, B. H. Belohradsky, U. Wintergerst, M. C. Velez, L. Leiva, R. Sorensen, N. Wulffraat, S. Blanche, F. D. Bushman, A. Fischer and M. Cavazzana-Calvo (2008). "Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1." J Clin Invest **118**(9): 3132-3142.

Hacein-Bey-Abina, S., C. Von Kalle, M. Schmidt, M. P. McCormack, N. Wulffraat, P. Leboulch, A. Lim, C. S. Osborne, R. Pawliuk, E. Morillon, R. Sorensen, A. Forster, P. Fraser, J. I. Cohen, G. de Saint Basile, I. Alexander, U. Wintergerst, T. Frebourg, A. Aurias, D. Stoppa-Lyonnet, S. Romana, I. Radford-Weiss, F. Gross, F. Valensi, E. Delabesse, E. Macintyre, F. Sigaux, J. Soulier, L. E. Leiva, M. Wissler, C. Prinz, T. H. Rabbitts, F. Le Deist, A. Fischer and M. Cavazzana-Calvo (2003). "LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1." Science **302**(5644): 415-419.

Hanna, J., S. Markoulaki, P. Schorderet, B. W. Carey, C. Beard, M. Wernig, M. P. Creyghton, E. J. Steine, J. P. Cassady, R. Foreman, C. J. Lengner, J. A. Dausman and R. Jaenisch (2008). "Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency." Cell **133**(2): 250-264.

Hanna, J., K. Saha, B. Pando, J. van Zon, C. J. Lengner, M. P. Creyghton, A. van Oudenaarden and R. Jaenisch (2009). "Direct cell reprogramming is a stochastic process amenable to acceleration." Nature **462**(7273): 595-601.

Hansel, M. C., R. Gramignoli, W. Blake, J. Davila, K. Skvorak, K. Dorko, V. Tahan, B. R. Lee, E. Tafaleng, J. Guzman-Lepe, A. Soto-Gutierrez, I. J. Fox and S. C. Strom (2013). "Increased Reprogramming of Human Fetal Hepatocytes Compared With Adult Hepatocytes in Feeder-Free Conditions." Cell Transplant.

Hatzistavrou, T., S. J. Micallef, E. S. Ng, J. Vadolas, E. G. Stanley and A. G. Elefanty (2009). "ErythRED, a hESC line enabling identification of erythroid cells." Nat Methods **6**(9): 659-662.

Heng, J. C., Y. L. Orlov and H. H. Ng (2010). "Transcription factors for the modulation of pluripotency and reprogramming." Cold Spring Harb Symp Quant Biol **75**: 237-244.

Hockemeyer, D., F. Soldner, C. Beard, Q. Gao, M. Mitalipova, R. C. DeKelver, G. E. Katibah, R. Amora, E. A. Boydston, B. Zeitler, X. Meng, J. C. Miller, L. Zhang, E. J. Rebar, P. D. Gregory, F. D. Urnov and R. Jaenisch (2009). "Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases." Nat Biotechnol **27**(9): 851-857.

Hockemeyer, D., H. Wang, S. Kiani, C. S. Lai, Q. Gao, J. P. Cassady, G. J. Cost, L. Zhang, Y. Santiago, J. C. Miller, B. Zeitler, J. M. Cherone, X. Meng, S. J. Hinkley, E. J. Rebar, P. D. Gregory, F. D. Urnov and R. Jaenisch (2011). "Genetic engineering of human pluripotent cells using TALE nucleases." Nat Biotechnol **29**(8): 731-734.

Hopkins, C. M., F. F. White, S. H. Choi, A. Guo and J. E. Leach (1992). "Identification of a family of avirulence genes from Xanthomonas oryzae pv. oryzae." Mol Plant Microbe Interact **5**(6): 451-459.

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

Howden, S. E., A. Gore, Z. Li, H. L. Fung, B. S. Nisler, J. Nie, G. Chen, B. E. McIntosh, D. R. Gulbranson, N. R. Diol, S. M. Taapken, D. T. Vereide, K. D. Montgomery, K. Zhang, D. M. Gamm and J. A. Thomson (2011). "Genetic correction and analysis of induced pluripotent stem cells from a patient with gyrate atrophy." Proc Natl Acad Sci U S A **108**(16): 6537-6542.

Hu, K., J. Yu, K. Suknuntha, S. Tian, K. Montgomery, K. D. Choi, R. Stewart, J. A. Thomson and Slukvin, II (2011). "Efficient generation of transgene-free induced pluripotent stem cells from normal and neoplastic bone marrow and cord blood mononuclear cells." Blood **117**(14): e109-119.

Hua, H., L. Shang, H. Martinez, M. Freeby, M. P. Gallagher, T. Ludwig, L. Deng, E. Greenberg, C. Leduc, W. K. Chung, R. Goland, R. L. Leibel and D. Egli (2013). "iPSC-derived beta cells model diabetes due to glucokinase deficiency." J Clin Invest **123**(7): 3146-3153.

Hussein, S. M., N. N. Batada, S. Vuoristo, R. W. Ching, R. Autio, E. Narva, S. Ng, M. Sourour, R. Hamalainen, C. Olsson, K. Lundin, M. Mikkola, R. Trokovic, M. Peitz, O. Brustle, D. P. Bazett-Jones, K. Alitalo, R. Lahesmaa, A. Nagy and T. Otonkoski (2011). "Copy number variation and selection during reprogramming to pluripotency." Nature **471**(7336): 58-62.

Hwang, W. Y., Y. Fu, D. Reyon, M. L. Maeder, P. Kaini, J. D. Sander, J. K. Joung, R. T. Peterson and J. R. Yeh (2013). "Heritable and Precise Zebrafish Genome Editing Using a CRISPR-Cas System." PLoS One **8**(7): e68708.

Hwang, W. Y., Y. Fu, D. Reyon, M. L. Maeder, S. Q. Tsai, J. D. Sander, R. T. Peterson, J. R. Yeh and J. K. Joung (2013). "Efficient genome editing in zebrafish using a CRISPR-Cas system." Nat Biotechnol **31**(3): 227-229.

Irion, S., H. Luche, P. Gadue, H. J. Fehling, M. Kennedy and G. Keller (2007). "Identification and targeting of the ROSA26 locus in human embryonic stem cells." Nat Biotechnol **25**(12): 1477-1482.

Jao, L. E., S. R. Wente and W. Chen (2013). "Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system." Proc Natl Acad Sci U S A **110**(34): 13904-13909.

Jasin, M. (1996). "Genetic manipulation of genomes with rare-cutting endonucleases." Trends Genet **12**(6): 224-228.

Jia, F., K. D. Wilson, N. Sun, D. M. Gupta, M. Huang, Z. Li, N. J. Panetta, Z. Y. Chen, R. C. Robbins, M. A. Kay, M. T. Longaker and J. C. Wu (2010). "A nonviral minicircle vector for deriving human iPS cells." Nat Methods **7**(3): 197-199.

Jiang, J., Y. Jing, G. J. Cost, J. C. Chiang, H. J. Kolpa, A. M. Cotton, D. M. Carone, B. R. Carone, D. A. Shivak, D. Y. Guschin, J. R. Pearl, E. J. Rebar, M. Byron, P. D. Gregory, C. J. Brown, F. D. Urnov, L. L. Hall and J. B. Lawrence (2013). "Translating dosage compensation to trisomy 21." Nature **500**(7462): 296-300.

Jiang, W., H. Zhou, H. Bi, M. Fromm, B. Yang and D. P. Weeks (2013). "Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice." Nucleic Acids Res.

Ju, Y. S., J. I. Kim, S. Kim, D. Hong, H. Park, J. Y. Shin, S. Lee, W. C. Lee, S. Kim, S. B. Yu, S. S. Park, S. H. Seo, J. Y. Yun, H. J. Kim, D. S. Lee, M. Yavartanoo, H. P. Kang, O. Gokcumen, D. R. Govindaraju, J. H. Jung, H. Chong, K. S. Yang, H. Kim, C. Lee and J. S. Seo (2011). "Extensive genomic and transcriptional diversity identified through massively parallel DNA and RNA sequencing of eighteen Korean individuals." Nat Genet **43**(8): 745-752.

Kaji, K., K. Norrby, A. Paca, M. Mileikovsky, P. Mohseni and K. Woltjen (2009). "Virus-free induction of pluripotency and subsequent excision of reprogramming factors." Nature **458**(7239): 771-775.

Kanno, H., H. Fujii and S. Miwa (1992). "Structural analysis of human pyruvate kinase L-gene and identification of the promoter activity in erythroid cells." Biochem Biophys Res Commun **188**(2): 516-523.

Kedar, P., T. Hamada, P. Warang, A. Nadkarni, K. Shimizu, H. Fujji, K. Ghosh, H. Kanno and R. Colah (2009). "Spectrum of novel mutations in the human PKLR gene in pyruvate kinase-deficient Indian patients with heterogeneous clinical phenotypes." Clin Genet **75**(2): 157-162.

Kim, D., C. H. Kim, J. I. Moon, Y. G. Chung, M. Y. Chang, B. S. Han, S. Ko, E. Yang, K. Y. Cha, R. Lanza and K. S. Kim (2009). "Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins." Cell Stem Cell **4**(6): 472-476.

Kim, J. B., H. Zaehres, M. J. Arauzo-Bravo and H. R. Scholer (2009). "Generation of induced pluripotent stem cells from neural stem cells." Nat Protoc **4**(10): 1464-1470.

Kim, Y. G., J. Cha and S. Chandrasegaran (1996). "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain." Proc Natl Acad Sci U S A **93**(3): 1156-1160.

Kobari, L., F. Yates, N. Oudrhiri, A. Francina, L. Kiger, C. Mazurier, S. Rouzbeh, W. El-Nemer, N. Hebert, M. C. Giarratana, S. Francois, A. Chapel, H. Lapillonne, D. Luton, A. Bennaceur-Griscelli and L. Douay (2012). "Human induced pluripotent stem cells can reach complete terminal maturation: in vivo and in vitro evidence in the erythropoietic differentiation model." Haematologica **97**(12): 1795-1803.

Kohn, B. and C. Fumi (2008). "Clinical course of pyruvate kinase deficiency in Abyssinian and Somali cats." J Feline Med Surg **10**(2): 145-153.

Kondo, T., M. Asai, K. Tsukita, Y. Kutoku, Y. Ohsawa, Y. Sunada, K. Imamura, N. Egawa, N. Yahata, K. Okita, K. Takahashi, I. Asaka, T. Aoi, A. Watanabe, K. Watanabe, C. Kadoya, R. Nakano, D. Watanabe, K. Maruyama, O. Hori, S. Hibino, T. Choshi, T. Nakahata, H. Hioki, T. Kaneko, M. Naitoh, K. Yoshikawa, S. Yamawaki, S. Suzuki, R.

Hata, S. I. Ueno, T. Seki, K. Kobayashi, T. Toda, K. Murakami, K. Irie, W. L. Klein, H. Mori, T. Asada, R. Takahashi, N. Iwata, S. Yamanaka and H. Inoue (2013). "Modeling Alzheimer's Disease with iPSCs Reveals Stress Phenotypes Associated with Intracellular Abeta and Differential Drug Responsiveness." Cell Stem Cell.

Kunisato, A., M. Wakatsuki, H. Shinba, T. Ota, I. Ishida and K. Nagao (2010). "Direct generation of induced pluripotent stem cells from human nonmobilized blood." Stem Cells Dev **20**(1): 159-168.

Kuroya, M. and N. Ishida (1953). "Newborn virus pneumonitis (type Sendai). II. The isolation of a new virus possessing hemagglutinin activity." Yokohama Med Bull **4**(4): 217-233.

Lapillonne, H., L. Kobari, C. Mazurier, P. Tropel, M. C. Giarratana, I. Zanella-Cleon, L. Kiger, M. Wattenhofer-Donze, H. Puccio, N. Hebert, A. Francina, G. Andreu, S. Viville and L. Douay (2010). "Red blood cell generation from human induced pluripotent stem cells: perspectives for transfusion medicine." Haematologica **95**(10): 1651-1659.

Laurent, L. C., I. Ulitsky, I. Slavin, H. Tran, A. Schork, R. Morey, C. Lynch, J. V. Harness, S. Lee, M. J. Barrero, S. Ku, M. Martynova, R. Semechkin, V. Galat, J. Gottesfeld, J. C. Izpisua Belmonte, C. Murry, H. S. Keirstead, H. S. Park, U. Schmidt, A. L. Laslett, F. J. Muller, C. M. Nievergelt, R. Shamir and J. F. Loring (2011). "Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture." Cell Stem Cell **8**(1): 106-118.

Li, H., M. Collado, A. Villasante, K. Strati, S. Ortega, M. Canamero, M. A. Blasco and M. Serrano (2009). "The Ink4/Arf locus is a barrier for iPS cell reprogramming." Nature **460**(7259): 1136-1139.

Li, T., S. Huang, X. Zhao, D. A. Wright, S. Carpenter, M. H. Spalding, D. P. Weeks and B. Yang (2011). "Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes." Nucleic Acids Res **39**(14): 6315-6325.

Liang, G. and Y. Zhang (2013). "Genetic and Epigenetic Variations in iPSCs: Potential Causes and Implications for Application." Cell Stem Cell **13**(2): 149-159.

Liu, G. H., K. Suzuki, J. Qu, I. Sancho-Martinez, F. Yi, M. Li, S. Kumar, E. Nivet, J. Kim, R. D. Soligalla, I. Dubova, A. Goebl, N. Plongthongkum, H. L. Fung, K. Zhang, J. F. Loring, L. C. Laurent and J. C. Izpisua Belmonte (2011). "Targeted gene correction of laminopathy-associated LMNA mutations in patient-specific iPSCs." Cell Stem Cell **8**(6): 688-694.

Liu, J., H. Sumer, J. Leung, K. Upton, M. Dottori, A. Pebay and P. J. Verma (2010). "Late passage human fibroblasts induced to pluripotency are capable of directed neuronal differentiation." Cell Transplant **20**(2): 193-203.

Liu, T., G. Zou, Y. Gao, X. Zhao, H. Wang, Q. Huang, L. Jiang, L. Guo and W. Cheng (2012). "High Efficiency of Reprogramming CD34(+) Cells Derived from Human Amniotic Fluid into Induced Pluripotent Stem Cells with Oct4." Stem Cells Dev.

Liu, Y., D. Cheng, Z. Li, X. Gao and H. Wang (2012). "The gene expression profiles of induced pluripotent stem cells (iPSCs) generated by a non-integrating method are more similar to embryonic stem cells than those of iPSCs generated by an integrating method." Genet Mol Biol **35**(3): 693-700.

Lombardo, A., P. Genovese, C. M. Beausejour, S. Colleoni, Y. L. Lee, K. A. Kim, D. Ando, F. D. Urnov, C. Galli, P. D. Gregory, M. C. Holmes and L. Naldini (2007). "Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery." Nat Biotechnol **25**(11): 1298-1306.

Loring, J. F. and M. S. Rao (2006). "Establishing standards for the characterization of human embryonic stem cell lines." Stem Cells **24**(1): 145-150.

Lu, S. J., Q. Feng, S. Caballero, Y. Chen, M. A. Moore, M. B. Grant and R. Lanza (2007). "Generation of functional hemangioblasts from human embryonic stem cells." Nat Methods **4**(6): 501-509.

Lu, S. J., Q. Feng, J. S. Park, L. Vida, B. S. Lee, M. Strausbauch, P. J. Wettstein, G. R. Honig and R. Lanza (2008). "Biologic properties and enucleation of red blood cells from human embryonic stem cells." Blood **112**(12): 4475-4484.

Luong, M. X., J. Auerbach, J. M. Crook, L. Daheron, D. Hei, G. Lomax, J. F. Loring, T. Ludwig, T. M. Schlaeger, K. P. Smith, G. Stacey, R. H. Xu and F. Zeng (2011). "A call for standardized naming and reporting of human ESC and iPSC lines." Cell Stem Cell **8**(4): 357-359.

Macarthur, C. C., A. Fontes, N. Ravinder, D. Kuninger, J. Kaur, M. Bailey, A. Taliana, M. C. Vemuri and P. T. Lieu (2012). "Generation of human-induced pluripotent stem cells by a nonintegrating RNA Sendai virus vector in feeder-free or xeno-free conditions." Stem Cells Int **2012**: 564612.

Machado, P., L. Manco, C. Gomes, C. Mendes, N. Fernandes, G. Salome, L. Sitoe, S. Chibute, J. Langa, L. Ribeiro, J. Miranda, J. Cano, J. Pinto, A. Amorim, V. E. do Rosario and A. P. Arez (2012). "Pyruvate kinase deficiency in sub-Saharan Africa: identification of a highly frequent missense mutation (G829A;Glu277Lys) and association with malaria." PLoS One **7**(10): e47071.

Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville and G. M. Church (2013). "RNA-guided human genome engineering via Cas9." Science **339**(6121): 823-826.

Manco, L., M. L. Ribeiro, H. Almeida, O. Freitas, A. Abade and G. Tamagnini (1999). "PK-LR gene mutations in pyruvate kinase deficient Portuguese patients." Br J Haematol **105**(3): 591-595.

Marion, R. M., K. Strati, H. Li, M. Murga, R. Blanco, S. Ortega, O. Fernandez-Capetillo, M. Serrano and M. A. Blasco (2009). "A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity." Nature **460**(7259): 1149-1153.

Martins-Taylor, K., B. S. Nisler, S. M. Taapken, T. Compton, L. Crandall, K. D. Montgomery, M. Lalande and R. H. Xu (2011). "Recurrent copy number variations in human induced pluripotent stem cells." Nat Biotechnol **29**(6): 488-491.

Martins-Taylor, K. and R. H. Xu (2012). "Concise review: Genomic stability of human induced pluripotent stem cells." Stem Cells **30**(1): 22-27.

Masaki, I., Y. Yonemitsu, K. Komori, H. Ueno, Y. Nakashima, K. Nakagawa, M. Fukumura, A. Kato, M. K. Hasan, Y. Nagai, K. Sugimachi, M. Hasegawa and K. Sueishi (2001). "Recombinant Sendai virus-mediated gene transfer to vasculature: a new class of efficient gene transfer vector to the vascular system." FASEB J **15**(7): 1294-1296.

Masaki, I., Y. Yonemitsu, A. Yamashita, S. Sata, M. Tanii, K. Komori, K. Nakagawa, X. Hou, Y. Nagai, M. Hasegawa, K. Sugimachi and K. Sueishi (2002). "Angiogenic gene therapy for experimental critical limb ischemia: acceleration of limb loss by overexpression of vascular endothelial growth factor 165 but not of fibroblast growth factor-2." Circ Res **90**(9): 966-973.

Meng, X., A. Neises, R. J. Su, K. J. Payne, L. Ritter, D. S. Gridley, J. Wang, M. Sheng, K. H. Lau, D. J. Baylink and X. B. Zhang (2012). "Efficient reprogramming of human cord blood CD34+ cells into induced pluripotent stem cells with OCT4 and SOX2 alone." Mol Ther **20**(2): 408-416.

Mentzer, W. C., Jr., R. L. Baehner, H. Schmidt-Schonbein, S. H. Robinson and D. G. Nathan (1971). "Selective reticulocyte destruction in erythrocyte pyruvate kinase deficiency." J Clin Invest **50**(3): 688-699.

Merling, R. K., C. L. Sweeney, U. Choi, S. S. De Ravin, T. G. Myers, F. Otaizo-Carrasquero, J. Pan, G. Linton, L. Chen, S. Koontz, N. L. Theobald and H. L. Malech (2013). "Transgene-free iPSCs generated from small volume peripheral blood nonmobilized CD34+ cells." Blood **121**(14): e98-107.

Meza, N. W., M. E. Alonso-Ferrero, S. Navarro, O. Quintana-Bustamante, A. Valeri, M. Garcia-Gomez, J. A. Bueren, J. M. Bautista and J. C. Segovia (2009). "Rescue of pyruvate kinase deficiency in mice by gene therapy using the human isoenzyme." Mol Ther **17**(12): 2000-2009.

Meza, N. W., O. Quintana-Bustamante, A. Puyet, P. Rio, S. Navarro, A. Diez, J. A. Bueren, J. M. Bautista and J. C. Segovia (2007). "In vitro and in vivo expression of human erythrocyte pyruvate kinase in erythroid cells: a gene therapy approach." Hum Gene Ther **18**(6): 502-514.

Miao, J., D. Guo, J. Zhang, Q. Huang, G. Qin, X. Zhang, J. Wan, H. Gu and L. J. Qu (2013). "Targeted mutagenesis in rice using CRISPR-Cas system." Cell Res.

Migliaccio, A. R., C. Whitsett, T. Papayannopoulou and M. Sadelain (2012). "The potential of stem cells as an in vitro source of red blood cells for transfusion." Cell Stem Cell **10**(2): 115-119.

Min-Oo, G., A. Fortin, M. F. Tam, P. Gros and M. M. Stevenson (2004). "Phenotypic expression of pyruvate kinase deficiency and protection against malaria in a mouse model." Genes Immun **5**(3): 168-175.

Molina-Estevez, F. J., M. L. Lozano, S. Navarro, Y. Torres, I. Grabundzija, Z. Ivics, E. Samper, J. A. Bueren and G. Guenechea (2013). "Impaired Cell Reprogramming in Non-Homologous End Joining Deficient Cells." Stem Cells.

Morimoto, M., H. Kanno, H. Asai, T. Tsujimura, H. Fujii, Y. Moriyama, T. Kasugai, A. Hirono, Y. Ohba, S. Miwa and Y. Kitamura (1995). "Pyruvate kinase deficiency of mice associated with nonspherocytic hemolytic anemia and cure of the anemia by marrow transplantation without host irradiation." Blood **86**(11): 4323-4330.

Muller, F. J., B. M. Schuldt, R. Williams, D. Mason, G. Altun, E. P. Papapetrou, S. Danner, J. E. Goldmann, A. Herbst, N. O. Schmidt, J. B. Aldenhoff, L. C. Laurent and J. F. Loring (2011). "A bioinformatic assay for pluripotency in human cells." Nat Methods **8**(4): 315-317.

Mussolino, C., R. Morbitzer, F. Lutge, N. Dannemann, T. Lahaye and T. Cathomen (2011). "A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity." Nucleic Acids Res **39**(21): 9283-9293.

Nakamura, N., K. Saeki, M. Mitsumoto, S. Matsuyama, M. Nishio, K. Saeki, M. Hasegawa, Y. Miyagawa, H. Ohkita, N. Kiyokawa, M. Toyoda, H. Akutsu, A. Umezawa and A. Yuo (2012). "Feeder-free and serum-free production of hepatocytes, cholangiocytes, and their proliferating progenitors from human pluripotent stem cells: application to liver-specific functional and cytotoxic assays." Cell Reprogram **14**(2): 171-185.

Natale, A., A. Frangipane di Regalbono, G. Zanellato, M. Cavalletto, P. Danesi, G. Capelli and M. Pietrobelli (2007). "Parasitological survey on stray cat colonies from the Veneto Region." Vet Res Commun **31 Suppl 1**: 241-244.

Nishimura, T., S. Kaneko, A. Kawana-Tachikawa, Y. Tajima, H. Goto, D. Zhu, K. Nakayama-Hosoya, S. Iriguchi, Y. Uemura, T. Shimizu, N. Takayama, D. Yamada, K. Nishimura, M. Ohtaka, N. Watanabe, S. Takahashi, A. Iwamoto, H. Koseki, M. Nakanishi, K. Eto and H. Nakauchi (2013). "Generation of rejuvenated antigen-specific T cells by reprogramming to pluripotency and redifferentiation." Cell Stem Cell **12**(1): 114-126.

Palis, J., J. Malik, K. E. McGrath and P. D. Kingsley (2010). "Primitive erythropoiesis in the mammalian embryo." Int J Dev Biol **54**(6-7): 1011-1018.

Papapetrou, E. P., G. Lee, N. Malani, M. Setty, I. Riviere, L. M. Tirunagari, K. Kadota, S. L. Roth, P. Giardina, A. Viale, C. Leslie, F. D. Bushman, L. Studer and M. Sadelain (2011). "Genomic safe harbors permit high beta-globin transgene expression in thalassemia induced pluripotent stem cells." Nat Biotechnol **29**(1): 73-78.

Papapetrou, E. P. and M. Sadelain (2011). "Generation of transgene-free human induced pluripotent stem cells with an excisable single polycistronic vector." Nat Protoc **6**(9): 1251-1273.

Paques, F. and P. Duchateau (2007). "Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy." Curr Gene Ther **7**(1): 49-66.

Park, I. H., N. Arora, H. Huo, N. Maherali, T. Ahfeldt, A. Shimamura, M. W. Lensch, C. Cowan, K. Hochedlinger and G. Q. Daley (2008). "Disease-specific induced pluripotent stem cells." Cell **134**(5): 877-886.

Pattanayak, V., C. L. Ramirez, J. K. Joung and D. R. Liu (2011). "Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection." Nat Methods **8**(9): 765-770.

Perez, E. E., J. Wang, J. C. Miller, Y. Jouvenot, K. A. Kim, O. Liu, N. Wang, G. Lee, V. V. Bartsevich, Y. L. Lee, D. Y. Guschin, I. Rupniewski, A. J. Waite, C. Carpenito, R. G. Carroll, J. S. Orange, F. D. Urnov, E. J. Rebar, D. Ando, P. D. Gregory, J. L. Riley, M. C. Holmes and C. H. June (2008). "Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases." Nat Biotechnol **26**(7): 808-816.

Popplewell, L., T. Koo, X. Leclerc, A. Duclert, K. Mamchaoui, A. Gouble, V. Mouly, T. Voit, F. Paques, F. Cedrone, O. Isman, R. J. Yanez-Munoz and G. Dickson (2013). "Gene correction of a duchenne muscular dystrophy mutation by meganuclease-enhanced exon knock-in." Hum Gene Ther **24**(7): 692-701.

Porteus, M. H. and D. Carroll (2005). "Gene targeting using zinc finger nucleases." Nat Biotechnol **23**(8): 967-973.

Qin, H., K. Blaschke, G. Wei, Y. Ohi, L. Blouin, Z. Qi, J. Yu, R. F. Yeh, M. Hebrok and M. Ramalho-Santos (2012). "Transcriptional analysis of pluripotency reveals the Hippo pathway as a barrier to reprogramming." Hum Mol Genet **21**(9): 2054-2067.

Qin, J., B. Song, H. Zhang, Y. Wang, N. Wang, Y. Ji, J. Qi, A. Chandra, B. Yang, Y. Zhang, G. Gong and Y. Xu (2013). "Transplantation of human neuro-epithelial-like stem cells derived from induced pluripotent stem cells improves neurological function in rats with experimental intracerebral hemorrhage." Neurosci Lett.

Rais, Y., A. Zviran, S. Geula, O. Gafni, E. Chomsky, S. Viukov, A. A. Mansour, I. Caspi, V. Krupalnik, M. Zerbib, I. Maza, N. Mor, D. Baran, L. Weinberger, D. A. Jaitin, D. Lara-Astiaso, R. Blecher-Gonen, Z. Shipony, Z. Mukamel, T. Hagai, S. Gilad, D. Amann-Zalcenstein, A. Tanay, I. Amit, N. Novershtern and J. H. Hanna (2013). "Deterministic direct reprogramming of somatic cells to pluripotency." Nature.

Ramos-Mejia, V., R. Montes, C. Bueno, V. Ayllon, P. J. Real, R. Rodriguez and P. Menendez (2012). "Residual expression of the reprogramming factors prevents differentiation of iPSC generated from human fibroblasts and cord blood CD34+ progenitors." PLoS One **7**(4): e35824.

Raya, A., I. Rodriguez-Piza, G. Guenechea, R. Vassena, S. Navarro, M. J. Barrero, A. Consiglio, M. Castella, P. Rio, E. Sleep, F. Gonzalez, G. Tiscornia, E. Garreta, T. Aasen, A. Veiga, I. M. Verma, J. Surralles, J. Bueren and J. C. Izpisua Belmonte (2009). "Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells." Nature **460**(7251): 53-59.

Richardson, C. and M. Jasin (2000). "Frequent chromosomal translocations induced by DNA double-strand breaks." Nature **405**(6787): 697-700.

Robbins, J. (1993). "Gene targeting. The precise manipulation of the mammalian genome." Circ Res **73**(1): 3-9.

Robinton, D. A. and G. Q. Daley (2012). "The promise of induced pluripotent stem cells in research and therapy." Nature **481**(7381): 295-305.

Rothberg, J. M., W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. A. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R.

Kasinskas, T. Sokolsky, J. A. Fidanza, E. Namsaraev, K. J. McKernan, A. Williams, G. T. Roth and J. Bustillo (2011). "An integrated semiconductor device enabling non-optical genome sequencing." Nature **475**(7356): 348-352.

Rouet, P., F. Smih and M. Jasin (1994). "Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells." Proc Natl Acad Sci U S A **91**(13): 6064-6068.

Rouger, H., C. Valentin, C. T. Craescu, F. Galacteros and M. Cohen-Solal (1996). "Five unknown mutations in the LR pyruvate kinase gene associated with severe hereditary nonspherocytic haemolytic anaemia in France." Br J Haematol **92**(4): 825-830.

Sebastiano, V., M. L. Maeder, J. F. Angstman, B. Haddad, C. Khayter, D. T. Yeo, M. J. Goodwin, J. S. Hawkins, C. L. Ramirez, L. F. Batista, S. E. Artandi, M. Wernig and J. K. Joung (2011). "In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases." Stem Cells **29**(11): 1717-1726.

Sedano, I. B., B. Rothlisberger, G. Deleze, C. Ottiger, M. A. Panchard, A. Spahr, M. Hergersberg, W. Burgi and A. Huber (2004). "PK Aarau: first homozygous nonsense mutation causing pyruvate kinase deficiency." Br J Haematol **127**(3): 364-366.

Seki, T., S. Yuasa and K. Fukuda (2011). "Derivation of induced pluripotent stem cells from human peripheral circulating T cells." Curr Protoc Stem Cell Biol **Chapter 4**: Unit4A 3.

Shijia Chen, G. O., Cindy N. Chiu, Brett J. Niles, Justin Liu, Daniel A. Lee, Igor Antoshechkin and David A. Prober (2013). "A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly
" Nucleic Acid Research **41**(4): 2769-2778.

Silva, E. A., F. Flexa and M. T. Zanella (2007). "Abdominal obesity, insulin resistance and hypertension: impact on left ventricular mass and function in women." Arq Bras Cardiol **89**(2): 77-82, 86-92.

Silva, G. H., M. Belfort, W. Wende and A. Pingoud (2006). "From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes." J Mol Biol **361**(4): 744-754.

Slobod, K. S., J. L. Shenep, J. Lujan-Zilbermann, K. Allison, B. Brown, R. A. Scroggs, A. Portner, C. Coleclough and J. L. Hurwitz (2004). "Safety and immunogenicity of intranasal murine parainfluenza virus type 1 (Sendai virus) in healthy human adults." Vaccine **22**(23-24): 3182-3186.

Smithies, O. (2001). "Forty years with homologous recombination." Nat Med **7**(10): 1083-1086.

Soldner, F., J. Laganiere, A. W. Cheng, D. Hockemeyer, Q. Gao, R. Alagappan, V. Khurana, L. I. Golbe, R. H. Myers, S. Lindquist, L. Zhang, D. Guschin, L. K. Fong, B. J. Vu, X. Meng, F. D. Urnov, E. J. Rebar, P. D. Gregory, H. S. Zhang and R. Jaenisch (2011). "Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations." Cell **146**(2): 318-331.

Somers, A., J. C. Jean, C. A. Sommer, A. Omari, C. C. Ford, J. A. Mills, L. Ying, A. G. Sommer, J. M. Jean, B. W. Smith, R. Lafyatis, M. F. Demierre, D. J. Weiss, D. L. French, P. Gadue, G. J. Murphy, G. Mostoslavsky and D. N. Kotton (2010). "Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette." Stem Cells **28**(10): 1728-1740.

Sommer, A. G., S. S. Rozelle, S. Sullivan, J. A. Mills, S. M. Park, B. W. Smith, A. M. Iyer, D. L. French, D. N. Kotton, P. Gadue, G. J. Murphy and G. Mostoslavsky (2012). "Generation of human induced pluripotent stem cells from peripheral blood using the STEMCCA lentiviral vector." J Vis Exp(68).

Sommer, C. A., C. Christodoulou, A. Gianotti-Sommer, S. S. Shen, B. S. Sailaja, H. Hezroni, A. Spira, E. Meshorer, D. N. Kotton and G. Mostoslavsky (2012). "Residual expression of reprogramming factors affects the transcriptional program and epigenetic signatures of induced pluripotent stem cells." PLoS One **7**(12): e51711.

Sommer, C. A., M. Stadtfeld, G. J. Murphy, K. Hochedlinger, D. N. Kotton and G. Mostoslavsky (2009). "Induced pluripotent stem cell generation using a single lentiviral stem cell cassette." Stem Cells **27**(3): 543-549.

Staerk, J., M. M. Dawlaty, Q. Gao, D. Maetzel, J. Hanna, C. A. Sommer, G. Mostoslavsky and R. Jaenisch (2010). "Reprogramming of human peripheral blood cells to induced pluripotent stem cells." Cell Stem Cell **7**(1): 20-24.

Stoddard, B. L. (2005). "Homing endonuclease structure and function." Q Rev Biophys **38**(1): 49-95.

Sun, N., N. J. Panetta, D. M. Gupta, K. D. Wilson, A. Lee, F. Jia, S. Hu, A. M. Cherry, R. C. Robbins, M. T. Longaker and J. C. Wu (2009). "Feeder-free derivation of induced pluripotent stem cells from adult human adipose stem cells." Proc Natl Acad Sci U S A **106**(37): 15720-15725.

Suvatte, V., V. S. Tanphaichitr, S. Visuthisakchai, C. Mahasandana, G. Veerakul, V. Chongkolwatana, D. Chandanayingyong and S. Issaragrisil (1998). "Bone marrow, peripheral blood and cord blood stem cell transplantation in children: ten years' experience at Siriraj Hospital." Int J Hematol **68**(4): 411-419.

Svendsen, C. N. (2013). "Back to the future: how human induced pluripotent stem cells will transform regenerative medicine." Hum Mol Genet.

Swaroop, M., N. Thorne, M. S. Rao, C. P. Austin, J. C. McKew and W. Zheng (2012). "Evaluation of cholesterol reduction activity of methyl-beta-cyclodextrin using differentiated human neurons and astrocytes." J Biomol Screen **17**(9): 1243-1251.

Taapken, S. M., B. S. Nisler, M. A. Newton, T. L. Sampsell-Barron, K. A. Leonhard, E. M. McIntire and K. D. Montgomery (2011). "Karotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells." Nat Biotechnol **29**(4): 313-314.

Takahashi, K., K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda and S. Yamanaka (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." Cell **131**(5): 861-872.

Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-676.

Tanaka, T., K. Takahashi, M. Yamane, S. Tomida, S. Nakamura, K. Oshima, A. Niwa, R. Nishikomori, N. Kambe, H. Hara, M. Mitsuyama, N. Morone, J. E. Heuser, T. Yamamoto, A. Watanabe, A. Sato-Otsubo, S. Ogawa, I. Asaka, T. Heike, S. Yamanaka, T. Nakahata and M. K. Saito (2012). "Induced pluripotent stem cells from CINCA syndrome patients as a model for dissecting somatic mosaicism and drug discovery." Blood.

Tanphaichitr, V. S., V. Suvatte, S. Issaragrisil, C. Mahasandana, G. Veerakul, V. Chongkolwatana, W. Waiyawuth and H. Ideguchi (2000). "Successful bone marrow transplantation in a child with red blood cell pyruvate kinase deficiency." Bone Marrow Transplant **26**(6): 689-690.

te Riele, H., E. R. Maandag and A. Berns (1992). "Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs." Proc Natl Acad Sci U S A **89**(11): 5128-5132.

Themeli, M., C. C. Kloss, G. Ciriello, V. D. Fedorov, F. Perna, M. Gonen and M. Sadelain (2013). "Generation of tumor-targeted human T lymphocytes from induced pluripotent stem cells for cancer therapy." Nat Biotechnol.

Thier, M., B. Munst and F. Edenhofer (2010). "Exploring refined conditions for reprogramming cells by recombinant Oct4 protein." Int J Dev Biol **54**(11-12): 1713-1721.

Thomson, J. A., J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall and J. M. Jones (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**(5391): 1145-1147.

Tilgner, K., I. Neganova, I. Moreno-Gimeno, J. Y. Al-Aama, D. Burks, S. Yung, C. Singhapol, G. Saretzki, J. Evans, V. Gorbunova, A. Gennery, S. Przyborski, M. Stojkovic, L. Armstrong, P. Jeggo and M. Lako (2013). "A human iPSC model of Ligase IV deficiency reveals an important role for NHEJ-mediated-DSB repair in the survival and genomic stability of induced pluripotent stem cells and emerging haematopoietic progenitors." Cell Death Differ **20**(8): 1089-1100.

Torres, R., A. Garcia, M. Paya and J. C. Ramirez (2011). "Non-integrative lentivirus drives high-frequency cre-mediated cassette exchange in human cells." PLoS One **6**(5): e19794.

Trobridge, G. D., B. C. Beard, R. A. Wu, C. Ironside, P. Malik and H. P. Kiem (2012). "Stem cell selection in vivo using foamy vectors cures canine pyruvate kinase deficiency." PLoS One **7**(9): e45173.

Tsutsumi, H., K. Tani, H. Fujii and S. Miwa (1988). "Expression of L- and M-type pyruvate kinase in human tissues." Genomics **2**(1): 86-89.

Tubsuwan, A., S. Abed, A. Deichmann, M. D. Kardel, C. Bartholoma, A. Cheung, O. Negre, Z. Kadri, S. Fucharoen, C. von Kalle, E. Payen, S. Chretien, M. Schmidt, C. J. Eaves, P. Leboulch and L. Maouche-Chretien (2013). "Parallel Assessment of Globin Lentiviral Transfer in Induced Pluripotent Stem Cells and Adult Hematopoietic Stem Cells Derived from the Same Transplanted beta-Thalassemia Patient." Stem Cells.

Valentine, W. N., K. R. Tanaka and S. Miwa (1961). "A specific erythrocyte glycolytic enzyme defect (pyruvate kinase) in three subjects with congenital non-spherocytic hemolytic anemia." Trans Assoc Am Physicians **74**: 100-110.

Valton, J., F. Daboussi, S. Leduc, R. Molina, P. Redondo, R. Macmaster, G. Montoya and P. Duchateau (2012). "5'-Cytosine-phosphoguanine (CpG) methylation impacts the activity of natural and engineered meganucleases." J Biol Chem **287**(36): 30139-30150.

van Dongen, J. J., A. W. Langerak, M. Bruggemann, P. A. Evans, M. Hummel, F. L. Lavender, E. Delabesse, F. Davi, E. Schuuring, R. Garcia-Sanz, J. H. van Krieken, J. Droese, D. Gonzalez, C. Bastard, H. E. White, M. Spaargaren, M. Gonzalez, A. Parreira, J. L. Smith, G. J. Morgan, M. Kneba and E. A. Macintyre (2003). "Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936." Leukemia **17**(12): 2257-2317.

van Wijk, R., E. G. Huizinga, A. C. van Wesel, B. A. van Oirschot, M. A. Hadders and W. W. van Solinge (2009). "Fifteen novel mutations in PKLR associated with pyruvate kinase (PK) deficiency: structural implications of amino acid substitutions in PK." Hum Mutat **30**(3): 446-453.

Vizcardo, R., K. Masuda, D. Yamada, T. Ikawa, K. Shimizu, S. Fujii, H. Koseki and H. Kawamoto (2013). "Regeneration of human tumor antigen-specific T cells from iPSCs derived from mature CD8(+) T cells." Cell Stem Cell **12**(1): 31-36.

Wakao, H., K. Yoshikiyo, U. Koshimizu, T. Furukawa, K. Enomoto, T. Matsunaga, T. Tanaka, Y. Yasutomi, T. Yamada, H. Minakami, J. Tanaka, A. Oda, T. Sasaki, R. Wakao, O. Lantz, T. Udagawa, Y. Sekiya, K. Higuchi, N. Harada, K. Nishimura, M. Ohtaka, M. Nakanishi and H. Fujita (2013). "Expansion of functional human mucosal-associated invariant T cells via reprogramming to pluripotency and redifferentiation." Cell Stem Cell **12**(5): 546-558.

Wang, H., H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng, F. Zhang and R. Jaenisch (2013). "One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering." Cell **153**(4): 910-918.

Wang, Y., C. G. Zheng, Y. Jiang, J. Zhang, J. Chen, C. Yao, Q. Zhao, S. Liu, K. Chen, J. Du, Z. Yang and S. Gao (2012). "Genetic correction of beta-thalassemia patient-specific iPS cells and its use in improving hemoglobin production in irradiated SCID mice." Cell Res **22**(4): 637-648.

Warren, L., P. D. Manos, T. Ahfeldt, Y. H. Loh, H. Li, F. Lau, W. Ebina, P. K. Mandal, Z. D. Smith, A. Meissner, G. Q. Daley, A. S. Brack, J. J. Collins, C. Cowan, T. M. Schlaeger and D. J. Rossi (2010). "Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA." Cell Stem Cell **7**(5): 618-630.

Weidauer, S., H. Vatter, J. Beck, A. Raabe, H. Lanfermann, V. Seifert and F. Zanella (2008). "Focal laminar cortical infarcts following aneurysmal subarachnoid haemorrhage." Neuroradiology **50**(1): 1-8.

Whitney, K. M. and C. D. Lothrop, Jr. (1995). "Genetic test for pyruvate kinase deficiency of Basenjis." J Am Vet Med Assoc **207**(7): 918-921.

Wiehe, J. M., P. Ponsaerts, M. T. Rojewski, J. M. Homann, J. Greiner, D. Kronawitter, H. Schrezenmeier, V. Hombach, M. Wiesneth, O. Zimmermann and J. Torzewski (2007). "mRNA-mediated gene delivery into human progenitor cells promotes highly efficient protein expression." J Cell Mol Med **11**(3): 521-530.

Woltjen, K., R. Hamalainen, M. Kibschull, M. Mileikovsky and A. Nagy (2011). "Transgene-free production of pluripotent stem cells using piggyBac transposons." Methods Mol Biol **767**: 87-103.

Xiao, A., Z. Wang, Y. Hu, Y. Wu, Z. Luo, Z. Yang, Y. Zu, W. Li, P. Huang, X. Tong, Z. Zhu, S. Lin and B. Zhang (2013). "Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish." Nucleic Acids Res **41**(14): e141.

Yang, H., H. Wang, C. S. Shivalila, A. W. Cheng, L. Shi and R. Jaenisch (2013). "One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering." Cell.

Yao, Y., B. Nashun, T. Zhou, L. Qin, S. Zhao, J. Xu, M. A. Esteban and X. Chen (2011). "Generation of CD34+ cells from CCR5-disrupted human embryonic and induced pluripotent stem cells." Hum Gene Ther **23**(2): 238-242.

Ye, L., M. O. Muench, N. Fusaki, A. I. Beyer, J. Wang, Z. Qi, J. Yu and Y. W. Kan (2013). "Blood cell-derived induced pluripotent stem cells free of reprogramming factors generated by Sendai viral vectors." Stem Cells Transl Med **2**(8): 558-566.

Ye, Z., H. Zhan, P. Mali, S. Dowey, D. M. Williams, Y. Y. Jang, C. V. Dang, J. L. Spivak, A. R. Moliterno and L. Cheng (2009). "Human-induced pluripotent stem cells from blood cells of healthy donors and patients with acquired blood disorders." Blood **114**(27): 5473-5480.

Yoshioka, N., E. Gros, H. R. Li, S. Kumar, D. C. Deacon, C. Maron, A. R. Muotri, N. C. Chi, X. D. Fu, B. D. Yu and S. F. Dowdy (2013). "Efficient Generation of Human iPSCs by a Synthetic Self-Replicative RNA." Cell Stem Cell **13**(2): 246-254.

Yu, D., T. Shioda, A. Kato, M. K. Hasan, Y. Sakai and Y. Nagai (1997). "Sendai virus-based expression of HIV-1 gp120: reinforcement by the V(-) version." Genes Cells **2**(7): 457-466.

Yu, J., K. Hu, K. Smuga-Otto, S. Tian, R. Stewart, Slukvin, II and J. A. Thomson (2009). "Human induced pluripotent stem cells free of vector and transgene sequences." Science **324**(5928): 797-801.

Yu, J., M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, Slukvin, II and J. A. Thomson (2007). "Induced pluripotent stem cell lines derived from human somatic cells." Science **318**(5858): 1917-1920.

Yuan, T., W. Liao, N. H. Feng, Y. L. Lou, X. Niu, A. J. Zhang, Y. Wang and Z. F. Deng (2013). "Human induced pluripotent stem cell-derived neural stem cells survive, migrate, differentiate, and improve neurological function in a rat model of middle cerebral artery occlusion." Stem Cell Res Ther **4**(3): 73.

Yusa, K., S. T. Rashid, H. Strick-Marchand, I. Varela, P. Q. Liu, D. E. Paschon, E. Miranda, A. Ordonez, N. R. Hannan, F. J. Rouhani, S. Darche, G. Alexander, S. J. Marciniak, N. Fusaki, M. Hasegawa, M. C. Holmes, J. P. Di Santo, D. A. Lomas, A. Bradley and L. Vallier (2011). "Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells." Nature **478**(7369): 391-394.

Zanella, A. and P. Bianchi (2000). "Red cell pyruvate kinase deficiency: from genetics to clinical manifestations." Baillieres Best Pract Res Clin Haematol **13**(1): 57-81.

Zanella, A., P. Bianchi, L. Baronciani, M. Zappa, E. Bredi, C. Vercellati, F. Alfinito, G. Pelissero and G. Sirchia (1997). "Molecular characterization of PK-LR gene in pyruvate kinase-deficient Italian patients." Blood **89**(10): 3847-3852.

Zanella, A., P. Bianchi and E. Fermo (2007). "Pyruvate kinase deficiency." Haematologica **92**(6): 721-723.

Zanella, A., E. Fermo, P. Bianchi, L. R. Chiarelli and G. Valentini (2007). "Pyruvate kinase deficiency: the genotype-phenotype association." Blood Rev **21**(4): 217-231.

Zanella, A., E. Fermo, P. Bianchi and G. Valentini (2005). "Red cell pyruvate kinase deficiency: molecular and clinical aspects." Br J Haematol **130**(1): 11-25.

Zaucha, J. A., C. Yu, C. D. Lothrop, Jr., R. A. Nash, G. Sale, G. Georges, H. P. Kiem, G. P. Niemeyer, M. Dufresne, Q. Cao and R. Storb (2001). "Severe canine hereditary hemolytic anemia treated by nonmyeloablative marrow transplantation." Biol Blood Marrow Transplant **7**(1): 14-24.

Zeuner, A., F. Martelli, S. Vaglio, G. Federici, C. Whitsett and A. R. Migliaccio (2012). "Concise review: stem cell-derived erythrocytes as upcoming players in blood transfusion." Stem Cells **30**(8): 1587-1596.

Zhou, L., D. L. Rowley, Q. S. Mi, N. Sefcovic, H. W. Matthes, B. L. Kieffer and D. M. Donovan (2001). "Murine inter-strain polymorphisms alter gene targeting frequencies at the mu opioid receptor locus in embryonic stem cells." Mamm Genome **12**(10): 772-778.

Zhou, W. and C. R. Freed (2009). "Adenoviral gene delivery can reprogram human fibroblasts to induced pluripotent stem cells." Stem Cells **27**(11): 2667-2674.

Zou, J., P. Mali, X. Huang, S. N. Dowey and L. Cheng (2011). "Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease." Blood **118**(17): 4599-4608.

Zou, J., C. L. Sweeney, B. K. Chou, U. Choi, J. Pan, H. Wang, S. N. Dowey, L. Cheng and H. L. Malech (2011). "Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting." Blood **117**(21): 5561-5572.

XI Annex

HUMAN GENE THERAPY 24:571–583 (June 2013)

DOI: 10.1089/hum.2012.251

Review

New Frontier in Regenerative Medicine: Site-Specific Gene Correction in Patient-Specific Induced Pluripotent Stem Cells

Zita Garate,[1] Brian R. Davis,[2] Oscar Quintana-Bustamante,[1] and Jose C. Segovia[1]

Abstract

Advances in cell and gene therapy are opening up new avenues for regenerative medicine. Because of their acquired pluripotency, human induced pluripotent stem cells (hiPSCs) are a promising source of autologous cells for regenerative medicine. They show unlimited self-renewal while retaining the ability, in principle, to differentiate into any cell type of the human body. Since Yamanaka and colleagues first reported the generation of hiPSCs in 2007, significant efforts have been made to understand the reprogramming process and to generate hiPSCs with potential for clinical use. On the other hand, the development of gene-editing platforms to increase homologous recombination efficiency, namely DNA nucleases (zinc finger nucleases, TAL effector nucleases, and meganucleases), is making the application of locus-specific gene therapy in human cells an achievable goal. The generation of patient-specific hiPSC, together with gene correction by homologous recombination, will potentially allow for their clinical application in the near future. In fact, reports have shown targeted gene correction through DNA-Nucleases in patient-specific hiPSCs. Various technologies have been described to reprogram patient cells and to correct these patient hiPSCs. However, no approach has been clearly more efficient and safer than the others. In addition, there are still significant challenges for the clinical application of these technologies, such as inefficient differentiation protocols, genetic instability resulting from the reprogramming process and hiPSC culture itself, the efficacy and specificity of the engineered DNA nucleases, and the overall homologous recombination efficiency. To summarize advances in the generation of gene corrected patient-specific hiPSCs, this review focuses on the available technological platforms, including their strengths and limitations regarding future therapeutic use of gene-corrected hiPSCs.

Introduction: Regenerative Medicine—Cell Plus Gene Therapy

Regenerative medicine aims to replace and/or to regenerate damaged cells, organs, or tissues in order to restore normal function. Cell therapy is an important regenerative medicine approach, in which either differentiated cells or stem cells capable of differentiation are transplanted into an individual with the objective of yielding specific cell types present in the damaged tissue and consequently restoring its function. The most successful example of cell therapy is bone marrow (BM) transplantation, in which the transplanted hematopoietic stem cells (HSCs) are able to regenerate the patient's blood. BM transplantation started in the 1950s and now is a widely established procedure for many hematopoietic diseases (Thomas *et al.*, 1977). Cell therapies for other tissues then followed in the footsteps of the hematopoietic experience. Nowadays, there are numerous ongoing clinical trials using various types of stem cells and some of them are U.S. Food and Drug Administration (FDA)-approved cell-based products (www.fda.gov/BiologicsBloodVaccines/CellularGeneTherapyProducts/ApprovedProducts/default.htm).

Cell replacement can be done with autologous or allogeneic stem cells. When performing allogeneic cell therapy, the risk of immune rejection usually requires the use of immunosuppressive drugs, which can induce toxicity and increase the risk of infections and cancer, which could be life-threatening.

[1]Differentiation and Cytometry Unit, Hematopoiesis and Gene Therapy Division, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBER-ER), 28040 Madrid, Spain.
[2]Center for Stem Cell and Regenerative Medicine, Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center, Houston, TX 77030.

This, together with the low availability of suitable donors, makes autologous cell therapy frequently the preferred option for regenerative medicine. However, in the case of monogenic diseases, in which all the cells from the body initially carry the disease-causing mutation in their genomic DNA, a gene correction approach should be considered to generate disease-free autologous cells. Thus, a combination of cell and gene therapy is used. Since the first gene therapy clinical trial in 1990 (Anderson *et al.*, 1990), much effort has been made to develop safer and more efficient approaches.

The first gammaretroviral vectors used in clinical trials were associated with enhancer-mediated *cis*- and *trans*-activation, which induced insertional mutagenesis and ended up in leukemia associated with the gene therapy procedure (Hacein-Bey-Abina *et al.*, 2003, 2008; Fischer *et al.*, 2010). In the ongoing clinical trials, lentiviral and retroviral vectors have deletions in their long terminal repeats (LTRs) to minimize *trans*-activation of the genes surrounding the integration. However, insertional mutagenesis still remains an issue. Other strategies, such as the use of nonintegrating viral vectors, are being studied, as in the case of integration-defective lentiviral vectors (Yanez-Munoz *et al.*, 2006; Matrai *et al.*, 2011), but procedures to maintain nonintegrated DNA in proliferating cells have not yet been developed. The introduction of genetic material in specific, known, and characterized loci of the genome via homologous recombination (HR) would be an ideal option. HR will allow, in principle, specific correction of the mutation without any additional modification in the genome, or introduction of the genetic material in a known and safe genome locus. HR is based on the natural DNA repair process, in which a double-strand break (DSB) is corrected with a homologous DNA sequence. The therapeutic application of HR involves exchanging the mutation for the correct sequence, or even introducing the correct version of the gene in the targeted locus.

An important consideration is the source of cells for autologous cell therapy. For some purposes, as is the case for hematological diseases, a hematopoietic multipotent stem/progenitor cell present in the adult body can be used. Other examples of these kinds of progenitors in humans include neural stem cells (Galli *et al.*, 2003), mesenchymal stem cells (Deans and Moseley, 2000), and intestinal stem cells (Yui *et al.*, 2012). In the majority of these adult stem cells, an important limitation is that correction of mutations by HR has rarely been described to occur in a manner that retains the multipotentiality of the stem cells. In addition, these kinds of progenitors have been described for only a few tissues in the body. Thus, an autologous stem cell source with wide expansion and differentiation potential is required for future clinical use of HR in the context of regenerative medicine. This issue has been solved with the generation of human induced pluripotent stem cells (hiPSCs) (Takahashi *et al.*, 2007). Human iPSCs offer a powerful novel technology in gene and cell therapies. Their essentially unlimited growth capability allows successfully targeted cell selection, with the possibility that 100% of potentially transplanted cells would be corrected. The fact that they represent a clonal cell population is also advantageous as we can completely interrogate the whole exome or the whole genome for any abnormality that could be accumulated during the entire manipulation procedure, as has been addressed in several reports (Table 1 and Fig. 1). We deal with some of these pioneer works involving hiPSCs and HR technologies in this review.

Generation of Patient-Specific Pluripotent Stem Cells

Choice of reprogramming platform

Since Yamanaka and colleagues first reported the generation of mouse iPSCs in 2006 (Takahashi and Yamanaka, 2006), and later the groups of Yamanaka (Takahashi *et al.*, 2007) and Thomson (Yu *et al.*, 2007) in human cells in 2007, many laboratories around the world have been able to reprogram a large variety of somatic cells into pluripotent stem cells, from neural stem cells (J.B. Kim *et al.*, 2009) to terminally differentiated B lymphocytes (Hanna *et al.*, 2008). The reproducibility and potentiality (unlimited self-renewal and ability to differentiate into any cell type) of these cells has caused the iPSC field to advance rapidly. hiPSC technology brings together all the potential of human embryonic stem cells (hESCs) in terms of self-renewal and pluripotency without the problems associated with hESC generation (i.e., ethical issues associated with embryo disruption and immunoincompatibility with the recipient of the cells). Therefore, hiPSC technology arises as one of the most promising fields for future cell therapies for many human diseases.

For the generation of hiPSCs, the first reports used gammaretroviruses to express the four defined factors required for reprogramming, *OCT4*, *SOX2*, *KLF4*, and c-*MYC* (Takahashi *et al.*, 2007), or *OCT4*, *LIN28*, *NANOG*, and *SOX2* (Yu *et al.*, 2007), separately in different viral vectors. Because of the nature of the vectors, expression of the factors was silenced after the endogenous pluripotent genes were activated at an adequate level. Safer and more efficient reprogramming approaches have since been developed and many patient-specific hiPSCs have been generated both to model human diseases and correct the diseased hiPSCs through gene therapy approaches. Depending on the cell type being reprogrammed, the number of factors used could be reduced and, more importantly, oncogenes or tumor-related proteins used for reprogramming, such as c-*MYC* or *KLF4*, could be removed from the original reprogramming cocktail. This, for example, was the case for reprogramming hematopoietic progenitors (Liu *et al.*, 2012; Meng *et al.*, 2012) or neural stem cells (J.B. Kim *et al.*, 2009), in which these two factors could be removed. Even reprogramming with *OCT4* alone was achieved (Thier *et al.*, 2010). However the reprogramming efficiency decreases after removing any reprogramming factor. Although avoiding tumor-related genes increases the safety of the reprogramming process, the safest reprogramming protocol will ultimately involve removable reprogramming transgenes or, even better, nonintegrative systems that will avoid potential adverse effects associated with the integration of the vector sequences in the cell genome (Sommer *et al.*, 2010). Several groups have developed Cre-mediated excisable polycistronic lentiviral vectors (Somers *et al.*, 2010; Papapetrou and Sadelain, 2011) or transposon-based reprogramming systems (Woltjen *et al.*, 2011), which could be removed after obtaining the hiPSC clones. The first truly nonintegrative reprogramming approach described in human cells was reported by the Thomson group using episomal plasmids for expression of the four Yamanaka transcription factors plus *NANOG*, *LIN28*, and SV40 large T antigen (SVLT) (Yu *et al.*, 2009). In

Table 1. Human Diseases from Which Human Induced Pluripotent Stem Cells Have Been Generated or Generated and Corrected

TF delivery method	*Efficiency/rounds*	*Integrative*	*Cell type*[a]	*Disease, phenotype correction*[a]
Lenti/retroviruses	0.01–1/single	Yes	Fibroblasts (1, 2) B lymphocytes (3) Human adipose stem cells (4) BM-MSCs (5) CD34 progenitors (6) BM-MNCs (7)	• Not corrected: Leopard syndrome (23), long QT syndrome (24), Timothy syndrome (25), ADA-SCID, Gaucher disease, Duchenne and Becker muscular dystrophy, type 1 diabetes mellitus, trisomy 21, Lesch–Nyhan syndrom (26), CINCA syndrome (27), familial hypercholesterolemia (28), Rett syndrome (29), dyskeratosis congenita (30, 31), familial dilated cardiomyopathy (32), chronic myelogenous leukemia (33), Alzheimer's disease (34), glucose-6-phosphatase deficiency, Crigler–Najjar syndrome, hereditary tyrosinemia type 1 (35) • Corrected by lentiviruses: Fanconi anemia (36) • Corrected by gene editing: Progeria (37), sickle cell anemia (13). X-linked chronie granulamatosis (5). Huntington's disease (38), β-thalassemia (39)
Adenoviruses	~0.0002/multiple	No	Fibroblasts (8)	
Sendai virus	~1/single	No	Fibroblasts (9) T cells (10) CD 34 progenitors (11)	• Corrected by gene editing: α^1-Antitrypsin deficiency (40)
Excisable lentivirus	0.01–1/single	Removable	Fibroblasts (12–14) CD34 progenitors (15)	• Not corrected: Cystic fibrosis (41) • Corrected by lentivirus: β- Thalassemia (12) • Corrected by gene editing: Parkinson's disease (42)
Excisable transposons	~0.1/single	Removable	Fibroblasts (16) BM-MSCs (5)	• Corrected by gene editing: Sickle cell anemia (5)
Minicircles	~0.005/multiple	No	Adipose-derived stromal cells (17)	
Episomal vectors	0.001/multiple	No	Fibroblasts (18) BM- and CB-MNCs (19)	• Not corrected: Neoplasic bone marrow (19) • Corrected by gene editing: Gyrate atrophy (43)
Proteins	0.001/multiple	No	Fibroblasts (20)	
miRNAs	0.1/single	Yes	Fibroblasts (21)	
mRNAs	1–4/multiple	No	Fibroblasts (22)	• Not corrected: Cystic fibrosis (22)

ADA-SCID, adenosine deaminase-deficient severe combined immunodeficiency; BM-MNCs, bone marrow-derived mononuclear cells; BM-MSCs, bone marrow-derived mesenchymal stem cells; CINCA syndrome, chronic infantile neurological cutaneous and articular syndrome; miRNAs, micro-RNAs, micro-RNAs; TF, transgene free.

[a]*References:* (1) Takahasi *et al.* (2007); (2) Yu *et al.* (2007); (3) Hanna *et al.* (2008); (4) Sun *et al.* (2009); (5) Zou *et al.* (2011a); (6) Ye *et al.* (2009); (7) Kunisato *et al.* (8) Zhou and Freed (2009); (9) Fusaki *et al.* (2009); (10) Seki *et al.* (2011); (11) Ban *et al.* (2011); (12) Papapetrou *et al.* (2011); (13) Sebastiano *et al.* (2011); (14) Soldner *et al.* (2009); (15) Ramos-Mejia *et al.* (2012); (16) Kaji *et al.* (2009); (17) Jia *et al.* (2010); (18) Yu *et al.* (2009); (19) Hu *et al.* (2011); (20) D. Kim *et al.* (2009); (21) Anokye-Danso *et al.* (2011); (22) Warren *et al.* (2010); (23) Carvajal-Vergara *et al.* (2010); (24) Moretti *et al.* (2010); (25) Yazawa *et al.* (2011); (26) Park *et al.* (2008); (27) Tanaka *et al.* (2012); (28) Cayo *et al.* (2012); (29) K.Y. Kim *et al.* (2011); (30) Agarwal *et al.* (2010); (31) Batista *et al.* (2011); (32) Sun *et al.* (2012); (33) Kumano *et al.* (2012); (34) Israel *et al.* (2012); (35) Rashid *et al.* (2010); (36) Raya *et al.* (2009); (37) Liu *et al.* (2011); (38) An *et al.* (2012); (39) Wang *et al.* (2012); (40) Yusa *et al.* (2011); (41) Somers *et al.* (2010); (42) Soldner *et al.* (2011); (43) Howden *et al.* (2011).

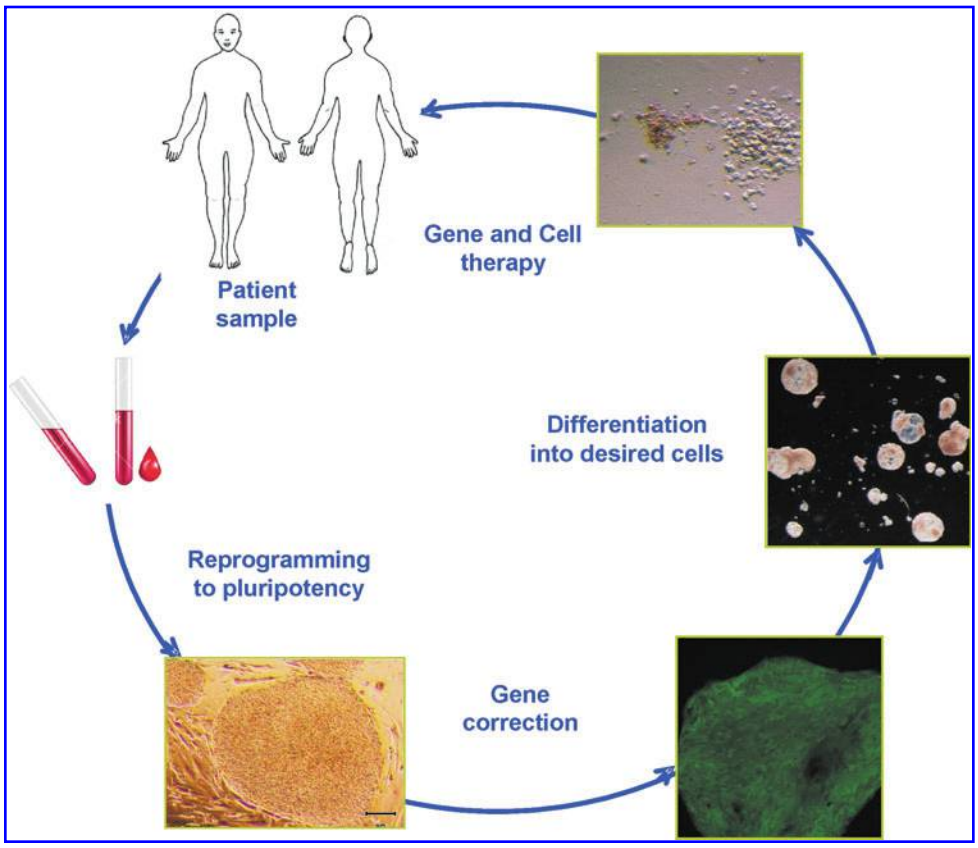


FIG. 1. Gene correction approach for a hematopoietic disease, using induced pluripotent stem cells. Color images available online at www.liebertpub.com/hum

the same year, reprogramming by recombinant proteins (D. Kim *et al.*, 2009), synthetic mRNAs (Warren *et al.*, 2010), and nonintegrating RNA Sendai viral vectors (Fusaki *et al.*, 2009) was also reported (Table 1). The majority of disease-specific hiPSCs reported until now have been generated with integrative systems, but an increasing number of disease-specific hiPSCs have been generated using these novel and potentially safer approaches (Table 1).

Choice of cell source for reprogramming

As reported by Hanna and colleagues, an adequate level of expression of the reprogramming factors in any cell type would likely allows the creation of an iPSC line (Hanna *et al.*, 2009). The preferred cell source for reprogramming will most likely be the most easily accessible and the one in which the reprogramming factors can be successfully delivered. That is why fibroblasts have been widely used by many groups (Takahashi *et al.*, 2007; Park *et al.*, 2008; Fusaki *et al.*, 2009; Carvajal-Vergara *et al.*, 2010; Liu *et al.*, 2010; Howden *et al.*, 2011; Papapetrou *et al.*, 2011; Sebastiano *et al.*, 2011; Tanaka *et al.*, 2012). Fibroblasts can be easily grown from a small human biopsy and can be efficiently transduced with viral vectors. Another cell source that can be easily obtained and presents several advantages are peripheral blood mononuclear cells (PB-MNCs) (Kunisato *et al.*, 2010). These cells can be obtained from routine blood tests or in patient follow-up, can be frozen and stored, and are easily cultured; in addition, stimulation of the preferred cell type within the PB-MNCs by cytokines is possible. Jaenisch's group (Staerk *et al.*, 2010) showed that by stimulation of PB-MNCs with granulocyte colony-stimulating factor (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), interleukin (IL)-3, and IL-6 before and during the first days of reprogramming, the induction of pluripotency could be prompted in progenitors and cells of myeloid origin, avoiding reprogramming of B or T cells.

However, there are important questions that remain unsolved. For example, is the cell source origin going to influence the characteristics of its corresponding hiPSCs? Several groups have compared hESCs and hiPSCs at the gene expression level and for their genome methylation status. Some authors have found significant differences between hiPSCs and hESCs, including an epigenetic memory of the original cell source in the hiPSCs (Chin *et al.*, 2009; Deng *et al.*, 2009; Doi *et al.*, 2009; Marchetto *et al.*, 2009; Ghosh *et al.*, 2010). Other groups attributed these differences to the intrinsic differences between various hiPSC clones of the same reprogramming experiment or to the various technological platforms that have been used for reprogramming (Guenther *et al.*, 2010; Newman and Cooper, 2010; Bock *et al.*, 2011). This issue has been deeply analyzed, and it has been pointed out that the small number of clones analyzed, in the studies in which a large difference is observed, could have negatively influenced the conclusions (Yamanaka, 2012). In the event that epigenetic memory is proven to be true, would the original cell source have an influence on the differentiation capacity of a specific hiPSC line? According to K. Kim and colleagues, hiPSCs derived from cord blood cells showed a hematopoietic differentiation advantage when compared with hiPSCs derived from keratinocytes (K. Kim *et al.*, 2011). On the other hand, other authors did not find epigenetic

memory in hiPSC lines derived from hepatocytes (Ohi *et al.*, 2011). This issue, clearly important for therapeutic applications, will require further study in order to determine to what extent the ultimate transplantable cell type should influence the source of patient-specific cells for reprogramming. As this issue remains unclear, we believe that the cell source should be, first, the most accessible and least invasive, and then, depending on the future use of the hiPSCs, an epigenetically related cell source should be considered if available. This is the case of hiPSCs for cell therapy of blood diseases, in which either fibroblasts or PB-MNCs could be used, based on their accessibility, but PB-MNCs may perhaps prove to be a better option provided they exhibit a differentiation advantage into the hematopoietic lineage.

Correction of Patient-Derived Pluripotent Stem Cells

At present, various strategies have been tested and proven for the correction of patient-specific hiPSCs. Although this review is focused on genetic correction directly in hiPSCs, this is not always achievable, because some genetic diseases imply a reprogramming barrier, as has been the case for Fanconi anemia (FA). In this case, genetic correction was carried out before the generation of FA-hiPSCs (Raya *et al.*, 2009). All the approaches described in this review could also be done before generating patient-specific hiPSCs if the cells of origin allow the culture needed for the genetic correction and selection of corrected cells.

Random integration

The first reports of correction of patient-derived hiPSCs used lentiviral vectors to correct the disease through transgene addition (Raya *et al.*, 2009). However, these vectors, because of their nearly random integration pattern in the genome, are susceptible to transcriptional silencing, depending on whether the integration site resides in a silent or active transcriptional region of chromatin. Furthermore, as mentioned previously, integrated vectors may show enhancer-mediated *cis*- and *trans*-activation and might consequently induce insertional mutagenesis. The identification of integrations in safe harbor genomic sites (e.g., far away from genes or coding information) could represent an alternative, safer mode of therapy. The self-renewal and almost indefinite growth properties of hiPSCs enable analysis of the integration sites of these vectors at a clonal level and the selection of those that could be potentially safer (Papapetrou *et al.*, 2011; Bedel *et al.*, 2012). However, the definition of a safe harbor site in the genome is challenging, and it will probably change as we get to know the genome in more depth. Future therapeutic applications of hiPSCs for cell therapy would benefit from a site-specific gene correction approach. The cooperation between hiPSC technology and homologous recombination (HR) has been extensively explored. HR is presented as an exciting and novel alternative to avoid insertional mutagenesis associated with integrative vector-mediated correction.

Site-specific gene editing

Gene editing is a process in which a DNA sequence is replaced or introduced into a specific locus at single-base pair resolution. This precise site-specific introduction requires an accurate recognition mechanism of the target site on the genome. Under normal conditions, maintenance of the integrity of the genome requires repair of the continuous cellular DNA damage with high fidelity. HR is a truly accurate DNA repair mechanism that is basically a "copy and paste" mechanism and is also used to resolve double-strand breaks (DSBs) in the DNA. This process uses an undamaged homologous segment of DNA as a template (conventionally, the sister chromatid) to copy the information across the DSB. Because it copies a normal copy of the undamaged DNA, HR is the most secure process by which to repair DSBs. The fidelity of HR gives the specificity and accuracy that gene editing requires.

The natural HR process has been exploited by researchers to achieve the desirable site-specific gene editing within a targeted locus by introducing exogenous genomic sequences, homologous to the target locus, flanking the desired DNA material to be inserted. These techniques have been widely used for the generation of knock-out and knock-in transgenic animals (Robbins, 1993). Routinely, homology arms are homologous DNA sequences that cover the target where HR will take place. Between these two arms, a therapeutic or correct sequence of the gene should be found. In addition, drug resistance genes can be introduced between both homology arms for positive selection or suicide genes outside of the homology arms for negative selection. The final structure and complexity of this construction, also called the repair matrix, will vary according to the needs of the researcher (Fig. 2). With the development of disease-specific hiPSCs, this methodology has already been used to correct mutations (Howden *et al.*, 2011; Liu *et al.*, 2011; Ohi *et al.*, 2011).

Gene editing via HR in human cells is inefficient and dependent on the generation of a DSB at the specific target site

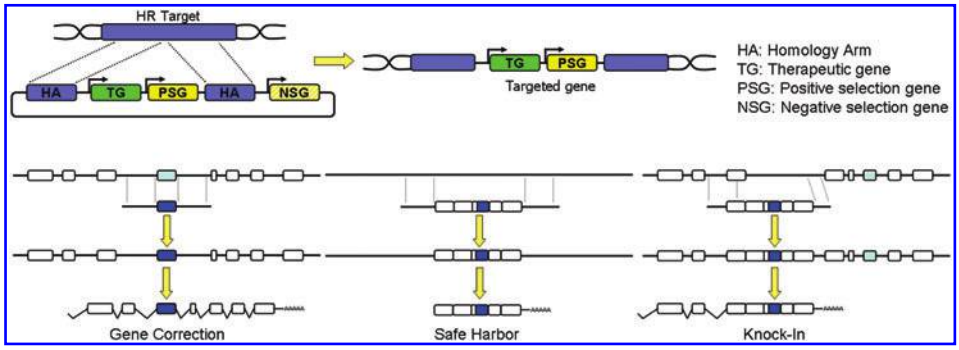


FIG. 2. Scheme of the repair matrix, pointing out the various required elements, and of the various disease correction strategies for homologous recombination (HR). Color images available online at www.liebertpub.com/hum

(Carroll, 2011). In the absence of a repair matrix, nonhomologous end joining (NHEJ) is the dominant pathway to solve these DNA lesions in human cells, and its resolution is highly error-prone (Grabarz *et al.*, 2012). In addition, HR varies in different cell types and requires transit through the $S–G_2$ phase of the cell cycle to take place (Delacote and Lopez, 2008). These limitations have typically made gene editing in human cells difficult to achieve. Various approaches have been used to improve gene editing by HR, such as the increase in the length of the DNA sequences homologous to the target site (homology arms) (Song *et al.*, 2010), the use of adeno-associated vectors to more efficiently introduce the repair matrix in the cells (Khan *et al.*, 2010), and the improvement of selection methods for the identification of correctly edited cells, or the stimulation of HR by inducing DSBs using specific DNA nucleases. The use of engineered DNA nucleases that recognize specific sites of the genome is an active area of investigation and is the most commonly reported method for correction of patient-specific hiPSCs.

Engineered DNA nucleases are enzymes that have been developed to induce DSBs specifically at a unique and defined sequence in the cell genome. The rationale for inducing the double-stranded DNA break in the immediate vicinity of the mutant sequences is that these DSBs have been shown to increase the efficiency of homology-directed repair (HDR) by 10^3- to 10^4-fold (Porteus and Carroll, 2005). Engineered DNA nucleases are formed by a nuclease domain and a DNA-binding domain, the sequence specificity of which can be artificially modified. The most widely used DNA nucleases are zinc finger nucleases (ZFNs), homing meganucleases (MNs), and transcription activator-like (TAL) effector nucleases (TALENs). They potentially identify a unique sequence within the genome and generate DSBs to induce the recruitment of the cell repair machinery to repair the DSBs, ideally by HR. The DNA-binding domain of ZFNs is derived from DNA-binding zinc finger proteins and is composed of a tandem repeat of Cys_2His_2 zinc fingers, each of which recognizes three nucleotides. The DNA-binding domain is linked to the nuclease domain of the restriction enzyme *Fok*I. ZFNs work as pairs of two monomers of ZFN in reverse orientation. This ZFN dimer can be designed to bind to a genomic sequence 18–36 nucleotides in length (Porteus and Carroll, 2005; Carroll, 2011). TALENs have a similar structure to ZFNs, but the DNA-binding domain comes from TAL effector proteins. The DNA-binding domain in TALENs is a tandem array of amino acid repeats. Each of these units is able to bind to one of the four possible nucleotides. Thus, the DNA-binding domain can be designed to recognize any desired genomic sequence. TALENs also cleave as dimers (Li *et al.*, 2011). In contrast to these synthetic DNA nucleases, natural MNs are a subset of homing endonucleases. MNs are monomeric proteins that have four DNA-binding domains that recognize a DNA sequence from 14 to 40 nucleotides in length. Directed mutagenesis can be applied to modify the DNA sequence specificity. A unique and specific MN recognition sequence can be found approximately every 300 bp (Paques and Duchateau, 2007).

ZFNs were first developed in 1996 by Y.G. Kim and colleagues and applied for disrupting gene expression by introducing mutations in the selected gene (Kim *et al.*, 1996). More recently, they have been widely used for gene editing in hESCs and hiPSCs. For example, Lombardo and colleagues showed the insertion of the gene encoding green fluorescent protein (GFP) into the *CCR5* safe harbor locus in hESCs after inducing HR by ZFN expression; targeted hESCs were able to differentiate into neurons keeping GFP expression (Lombardo *et al.*, 2007). TALENs have been also tested in hESCs and hiPSCs (Hockemeyer *et al.*, 2011). One of the most important potential disadvantages of engineered nucleases is the possibility of their cutting other, related sequences of the genome, the so-called off-target sites. After targeting several loci and comparing HR efficiencies and the presence of off-targets with both types of nucleases, these authors concluded that both ZFNs and TALENs show similar efficiencies and accuracy (Hockemeyer *et al.*, 2011). Nevertheless, an *in vitro* gene disruption comparison between ZFNs and TALENs showed that TALENs were more efficient and less cytotoxic in this assay (Mussolino *et al.*, 2011). Other authors have reported that targeted efficiency by various nucleases seems to be affected by the epigenetic status of the locus to be targeted (Daboussi *et al.*, 2012). The presence of methylated CpGs (mCpGs) in the TALEN recognition site can dramatically decrease its efficiency (Valton *et al.*, 2012b) and in the case of MNs this also happens if the mCpG is in the central tetrabase of the recognition site (Valton *et al.*, 2012a)

The proof of principle for the clinical application of nuclease-mediated gene editing was tested in hiPSCs from patients affected by various genetic diseases some time later. To correct or insert/express a transgene by HR, three different strategies can be considered (Fig. 2 and Table 2). In the following sections we discuss the various attempts applied for nuclease-based correction.

Targeted safe harbor integration. For safe harbor integration, a complete expression cassette (the therapeutic transgene, promoter, and possibly additional regulatory signals [e.g., enhancer]) is inserted into a specific genome locus that is not susceptible to transgene silencing via epigenetic mechanisms. Ideally, the targeted integration will either not affect expression of the neighboring genes or at least allow modified cells to function normally if the targeting results in disruption of the safe harbor locus. This seems to be the case for *AAVS1*, *CCR5*, and *ROSA26* loci (Irion *et al.*, 2007; Torres *et al.*, 2011; Yao *et al.*, 2011). Special attention should be taken to avoid targeting loci previously considered nonfunctional but now known to fulfill important regulatory functions as per ENCODE (the *Enc*yclopedia *o*f *D*NA *E*lements). One potential advantage of the safe harbor strategy is that there should be significantly less cell-to-cell variation in transgene expression than that resulting from random integration. Although the addition of a promoter to express the therapeutic gene is needed, the main advantage of the safe harbor strategy is the wide variety of diseases that could be treated with a similar repair matrix, only exchanging the therapeutic gene for each disease. Examples are as follows:

X-Linked Chronic Granulomatous Disease: Seminal work published by Malech's group in March 2011 showed, for the first time, ZFN-mediated phenotype correction in neutrophils generated from X-linked chronic granulomatous disease (X-CGD) hiPSCs by inserting a wild-type copy of the *CYBB* gene (encoding the $gp91^{phox}$ protein) driven by the CAG (cytomegalovirus early enhancer/chicken β-actin) chimeric promoter in the previously described *AAVS1* safe

Table 2. Summary of the Various Gene Correction Strategies in Terms of Safety and Applicability

		Safety	*Recommended scenario*	*Contraindicated scenario*	*References*
Random integration	Random integration	Risk of insertional mutagenesis	Simple methodology Avoids DSB Promoter mutations	Negative effect of mutated gene product	Raya *et al.* (2009)
	Random integration with safe harbor (SH) clone selection	Definition of safe harbors not clear	Avoids DSB Promoter mutations	Negative effect of mutated gene product	Bedel *et al.* (2012); Papapetrou *et al.* (2011)
Gene editing	Targeted safe harbor integration	Safe for AAVS1 Risks for newly defined safe harbors	Promoter mutations Many described mutations	Negative effect of mutated gene product Transgene needs to be tightly regulated	Chang and Bouhassira (2012); Zou *et al.* (2011b)
	Targeted knock-in	Safe	Many described mutations within an area of the gene Transgene needs to be tightly regulated	In some cases incompatible with correction of many isoforms	—
	Targeted gene correction	Safe	Unique common mutation in many patients Negative effect of mutated gene product Transgene needs to be tightly regulated Many gene isoforms	Many described mutations	An *et al.* (2012); Howden *et al.* (2011); Liu *et al.* (2011); Sebastiano *et al.* (2011); Soldner *et al.* (2011); Wang *et al.* (2012); Yusa *et al.* (2011); Zou *et al.* (2011a)

AAVS1, adeno-associated virus integration site 1; DSB, double-strand break.

harbor locus (Zou *et al.*, 2011b). Puromycin selection was also included in the inserted DNA to select recombined clones. In addition, in some of the *AAVS1* alleles that were not targeted, there were mutations associated with NHEJ correction, evidence for cleavage by ZFNs at this site. Having a high number of targeted clones makes it possible to select and grow just those that show no off-target integrations or new mutations. Importantly, after differentiation of the corrected X-CGD hiPSCs, the resulting neutrophils showed equal levels of therapeutic reactive oxygen species (ROS) to neutrophils derived from wild-type hiPSCs.

β-Thalassemia: To achieve a more physiological expression level of the transgene, Chang and Bouhassira (2012) used the specific *β*-globin promoter for directing expression of the transgene when targeted into the AAVS1 locus. After puromycin selection, all the clones analyzed were targeted at the *AAVS1* locus and 50% represented homozygous targeting (i.e., targeting into both *AAVS1* loci) as assessed by PCR and Southern blot. Erythroid differentiation of corrected clones showed restoration of hemoglobin quantity and quality without disturbing any *AAVS1* locus-neighboring genes.

Targeted correction. Targeted correction typically uses site-specific nucleases designed to recognize a site in the immediate vicinity of the mutation targeted for correction together with a repair matrix precisely matching that of the targeted endogenous sequences, with the exception of the base or bases intended for alteration. The mutant target bases are substituted for by the wild-type bases present in the introduced repair matrix, thus correcting or repairing the gene. In repairing the defective sequence within the endogenous gene locus, the corrected genetic material is maintained within its normal chromatin environment. This ensures the appropriate genetic regulation and expression in the cell. In situations in which the mutant gene product exercises a dominant negative influence over the normal gene product, gene correction may be the only suitable strategy. Gene correction is especially useful for diseases in which the majority of patients have the same well-defined, limited alterations in the DNA sequence, such as sickle cell anemia or cystic fibrosis. When different mutations for the same gene have been reported, gene correction would turn into a patient-specific therapy and therefore the repair matrix, and also perhaps the site-specific nucleases, should be tailor-made for each patient or set of patients. Examples of this approach are as follows:

Parkinson's Disease: hiPSCs from a patient with the A53T mutation in the α-synuclein gene were corrected by ZFN-assisted HR in the mutated locus. In this case, the targeting sequence in the donor vector was approximately 1 kb of the wild-type sequence of the α-synuclein gene with the targeted mutant base in the middle, close to the ZFN cleavage site. As there was no selection cassette in the donor, a selection-free approach was mandatory and the number of clones that had to be analyzed to obtain a correctly targeted clone was higher than with selection-based approaches. This procedure could also be seen as an advantage, as there was just one clonal step instead of two or three, therefore reducing the manipulation steps and the probability of additional genetic alterations.

Sickle Cell Anemia: Two studies have been published for the genetic correction of hiPSCs from patients with sickle cell anemia (Sebastiano *et al.*, 2011; Zou *et al.*, 2011a). The first report showed specific ZFN-mediated gene correction of the β^s (A > T) mutation in the *HBB* locus (Zou *et al.*, 2011a). The authors used a donor vector with a *loxP*-flanked ("floxed") hygromycin resistance selection cassette that, after nucleofection of the ZFNs and the repair matrix, allowed the detection of hygromycin-resistant clones. PCR analysis and Southern blotting verified the presence of hiPSC clones correctly targeted within the *HBB* locus and in no additional loci. After erythrocyte differentiation of the corrected hiPSCs, the authors suspected that the presence of the selection cassette affected expression of the corrected transcript. To avoid possible interference, the selection cassette was excised in 4 of 24 clones by Cre recombinase. Surprisingly, in those clones in which the selection cassette was excised, expression of the corrected gene was still only 25–40% of the expression of the uncorrected allele. The authors speculate that the reduced expression level could be due to two main reasons: either the presence of the remaining *loxP* sequences after excision of the selection cassette, or the presence of a nucleotide variant (A > G) affecting a GATA-containing 3′ enhancer that may have been generated during HR. This study points out that a selectable cassette could have clear benefits in reducing the number of clones to be analyzed, but it could potentially adversely affect the intended correction by repressing the expression of the transgene unless excised. This work also highlights the importance of investigating the possible acquisition of genetic modifications during reprogramming and/or HR because such mutations could influence the behavior of the corrected hiPSCs. The other gene correction approach for the sickle cell mutation β^s was published by Sebastiano and colleagues, following a similar selection-based approach (Sebastiano *et al.*, 2011). They achieved efficient targeting and showed no additional modifications in the nontargeted allele due to NHEJ and no off-target modifications.

β-Thalassemia: The correction of mutations in the *β*-globin gene was also addressed by Wang and colleagues, who performed genetic correction by ZFN-assisted HR (Wang *et al.*, 2012) and also applied a drug selection procedure to increase targeting efficiency. They were able to differentiate the corrected hiPSCs, as well as uncorrected hiPSCs, to hematopoietic progenitors. Moreover, human *β*-globin was detected in the peripheral blood of immunodeficient mice transplanted with the corrected hiPSC-derived hematopoietic progenitors, confirming the genetic correction of *β*-thalassemia.

α_1-Antitrypsin Deficiency: ZFN-mediated gene correction was also performed at the α_1-antitrypsin (*A1AT*) locus to correct *A1AT* deficiency (A1ATD) in hiPSCs derived from a patient with the Glu342Lys point mutation. This approach used a puromycin resistance cassette flanked by piggyBac inverted repeats. Subsequently, the selection cassette was removed from the homozygously targeted clones by piggyBac transposase, obtaining corrected hiPSC clones without any residual sequence footprint. Corrected, excised hiPSC clones were subsequently differentiated into hepatocyte-like cells, confirming the successful correction of A1ATD (Yusa *et al.*, 2011).

Targeted knock-in. In the targeted knock-in strategy, a full or partial cDNA of the therapeutic transgene gene is

directly introduced into the endogenous mutant gene locus, generally near the start of the gene in order to precede all or the majority of the mutant exons. Typically, splicing signals are incorporated into the transgene sequences such that expression of the introduced cDNA is regulated by the endogenous regulatory elements of the locus where it is inserted. In principle, this strategy maintains the genetic regulation of the gene and it is applicable to diseases in which a large number of distinct gene mutations occur (in contrast to a single mutant genotype responsible for a significant majority of patients). The knock-in strategy is a highly versatile HR strategy capable, in principle, of treating a large number of patients using a single set of gene-modifying tools (i.e., site-specific nucleases and repair matrices), while preserving the endogenous regulation of the therapeutic gene. Although this strategy has been used to express marker genes led by endogenous promoters (Hockemeyer *et al.*, 2009, 2011; Wang *et al.*, 2011), there is not yet any reported example of this strategy for the correction of patient-specific disease hiPSCs.

We have successfully generated, using Sendai vectorized reprogramming factors, hiPSCs from patients with pyruvate kinase deficiency (PKD), who suffer from nonspherocytic hemolytic anemia. We are pursuing a correction strategy that is capable, in principle, of treating all PKD patients with mutations from exon 3 to the end of the *PKLR* gene by developing an appropriate repair matrix. Moreover, expression of the corrected R-type pyruvate kinase (RPK) transcript will be regulated under the control of the endogenous *PKLR* promoter after the knock-in of the partial RPK cDNA into intron 2. If successful, only the corrected RPK protein should be expressed in red blood cells.

Selection of one of these previously described strategies (see the sections Targeted Safe Harbor Integration, Targeted Correction, and Targeted Knock-In, above) requires consideration of both the disease and the number of patients in whom this strategy could be used. For each of these strategies, the type of therapeutic matrix to be used will be different (see Fig. 2).

Risks of Genome Alteration

One of the most important issues in using the aforementioned novel methodologies (i.e., epigenetic reprogramming and site-specific gene correction) will concern ensuring the integrity of the chromosomal DNA. Even though these methodologies have been employed only in the limited number of studies cited previously, it is already clear that genetic abnormalities may be introduced into the hiPSCs either through the reprogramming process, the tissue culture expansion, and/or the gene correction process itself (Blasco *et al.*, 2011; Gore *et al.*, 2011; Pera, 2011).

In the study of gene correction of A1ATD hiPSCs, a complete genome integrity study was performed. Comparative genomic hybridization confirmed that reprogramming and prolonged culture generated amplifications or deletions ranging from 20 kb to 1.3 Mb. But, importantly, there was one corrected line of three lines examined that retained a normal genome. These authors also detected genetic alterations in 2 of 6 lines after HR correction and in 4 of 16 lines after the excision process. They concluded that more genetic alterations were generated during the reprogramming process and the extensive culture of the hiPSCs than during the HR correction (Yusa *et al.*, 2011).

In the study of α-synuclein gene correction (Soldner *et al.*, 2011), the authors examined the hiPSC lines for copy number variation (CNV), because CNVs were previously reported to commonly result from reprogramming as well as from prolonged pluripotent stem cell culture. The authors saw on average 77 CNVs per cell line with an average size of 158 kb. These genetic alterations were most likely generated during reprogramming and culture as there were no substantial differences after gene editing and excision. They also performed whole genome expression array analysis before and after correction and did not detect any expression pattern differences related to gene targeting, indicating that reprogramming itself had a greater impact on genome integrity than the gene-targeting procedures.

In addition to the risks of genomic alteration, it should also be taken into account that a considerable number of hiPSC differentiation protocols include the forced expression of tissue-specific transcription factors (Hanna *et al.*, 2007; Karumbayaram *et al.*, 2009; Belay *et al.*, 2010; Takayama *et al.*, 2012). These procedures constitute an additional step of genome manipulation. Similar procedures of transient expression or genome excision by means of the Cre–*loxP* system, as done for the expression of hiPSC reprogramming factors, should be used to avoid additional side effects.

Although not related to genome alterations, another potential risk of hiPSC use is their potential immunogenic properties. Some authors have argued the possibility that despite being autologous, hiPSCs could trigger an immune reaction after transplantation. The latest reports regarding this issue have shown that differentiated hiPSCs are not immunogenic at these stages (Araki *et al.*, 2013).

Concluding Remarks

The number of disease-specific hiPSC lines is increasing rapidly. Until now, only a few of them have been genetically corrected by gene-editing approaches. The unlimited proliferating capacity of hiPSCs, while maintaining pluripotent properties, allows for the application of HR techniques and the subsequent selection of properly corrected clones. Selection of the best gene-editing strategy depends on the disease to be corrected (Fig. 2 and Table 2). Targeted correction is the cleanest option. The patient mutation is corrected while leaving no exogenous elements, with the sequence of the corrected locus being indistinguishable from that of a wild-type locus. However, this approach is suitable only for a specific patient or group of patients carrying the same mutation, which limits its use. On the other hand, safe harbor integration is applicable to treat all the genetic diseases already addressed by genetic therapies with retro/lentiviral vectors. However, because the therapeutic gene loses its endogenous regulation, a specific promoter may be required to regulate its expression; in addition, the definition of a safe harbor locus might not be accurate or complete until we have a more in-depth knowledge of regulatory elements within the human genome. The knock-in strategy is an intermediate possibility in which a large number of patients with a defined disease might benefit from this strategy, reducing its development costs. Moreover, the endogenous elements of the locus will regulate expression of the therapeutic gene.

However, one concern regarding the knock-in strategy, shared with the safe harbor approach, is that the use of a specific cDNA transgene may exclude the coexpression of various splicing variants.. The election among them will depend of knowledge of the targeted locus.

Gene-editing procedures need improvements in terms of efficiency and safety before being applied in humans. The synergy between reprogramming and gene editing is prompting progress in this field of research, in which a wide spectrum of genetic diseases could be treated. Moreover, patient-specific hiPSCs are an ideal platform to improve gene-editing techniques in order to achieve the high efficiency and specificity that gene therapy needs for its future clinical use. There are still bottlenecks for their clinical application. Gene-corrected hiPSCs currently lack robust differentiation procedures to generate a variety of transplantable cells. For example, in the hematology field, the generation of hematopoietic stem cells capable of long-term reconstitution of the whole hematopoietic system has been reported (Amabile *et al.*, 2013). Unfortunately, the need for teratoma formation to obtain functional HSCs in this report avoids its potential clinical application. Another possibility is the transplantation of more mature progenitor cells or terminally differentiated cells capable of long-term survival after infusion such as T cells, erythrocytes, or platelets. Similar strategies could be followed for other tissues. In addition, the development of homologous recombination technology in hiPSCs has broken new ground for its application to other stem cells already used in clinics, such as HSCs (Lombardo *et al.*, 2007). We fully expect that future gene therapy protocols using the aforementioned methodologies will emerge.

Acknowledgments

The authors thank Dr. J.A. Bueren and V.K. Quiceno for careful reading and correction of the manuscript. This work has been supported by grants from the Ministerio de Economía y Competitividad (SAF2011-30526-C02-01), the Fondo de Investigaciones Sanitarias (RD06/0010/0015 and RD12/0019/0023), the Comunidad Autónoma de Madrid (CellCAM S2010/BMD-2420), the PERSIST European Project, the Cystic Fibrosis Foundation (B.R.D.), and the National Heart, Lung, and Blood Institute (B.R.D.). The authors also thank the Fundación Botín for promoting translational research at the Hematopoiesis and Gene Therapy Division-CIEMAT/CIBERER. Z.G. is supported by a fellowship of the Gobierno Vasco (Spain).

Author Disclosure Statement

No competing financial interests exist.

References

Agarwal, S., Loh, Y.H., McLoughlin, E.M., *et al.* (2010). Telomere elongation in induced pluripotent stem cells from dyskeratosis congenita patients. Nature 464, 292–296.

Amabile, G., Welner, R.S., Nombela-Arrieta, C., *et al.* (2013). *In vivo* generation of transplantable human hematopoietic cells from induced pluripotent stem cells. Blood 121, 1255–1264.

An, M.C., Zhang, N., Scott, G., *et al.* (2012). Genetic correction of Huntington's disease phenotypes in induced pluripotent stem cells. Cell Stem Cell 11, 253–263.

Anderson, W.F., Blaese, R.M., and Culver, K. (1990). The ADA human gene therapy clinical protocol: Points to consider: Response with clinical protocol, July 6, 1990. Hum. Gene Ther. 1, 331–362.

Anokye-Danso, F., Trivedi, C.M., Juhr, D., *et al.* (2011). Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. Cell Stem Cell 8, 376–388.

Araki, R., Uda, M., Hoki, Y., *et al.* (2013). Negligible immunogenicity of terminally differentiated cells derived from induced pluripotent or embryonic stem cells. Nature 494, 100–104.

Ban, H., Nishishita, N., Fusaki, N., *et al.* (2011). Efficient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus vectors. Proc. Natl. Acad. Sci. U.S.A. 108, 14234–14239.

Batista, L.F., Pech, M.F., Zhong, F.L., *et al.* (2011). Telomere shortening and loss of self-renewal in dyskeratosis congenita induced pluripotent stem cells. Nature 474, 399–402.

Bedel, A., Taillepierre, M., Guyonnet-Duperat, V., *et al.* (2012). Metabolic correction of congenital erythropoietic porphyria with iPSCs free of reprogramming factors. Am. J. Hum. Genet. 91, 109–121.

Belay, E., Matrai, J., Acosta-Sanchez, A., *et al.* (2010). Novel hyperactive transposons for genetic modification of induced pluripotent and adult stem cells: A nonviral paradigm for coaxed differentiation. Stem Cells 28, 1760–1771.

Blasco, M.A., Serrano, M., and Fernandez-Capetillo, O. (2011). Genomic instability in iPS: Time for a break. EMBO J. 30, 991–993.

Bock, C., Kiskinis, E., Verstappen, G., *et al.* (2011). Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell 144, 439–452.

Carroll, D. (2011). Zinc-finger nucleases: A panoramic view. Curr. Gene Ther. 11, 2–10.

Carvajal-Vergara, X., Sevilla, A., D'Souza, S.L., *et al.* (2010). Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome. Nature 465, 808–812.

Cayo, M.A., Cai, J., Delaforest, A., *et al.* (2012). "JD" iPS cell-derived hepatocytes faithfully recapitulate the pathophysiology of familial hypercholesterolemia. Hepatology 56, 2163–2171.

Chang, C.J., and Bouhassira, E.E. (2012). Zinc-finger nuclease mediated correction of α-thalassemia in iPS cells. Blood 120, 3906–3914.

Chin, M.H., Mason, M.J., Xie, W., *et al.* (2009). Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. Cell Stem Cell 5, 111–123.

Daboussi, F., Zaslavskiy, M., Poirot, L., *et al.* (2012). Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. Nucleic Acids Res. 40, 6367–6379.

Deans, R.J., and Moseley, A.B. (2000). Mesenchymal stem cells: Biology and potential clinical uses. Exp. Hematol. 28, 875–884.

Delacote, F., and Lopez, B.S. (2008). Importance of the cell cycle phase for the choice of the appropriate DSB repair pathway, for genome stability maintenance: The trans-S double-strand break repair model. Cell Cycle 7, 33–38.

Deng, J., Shoemaker, R., Xie, B., *et al.* (2009). Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat. Biotechnol. 27, 353–360.

Doi, A., Park, I.H., Wen, B., *et al.* (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat. Genet. 41, 1350–1353.

Fischer, A., Hacein-Bey-Abina, S., and Cavazzana-Calvo, M. (2010). 20 years of gene therapy for SCID. Nat. Immunol. 11, 457–460.

Fusaki, N., Ban, H., Nishiyama, A., *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. Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. 85, 348–362.

Galli, R., Gritti, A., Bonfanti, L., and Vescovi, A.L. (2003). Neural stem cells: An overview. Circ. Res. 92, 598–608.

Ghosh, Z., Wilson, K.D., Wu, Y., *et al.* (2010). Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. PLoS One 5, e8975.

Gore, A., Li, Z., Fung, H.L., *et al.* (2011). Somatic coding mutations in human induced pluripotent stem cells. Nature 471, 63–67.

Grabarz, A., Barascu, A., Guirouilh-Barbat, J., and Lopez, B.S. (2012). Initiation of DNA double strand break repair: Signaling and single-stranded resection dictate the choice between homologous recombination, non-homologous end-joining and alternative end-joining. Am. J. Cancer Res. 2, 249–268.

Guenther, M.G., Frampton, G.M., Soldner, F., *et al.* (2010). Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell 7, 249–257.

Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., *et al.* (2003). *LMO2*-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science 302, 415–419.

Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., *et al.* (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J. Clin. Invest. 118, 3132–3142.

Hanna, J., Wernig, M., Markoulaki, S., *et al.* (2007). Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. Science 318, 1920–1923.

Hanna, J., Markoulaki, S., Schorderet, P., *et al.* (2008). Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. Cell 133, 250–264.

Hanna, J., Saha, K., Pando, B., *et al.* (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature 462, 595–601.

Hockemeyer, D., Soldner, F., Beard, C., *et al.* (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat. Biotechnol. 27, 851–857.

Hockemeyer, D., Wang, H., Kiani, S., *et al.* (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat. Biotechnol. 29, 731–734.

Howden, S.E., Gore, A., Li, Z., *et al.* (2011). Genetic correction and analysis of induced pluripotent stem cells from a patient with gyrate atrophy. Proc. Natl. Acad. Sci. U.S.A. 108, 6537–6542.

Hu, K., Yu, J., Suknuntha, K., *et al.* (2011). Efficient generation of transgene-free induced pluripotent stem cells from normal and neoplastic bone marrow and cord blood mononuclear cells. Blood 117, e109–e119.

Irion, S., Luche, H., Gadue, P., *et al.* (2007). Identification and targeting of the *ROSA26* locus in human embryonic stem cells. Nat. Biotechnol. 25, 1477–1482.

Israel, M.A., Yuan, S.H., Bardy, C., *et al.* (2012). Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. Nature 482, 216–220.

Jia, F., Wilson, K.D., Sun, N., *et al.* (2010). A nonviral minicircle vector for deriving human iPS cells. Nat. Methods 7, 197–199.

Kaji, K., Norrby, K., Paca, A., *et al.* (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature 458, 771–775.

Karumbayaram, S., Novitch, B.G., Patterson, M., *et al.* (2009). Directed differentiation of human-induced pluripotent stem cells generates active motor neurons. Stem Cells 27, 806–811.

Khan, I.F., Hirata, R.K., Wang, P.R., *et al.* (2010). Engineering of human pluripotent stem cells by AAV-mediated gene targeting. Mol. Ther. 18, 1192–1199.

Kim, D., Kim, C.H., Moon, J.I., *et al.* (2009). Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. Cell Stem Cell 4, 472–476.

Kim, J.B., Zaehres, H., Arauzo-Bravo, M.J., and Scholer, H.R. (2009). Generation of induced pluripotent stem cells from neural stem cells. Nat. Protoc. 4, 1464–1470.

Kim, K., Zhao, R., Doi, A., *et al.* (2011). Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. Nat. Biotechnol. 29, 1117–1119.

Kim, K.Y., Hysolli, E., and Park, I.H. (2011). Neuronal maturation defect in induced pluripotent stem cells from patients with Rett syndrome. Proc. Natl. Acad. Sci. U.S.A. 108, 14169–14174.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: Zinc finger fusions to *Fok* I cleavage domain. Proc. Natl. Acad. Sci. U.S.A. 93, 1156–1160.

Kumano, K., Arai, S., Hosoi, M., *et al.* (2012). Generation of induced pluripotent stem cells from primary chronic myelogenous leukemia patient samples. Blood 119, 6234–6242.

Kunisato, A., Wakatsuki, M., Shinba, H., *et al.* (2010). Direct generation of induced pluripotent stem cells from human nonmobilized blood. Stem Cells Dev. 20, 159–168.

Li, T., Huang, S., Zhao, X., *et al.* (2011). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res. 39, 6315–6325.

Liu, G.H., Suzuki, K., Qu, J., *et al.* (2011). Targeted gene correction of laminopathy-associated LMNA mutations in patient-specific iPSCs. Cell Stem Cell 8, 688–694.

Liu, J., Sumer, H., Leung, J., *et al.* (2010). Late passage human fibroblasts induced to pluripotency are capable of directed neuronal differentiation. Cell Transplant. 20, 193–203.

Liu, T., Zou, G., Gao, Y., *et al.* (2012). High efficiency of reprogramming $CD34^+$ cells derived from human amniotic fluid into induced pluripotent stem cells with Oct4. Stem Cells Dev. 21, 2322–2332.

Lombardo, A., Genovese, P., Beausejour, C.M., *et al.* (2007). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat. Biotechnol. 25, 1298–1306.

Marchetto, M.C., Yeo, G.W., Kainohana, O., *et al.* (2009). Transcriptional signature and memory retention of human-induced pluripotent stem cells. PLoS One 4, e7076.

Matrai, J., Cantore, A., Bartholomae, C.C., *et al.* (2011). Hepatocyte-targeted expression by integrase-defective lentiviral vectors induces antigen-specific tolerance in mice with low genotoxic risk. Hepatology 53, 1696–1707.

Meng, X., Neises, A., Su, R.J., *et al.* (2012). Efficient reprogramming of human cord blood $CD34^+$ cells into induced pluripotent stem cells with *OCT4* and *SOX2* alone. Mol. Ther. 20, 408–416.

Moretti, A., Bellin, M., Welling, A., *et al.* (2010). Patient-specific induced pluripotent stem-cell models for long-QT syndrome. N. Engl J. Med. 363, 1397–1409.

Mussolino, C., Morbitzer, R., Lutge, F., *et al.* (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res. 39, 9283–9293.

Newman, A.M., and Cooper, J.B. (2010). Lab-specific gene expression signatures in pluripotent stem cells. Cell Stem Cell 7, 258–262.

Ohi, Y., Qin, H., Hong, C., *et al.* (2011). Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nat. Cell Biol. 13, 541–549.

Papapetrou, E.P., and Sadelain, M. (2011). Generation of transgene-free human induced pluripotent stem cells with an excisable single polycistronic vector. Nat. Protoc. 6, 1251–1273.

Papapetrou, E.P., Lee, G., Malani, N., *et al.* (2011). Genomic safe harbors permit high β-globin transgene expression in thalassemia induced pluripotent stem cells. Nat. Biotechnol. 29, 73–78.

Paques, F., and Duchateau, P. (2007). Meganucleases and DNA double-strand break-induced recombination: Perspectives for gene therapy. Curr. Gene Ther. 7, 49–66.

Park, I.H., Arora, N., Huo, H., *et al.* (2008). Disease-specific induced pluripotent stem cells. Cell 134, 877–886.

Pera, M.F. (2011). Stem cells: The dark side of induced pluripotency. Nature 471, 46–47.

Porteus, M.H., and Carroll, D. (2005). Gene targeting using zinc finger nucleases. Nat. Biotechnol. 23, 967–973.

Ramos-Mejia, V., Montes, R., Bueno, C., *et al.* (2012). Residual expression of the reprogramming factors prevents differentiation of iPSC generated from human fibroblasts and cord blood $CD34^+$ progenitors. PLoS One 7, e35824.

Rashid, S.T., Corbineau, S., Hannan, N., *et al.* (2010). Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. J. Clin. Invest. 120, 3127–3136.

Raya, A., Rodriguez-Piza, I., Guenechea, G., *et al.* (2009). Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. Nature 460, 53–59.

Robbins, J. (1993). Gene targeting: The precise manipulation of the mammalian genome. Circ. Res. 73, 3–9.

Sebastiano, V., Maeder, M.L., Angstman, J.F., *et al.* (2011). *In situ* genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. Stem Cells 29, 1717–1726.

Seki, T., Yuasa, S., and Fukuda, K. (2011). Derivation of induced pluripotent stem cells from human peripheral circulating T cells. Curr. Protoc. Stem Cell Biol. Chapter 4, Unit4A 3.

Soldner, F., Hockemeyer, D., Beard, C., *et al.* (2009). Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. Cell 136, 964–977.

Soldner, F., Laganiere, J., Cheng, A.W., *et al.* (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell 146, 318–331.

Somers, A., Jean, J.C., Sommer, C.A., *et al.* (2010). Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette. Stem Cells 28, 1728–1740.

Sommer, C.A., Sommer, A.G., Longmire, T.A., *et al.* (2010). Excision of reprogramming transgenes improves the differentiation potential of iPS cells generated with a single excisable vector. Stem Cells 28, 64–74.

Song, H., Chung, S.K., and Xu, Y. (2010). Modeling disease in human ESCs using an efficient BAC-based homologous recombination system. Cell Stem Cell 6, 80–89.

Staerk, J., Dawlaty, M.M., Gao, Q., *et al.* (2010). Reprogramming of human peripheral blood cells to induced pluripotent stem cells. Cell Stem Cell 7, 20–24.

Sun, N., Panetta, N.J., Gupta, D.M., *et al.* (2009). Feeder-free derivation of induced pluripotent stem cells from adult human adipose stem cells. Proc. Natl. Acad. Sci. U.S.A. 106, 15720–15725.

Sun, N., Yazawa, M., Liu, J., *et al.* (2012). Patient-specific induced pluripotent stem cells as a model for familial dilated cardiomyopathy. Sci. Transl. Med. 4, 130ra47.

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

Takahashi, K., Tanabe, K., Ohnuki, M., *et al.* (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861–872.

Takayama, K., Inamura, M., Kawabata, K., *et al.* (2012). Efficient generation of functional hepatocytes from human embryonic stem cells and induced pluripotent stem cells by HNF4α transduction. Mol. Ther. 20, 127–137.

Tanaka, T., Takahashi, K., Yamane, M., *et al.* (2012). Induced pluripotent stem cells from CINCA syndrome patients as a model for dissecting somatic mosaicism and drug discovery. Blood.

Thier, M., Munst, B., and Edenhofer, F. (2010). Exploring refined conditions for reprogramming cells by recombinant Oct4 protein. Int. J. Dev. Biol. 54, 1713–1721.

Thomas, E.D., Fefer, A., Buckner, C.D., and Storb, R. (1977). Current status of bone marrow transplantation for aplastic anemia and acute leukemia. Blood 49, 671–681.

Torres, R., Garcia, A., Paya, M., and Ramirez, J.C. (2011). Non-integrative lentivirus drives high-frequency Cre-mediated cassette exchange in human cells. PLoS One 6, e19794.

Valton, J., Daboussi, F., Leduc, S., *et al.* (2012a). 5′-Cytosine-phosphoguanine (CpG) methylation impacts the activity of natural and engineered meganucleases. J. Biol. Chem. 287, 30139–30150.

Valton, J., Dupuy, A., Daboussi, F., *et al.* (2012b). Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. J. Biol. Chem. 287, 38427–38432.

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

Wang, Y., Zheng, C.G., Jiang, Y., *et al.* (2012). Genetic correction of β-thalassemia patient-specific iPS cells and its use in improving hemoglobin production in irradiated SCID mice. Cell Res. 22, 637–648.

Warren, L., Manos, P.D., Ahfeldt, T., *et al.* (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell 7, 618–630.

Woltjen, K., Hamalainen, R., Kibschull, M., *et al.* (2011). Transgene-free production of pluripotent stem cells using piggyBac transposons. Methods Mol. Biol. 767, 87–103.

Yamanaka, S. (2012). Induced pluripotent stem cells: Past, present, and future. Cell Stem Cell 10, 678–684.

Yanez-Munoz, R.J., Balaggan, K.S., MacNeil, A., *et al.* (2006). Effective gene therapy with nonintegrating lentiviral vectors. Nat. Med. 12, 348–353.

Yao, Y., Nashun, B., Zhou, T., *et al.* (2011). Generation of $CD34^+$ cells from CCR5-disrupted human embryonic and induced pluripotent stem cells. Hum. Gene Ther. 23, 238–242.

Yazawa, M., Hsueh, B., Jia, X., *et al.* (2011). Using induced pluripotent stem cells to investigate cardiac phenotypes in Timothy syndrome. Nature 471, 230–234.

Ye, Z., Zhan, H., Mali, P., *et al.* (2009). Human-induced pluripotent stem cells from blood cells of healthy donors and patients with acquired blood disorders. Blood 114, 5473–5480.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., *et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science 318, 1917–1920.

Yu, J., Hu, K., Smuga-Otto, K., *et al.* (2009). Human induced pluripotent stem cells free of vector and transgene sequences. Science 324, 797–801.

Yui, S., Nakamura, T., Sato, T., *et al.* (2012). Functional engraftment of colon epithelium expanded *in vitro* from a single adult $Lgr5^+$ stem cell. Nat. Med. 18, 618–623.

Yusa, K., Rashid, S.T., Strick-Marchand, H., *et al.* (2011). Targeted gene correction of α_1-antitrypsin deficiency in induced pluripotent stem cells. Nature 478, 391–4.

Zhou, W., and Freed, C.R. (2009). Adenoviral gene delivery can reprogram human fibroblasts to induced pluripotent stem cells. Stem Cells 27, 2667–2674.

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

Zou, J., Sweeney, C.L., Chou, B.K., *et al.* (2011b). Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: Functional correction by zinc finger nuclease-mediated safe harbor targeting. Blood 117, 5561–5572.

Address correspondence to:
Dr. Jose C. Segovia
Differentiation and Cytometry Unit
Hematopoiesis and Gene Therapy Division
CIEMAT-CIBERER
Av. Complutense 22
28040 Madrid
Spain

E-mail: jc.segovia@ciemat.es

Received for publication December 23, 2012;
accepted after revision May 3, 2013.

Published online: May 15, 2013.