



UNIVERSITY OF MILANO

BICOCCA SCHOOL OF MEDICINE AND

FACULTY OF SCIENCE

PhD PROGRAM IN TRANSLATIONAL AND MOLECULAR  
MEDICINE-DIMET

**Targeted Gene Correction and Reprogramming of SCID-X1  
Fibroblasts to Rescue IL2RG Expression in iPSC-derived  
Hematopoietic Cells**

Coordinator: Ch.mo Prof. Andrea Biondi

Tutor: Ch.ma Prof. Giuliana Ferrari

Co-Tutor: Ch.mo Prof. Luigi Naldini

Dottoranda: Claudia Firrito

XXVIII CYCLE

ACADEMIC YEAR 2014-2015

The work presented in this thesis was performed at the Telethon Institute for Gene therapy (HSR-TIGET), Department of Regenerative Medicine, Stem Cells and Gene Therapy, San Raffaele Scientific Institute, in the laboratory directed by Prof. Luigi Naldini, with the supervision of Dr. Angelo Lombardo.

Era necessario filtrare via quanto di personale e casuale vi era in tutte quelle sensazioni per poter raggiungere il puro fluido, l'olio essenziale della verità.

Virginia Woolf



## TABLE OF CONTENTS

|                                                                                    |    |
|------------------------------------------------------------------------------------|----|
| CHAPTER 1.....                                                                     | 7  |
| 1.1 Gene therapy.....                                                              | 8  |
| 1.1.1 Vectors for gene-based therapies.....                                        | 9  |
| 1.1.1.1 Non-viral vectors.....                                                     | 9  |
| 1.1.1.2 Viral vectors.....                                                         | 11 |
| 1.1.1.3 Expression and regulation of the transgene.....                            | 18 |
| 1.2 Hematopoietic stem cell therapy .....                                          | 19 |
| 1.3 Integrating vectors-based clinical trials.....                                 | 21 |
| 1.3.1 Gamma-retroviral based clinical trials.....                                  | 21 |
| 1.3.2 Lentiviral based clinical trials.....                                        | 26 |
| 1.3.3 Insertional genotoxicity .....                                               | 32 |
| 1.4 Tailored genome editing.....                                                   | 35 |
| 1.4.1 Gene targeting: an historical perspective.....                               | 36 |
| 1.4.2 The DNA damage response: the DNA Double strand<br>breaks repair pathway..... | 40 |
| 1.4.2.1 Non homologous end joining pathway.....                                    | 45 |
| 1.4.2.2 Homologous recombination pathway.....                                      | 46 |
| 1.4.3 Inducing site-specific DSBs.....                                             | 48 |
| 1.4.3.1 The Zinc Finger Nuclease technology.....                                   | 49 |
| 1.4.3.2 The Transcription Activator Like Effector<br>nuclease technology.....      | 59 |
| 1.4.3.3. The RNA-guided engineered nucleases.....                                  | 61 |
| 1.4.5 Therapeutic applications of genome editing.....                              | 64 |

|                                                                                     |                |
|-------------------------------------------------------------------------------------|----------------|
| <b>1.5 Induced pluripotent stem cells: a new era for regenerative medicine.....</b> | <b>70</b>      |
| <b>1.5.1 Advances in cell reprogramming technologies.....</b>                       | <b>72</b>      |
| <b>1.5.2 Molecular and epigenetic reprogramming.....</b>                            | <b>73</b>      |
| <b>1.5.3 Road to the future: iPSCs clinical applications.....</b>                   | <b>75</b>      |
| <b>1.5.4 IPSC as alternative source of HSCs.....</b>                                | <b>81</b>      |
| <br><b>Scope of the thesis.....</b>                                                 | <br><b>85</b>  |
| <b>References.....</b>                                                              | <b>87</b>      |
| <br><b>CHAPTER 2.....</b>                                                           | <br><b>112</b> |
| <b>Summary.....</b>                                                                 | <b>113</b>     |
| <b>Introduction.....</b>                                                            | <b>114</b>     |
| <b>Results.....</b>                                                                 | <b>117</b>     |
| <b>Discussion.....</b>                                                              | <b>143</b>     |
| <b>Experimental procedure.....</b>                                                  | <b>147</b>     |
| <b>Supplementary figures.....</b>                                                   | <b>161</b>     |
| <b>References.....</b>                                                              | <b>169</b>     |
| <br><b>CHAPTER 3.....</b>                                                           | <br><b>173</b> |
| <b>Final Discussion.....</b>                                                        | <b>175</b>     |
| <b>Publications.....</b>                                                            | <b>179</b>     |
| <b>Acknowledgments.....</b>                                                         | <b>181</b>     |

## **CHAPTER 1**

## **1.1 Gene-Therapy**

Thanks to the substantial progress that has been made in the development of gene transfer technologies, gene therapy is emerging as promising therapeutic option for the treatment of several genetic disorders, including monogenic and non-monogenic disorders, and, more recently, for acquired diseases. Gene therapy is defined as the introduction of nucleic acids within the cells by means of vectors, with the aim of altering gene expression to prevent, halt or reverse a pathological process and can be accomplished through different gene-modifications: gene replacement or gene addition, gene correction/alteration, gene knockdown and gene knockout (Kay, 2011). Gene replacement, the most common adopted approach, allows for the replacement of a functional copy of a gene whose endogenous counterpart is functionally inactive, while gene addition usually refers to the transfer of a therapeutic gene or a selectable marker for cell isolation, proliferation or conditional elimination. Gene knockdown exploits microRNA-mediated gene regulation to module or downregulate a gene product using RNAi, which can be induced both by short hairpin RNAs (shRNAs) and by short dsRNAs with complete complementarity to the target mRNA (Kay, 2011). Gene knockout, instead, allows for permanent inactivation of a gene function. To achieve the final goal, the treatment of a disease, several hurdles must be overcome. First of all the vector of choice must ensure good level of gene transfer, though the delivery of the

genetic information into the cells, without any interference with the endogenous transcriptional and regulatory machinery.

Nucleic acids are delivered within the cells through vectors, distinguished into viral vectors, which harness the viral infection pathway, and non-viral vectors.

Moreover, vector administration may be carried out *in vivo*, through a local or systemic delivery, or *ex vivo*, using cells isolated either from the patient or from normal donors. *Ex vivo* therapies are often preferable, since they allow for the selection, expansion and/or differentiation of the transduced cells, while avoiding direct exposure of the patient to the gene transfer vector (Naldini, 2011).

Once the gene transfer has been achieved, the genetically modified cells should ensure the achievement of a therapeutic threshold to revert the disease, the gene-transmission to the cell-progeny, long-term survival and immunological escape (Naldini, 2015).

### **1.1.1 Vectors for gene-based therapy**

In order to achieve therapeutic success, transfer vehicles for gene therapy must be able to transduce target cells while avoiding impact on non-target cells. Two main vector systems have been used so far to deliver the genetic material within the cells: viral vectors and non-viral vectors.

#### **1.1.1.1 Non-viral vectors**

Non-viral vectors are composed of plasmidic DNA molecule and the delivery platform. Since they do not contain any viral component, they do not activate pre-existing immune response;

furthermore there is no limit in the size of the DNA molecule they can allocate. Although non-viral vectors are able to achieve *ex vivo* gene transfer, they are quite inefficient for gene-delivery in most of the *in vivo* applications (Kay, 2011).

Nucleic acids, to be delivered, have been administered in association with polycationic-lipid-based systems or with polylysine-, polyethylenimine-based macromolecules or polysaccharides. However administration of methylated CpG DNA molecules can induce toxicity by activating immune responses, while DNA-macromolecular complexes may be too large to cross vascular endothelial barriers or may be degraded by the endosomal compartments (Kay, 2011).

Many routes have been exploited to deliver DNA-complexes, as oral ingestion or intravenous infusion; more recently other delivery technologies have been applied, as hydrodynamic transfection for hepatic delivery or electroporation of ultrasound-guided DNA, but the relevance for clinical application is not yet well established (Suda, 2007).

Although plasmids exist in an episomal form, they may integrate into the host cell-genome by two different mechanisms. The first, based on DNA transposons or class II transposable elements, exploits a transiently expressed transposase to allow integration into the host chromosome. A second mechanism relies on the use of bacteriophage recombinases, transiently provided, to achieve integration into pseudosites which resemble the prokaryotic integration sequences. The main disadvantage of DNA vectors is the low transgene expression level, which decreases over weeks.

For this purpose minicircle DNAs are often used to achieve higher level of transgene expression, which is 100 to 1000 higher than those achieved with standard plasmids. Thanks to this feature, they will likely replace routine plasmids for clinical applications (Kay, 2011).

#### **1.1.1.2 Viral vectors**

Viruses have evolved a specific machinery to deliver their genes into cells or to integrate into the host cell genome. As such, viral vectors are the most suitable vehicles for highly efficient gene transfer and for sustained expression of the transgene in primary cells. At present, there are two main vector systems that can mediate delivery of the genetic material into recipient cells: integrating vectors such as retroviral vector (RV) (Maier et al., 2010) and lentiviral vectors (LV) (Matrai et al., 2011; Naldini, 2011), and non-integrating vectors as adenoviral (AdV) (Segura et al., 2008) and adeno-associated vectors (AAV) (Deyle and Russell, 2009), that are essentially maintained as episomal genomes and eventually diluted upon cell proliferation. According to these properties, non-integrating vectors have been exploited to achieve persistent transgene expression in non-dividing cells.

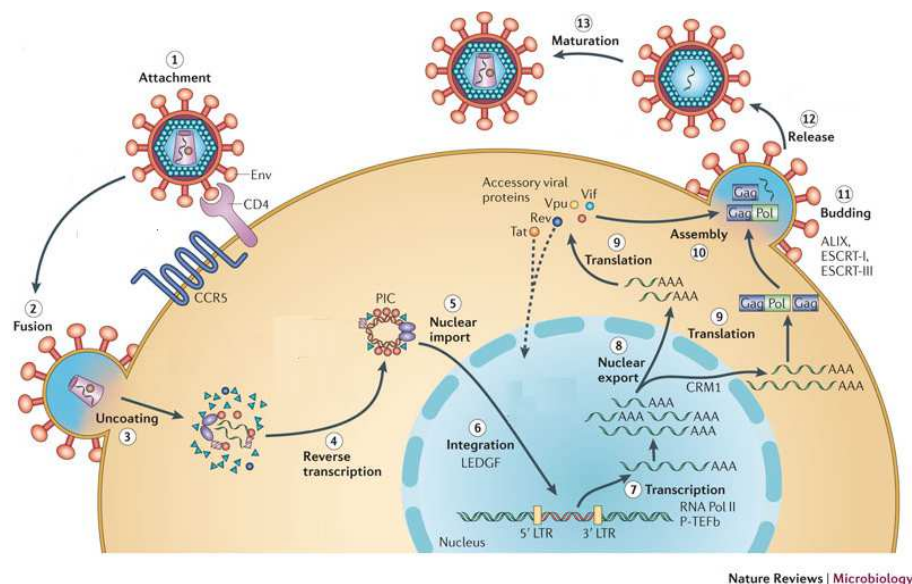
Adenoviral vectors, with their 36 kb-vector genome, which remains episomal, can allocate very large sequences and can achieve good levels of systemic gene transfer. They have been extensively modified by deleting the early genes to reduce toxicity due to an innate immune response, which is induced especially when multiple administrations are required.

Single-stranded, non-pathogenic Adeno-Associated viral vectors (AAVs) are derived from small simple DNA viruses composed of two genes: rep, required for viral replication, and cap, responsible for the packaging of the viral genome. Eight different serotypes have been isolated, capable to transduce different cell types with variable efficiencies, but almost all the rAAV vectors used so far have been derived from AAV2. To allocate the transgene cassette, whose size is restricted to 4 kb, rep and cap have been removed, while the viral inverted terminal repeats are retained. Moreover, by exploiting in vivo concatamerization of AAV, the packaging capacity has been increased. The vector can be pseudotyped, produced at high titres and used to transduce both dividing and non-dividing cells, always remaining in episomal form as linear monomers, circular monomers and linear concatamers (Key, 2011). HSV-1, with its high capacity, is the largest of all viruses to be adopted for gene therapy. In fact it can allocate foreign sequences up to 40 kb, thus allowing the delivery of multiple single cassette or very large cassette. Moreover, because of its neurotropism, it has been applied for the treatment of neuropathological disorders (Thomas, 2003).

Integrating vectors as RVs and LVs, thanks to their ability to stably integrate in the host genome, are the best tool for gene transfer into actively proliferating cells as hematopoietic stem cells (HSCs), potentially allowing a lifelong expression of the corrected gene product in the cells and its progeny.

Both RVs and LVs are derived from viruses belonging to the *retroviridae* family. This is a large family of enveloped RNA

viruses found in all vertebrates. The retroviral genome is composed by a homodimer of linear, positive-sense, single-stranded RNAs of 7 to 11 kilobases (Kb), surrounded by a cone-shaped protein core. After infection of the host cell, the RNA genome is reverse transcribed into double-stranded DNA and then becomes integrated in the host genome in a semi-random manner (Lewinski and Bushman, 2005). In the retroviral life-cycle, the genetic information goes from RNA to DNA and exists in two different forms, as genomic RNA when inside the viral particle and as proviral double-stranded DNA when integrated in the host genome (Engelman and Cherepanov, 2012) (Figure 1.1).



**Figure 1.1. HIV-1 life cycle.** The incoming virion particle contains two single stranded RNA surrounded by a lipid bilayer. Upon receptor binding and fusion of the viral and cellular membrane the core is release into the cytoplasm (Entry). The core is then disassembled (uncoating) and the RNA genome is retrotranscribed in linear dsDNA (Reverse Transcription), giving rise to the viral pre-integration complex (PIC). The PIC is transported through the cytoplasm and

through the nuclear membrane. The viral DNA integrates into the host chromatin to form the provirus. (A) In the wild type HIV the integrated provirus lead expression of the viral genes that generate the new infective particles (Engelman and Cherepanov, 2012).

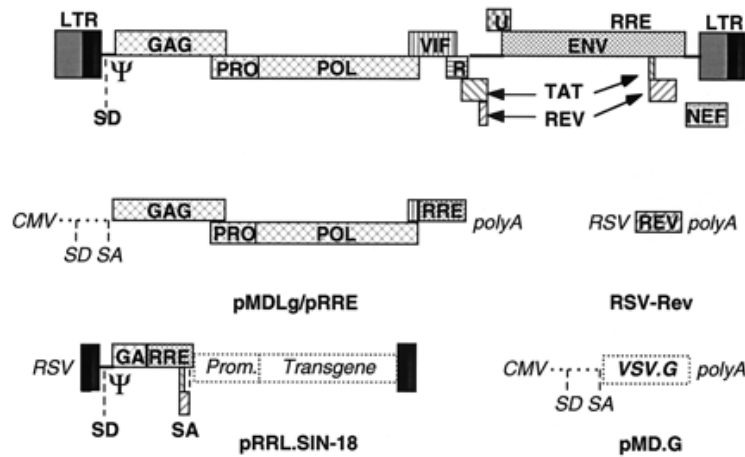
Taxonomic classification of retroviruses distinguishes three groups: C-type retroviruses (oncoretrovirus), lentiviruses and spumaviruses (or foamyviruses). While the oncoretroviral genome encodes only for viral structural proteins, lentiviruses and foamyviruses present a more complex structure, encoding for additional regulatory and accessory proteins. To date, vectors for gene transfer have been developed from viruses belonging to each of these subfamilies. However, while oncoretroviral and lentiviral vectors have been long studied, spumavirus-derived vectors have been applied only recently in the field (Bauer et al., 2008; Taylor et al., 2008).

The basic principle of turning these viruses into gene delivery systems relies on the spatial segregation of cis- and trans-acting sequences in different constructs during the vector production. The development of oncoretroviral and lentiviral vectors has followed the same three prototypic stages, referred to as generations.

Cis-acting sequences are required for genome-packaging, reverse transcription, nuclear translocation and provirus integration into recipient cell's genomic DNA, and are maintained in the transfer construct together with the transgene expression cassette.

Trans-acting sequences are included in the packaging construct and complement the transfer construct in the producer cells, encoding both structural and enzymatic viral proteins. This allows the packaging of a virion similar to the parental virus, but containing only the transfer construct-derived viral genome (transfer vector),

which is the only one carrying the packaging signal. While the gene of interest, together with its transcriptional regulatory elements, is inserted into the vector construct, all pathogenic and dispensable genes are deleted from the viral genome (Figure 1.2).



**Figure 1.2. Four plasmids system for the production of third generation LVs.** Schematic drawing of the HIV provirus and the four constructs used to make a lentivirus vector of the third generation. The viral LTRs, the reading frames of the viral genes, the major 5' splice donor site (SD), the packaging sequence (Ψ), and the RRE are boxed and indicated in bold type. The conditional packaging construct, pMDLg/pRRE, expresses the gag and pol genes from the CMV promoter and intervening sequences and polyadenylation site of the human β-globin gene. As the transcripts of the gag and pol genes contain cis-repressive sequences, they are expressed only if Rev promotes their nuclear export by binding to the RRE. All tat and rev exons have been deleted, and the viral sequences upstream of the gag gene have been replaced. A non-overlapping construct, RSV-Rev, expresses the rev cDNA. The transfer construct, pRRL.SIN-18, contains HIV-1 cis-acting sequences and an expression cassette for the transgene. It is the only portion transferred to the target cells and does not contain wild-type copies of the HIV LTR. The 5' LTR is chimeric, with the enhancer/promoter of RSV replacing the U3 region (RRL) to rescue the transcriptional dependence on Tat. The 3' LTR has an almost complete deletion of the U3 region, which includes the TATA box (from nucleotides -418 to -18 relative to the U3/R border). As the latter is the template used to generate both copies of the LTR in the integrated provirus, transduction of this vector results in transcriptional inactivation of both LTRs; thus, it is a self-inactivating vector (SIN-18). The fourth construct, pMD.G, encodes a heterologous envelope to

pseudotype the vector, here shown coding for VSV G. Only the relevant parts of the constructs are shown (Dull et al., 1998).

RV vectors based on the Moloney-murine Leukemia Virus (MLV) were the first viral delivery systems to be developed for gene therapy applications (Kohn et al., 1987). These vectors have a good infectivity, however, since they can enter in the cell nucleus only when nuclear membrane breaks down, their application is limited to *ex vivo* gene transfer approaches for dividing cells. Moreover, because only a fraction of the cells divide in a given time in culture, multiple hits of transduction or cytokine stimulation are required to infect most of the cells (Nolta and Kohn, 1990). This has been shown to be detrimental for HSCs, whose prolonged maintenance in culture can induce their differentiation and exhaustion (Dorrell et al., 2000).

Many efforts have been focusing on the development of lentiviral vector, which have been derived from human immunodeficiency virus (HIV) and primate lentiviruses. Unlike retroviruses, lentiviral cDNA interacts with the viral pre-initiation complex (PIC) to translocate across the intact nuclear membrane, thus being able to transduce also non-dividing cell (Follenzi et al., 2000; Miyoshi et al., 1999; Suzuki and Craigie, 2007). Because of this property, they have been used to *ex vivo* transduce HSCs without the need for cytokines stimulation, as well as cells from muscles, liver and central nervous system (CNS) (Thomas, 2003).

In order to improve LV biosafety, deletions have been made in the viral long terminal repeats (LTRs) to create self-inactivating (SIN) LV (Zufferey et al., 1998). This modification abrogates the

function of the native viral promoter/enhancer and diminishes the possibility of vector mobilization. It further reduces the transactivation of neighbouring genes, thus increasing vector safety (Montini et al., 2006). In this context, works on the safety of viral vector integration have suggested that LV with SIN LTRs and with a moderate internal promoter are less genotoxic in respect to a prototypic MLV-based RV (Modlich et al., 2009; Montini et al., 2009; Montini et al., 2006; Zychlinski et al., 2008). Montini et al. used a sensitive tumour-prone mouse model to compare the *in vivo* genotoxic potential of RV and LV in murine HSCs and showed reduced genotoxic potential of SIN LV respect to RV, endorsing its use for clinical purposes. In a more recent work the same authors demonstrated that the integration site selection and the vector features, such as active LTRs, are the major determinants in defining the genotoxic potential of an integrating vector, and that LV and RV with inactive LTRs results to be significantly less genotoxic than their counterparts with active LTRs (Montini et al., 2009). Because of these and other features, LVs have been used to correct a variety of genetic diseases in animal models and have now moved to application in clinical trials (Biffi et al, 2014; Aiuti et al. 2014).

Viral cellular tropism, dictated by the interactions of the viral envelope glycoproteins with the cellular receptors on the target cells, is another important aspect for vector development. Indeed both RVs and LVs are able to incorporate proteins, provided in trans during vector production, from related and unrelated viruses in their envelope, in a process referred as pseudotyping. This

procedure can be exploited to allow broadening or selectively restricting the vector tropism (Anliker et al., 2010; Buchholz et al., 2009; Cronin et al., 2005; Frecha et al., 2008).

One of the major hurdles of gene therapy the immune response is that can be elicited by the vector itself or by the transgene product, when is recognized as foreign. Vectors can be neutralized by circulating antibodies, thus precluding efficient transduction while transgene or viral gene products can induce cytotoxic T-lymphocytes. Also the capsid itself, as in the case of AdVs, may cause a cytokine-mediated inflammatory response as well as a humoral response (Thomas, 2003).

#### **1.1.1.3 Expression and regulation of the transgene**

In the vast majority of the cases, the most important goal is to obtain the appropriate transgene expression pattern. To this regard, it is often required that the correct transgene expression level is achieved in a tissue/stage-specific manner. Thus, many efforts have been made to optimize the reconstitution of the promoter region (Goverdhana et al., 2005; Maston et al., 2006; Reynolds et al., 2001) or to introduce post-transcriptional levels of regulation.

Moreover each integration site may influence the level of transgene expression through local epigenetic mechanisms, while cell differentiation may lead to chromatin remodelling. Expression of the transgene by  $\gamma$ -retroviral vectors can be altered by methylation, which can occur in the LTR, while SIN-lentiviral vector seem to be more resistant to this phenomenon (Ellis, 2005).

For instance, by including a microRNA- target sequence at the 3' UTR of the transgene, it is possible to suppress its expression in the cells that express that microRNA through post-transcriptional mechanisms, thus de-targeting transgene expression in the unwanted cell types (Brown et al., 2007; Brown et al., 2006; Gentner et al., 2010). However, these strategies are not always feasible because it might be difficult or impracticable to identify and reconstruct all the regulatory elements needed for tissue specific transcription of the transgene.

## **1.2 Hematopoietic stem cell therapy**

Hematopoietic stem cells (HSCs) with their self-renewing properties and the gene transmission to their progeny, are the best target for ex-vivo gene therapy. For HSC gene therapy, hematopoietic progenitors from the bone marrow, cord blood or mobilized peripheral blood are purified from leukocytes by CD34 surface marker expression, maintained in culture for few days in presence of growth-stimulating cytokines and exposed to the vector containing the therapeutic transgene. Before administration of the cells, the recipient may undergo to pre-conditioning regimen, to deplete both endogenous progenitors and mature cells in the bone marrow, thus promoting gene transferred cell-engraftment. This treatment can expose the patients to toxicity related to mucosal damage, hematopoietic cells depletion and transient immunodeficiency; in some cases are also able to induce microglia

progenitors depletion. The major issues for HSCs-gene therapy include efficient gene transfer into long-term repopulating HSCs, their ex vivo expansion, which may impact on the reconstitution potential, and the risk of haematological malignancies, related to insertional oncogenesis and unregulated transgene expression (Naldini, 2011). Indeed the level of hematopoietic reconstitution achieved depends also on the disease, the correction of the lineages, the preconditioning regimen and the vector adopted for gene transfer. Past trials of  $\gamma$ -RV-based HSC gene therapy have showed so far low level and transient hematopoietic reconstitution, suggesting the limited ability of  $\gamma$ -RVs to transduce the most primitive progenitor cells. Conversely, LVs-based clinical trials have shown high and stable-level of reconstitution, achieving up to 90% in some recipients. Moreover, since transfer has been observed in all hematopoietic lineages, gene correction has likely occurred in self renewing multipotent hematopoietic progenitor. Despite the high number of patients treated with LV-based HSC gene therapy, no adverse events have been reported up to now, although follow-up is still ongoing (Naldini, 2015).

Among the application of hematopoietic cell-based gene therapy, T-cells represent another cell type of interest especially for cancer therapies or treatment of infectious diseases. These cells can be harvested from the peripheral blood of the patient, culture and transduced with vector expressing antiviral proteins, T-cell antigen receptors (TCR) or chimeric antigen receptors, thus conferring new specificities to T-cells to target cancer or infected cells (Kochenderfer et al., 2015).

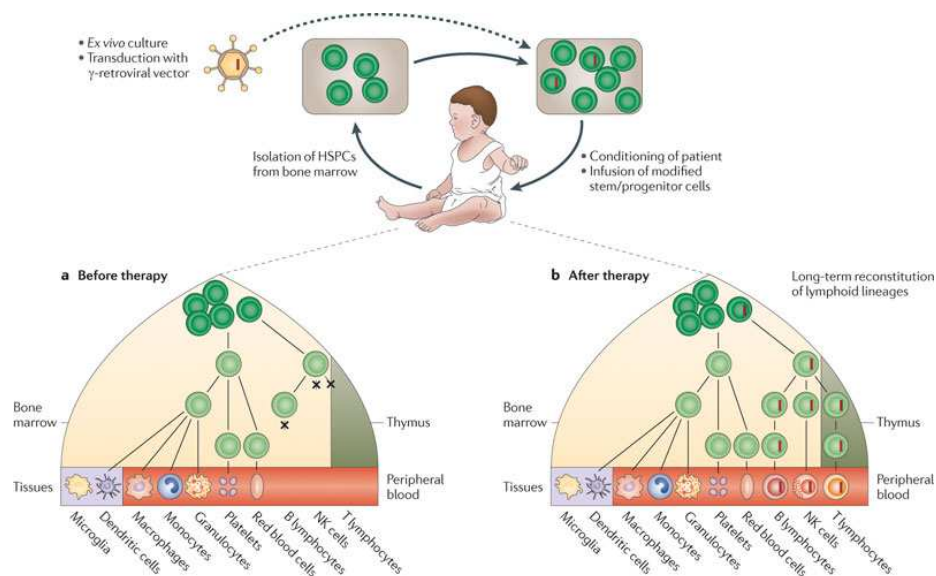
## **1.3 Integrating-vectors based clinical trials**

From 1989, when the first clinical trial for gene therapy was approved, hundreds of clinical studies have been performed all over the world, between success and failure.

### **1.3.1 Gamma-RV based clinical trials**

The first successful clinical trial based on a gene-therapy approach was reported by Fischer and Cavazzana's group for the treatment of X-linked Severe combined Immunodeficiency (SCID-X1), a rare immunological disorder characterized by complete lack of T-cells and Natural Killer (NK) cells and profound B-cells abnormalities. SCID-X1 is caused by naturally occurring mutations in the cytokine receptor common gamma chain. Because this receptor, encoded by *IL2RG* gene, is part of several interleukin receptors for IL-2, IL-4, IL-7, IL-9, IL-15, IL-21, its genetic inactivation leads to both cellular and humoral immune deficiency. SCID-X1 is invariably fatal, unless treated, and the only life-saving option is HLA-matched bone marrow transplantation. In 1999 nine paediatric patients, who lacked an HLA-identical donor, were enrolled for ex vivo  $\gamma$ -RV mediated gene replacement with autologous bone marrow (BM) derived-stem cells (Cavazzana-Calvo et al., 2000; Cavazzana-Calvo and Fischer, 2007). In 2004 a similar SCID-X1 clinical trial started in UK conducted by Trasher and his group (Gaspar et al., 2004, Trasher et al., 2005). Briefly patients underwent transplantation of autologous CD34<sup>+</sup> cells transduced with retroviral vector containing *IL2RG* cDNA without

pre-conditioning regimen. In both studies reconstitution of cellular and humoral immunity was achieved in most of the patients. The key for the successful achievements was clearly related to the selective growth advantage of the genetically modified cells, thus allowing reconstitution of the immune function and correction of the disease phenotype (Figure 1.3).



**Figure 1.3 Schematics of a prototypic protocol for *ex vivo*-based HSC gene therapy.** HSC isolated from the bone marrow (or mobilized peripheral blood) of a patient that, in this cartoon, is affected by a primary immunodeficiency, are cultured *ex vivo* in conditions to stimulate cell proliferation, and then exposed to a retroviral vector expressing a functional copy of the defective gene. After this procedure, the corrected cells are transplanted back into the patient. Usually, cell infusion is preceded by a pharmacological conditioning regimen that creates space in the BM niche of the patient. This treatment eliminates the endogenous HSCs and favours the engraftment of the transplanted cells, that can long-term 30 correct the disease phenotype. The engrafted, gene-corrected HSCs generate functional progenies that reconstitute all lineages and restore immune functions of the patient. If the gene-corrected cells have a selective growth advantage compared to the unmodified cells, full reconstitution of the immune cell compartments is obtained even from a few engrafted transduced progenitor cells,

as depicted in the figure, and this may occur even without conditioning (Naldini, 2011).

In 2002 the group of Alessandro Aiuti reported successful results for the treatment of another immunodeficiency, the adenosine deaminase (ADA)-SCID by HSC-gene therapy (Aiuti, Slavin et al, 2002; Aiuti and Roncarolo, 2010). ADA-SCID is a fatal autosomal recessive disorder characterized by impaired immune responses, which is caused by deficiency of adenosine deaminase. Given the role of this enzyme for nucleic acid turnover, its deficiency leads to accumulation of toxic level of purine metabolites, resulting in sensorineural deafness and in skeletal, hepatic, neurological and behavioural alterations. In this study children affected by ADA-SCID, who lacked HLA-sibling donor, underwent non-myeloablative conditioning followed by infusion of bone marrow derived-CD34+ cells previously transduced with ADA-containing  $\gamma$ -retroviral vector. Treated patients experienced long-term immune reconstitution and metabolic detoxification with substantial clinical benefits and no evidence of adverse events. One of the crucial factors in the successful outcome of this trial was the use of nonmyeloablating conditioning, which allowed engraftment of transduced stem cells and immune reconstitution (Aiuti, Cattaneo et al., 2009). Encouraged by these promising results, also patients affected by Wiskott-Aldrich syndrome (WAS), a complex primary immunodeficiency caused by WAS mutation and characterized by autoimmunity, thrombocytopenia, eczema and recurrent infections, were treated with gene therapy. Because of the pleiotropic role of the affected WAS gene, which is involved in signalling,

immunological-synapse formation and cell locomotion, its deficiency results in multiple leukocyte dysfunctions. In patients that were treated with autologous genetically corrected hematopoietic stem cells considerable clinical benefit was observed, with marked improvement in susceptibility to infections, autoimmunity and bleeding. In this setting, the use of nonmyeloablating conditioning, which promotes the engraftment of hematopoietic progenitor, and the proliferative growth advantage of the WAS protein (WASP) -positive cells could explain the favourable result observed.

Although preliminary encouraging evidence arose from these clinical trials, therapeutic benefit has been tempered by the occurrence of leukemia reported in 5 out of 20 SCID-X1 treated patients, in the two studies combined (Hacein-Bei-Abina, 2003, Howe, 2008).

Genetic analysis of the malignant cells showed that in three patients the retroviral vector had inserted into, and activated an oncogene called LIM domain only 2 (*LMO-2*) (Hacein-Bey-Abina, 2008, Howe, 2008). Vector-mediated transactivation of the oncogene, called also insertional mutagenesis, was not the only cause of malignancy, but most likely the event that triggered it. Indeed, some reports suggest that cell transformation may have been facilitated by the constitutive and unregulated expression of the *IL2RG* therapeutic transgene (Ginn, 2010; Scobie, 2009; Woods, 2006). In the WAS trial, one case of leukaemia was reported and, even in this patient, molecular analyses showed vector-mediated up-regulation of the proto-oncogene *LMO-2*.

Insertional mutagenesis was also reported in another RV-based clinical trial by Grez and colleagues (Ott, 2006). In this study RV vectors were used to replace the non-functional enzyme gp91phox of the chronic granulomatous disease (CGD) into patient's HSPCs. CGD results from mutation in the NADPH oxidase complex, an enzyme normally used by phagocytes to kill microbes. Although the restoration of this function in the myeloid progenitors should not confer a growth advantage to the corrected progenitors cells, an unexpected sustained engraftment of functionally corrected cells with therapeutically relevant levels of superoxide production was shown in the two treated patients. This expansion was caused by vector insertional activation of some oncogenes (*MDS1-EVI1*, *PRDM16* or *SETBP1*) that prompted proliferation of cell clones that were able to reconstitute the immune function (Métais and Dunbar, 2008; Ott et al., 2006). Unfortunately, this clinical benefit was lost after some months by the inactivation of the transgene expression due to methylation of the retroviral vector promoter used to express the therapeutic transgene (the Spleen Focus Forming Virus –SFFV\_LTR). Moreover, this silencing did not affect the enhancer sequences responsible for up-regulation of the oncogenes and, more recently, it was reported that both patients in this trial progressed to myelodysplasia, a pre-malignant condition and, for one of them, this clonal expansion progressed to acute myeloid leukemia (AML) (Ott, 2006; Stein, 2010). It must be noted that no adverse effects have been reported in the ADA-SCID clinical trials based on RV-transduced HSPCs, suggesting that disease-, vector-, or transgene specific factors may cooperate

together with insertional gene activation in inducing malignant or premalignant transformation.

Thus, both the SCID-X1, the WAS and the CGD trials have raised serious concerns about the safety of integrating vector-mediated gene therapy in HSPCs and the consequences of insertional mutagenesis. Indeed, the unexpected frequency of these adverse events fuelled debate over the future of gene therapy and led investigators and gene therapy societies to critically re-examine the risk/benefit ratio of gene therapy as a clinical treatment for diseases (Gansbacher, 2003). In this perspective, a lot of investigations are being conducted in order to determine the safety of the vectors with particular attention to the integration site analysis of retroviral and lentiviral vectors, and to improve vector design. Indeed, although the mechanisms of proto-oncogene activation by retroviral insertional mutagenesis and oncogene cooperation in tumorigenesis have been established (Collier and Largaespada, 2005), its occurrence to a relatively high frequency in a clinical trial using replication defective gene transfer vector had not been anticipated in the preclinical studies.

### **1.3.2 Lentiviral vector based clinical trials**

LVs have been now proposed as an alternative gene transfer delivery platform to RVs for in HSPCs-based gene therapy. The major advantages of the LV system is the high efficiency by which it can transduce HSPCs with minimal manipulation, leading to robust and stable gene expression in their progeny.

In 2009 Cartier et al. reported for the first time the use of lentiviral vectors for gene transfer in hematopoietic stem cell. In this clinical trial patients affected by X-linked Adrenoleukodystrophy (ALD), a severe neurodegenerative disease of the central nervous system (CNS), were treated with lentiviral vector-mediated gene transfer of *ABCD1* gene, which encodes for a peroxisomal adenosine triphosphate-binding cassette transporter. Given its role in myelin turnover in microglia and oligodendrocytes, *ABCD1* deficiency results in demyelination and nervous system dysfunctions. The use of lentiviral vectors, superior to  $\gamma$ -retroviral vectors to achieve gene transfer, allowed both high level of transduction of stem and progenitor cells as well as stable transgene expression and its combination with total myeloablating regimen resulted in effective engraftment of transduced HSCs. Thus, despite the low level of gene correction and the absence of a selective growth advantage over the non-corrected counterpart, genetically corrected HSCs were able to reconstitute CNS microglia and to cross-correct the disease, paving the way for HSC-gene therapy for other genetic CNS diseases.

More recently metachromatic leukodystrophy (MLD), another lysosomal storage disorder, was treated with LV-based gene therapy. MLD is an autosomal recessive disease caused by the deficiency of the lysosomal enzyme arylsulfatase A (*ARSA*). The enzymatic defect results in the accumulation of the *ARSA* substrate galactosylceramide (sulfatide), a major sphingolipid of myelin. Despite the fact that the enzymatic deficiency is systemic, disease manifestations are restricted to the nervous system and are

characterized by myelin degeneration in both the central and peripheral nervous system. Children affected by MLD display progressive neurologic symptoms, including ataxia, seizures, and quadriplegia, culminating in decerebration and eventual death early in infancy. Currently, no available treatment can reverse the fatal outcome of this devastating disease. HSPCs transplantation combined with gene therapy represents a candidate strategy because, as demonstrated by pre-clinical studies, combines advantages of an autologous HSPCs source with the benefits of enzyme over-expression in transplanted cells. A critical role for the success of this strategy is the expression at supra-normal levels of the therapeutic enzyme by hematopoietic cells that replace brain microglial cells and become a quantitatively effective source of functional enzyme for neural cells (Biffi, 2004). In 2013 Biffi and colleagues reported the short-term clinical benefits observed in three pre-symptomatic patients treated with LV-based gene therapy for MLD. Briefly, these patients with genetic, biochemical, and neurophysiological evidence of late infantile MLD, were treated with HSCs transduced with LV carrying functional copy of the *ARSA* gene. After reinfusion, the patients showed stable and extensive *ARSA* gene replacement, which led to high enzyme expression in cerebrospinal fluid and throughout hematopoietic lineages. There were reported measurable and objective improvements in brain imaging, electrophysiological measurements and biochemical studies; the result was an impressive arrest in the disease progression. Up to 80% of the cells in the blood contained vector sequences that led to expression of

ARSA at >10- fold the levels in healthy controls, and functional protein was found in the cerebral spinal fluid at levels comparable to those in normal controls at 1 and 2 years post-infusion of genetically modified HSPCs (Biffi, 2013).

In the same year Aiuti et al. reported another study with third generation LV to achieve gene transfer into HSPCs derived from patients with Wiskott-Aldrich syndrome. Although the efficacy of WAS gene-therapy was already provided in a previous clinical trial, one subject developed leukemia from insertional mutagenesis. To further increase the safety of this LV approach, expression of the WAS cDNA was controlled by cis-regulatory elements of the WAS gene in an attempt to restrict transgene expression to the target cells required for phenotypic correction, thus potentially reducing enhancer effects in other lineages. Three patients received infusion of autologous genetically modified HSPCs after reduced chemotherapy conditioning to enhance engraftment of these cells. The treatment resulted in improved platelet counts with increased platelet size and reduced bleeding, allowing the patients to become transfusion independent. Moreover, improved immune function and reduced autoimmune symptoms were observed. Vector integration analyses of the common integration sites (CIS) showed a high polyclonal engraftment derived from the corrected cells. Up to now, no evidence of vector genotoxicity such as selection of integrations near oncogenes or aberrant clonal expansion is reported (Aiuti et al., 2013).

Completely different diseases are  $\beta$ -thalassemias. These are a group of hereditary blood disorders characterized by anomalies in the

synthesis of the  $\beta$  chains of haemoglobin resulting in variable phenotypes ranging from severe anemia to clinically asymptomatic individuals.  $\beta$ -thalassemias are caused by mutations in the  $\beta$ -globin gene, leading to decreased ( $\beta^+$ ) or absent ( $\beta^0$ ) synthesis of the  $\beta$ -chains of hemoglobin. Correction of the  $\beta$ -globinopathies using LV carrying the  $\beta$ -globin genes and elements of the locus control region (LCR) has been well established in murine models. A clinical trial started in France exploiting a SIN LV that encodes for the  $\beta$ -globin gene under the control of human  $\beta$ -globin promoter and the  $\beta$ -LCR elements. The first patient treated failed to engraft because the HSCs had been compromised by the technical handling of the cells without relation to the gene therapy vector. A second patient,  $\beta^+/\beta^0$  thalassemic dependent on transfusions since early childhood received LV gene therapy and had become transfusion independent with a significant improvement in life's quality. However at the moment it is not clear if the therapeutic effect obtained in this patient is either coincidental or the result of a benign cell expansion caused by vector-induced deregulation of the *HMGA2* gene in stem/progenitor cells. Molecular analyses revealed that the predominant form of the HMGA2 protein was truncated by alternative splicing of the third intron with a cryptic splice signal located within the cHS4 insulator core and polyadenylation within the adjacent R region of the left LTR. This leads to loss of let-7 miRNAs target sites that are located in distal exons four and five and consequent accumulation of the aberrant protein (Cavazzana-Calvo, 2010).

A crucial issue for the success of gene therapy is the regulation of the expression of the therapeutic gene inserted by viral vectors in the host genome. In some clinical settings, as in the case of MLD, a stable and strong level of transgene expression, even to supra-normal levels, is essential to partially supply the deficiency of the non-corrected cells, thus allowing the correction of the disease phenotype (Biffi, 2006), while in other settings, constitutive and non-physiological expression might be detrimental. Indeed for certain application, as in the case of globin transgenes used for HSC gene therapy of  $\beta$ -thalassemia, tight regulation of the therapeutic gene is required. In those cases, transcriptional regulatory elements derived from the endogenous locus or from other genes with a similar expression pattern may be incorporated into the vector (Cavazzana-Calvo, 2010).

When instead forced ectopic expression of the therapeutic gene transgene in the target cell type is reported to be toxic and results in the counterselection of the gene-modified cells, as in the case of GALC in HSCs for the treatment of globoid cell leukodystrophy (GLD), the inclusion of the target sequence for a microRNA in the 3' UTR of the transgene, which is differentially expressed across different cell types, will result in the suppression of the transgene expression specifically in the cells that express that microRNA (Gentner, 2010). Moreover, each integration site can influence the level of transgene expression, probably through local epigenetic features and the underlying endogenous gene expression pattern, resulting in unpredictable and unreliable transgene expression (Naldini, 2011; Bushman and Lewinski, 2005).

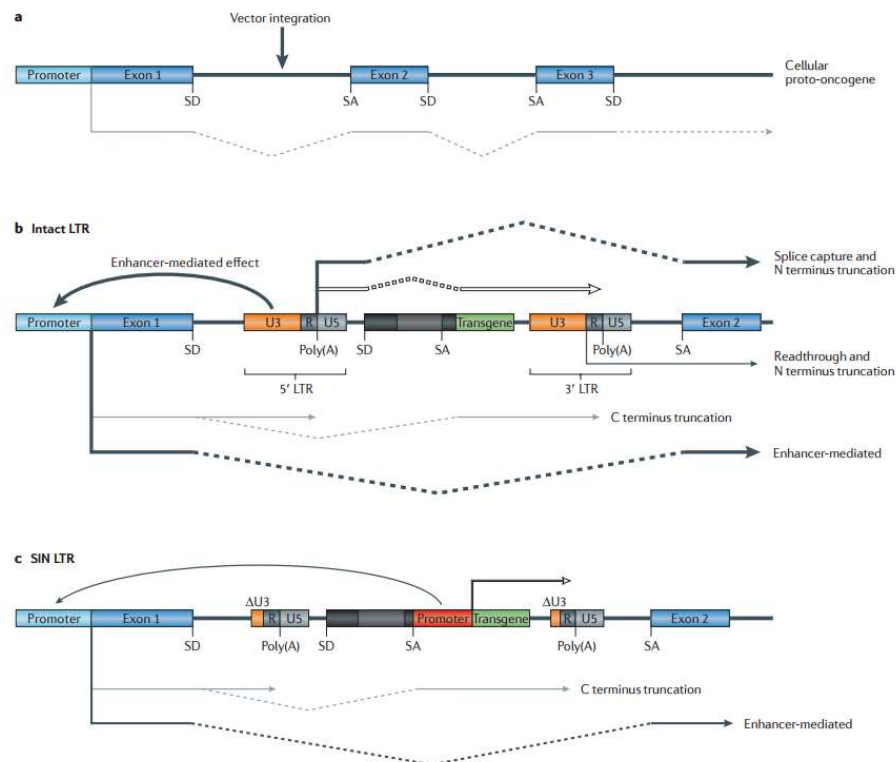
### **1.3.3 Insertional genotoxicity**

When facing the choice of the vector to be used, insertional mutagenesis is one of the factors to take into account for clinical applications. Retroviruses contain in their long terminal repeats (LTRs) two sets of promoters and enhancers and a strong splice donor site downstream the 5' LTR. This configuration allows activation of the proto-oncogenes near the integration site through the enhancer /promoter of the LTRs or through aberrant splicing from the vector transcripts, mediated by capture of flanking genes by splicing of the 5' originating vector transcripts or by transcriptional readthrough from the 3' LTR. Several studies aimed at defining the integration target site selection of  $\gamma$ -RV showed that they preferentially integrate near transcriptional start site, CpG islands and DNase I hypersensitive sites (Bushman, Lewinski et al., 2005). This has been confirmed in both SCID-X1 and ADA trials, in which vector integrations in those regions led viral LTR-associated (LTR) enhancer to alter the expression of the nearby genes, including proto-oncogenes.

However, despite both RV and LV vectors potentially integrate at or near expressed genes, only RV show a strong bias for integration near promoter region. This behaviour may increase the risk of transcriptional crosstalk between the two transcription units, i.e. the provirus and the endogenous gene, leading to vector-mediated deregulation of the targeted endogenous gene. This mechanism is mediated by promoter/enhancer sequences present in the vector, which may recruit ubiquitous and cell type-specific transcription factors to upregulate expression of flanking genes independently on

the orientation of the integration (Figure 1.4). Indeed, it has been shown that LTR insertions are able to upregulate expression of proto-oncogenes as far as 300 Kb from the insertion site.

In addition, proviral integration may occur within the transcription unit of a gene, causing different effects on the resulting protein (Baum, 2004; Naldini, 2011). The R region of the LTRs, which is also present in the SIN vectors, contains both a canonical and a cryptic polyA signal. Because these two polyAs act independently on either genomic strand, integrations within a transcription unit can elicit premature termination of endogenous transcription independently on the orientation of the integrated proviruses (C terminus truncation in Figure 1.4).



**Figure 1.4. Mechanisms of insertional mutagenesis of random integrating vectors.** Certain genotoxic events could lead to activation of the transforming potential of a cellular proto-oncogene after integration of a retroviral vector within or near the proto-oncogene. A vector integration site is depicted in the first intron of a proto-oncogene. The grey arrow beneath the gene indicates the normal transcript, with the broken segments indicating intronic sequences that are removed by splicing. SD indicates a splice donor site and SA indicates a splice acceptor site. (Up) A conventional  $\gamma$ -retroviral vector is integrated. The two LTRs contain strong enhancers and promoter elements in the U3 region; the R and U5 regions are also indicated. The white arrow indicates the vector transcript, which encodes the transgene. This integration could lead to up-regulated transcription of the proto-oncogene from its cellular promoter by enhancer-mediated effects. Alternatively, splice capture from the promoter in the vector 5' LTR could give rise to a chimeric transcript encoding an N-terminally truncated form of the oncogene with constitutive activity and transforming potential. Another possible mechanism giving rise to an N-terminally truncated form of the oncogene is readthrough transcription originating from the vector 3' LTR. This event is less likely when the vector encodes a transgene because of promoter interference (occlusion) between the upstream and downstream LTRs. Finally, truncation of the endogenous transcript might occur as a result of transcription termination at the polyA sites contained in the vector LTR, with or without aberrant splicing between the cellular and the vector splice sites. This transcript gives rise to a C-terminally truncated form of the oncogene that may have constitutive activity and transforming potential. (Bottom) All LTR-dependent events previously shown are abrogated by the use of a vector with self-inactivating (SIN) LTRs. Here, the U3 transcriptional control elements are deleted ( $\Delta$ U3) from both LTRs during the transduction process, and the vector expresses the therapeutic gene from an internal promoter. A residual concern is long-range transcriptional activation of the oncogene, which could be mediated by the enhancer elements of the internal promoter. However, this concern can be alleviated if an exogenous promoter with only moderate activity is used in the vector, resulting in lower proto-oncogene expression than given by intact LTRs (Naldini, 2011).

More in detail, insertion in the 3'UTR of a gene may remove mRNA-destabilizing motifs, such miRNA target sites, from the endogenous transcript. This would result in increased stability of both the endogenous mRNA and its protein. With the same mechanism and moving from the 3' to the 5' of the targeted transcriptional unit, insertion may induce the formation of a 3'-truncated mRNA. On the contrary, transcription may start from the

viral promoter (being this either an LTR or an internal promoter) and enter into the endogenous transcriptional unit, generating a 5'-truncated mRNA (N terminus truncation in Figure 1.4). The resulting C-terminally or N-terminally truncated proteins may possess oncogenic properties and induce tumorigenesis (Girard et al., 1996; Uren et al., 2005; van Lohuizen et al., 1989). In the case of SIN LVs, the strength of the internal promoter influences the level of proto-oncogene transactivation (Cesana et al., 2014). Moreover, aberrant vector/genome splicing events may generate mutant protein with enhanced or reduced activity. Also in this case, the aberrant protein may display oncogenic activity. It has to be noticed that all this mechanisms may synergistically act in the generation of high levels of a mutant protein.

## **1.4 Tailored genome editing**

Although integrating vectors have clearly shown therapeutic potential for stable gene transfer, their semi-random integration into the host cellular genome may alter the function of the nearby cellular genes found, resulting in deleterious consequences, thus calling for the development of safer therapeutic approaches for gene transfer.

Gene targeting is now emerging as a powerful technology for designed genomic modifications. This strategy, which exploits homologous recombination (HR), a process that evolved to resolve stalled DNA replication forks, enables incorporation of the site-

specific genetic changes in the chromosome of target cells (Ciccia and Elledge, 2010; Moynahan and Jasin, 2010).

As such, when DNA molecules with homology to the target locus, including also mismatches or intervening sequences, are delivered to the target cells, HR can mediate the incorporation of novel DNA sequences in site-specific fashion. Furthermore, by including selector genes within the donor sequence, it will be possible to select and expand of the genetically modified cells.

#### **1.4.1 Gene targeting: a historical perspective**

In the eighties Smithies and colleagues reported the first chromosomal gene targeting in mammalian cells using a construct containing a neomycin-resistance (neoR) gene in the human  $\beta$ -globin locus, delivered by transfection into human EJ bladder cells or by electroporation into the mouse erythroleukemia/human fibroblast hybrid cell line (Smithies et al., 1985). In this study the absolute targeting frequency achieved was one targeted event per  $4 \times 10^6$  transfected bladder cells. One year later Capecchi's group reported an increase in targeting efficiency up to 1 of  $10^3$  cells that received plasmid DNA, by using microinjection techniques (Thomas and Capecchi, 1986; Thomas et al., 1986). Despite being labour-intensive and time consuming for the selection of the poor fraction of edited cells, gene targeting has allowed the knock-in and knock-out of several genes in rodent embryonic stem cells, thus enabling the generation of mutant mice of relevant interest studying both for gene function and pathogenesis of human genetic diseases

(Capecchi, 2005; Evans, 2001; Smithies, 2001). Moreover, conditional mutations can be generated by coupling gene targeting with a designated site-specific recombination system (Branda and Dymecki, 2004), such as Cre-LoxP or Flp-FRT, to allow evaluation of a gene's function in specific tissues and/or at a restricted time (Capecchi, 2005).

These studies provided evidence of the potential of gene targeting technology, which has allowed both the knock-in and the knock-out of several genes in rodent embryonic stem cells, thus proving to be an invaluable tool for gene discovery; however the very low frequency, ranging from 1 to  $10^6$  to 1 to  $10^7$ , and the need for selection of the edited cells have hampered its therapeutic application (Porteus and Carroll, 2005; Capecchi, 2005).

While vector-mediated gene addition results in random integration of the transgene in non-homologous chromosomal sites, gene targeting allows for site-specific editing of a chosen gene. According to this, the gene targeting technology allows for direct correction of the mutation(s) occurring in a disease-causing gene. This approach, called gene correction, results in restoration of both function and physiologic expression of the corrected gene, thus overcoming one of the major limitations of current gene replacement strategies (Figure 1.5) (de Semir and Aran, 2006; Hanna et al., 2007).

Transgene insertion may be also targeted to “genomic safe harbours” (GSHs), regions in which insertion of a therapeutic cassette leads to transgene expression without adverse effects on the flanking endogenous genes. Safe genomic regions can be

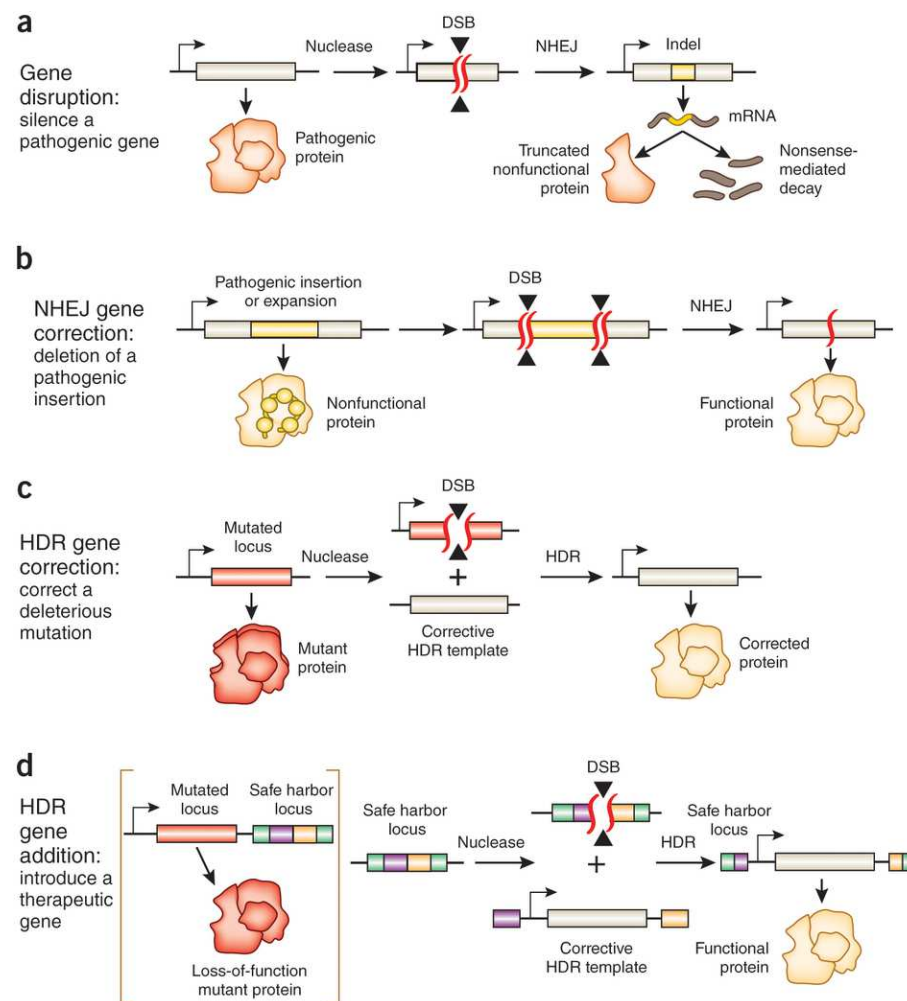
intragenic or extragenic regions. Extragenic regions, as heterochromatin, may be preferable, however they often result in unreliable and unstable expression. In contrast intragenic regions, as non-essential genes, may provide more robust and predictable transgene expression levels.

An ideal site for gene transfer intent would ensure: i) stable and robust transgene expression in different cell types, ii) no transcriptional perturbation due to integration of the transgene cassette, iii) no disruption of coding or regulatory sequences (Figure 1.5).

In humans, indication of such transgene acceptor sites may be obtained by combining the emerging knowledge on sequence variation among genomes (Frazer, 2009) and on clinically silent homozygous gene deficiencies, with the information available from gene expression atlases (de Boer et al., 2009), from chromatin remodelling analyses (Ho and Crabtree, 2010) and from the few identified functional vector insertions associated with a benign outcome in gene therapy clinical trials (Aiuti and Roncarolo, 2009; Cartier, 2009).

In 2011 a comprehensive study by Lombardo et al. validated the *AAVS1* genomic locus as a safe harbour for genomic modification. In this study the transcriptional and epigenetic impact of different transgene expression cassettes, targeted into *AAVS1* by ZFNs were analysed in many cell types, including primary human lymphocytes, providing evidence of robust transgene expression without detectable transcriptional perturbation of the targeted locus and its flanking genes. Genomic modification within *AAVS1* can be

exploited to achieve efficient and neutral transgene insertions, thus further improving the safety of ZFNs-mediated gene targeting for therapeutic applications ((Lombardo, 2011; Sadelain and Bushman, 2012)



**Figure 1.5. Types of therapeutic genome modifications.** The specific type of genome editing therapy depends on the nature of the mutation causing disease. (a) In gene disruption, the pathogenic function of a protein is silenced by targeting the locus with NHEJ. Formation of indels in the gene of interest often results in frameshift mutations that create premature stop codons resulting in a

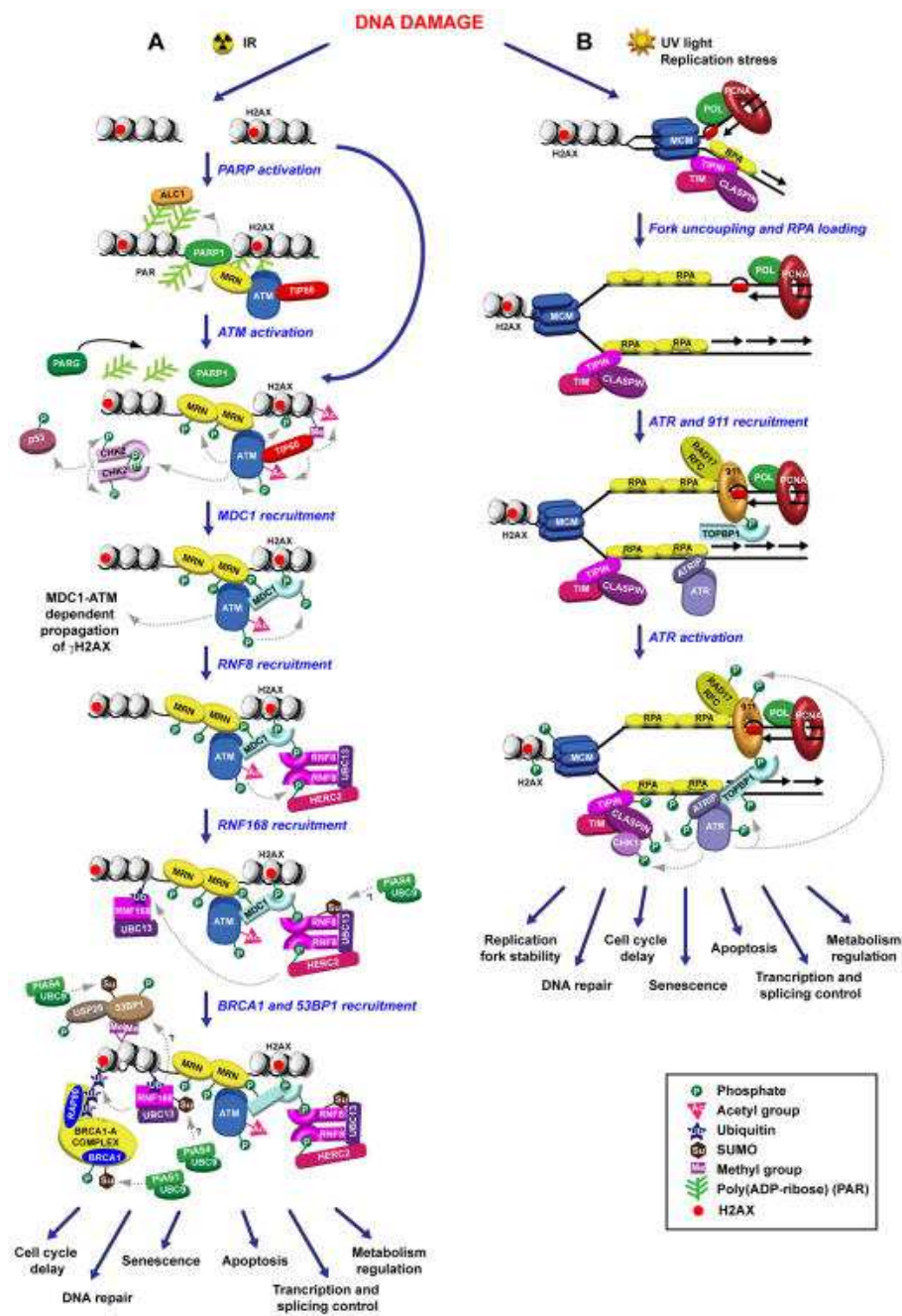
non-functional protein product or nonsense-mediate decay of transcripts, suppressing gene function. Gene disruption may also be used to introduce protective loss of- function mutations into wild-type genes to generate a therapeutic effect. (b) In NHEJ gene correction, two DSBs targeted to both sides of a pathogenic expansion or insertion may be resolved by NHEJ, causing a therapeutic deletion of the intervening sequences. This form of treatment would require multiplexed targeting of disease-causing mutations. (c) HDR gene correction can be used to correct a deleterious mutation. A DSB is induced near the mutation site in the presence of an exogenously provided, corrective HDR template. HDR repair of the break site with the exogenous template corrects the mutation, restoring gene function. (d) An alternative to gene correction is HDR gene addition, which introduces a therapeutic transgene into a predetermined locus. This may be the native locus, a safe harbour locus or a non-native locus. A DSB is induced at the desired locus, and an HDR template containing sequence similarity to the break site, a promoter, a transgene and a polyadenylation sequence is introduced to the nucleus. HDR repair restores gene function in the target locus, albeit without true physiological control over gene expression (Turitz-Cox et al., 2015).

#### **1.4.2 The DNA-damage response: a focus on the DNA Double Strand Break (DSBs) repair pathways**

To preserve the genetic information, cells have evolved different mechanisms to counteract DNA insults. Those can be spontaneous (dNTPs misincorporation, loss of bases due to depurination, deamination, alkylation, oxidation of DNA bases) or environmental, as those induced by physical sources (ionizing radiation, ultraviolet light) and chemical sources (alkylating agents, crosslinking agents, topoisomerase-inhibitors); altogether, they accounts for up to  $10^5$  spontaneous DNA lesions per day.

DNA lesions can occur in a single strand (single-strand breaks-SSB) or involve both DNA strands (double-strand breaks-DSBs). In the latter case, these lesions, if left unrepaired, would quickly cause chromosome breakdown and loss of genes during cellular division.

To ensure that cells pass accurate copies of their genomes to the next generation, evolution has overlaid the core cell-cycle machinery with cell-cycle checkpoints, a series of surveillance pathways that detect damaged or abnormally structured DNA and coordinate cell-cycle progression (typically by slowing or arresting the cell-cycle) with DNA repair (Branzei and Foiani, 2010).



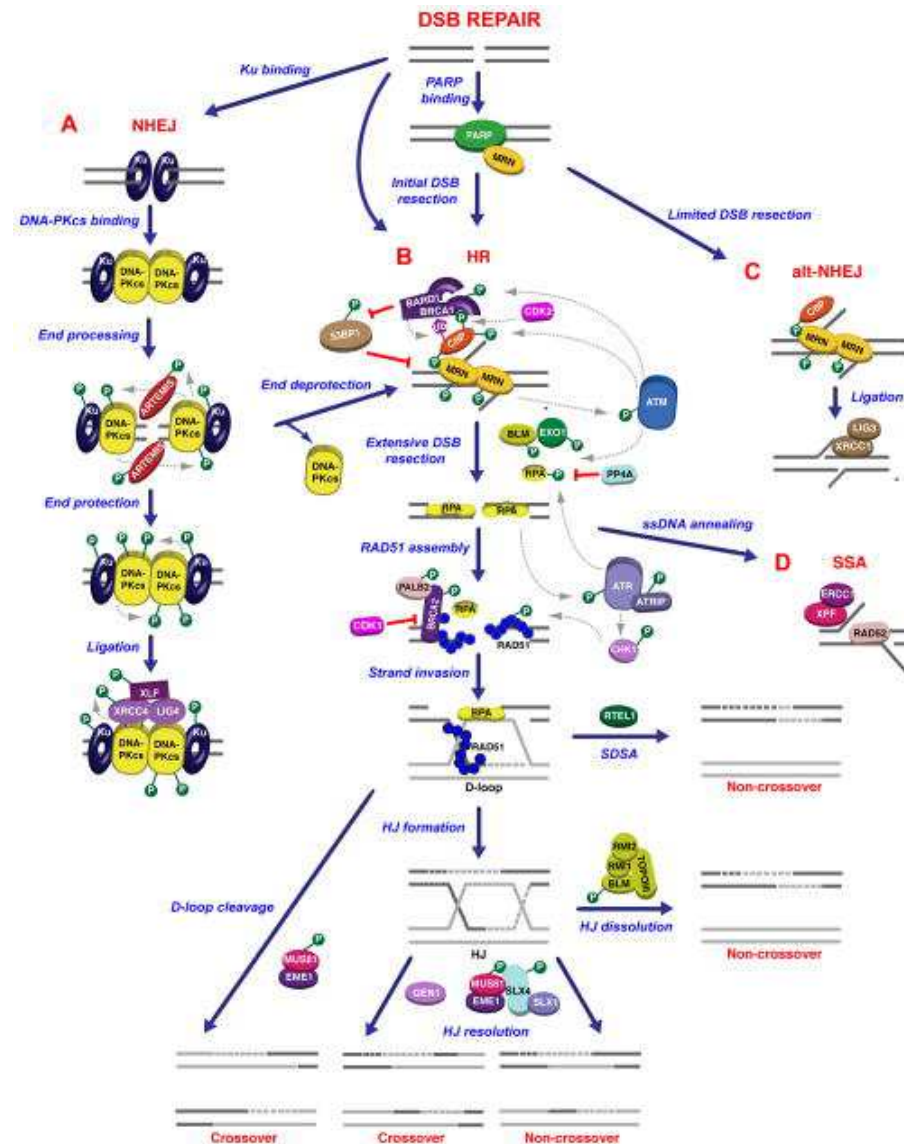
**Figure 1.6. Schematic model for ATM and ATR activation in response to DNA damage.** A) Formation of DSBs following IR treatment activates PARP1, which mediates the initial recruitment of the MRN/ATM complex at DSBs.

Activation of the ATM kinase activity by MRN and TIP60 leads to the phosphorylation of CHK2 and p53, in addition to a wide number of other DDR factors, and the induction of the  $\gamma$ H2AX-dependent signalling cascade, which results in the recruitment of MDC1, RNF8, RNF168, BRCA1 and 53BP1 to DSBs, as described in greater details in the main text. (B) DNA lesions induced by UV light or replication stress (denoted by red rectangular shapes) lead to replication fork stalling and accumulation of RPA-coated ssDNA regions, which recruit the ATR/ATRIP and the RAD17/RFC2-5 complexes. When the 9-1-1 complex is loaded by RAD17/RFC2-5 and ATR kinase activity is stimulated by the 9-1-1 associated protein TOPBP1, activation of the ATR signalling cascade and CHK1 phosphorylation will occur. Post-translational modifications of the DDR factors here depicted are represented by different coloured shapes, as indicated by the legend at the bottom of the figure (Ciccina and Elledge, 2010).

Among these checkpoint proteins, the RAD group, including RAD17, RAD1, RAD9, RAD26 and Hus1, is widely expressed in all eukaryotic cells and is thought to play critical role in DNA-damage sensing (Abraham, 2001; Ciccina and Elledge, 2010). Other proteins responsible for early signal transmission through cell cycle checkpoints include member of the family of phosphoinositide 3-kinase related kinases (PIKK). In mammalian cells, two PIKK family members, ATM (Ataxia- Telangiectasia Mutated) and ATR (ATM and RAD 3-related), mediate the activation of the pathway for the cell cycle arrest, the DNA repair or, if the damage cannot be repaired, the apoptosis (Figure 1.6) (Abraham, 2001; Ciccina and Elledge, 2010).

Mammalian cells have evolved different mechanisms to repair DNA damage: SSBs are repaired by single-strand breaks-repair (SSBR), including mismatch repair (MMR) and single base-excision repair (BER), while DSBs are repaired mainly by non-homologous end-joining (NHEJ), by homologous recombination (HR) and, to less extent, by alternative NHEJ /altNHEJ and single-strand annealing (SSA) (Figure 1.7). Among them, only HR, which

uses sister chromatids as donor template can precisely restore the genomic sequence of the broken DNA.



**Figure 1.7 Alternative DNA repair pathways involved in the repair of double-strand breaks.** A, Rapid association of Ku to DSBs promotes NHEJ by recruiting DNA-PKcs. Sequential phosphorylation events on multiple DNA-

PKcs amino acid clusters favours the initial processing of DNA ends by ARTEMIS, followed by DNA-PKcs-dependent protection of DNA ends required for DNA ligation. **B**, Alternatively to NHEJ, MRN, which is initially recruited to DSBs by PARP in competition with Ku, mediates the initial stages of DSB resection together with CtIP and BRCA1 to promote homologous recombination in S and G2. 53BP1 has an inhibitory role on DSB resection and is negatively regulated by BRCA1 by unknown mechanisms. The MRN/CtIP/BRCA1 complex can also promote DSB resection following deprotection of DNA ends when NHEJ fails. Extensive DSB resection and formation of RPA-coated 3'-ssDNA ends is induced by EXO1 and BLM. ATM plays a central role in the regulation of DSB resection as described in the main text. Displacement of RPA from the 3'-ssDNA ends and assembly of RAD51 filaments mediated by BRCA2 leads to strand invasion into homologous DNA sequences. Recruitment of RAD51 to ssDNA ends is regulated by the ATR pathway, which is activated following DSB resection. D-loop structures formed after strand invasion can be cleaved by MUS81/EME1 or displaced by RTEL1 during SDSA to generate crossover or non-crossover events, respectively. Non-crossovers are generated also by dissolution of Holliday junctions (HJs) by the BLM/TOPOIII complex, whereas HJ resolution by the nucleases GEN1 and SLX1/SLX4, which associates with MUS81/EME1, can generate both crossover and non-crossover events. **C**, Limited DSB resection carried out by CtIP and MRN in G1 results in alternative NHEJ. **D**, Following DSB resection, 3'-ssDNA ends with homologous sequences can be directly annealed by RAD52 (Ciccia and Elledge, 2010).

The DSBs repair by HDR and NHEJ can vary significantly according to cell type and cell state. Moreover, while NHEJ has been observed in several cell types, including dividing and post-mitotic cells, HDR occurs primarily during S/G2 phase, thus limiting its application to mitotic cells.

#### **1.4.2.1 The DSBs repair: Non-homologous end joining pathway**

Once a DSB has been made, non-homologous end joining can rapidly repair the DNA damage by simply re-join broken ends, when proximal, by DNA ligation. Although there is a tendency,

known as microhomology, to re-join sites with 1-4 nucleotides complementary between the two ends, NHEJ often results in the formation of small insertion either deletion mutation (indels). The activation of the repair pathway requires Ku, the catalytic subunit of the DNA-dependent protein kinase (PK), which stabilizes DSB ends and activates DNA ends processive enzymes, such as ARTEMIS and the XRCC4/DNA ligase IV heterodimer. Artemis and DNA-PKcs interact together to form a physical complex that acts as an endonuclease at both 5' and 3' overhangs. Ligation, which requires a ligatable nick on each strand, is catalysed by XRCC4 (X-ray cross complementation 4)/DNA-ligase IV complex (Figure 1.7) (Lieber et al., 2003). Although inaccurate, this is a common repair mechanism in somatic cells, where a change in the DNA sequence may be considered as an acceptable solution to a more dangerous DNA damage.

#### **1.4.2.2 Homologous recombination repair**

A much more accurate mechanism to DSBs repair is homologous recombination (HR), which occurs in newly replicated DNA. This process, which represents a genetic exchange between a pair of homologous DNA sequences, may take place shortly after the DNA replication, in S or G2 phase of the cellular division, when sister chromatids are available as template to the DNA repair machinery.

DSB repair through HR starts with the DNA ends resection to form 3' single-stranded overhangs (Mimitou and Symington, 2009), that scan the genome for homologous sequences, then one tail invades a

homologous DNA strand and acts as a primer for copy synthesis. In the DSBs- repair model proposed by Szostak and colleagues (Szostak et al., 1983), this strand invasion causes the displacement of the (D)-loop, leading to capture of the second 3' overhang and formation of a double Holliday junction (HJ). Resolution of this HJ results in a gene conversion, with or without crossovers. To explain the low number of associated crossovers in non-meiotic cells, the synthesis dependent strand annealing (SDSA) model was proposed. In this model, DSB repair is also initiated by strand invasion, but after copy synthesis the newly synthesized strand is displaced from the template and returned to the broken DNA molecule, resulting in repair of the break without associated crossover (Chen et al., 2007). Central to all HR pathways is the strand exchange reaction catalysed by RAD51 that mediate the D-loop formation (Holthausen et al., 2010). Resolution of the D-loop is accomplished to yield recombinants that either entails a reciprocal exchange of genetic information flanking the initiation site (crossover recombinants) or not (non-crossover recombinants). When a DSB is formed between two nearly (400bp to 15kb) homologous repeats, an error-prone pathway called single-strand annealing (SSA) can be eventually involved in the repair, thus resulting in the deletion of the sequence between the homologous sequences flanking the DSB. SSA is a particular type of NHEJ that starts with the resection of the DSB ends by an exonuclease that produces long single-stranded DNA overhangs. When two complementary sequences are exposed, they can anneal, leaving long single-stranded non-homologous DNA flaps. The Removal of these flaps

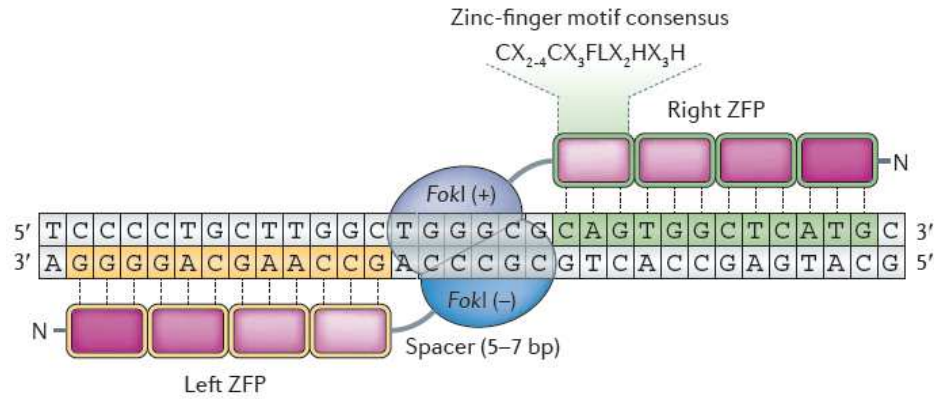
is dependent on the RAD1/RAD10 flap endonucleases (also involved in nucleotide excision repair). When a DSB is formed between two homologous repeats, SSA is the predominant repair pathway and is independent of RAD51 (Figure 1.7) (Ciccia and Elledge, 2010).

### **1.4.3 Inducing site-specific DSBs**

The seminal discovery, by Jasin and colleagues (Jasin, 1999), that SceI-induced double strand breaks are able to stimulate homologous recombination by 2-3 orders of magnitude in the yeast, encouraged the development of gene editing technologies. In fact, it has been demonstrated that, both in yeast and mammals, by inducing site-specific double strand breaks, it is possible to increase chromosomal interaction with an exogenous donor DNA, thus improving the gene targeting frequency by several orders of magnitude. To this aim artificial nucleases, which harness the effect of DSBs to stimulate homologous recombination, have been developed to enable precise genome editing. To date, four main classes of nucleases have been developed: Zinc Finger Nucleases (ZFNs), meganucleases and transcription activator-like effector nucleases (TALENs), which bind to DNA through protein-DNA interaction, and CRISPR-associated nuclease Cas9, which targets specific DNA sequence through a short RNA guide.

#### **1.4.3.1 The Zinc Finger Nucleases technology**

ZFNs were the first artificial nucleases to be developed in the 1996 (Kim, Cha et al., 1996). These are chimeric proteins composed of tandem arrays (from 3-6) of C2H2 zinc proteins, which binds to DNA, fused through a flexible peptide linker to a non-specific nuclease domain from the type IIS restriction enzyme *FokI* (Urnov, Miller, 2005; Carroll, 2011). Studies on the mechanism of ZFNs-mediated double-strand cleavage have shown that dimerization of the nuclease domain is required in order to cut the DNA substrate (Bitinaite et al., 1998). Indeed, it has been shown that two ZFNs with different sequence-specificities collaborate as a heterodimer to produce a DSB when their binding sites are appropriately placed and oriented with respect to each other (the binding sites are positioned in a head to tail orientation on the top and bottom strands of the DNA). Given that each zinc finger protein, which is composed of 30 amino acids, is able to bind to 3 base pair sequence in the major groove, total target DNA sequences are 18-36 bps length. Between the two binding site there is a nucleotide spacer, whose sequence does have a big impact on the activity of the ZFNs, but a minimal space of 4-6 bp must be maintained for efficient cleavage (Durai et al., 2005). In contrast, both sequence and length of the flexible peptide linker that interconnects the *FokI* domain with the ZFP array influence ZFNs activity and specificity (Handel et al., 2009) (Figure 1.8).



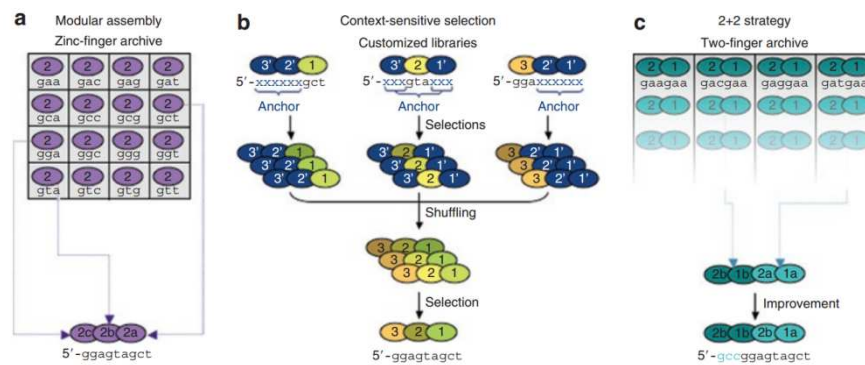
**Fig. 1.8. Structure of ZFNs.** A schematic representation of a zinc-finger nuclease (ZFN) pair is shown. Each ZFN is composed of a zinc-finger protein (ZFP) at the amino terminus and the FokI nuclease domain at the carboxyl terminus. In the zinc-finger motif consensus, X represents any amino acid. Target sequences of ZFN pairs are typically 18–36 bps in length, excluding spacers (Kim and Kim, 2014).

These features explain the high specificity of these proteins. Indeed custom-made ZFNs with desired specificity can be generated by modular assembly of pre-characterized zinc fingers. However, although each zinc-finger can bind to 3 bp DNA target sequence, there is no coverage of all the possible combinations of triplets; moreover newly assembled ZFNs sets need to be tested for efficient DNA cleavage at the intended genomic site (Kim and Kim, 2014).

Zinc Fingers Proteins (ZFPs) of the classical Cys2His2 type are the most frequently used class of transcription factors and account for about 3% of genes in the human genome (Klug, 2010; Pabo et al., 2001). The C2H2 motif consists of a sequence of about 30 amino acids containing two histidines, two cysteines and three hydrophobic residues, all at conserved positions. It forms a small, independently folded domain stabilized by  $\text{Zn}^{2+}$ , which can be used repeatedly in a modular tandem fashion to achieve sequence-

specific recognition of DNA (Miller et al., 1985). The domains, which share the same structural framework, achieve chemical distinctiveness through variations in key residues. This modular design thus offers a large number of combinatorial possibilities for DNA recognition.

A variety of strategies have been described to generate ZFPs with new binding specificities (Cathomen and Joung, 2008; Durai et al., 2005; Urnov et al., 2010). The first approach, which has been termed 'modular assembly' (Segal et al., 2003), generates candidate ZFPs for a given target sequence by identifying fingers for each component triplet and linking them into a multifinger peptide (Figure 1.9). Fingers used for modular assembly have been developed for most triplet sequences (Bae et al., 2003; Choo and Klug, 1994; Segal et al., 1999).



**Figure 1.9 The most common strategies for the design of ZFN with novel specificity.** **A**, The modular assembly' approach: simple addition of discrete modules with known DNA affinity. **B**, The "OPEN" approach: context-dependent preliminary selection strategy (aimed at the identification of 61 zinc-fingers sets that work well together). The relative position of each zinc-finger (1, 2, or 3) affects the outcome of the first selection step; the impact of the respective neighbour(s) on the overall affinity is determined in a following selection step. **C**, The "2 + 2 strategy": a proprietary approach from Sangamo BioScience. Most

likely, the four-finger domains are assembled from a pool of two-finger units with known DNA-binding specificities, followed by minor further optimization (Cathomen and Joung, 2008).

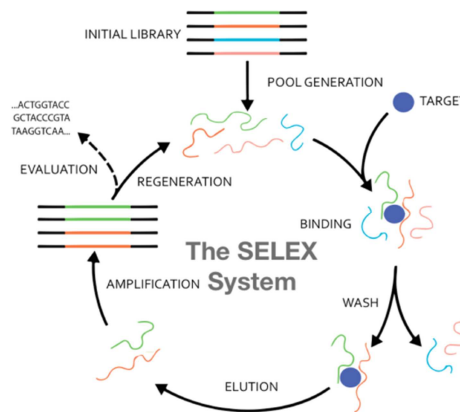
Although this modular assembly approach is relative simple, the designed ZFPs appear to have the highest affinity and sequence-specificity for their targets only when the individual ZFP motif designs are chosen in the context of their neighbouring fingers. In addition, the presence of Asp2 at position 2 of the  $\alpha$ -helix of the preceding ZFP motif, that promotes a cross-strand contact to a base outside the canonical triplet site, results in a target site overlap. While this increases the affinity of the ZFP to the target site, it also precludes the presence of a simple general recognition code for easy rational design of ZFP motifs-based DNA-binding proteins (Durai et al., 2005). However, many studies have shown that a number of ZFNs with sufficient affinity and specificity could be engineered using the known ZFP motif designs by an assembly strategy (Kim et al., 2009). Several alternatives to modular assembly have been developed to address the problems of finger–finger cooperativity and cross-strand contact. One approach, called the 'OPEN' system (Foley et al., 2009; Hurt et al., 2003; Maeder et al., 2008) uses bacterial selections to identify finger combinations that will work well together (Figure 1.9). A second approach for identifying ZFPs with new specificities uses a bacterial selection system that is similar to OPEN but a different strategy for library construction (Gupta et al., 2011; Meng et al., 2008). Another strategy is to use two-finger modules (instead of individual fingers) as the principle unit of DNA recognition (Hockemeyer et al., 2009;

Lombardo et al., 2007; Moore et al., 2001) (Figure 1.9). The library of 2- finger or 3-finger peptides was developed at Klug's MRC Laboratory in Cambridge, and then supplied to the Californian Biotech Company Sangamo BioSciences Inc. Advantage of this approach is that enables optimization of finger junctions within each module for more cooperative and specific base recognition. Moreover, it reduces the number of untested finger–finger junctions in any new ZFP design and therefore the risk of a poor interaction between newly joined fingers. A four-finger ZFP, for example, will contain just one new junction instead of three if assembled from one-finger units. Even if each of these methods has been successfully used to generate endogenously active ZFNs, they differ substantially from each other in terms of time and cost for reagent development and success rate.

Another hurdle in clinical exploitation of ZFNs concerns their specificity of inducing site-specific DNA DSBs. DNA DSBs are rapidly resolved in live cells and, according on the pathway, can be repaired perfectly leaving no marks or indels to indicate the transient presence of the DSB (Ciccia and Elledge, 2010). Consequently, comprehensive identification of ZFN cleavage sites *in vivo* has remained an open challenge.

To analyse ZFN activity, several independent assays have been developed, which can be broadly divided into methods guided by a biochemical determination of the specificity of the two ZFP DNA binding domains that encompass a given ZFN pair and methods that are independent of such prior knowledge. The first approach relies upon the *in vitro* determination of the consensus DNA

binding site for a given ZFN pair by Systematic Evolution of Ligands by Exponential Enrichment (SELEX) (Blackwell and Weintraub, 1990; Phillips et al., 2009; Tuerk and Gold, 1990) (Figure 1.10).

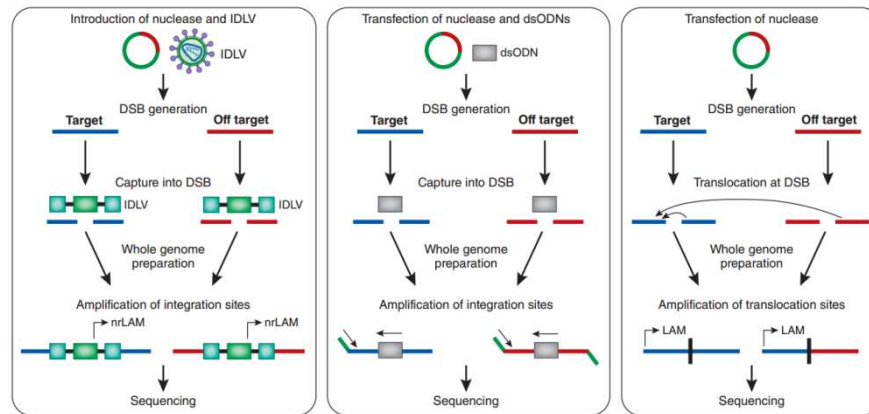


**Figure 1.10. SELEX system.** SELEX is employed for the identification of RNA or DNA molecules that bind to their target molecule with high affinity. Starting with combinatorial libraries with up to  $10^{15}$  different molecules, the specific binders are isolated by an iterative process of ligand binding, washing, recovery (elution), and amplification.

This consensus can be used to bioinformatically interrogate the genome and generate a rank order of potential off-target sites with highest sequence similarity to the consensus. The ZFN activity at these *in silico* identified sites can thus be studied by direct sequencing of these loci (typically 5–15) in ZFN-treated cells. This analysis has been performed in primary human T cells (Perez et al., 2008) and human ESC and induced Pluripotent Stem Cells (Hockemeyer et al., 2009) showing that all putative ZFN off-target sites (with one exception) were found to be wild type. Moreover a cytogenetic analysis of ZFN-edited cells to look for gross

chromosomal changes showed that all possessed a wild-type karyotype (Hockemeyer et al., 2009; Santiago et al., 2008).

Even if all these approaches provide a first line of evidence of the exquisite specificity of these engineered nucleases in the context of the human genome, an unbiased genome-wide method for directly detecting cleavage events at on- and off-target sites *in vivo* will be preferable. Thus, to assess ZFN specificity genome-wide, our group developed a strategy of DSB identification based on Integrase Defective lentiviral Vector (IDLV) trapping. The hypothesis was that linear double-stranded IDLV genomes present in the nucleus of transduced cells (Li et al., 2001; Banasik and McCray, 2010) could be ligated into DSBs by NHEJ, thereby stably tagging transient, otherwise undetectable DSBs. We comprehensively mapped the locations of IDLV integration sites in ZFN-treated cells by linear amplification-mediated (LAM)-PCR (Schmidt et al., 2007) and non-restrictive (nr)LAM-PCR (Gabriel et al., 2009; Paruzynski et al., 2010). Using this strategy we found that with all ZFN pairs tested, the largest cluster of tagged IDLV integration sites occurred at the intended genomic target site, but clustered integration sites were also identified at other locations that bear substantial homology to the intended ZFN target site. Notably, however, target homology alone was not sufficient to determine whether an *in silico*-predicted off-target site was actually cleaved *in vivo*. Thus, we provide for the first time a snapshot of ZFN action in a living cell using a method which requires no a priori knowledge of ZFN specificity and is unbiased by *in silico* homology-based predictions (Gabriel et al., 2011).



**Figure 1.11 Methods for detection of off-target activity of engineered nucleases on a genome-wide scale. a)** A donor sequence together with an integrase-defective lentiviral vector (IDLV) as a foreign DNA bait is captured at nuclease cleavage sites. The results are analysed by nrLAM-PCR. DSBs. **(b)** double-stranded oligodeoxynucleotides (dsODNs) are used as foreign DNA baits and analyse captured sites by single-tail adapter/tag (STAT)-PCR. **(c)** A LAM-PCR HTGTS is used to identify translocations of endogenous genome sequences (Gabriel, von Kalle and Schmitt, 2015).

Very recently, three methods to identify off-target double strand-breaks were reported; Tsai et al., 2015; Frock et al., 2015). Wang and colleagues reported the use of non-restrictive (nr) linear amplification-mediated PCR (LAM-PCR) to identify the trapping of an integrase –defective lentiviral vector into double strand-breaks, with a sensitivity close to 1% of the mutated sequence (Wang et al., 2015). Another method, called “genome-wide unbiased identification of double-strand breaks enabled by next-generation sequencing (GUIDE-seq), was used by Tsai et al. to identify off-target profiles of Cas9 nucleases, by using passive capturing of a double-strand oligodeoxynucleotide molecule into nuclease-induced DSBs. Frock et al., instead, used the LAM-PCR

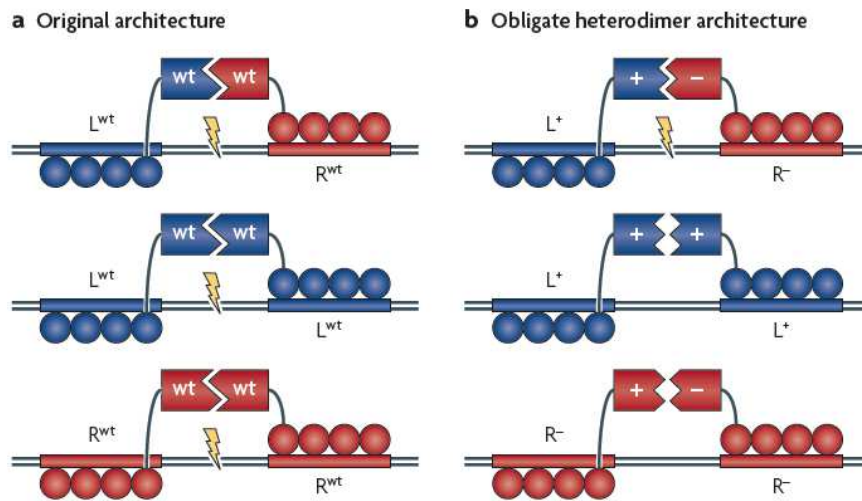
high-throughput genome-wide (HTGTS) to detect off-target DSBs (Figure 1.11) (Frock et al., 2015).

The major consequences of off-target ZFNs cleavage are: i) reduced efficiency of on-target modification and ii) cytotoxicity. An off-target cleavage site may confound the interpretation of the intended genome editing event or, worse, lead to an adverse event in a therapeutic setting. To address these concerns, three approaches were taken. The

first approach consists in the assembly of ZFNs with longer (12–18 bp) DNA recognition sites. Such long sites are potentially rare even in complex genomes. However, it has been recently demonstrated that ZFNs with long recognition site (15-18 bp) display a decreased activity as compared to ZFNs with 9-12 bp long recognition site, and that long arrays of ZFP have the potential to form complex binding behaviours (Shimizu et al., 2011), thus increasing the possibility of off-site recognition. The second approach consists in the replacement of certain residues or even entire  $\alpha$ -helices in the ZFP with those that are likely to perform better in vivo (Urnov et al., 2005). The third approach, which is complementary to the previous two, is to engineer the FokI domain so that only two heterodimeric ZFNs can induce a DSB. Recently, two independent works addressed this issue (Miller et al., 2007; Szczepek et al., 2007) (Figure 1.12).

Using in silico protein modelling and energy calculation, or using a randomized mutagenesis method followed by an in vivo activity selection, both groups obtained two mutated *FokI* domains that could only heterodimerize, because of electrostatic

attraction/repulsion force, but that retain their enzymatic activity. In addition it was shown that forced heterodimerization of the ZFNs reduced the levels of genome-wide cleavage and then improve the specificity of the system. Several groups have recently reported further refinements in the ZFNs engineering, resulting in lower off-target activity-related toxicity (Ramalingam et al., 2011) and, eventually, to the possibility of co-expressing two distinct ZFN sets designed to target different loci without cross-reactivity (Sollu et al., 2010). More recently Doyon and colleagues showed similar improvements in the design of obligate-heterodimeric architectures by means of a peculiar temperature-sensitive screening in yeast, which allowed them to identify cold-sensitive *FokI* variants and thus new important residues for dimerization at the protein-protein interface. Introduction of oppositely charged amino acids in these positions in already established obligate heterodimeric backbones resulted in a 2 fold increase in NHEJ and stimulation of HR at targeted loci. Moreover co-expression in the same cells of two different optimized sets of ZFN (orthogonal ZFN pairs) resulted both in increased cutting activity and specificity (Doyon et al., 2011).



**Figure 1.12 Schematics of the possible combination of two ZFNs to give hetero- and homodimerization.** A, ZFN heterodimer bound to its intended target (top), and two corresponding ZFN homodimers binding to alternative target sites in an architecture that permits (A) or prevents (B) heterodimerization. If the ZFNs carry FokI domains engineered to function solely as heterodimers (indicated as shapes labelled with plus and minus signs in B, binding to the intended target (top) will lead to DNA cleavage (indicated by the lightning symbol), but homodimerization (induced in the example shown by the proximity of two R<sup>-</sup> or two L<sup>+</sup> binding sites) will be impeded by the inability of the FokI domains they carry to form a productive dimer. wt, wild type (Urnov et al., 2010).

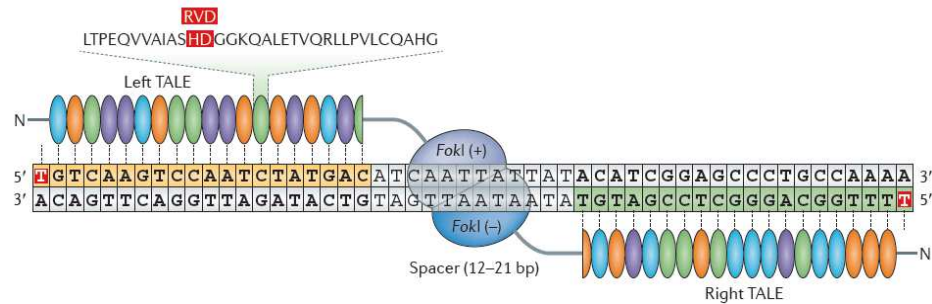
#### 1.4.3.2 The Transcription activator-like effector nucleases technology

Shortly after the coming of ZFNs, other DNA binding proteins were discovered in plant pathogenic *Xanthomonas* spp. bacterium (Boch et al., 2009; Bogdanove et al., 2009). Their DNA binding domain, the Transcription Activator Like Effectors (TALEs), has been engineered to be fused with a FokI nuclease domain at their carboxyl termini. Each TALE protein usually contains tandem arrays of 33-35 amino acids repeat and each of these binds to a

specific single nucleotide in the major groove. Two amino acids residues found in 12<sup>th</sup> and 13<sup>th</sup> position, known as repeat variable diresidues (RVD), were found to be specific for single nucleotide recognition. Given the one-to-one correspondence between the four bases and the four RVD modules, it is possible to design new TALENs with any desired sequence specificities. The only requirement for TALEN design is a thymine at the 5' end of the target sequence, which is recognized by two cryptic repeat folds at the amino-terminus. However their construction can be time and effort consuming, both because each TALEN often contains up to 20 RVDs, each of those 3–4 times bigger than ZFPs, and because of possible recombination within the cells due to high levels of sequence homology between each RDV. Moreover standard TALEN cannot cleave DNA that contains methylated cytosines, when located in the minor groove (Kim and Kim, 2014).

In 2011 two independent groups performed a rigorous side-by-side comparison between the gene editing efficiencies of ZFNs and TALE nucleases in transfected cell line and human ES cells and, by assessing the activities of these two platforms against the same genomic sites, they showed similar gene-editing activity for the two platforms (Hockemeyer et al., 2011; Mussolino et al., 2011). However, these two studies did not address one of the major concerns associated with a possible therapeutic use of these new technologies, i.e. their genome wide specificity. While these information are reported for the ZFNs (Gabriel et al., 2011; Pattanayak et al., 2011), they are still lacking for the TALE

nucleases, making difficult to understand the clinical value of the latter technology.



**Figure 1.13. Structure of TALENs.** A schematic representation of a transcription activator-like effector nuclease (TALEN) pair is shown. Each TALEN is composed of transcription activator-like effectors (TALEs) at the amino terminus and the *FokI* nuclease domain at the carboxyl terminus. Each TALE repeat is comprised of 33–35 amino acids and recognizes a single base pair through the amino acids at positions 12 and 13, which is called the repeat variable diresidues (RVD; shown in red). Target sequences of TALEN pairs are typically 30–40 bp in length, excluding spacers (Kim and Kim, 2014)

#### 1.4.3.3 RNA-guided engineered nucleases

RNA-guided nucleases are derived from CRISPR/Cas system, present in archaea and bacteria, which provides adaptive immunity against plasmids or phages. In fact, when foreign DNA is detected, this is cleaved into small fragments or *protospacers*, which are inserted into the bacterial genome to form a clustered regularly interspaced short palindromic repeats (CRISPR). CRISPR regions, together with a trans-activating crRNA (tracrRNA), are transcribed as pre-CRISPR RNA (pre-crRNA) and then processed as target-specific crRNA. The so generated tracrRNA and crRNA bind to CRISPR-associated protein 9 (Cas9), an enzymes contain two conserved nuclease domains, HNH and RuvC, capable of cleavage

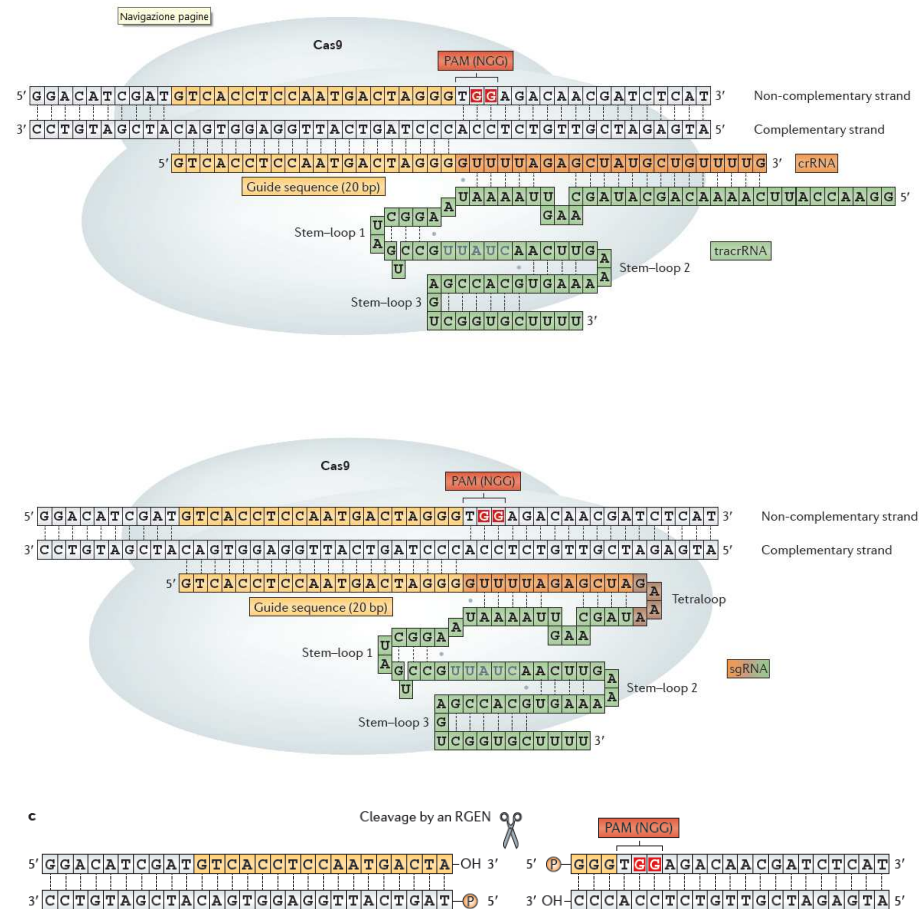
of DNA strand complementary and non-complementary to the guide RNA, respectively.

The so-generated dualRNA-Cas9 complex cleaves 23-bp target DNA sequence, composed of 20-bp guide sequence in the crRNA (the protospacer) and a sequence known as protospacer adjacent motif (PAM, 5'-NGG-3'). tracrRNA and crRNA can be engineered to form a single-chain guide RNA (sgRNA), while Cas9 with different PAM specificities can be retrieved by different bacterial sources than *Streptococcus pyogenes* Cas9 (Figure 1.14).

The main advantage of the RNA-guided engineered nucleases (RGENs) is their easy design and construction. In fact, while keeping the same Cas9 protein, RGENs with different target sequence can be simply constructed by cloning a 20-bp sequence (protospacer) into a vector encoding crRNA or sgRNA. However targetable sites are limited by the Cas9 requirement for PAM sequence.

The main drawback is the potential off-target activity, which may hamper clinical application of RGENs. In fact RGENs can recognize off-target sequences that differ up to 5-bps from the target sequence, thus potentially resulting in thousands of off-target mutations in the human genome which may lead to chromosomal translocation. Finally, their effective form as monomers further results in reduction of specificity (Kim and Kim, 2014). To reduce the frequency of off-target activity, the conversion of the Cas9 into a single-strand DNA nickase that generates DSBs by acting on two single DNA strands with two separate sgRNA, allowed reduction

of the nuclease activity in computationally predicted off-target sites (Cho et al., 2014; Shen et al., 2014).



**Figure 1.14 Structure of RGENs.** Schematic representations of RNA-guided engineered nucleases (RGENs) are shown. A, an RGEN is comprised of CRISPR (clustered regularly interspaced short palindromic repeat)-associated protein 9 (Cas9), a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA), which form the dualRNA-Cas9. B, alternatively, an RGEN can contain Cas9 and a single-chain guide RNA (sgRNA). The guide sequence in the crRNA (part A) or sgRNA (B) is complementary to a 20-bp target DNA sequence known as protospacer, which is next to the 5'-NGG-3' (where N represents any nucleotide) sequence known as protospacer adjacent motif (PAM). Grey dots indicate weak

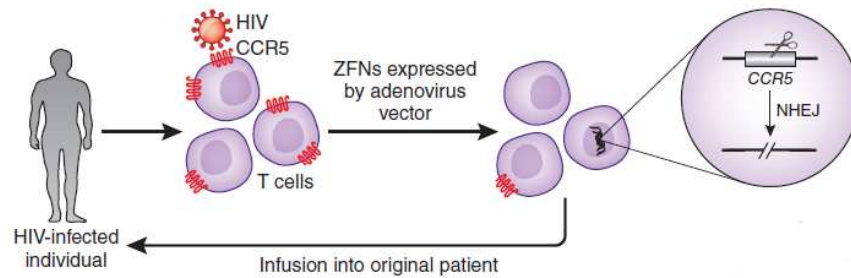
bonding. C, Target DNA cleaved by an RGEN yielding blunt ends is shown (Kim and Kim, 2014).

#### **1.4.4 Therapeutic application of genome editing**

Therapeutic Genome editing can be achieved through different strategies, including insertion of therapeutic gene, correction or inactivation of mutations. Artificial nucleases can be exploited to achieve genetic disruption of an intended locus. This application takes advantage of small insertions and deletions (indels) that are introduced during NHEJ-mediated DNA repair, to disrupt or abolish the function of a gene or genomic region. In the case of pathogenic mutations with gain-of-function (GOF), as those responsible for achondroplasia or Huntington disease, NHEJ-based approaches could be also useful to specifically inactivate the mutated allele, leaving intact the wild type allele on the homologous chromosome (Turizt-Cox).

In 2008 was reported the first ZFNs-based clinical trial, in which the ZFNs-mediated gene disruption was exploited to genetic inactivate *CCR5*, which encodes for the C-C chemokine receptor type 5. This protein, which acts as a co-receptor of HIV-1 on CD4+ cells, was found to be inactive in some individuals of Northern Europe, who appear protected from HIV infection (Liu et al., 1996). With this in mind, Sangamo BioSciences developed a gene mutation approach aimed at permanently disrupting the endogenous *CCR5* gene, based on *ex vivo* transduction of primary CD4+ T cells from healthy donors with an adenoviral ZFNs

expression vector to knock-out the *CCR5* gene located on chromosome 3 (Perez et al., 2008).



**Figure 1.15. Exploiting targeted gene editing for therapeutic applications.** (a) T cells from an HIV-infected individual are transduced by an adenoviral vector expressing ZFNs targeting the *CCR5* gene, a chemokine receptor essential in HIV infection. Repair of the DNA DSB by the NHEJ repair pathway inserts or deletes nucleotides, thus disrupting the coding sequence of the gene. The treated cells are then infused back into the patient—those lacking *CCR5* are resistant to HIV infection (Lombardo and Naldini, 2014).

Currently, the Sangamo *CCR5*-specific ZFNs are being assessed in two clinical trials to treat HIV/AIDS patients (NCT00842634; NCT01044654). Another phase I clinical trial for the treatment of glioblastoma (NCT01082926) started in 2010: here allogeneic CD8<sup>+</sup> Cytolytic T-Cell Line (CTL) were genetically modified to express the IL 13-Zetakine to specifically target tumour cells and treated with ZFN to disrupt the glucocorticoid receptor and confer resistance to anti-inflammatory glucocorticoids, which are often used in post-surgery profilaxis (Reik, ASGCT 2008).

A recent study, by combining genome-wide associations (GWAS) studies and chromatin immunoprecipitation sequencing (ChIP-seq), identified a non-coding mutation in *BCL11A* gene, whose product acts as a negative regulator of HbF in erythroid lineages (Xu et al,

2011). The mutation resulted in *BCL11A* enhanced expression, due to promotion of transcription factor binding within an intron in *BCL11A*, thereby decreasing HbF expression in red blood cells.

A therapeutic strategy for sickle cell anemia, whose severity decreases with HbF increased expression, would therefore aim at disrupting the intronic region in erythroid cells, thus augmenting the HbF expression levels and alleviate clinical symptoms (Bauer, 2013).

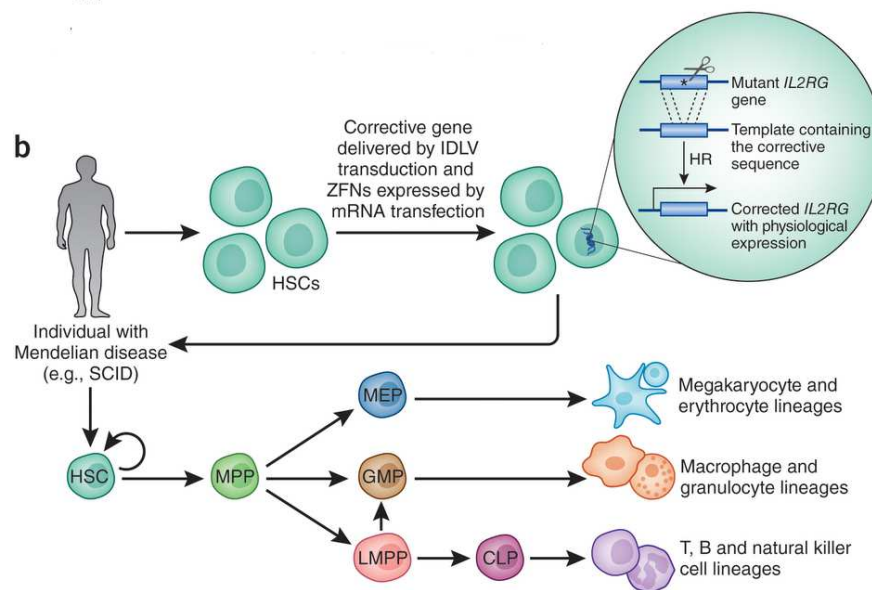
HDR instead, by introducing site specific mutations, can be exploited to restore the wild-type sequence and the gene function. HDR-based approaches are firstly required to treat loss-of-function mutations (LOF). Therapeutic effect may be achieved also by inserting a gain-of function mutant, when this exerts a protective effect (Turitz Cox, 2015).

To achieve gene insertion, nucleases are co-delivered with a targeting vector, in which the transgene cassette has been flanked by homology arms with nucleotide sequence identical to the region nearby the endogenous target site, as a plasmid, a non-integrase defective lentiviral vector (IDLV) or an adeno associated viral vector (AAVs). For point mutation, single-strand oligodeoxynucleotides (ssODNs), which can be synthesized in few days, can be used instead of targeting vectors.

In 2005 Urnov and colleagues reported ZFN-mediated targeted gene conversion into the exon 5 of the IL-2 Receptor common  $\gamma$ -chain (*IL2RG*). This gene encodes for a cytokine receptor that is required for T-, NK-, and B-cell development and its mutations cause the most common form of severe combined

immunodeficiency, SCID-X1. By this approach, the authors were able to achieve up to 18% of gene modification in K-562 cells and 5.3% in primary T-lymphocytes (Urnov, 2005).

In 2007 Lombardo et al. showed that the ZFN-technology could be used not only to edit a single nucleotide but also to integrate transgene expression cassettes into desired genomic sites (Lombardo, 2007). To achieve this goal, they developed a delivery platform based on Integrase- Defective Lentiviral Vectors (IDLV) to both introduce the donor template DNA and transiently express ZFNs in human cells. Using this means of delivery, they achieved up to 30% gene conversion of the *IL2RG* gene in hematopoietic cell lines and up to 40% targeted integration of transgene expression cassettes into the *IL2RG* or the *CCR5* gene across a panel of different human cell lines. Importantly, they broadened the application of ZFNs technology by exploiting these findings to knock-in a functional cDNA downstream of its endogenous promoter, thus achieving physiological expression of the corrected gene. In this setting, the same pair of ZFNs and donor vector can be used to correct most mutations (including deletions) occurring within and downstream the insertion site (Figure 1.16). IDLV-mediated delivery was also successful, albeit with low efficiencies, in human ESCs and cord blood-derived HSPCs, where they observed targeted integration in *CCR5* locus and stable transgene expression in up to 5% and 0.1% of the cells, respectively (Lombardo, 2007).



**Figure 1.16. Exploiting targeted gene editing for therapeutic applications.** In a proposed strategy currently only demonstrated in human cells transplanted into SCID mice to correct a Mendelian disease, HSCs from an individual affected by SCID-X1, a primary immunodeficiency caused by mutations in the *IL2RG* gene, would be transduced by an integrase-defective lentiviral vector (IDLV) containing a corrective *IL2RG* transgene and then transfected with mRNAs encoding for *IL2RG*-targeting ZFNs. Repair of the DNA DSB using the transduced IDLV as a template corrects the coding sequence of the gene and maintains its physiologic expression pattern. The treated HSCs could be infused back into the patient, where they would give rise to functional lymphocytes that would restore immunity long term. MPP, multipotent progenitor; MEP, megakaryocyte-erythroid progenitor; GMP, granulocyte-macrophage progenitor; LMPP, lymphoid-primed multipotent progenitor; CLP, common lymphoid progenitor (Lombardo and Naldini, 2014).

Different groups provided additional evidences that the ZFN technology could be used to insert expression cassettes into the genome of human cells. Indeed, the ZFN technology was successfully used for targeting six distinct loci in human ESCs and induced Pluripotent Stem (iPS) cells, without detectable alterations in stem cell

karyotype or in their pluripotency potential. The first proof-of-principle of targeted integration in human iPSCs came from the work of Zou and colleagues (Zou et al., 2009), where they disrupted by HR-mediated insertional mutagenesis the coding sequence of a disease-related gene, *PIGA*. This gene encodes for an enzyme mutated in the hematopoietic stem cells from patients suffering from paroxysmal nocturnal hemoglobinuria. This study was then extended by Hockemeyer and colleagues to other three genomic loci of human ESCs and iPSCs (Hockemeyer et al., 2009). First, using ZFNs specific for the *OCT4* locus, they generated OCT4-eGFP reporter cells to monitor the pluripotent state of ES cells. Second, by targeting the *PITX3* gene, they showed that ZFNs can be used to generate reporter cells by targeting non-expressed genes in ES and iPS cells.

Recently, Zou and colleagues exploited the *AAVS1* locus to insert an expression cassette for the gp91<sup>phox</sup> therapeutic minigene in iPS cells derived from X-linked chronic granulomatous disease (X-CGD) patients (Zou et al., 2011). This disease causes a reduced reactive oxygen species (ROS) production from neutrophils, which impairs their microbicidal activity. Upon neutrophil differentiation of the gene corrected iPSCs the authors observed sustained expression of gp91<sup>phox</sup> and functional restoration of ROS production. However, the overall gene targeting efficiencies reported for ES and iPS cells in all these studies were lower (up to 0.24%) as compared to what has been achieved in cell lines, but still is open the possibility to select and expand the clones carrying the desired genetic modification.

Genome editing is now established as a very versatile therapeutic approach, although several factors could influence the clinical efficacy. First of all, correction of the disease phenotype would rely on the genome modification threshold achieved. Secondly, if the therapeutic gene imparts a proliferative advantage over the unmodified counterpart, low numbers of edited cells will be sufficient to reverse the symptoms. For example, in the SCID-X1 clinical trials, corrected hematopoietic progenitors were able to reconstitute the affected lymphoid compartment, thus treating the disease. For other diseases, in which the mutated genes works in a non-cell autonomous fashion, a small number of corrected cells may be sufficient to product the restored protein to rescue the disease. Other factors that may influence the HDR outcome are the nature of the genome modification, the extent of complementarity between the target loci and the DNA template, the topology of the delivered donor DNA. Moreover, inhibition of the competing NHEJ pathway, by small molecules or viral proteins, has been shown to have positive effect in improving HDR levels (Maruyama et al., 2015).

## **1.5 Induced pluripotent stem cells: a new era for regenerative medicine**

One of the most important advances of the past decade was the discovery, by Yamanaka and Takahashi, that somatic cells can be reverted to a pluripotent state by forced expression of defined

transcription factors. In the past, cell reprogramming had been achieved by transfer of differentiated cell nuclei into enucleated oocytes (Gurdon, 1962) and by somatic cloning of adult differentiated cells to generate mammals (Wilmut, 1997), demonstrating that somatic cells contain all the information required to specify for the entire organism and that oocytes contain factors able to reprogram adult cells. These findings, together with the identification of transcriptional factors defined as “master regulators” of the fate of a given lineage and of the factors, as the leukemia inhibitory factor (LIF), responsible for ESC maintenance in culture enabled to define the transcription factors required for pluripotency induction in somatic cells (Takahashi and Yamanaka, 2006).

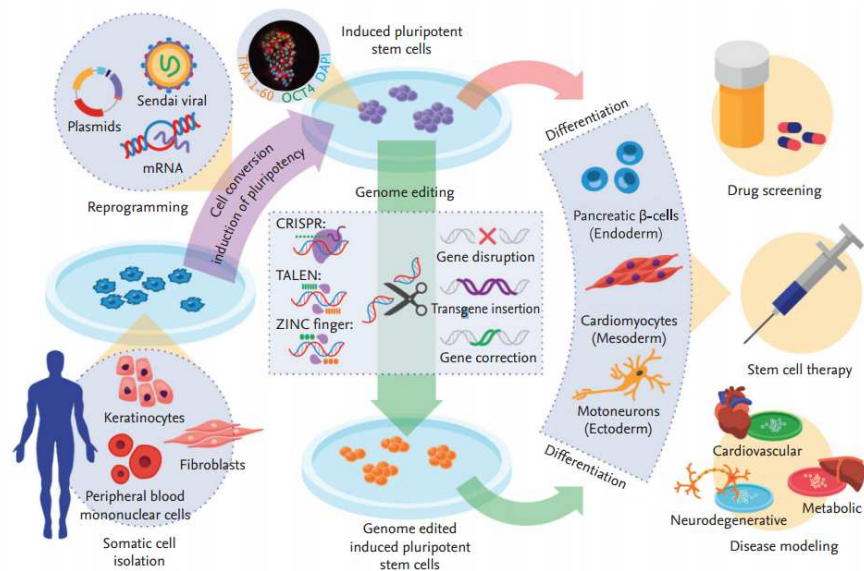
Stem cells include embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs). Both ESCs and iPSCs can be propagated indefinitely in culture and can be differentiated into several cell types. Importantly, while the use of ESCs, which are derived from the inner mass of preimplantation embryos, requires oocytes and embryo destruction, thus raising many ethical debates, iPSCs can be generated from many easily accessible cell types, rendering them an attractive resource for regenerative medicine and many other applications. As such, iPSCs became the object of different research areas, not only those focusing on the mechanisms of the stemness establishment and maintenance, but also those studying differentiation and specification of human somatic tissues, or to test phenotypic effects of small molecules and for preclinical toxicology screens for drug development.

Moreover, given the possibility to generate iPSCs from any individual, including patients with any disease, they hold invaluable relevance to study pathophysiology of human diseases (Yamanaka, 2012; Wu et al., 2011).

### **1.5.1 Advances in cell-reprogramming technologies**

The first reprogramming method was reported by Takahashi and Yamanaka to convert mouse embryonic fibroblasts (MEFs) and mouse-tail tip fibroblasts to iPSCs by overexpressing the four reprogramming factors (*Sox2*, *Oct4*, *c-Myc* and *Klf4*; OSKM) through a retroviral vector (Takahashi and Yamanaka, 2006). One year later, Yu et al. exploited lentiviral vectors to overexpress *OCT4*, *SOX2*, *NANOG*, and *LIN28* in human fibroblasts, thus converting them into a pluripotent state (Yu et al., 2007). During reprogramming, integrated proviruses are silenced, while endogenous pluripotency-associated genes are activated. It is worth notice that viral integration often occurs within endogenous genes, thus potentially resulting in adverse events, due to insertional mutagenesis. Moreover, reactivation of transgenes carried by retroviruses or lentiviruses can lead to tumours, while transgene overexpression may impact on differentiation potential of iPSCs (Zhao et al. 2011, Nakagawa 2008; Okita, 2007). Because of this, integration-free reprogramming systems were developed, as those based on plasmids, synthesized RNAs, proteins, adenovirus, Sendai virus or even small molecules, eliminating the risk of virally-induced tumour formation. Although safer, their main drawback is

the poor reprogramming efficiency, if compared to integrating systems.



**Figure 1.17. Generation of patient-specific induced pluripotent stem cells (iPSCs) and clinical applications thereof.** Somatic cells isolated from a patient are reprogrammed into iPSCs by transduction with the four reprogramming factors, octamer-binding transcription factor 4 (Oct4), sex determining region Y-box 2, Kruppel-like factor 4, and c-Myc. Genetic defects in iPSCs can be corrected via gene editing with zinc finger nucleases (ZFNs), activator-like effector nucleases (TALENs), and the clustered regularly interspaced short palindromic repeats (CRISPR) system. Next, iPSCs with or without edited modifications are differentiated into various target cells for disease modeling, drug screening, and stem cell therapy. DAPI, 4',6-diamidino-2- phenylindole (Diecke, Jung, Lee and Ju, 2014).

### 1.5.2 Molecular and epigenetic reprogramming.

Although iPSCs can be generated ideally from any somatic cell, the relative reprogramming efficiency is quite low and less than 1% of the transfected cells are efficiently converted to iPSC. Several studies have reported the dedifferentiation to pluripotency of terminally differentiated cells, including post-mitotic neurons or

lymphocytes, leading to the idea that, if not all, most of the somatic cells hold the potential to become iPSC. The bottleneck of the process seems to be, not the initiation of the reprogramming process, rather than its completion. The molecular events that govern reprogramming are: the epigenetic and transcriptional changes upon ectopic overexpression of the nuclear factors, the transition through intermediate cell states and, finally, the activation of a self-endogenous pluripotency circuit (Theunissen and Jaenisch, 2014). To dissect the heterogeneous stages of the reprogramming process, cell surface markers have been exploited to follow the transition of somatic cell to pluripotency.

A first attempt described the reprogramming process as the sequence of three waves of gene regulation (Hansson, 2012; Polo, 2012). The first wave, from day 0 to day 3, is characterized by upregulation of genes involved in metabolism, cell proliferation and cytoskeletal organization. The second wave takes place from day 3 to day 9, with an increase of activity of pluripotency-associated genes in those cells positive for the stage-specific embryonic antigen 1 (*SSEA-1*) while, from day 9 starts the third wave, with an increase in gene involved in stem cell maintenance and in pluripotency associated-promoters demethylation. The main limit of this roadmap is that most of the cells *SSEA-1*-positive never achieve pluripotency. To overcome this limitation, other markers have been considered, as CD44 or ICAM, which is homogenously expressed in iPSCs, although other studies have suggested that the expression of one or more transcriptional

determinants as *Esrrb*, *Utf1*, *Lin28* or *Dppa2*, would better predict progression toward pluripotency (Buganim et al., 2012).

To explain the kinetic of the reprogramming, several models have been proposed. The stochastic model assumes that pluripotency is the final result of the combination of one or more rate-limiting steps, while the deterministic model describes the reprogramming process as serial ordered events with fixed latency. Other studies, however, supported the stochastic nature of the early changes in gene expression, while later stages, with the reactivation of the endogenous pluripotency circuit, are in line with a deterministic fate.

Furthermore, reprogramming efficiency can be influenced by culture conditions, by levels and stoichiometry of the reprogramming factors, and by the gene delivery system adopted. In addition, stochastic events, which are far beyond experimental control, can impact the reprogramming effectiveness. The efficiency of this process, however, can be increased by specifically removing barriers that hamper induced pluripotency.

Despite preliminary studies, in which iPSCs were shown to be very similar to ESCs, there is substantial evidence of differences in gene expression, DNA methylation, teratomas propensity and *in vitro* differentiation potential. Indeed, the clonal origin of iPSCs may account for copy number variations, while some other alterations may be the result of culturing. However it is to be clarified whether these differences are associated to the reprogramming process, if they are solely the result of biological variations or handlings of the

cells or if they represent inherent differences between ESCs and iPSCs (Wu and Hochedliger, 2011).

In the original Yamanaka's combination, *Oct4*, *Sox2*, *Klf4* and *c-Myc* (OSKM) were used to induce pluripotency, but, in theory, each factor can be replaced by other transcription factors. In the 2008, a three-factor OSK reprogramming combination was used to reprogram fibroblasts (Nakagawa et al., 2008; Wernig et al., 2008); iPSCs were still generated but the reprogramming process was significantly delayed, respect to that in which *c-Myc* was included. This gene is thought to promote open chromatin, to transcriptionally amplify genes involved in proliferation and to promote the initial engagement of OSK with several chromatin sites (Buganim, 2012). Given the high redundancy of the nuclear factors required for pluripotency, *Klf-4* and *c-Myc* can be replaced with *Esrrb* in mouse fibroblasts (Fheng et al., 2009), and with *NANOG* and *LIN28* in human fibroblasts (Yu et al. 2007). *Oct4* can be as well replaced by *E-cadherin*, a master regulator of the epithelial phenotype that prevents the nuclear localization of the  $\beta$ -catenin, known to negatively regulate the early-reprogramming process phase (Ho, 2013). Buganim and colleagues reported that *Sox2* activation is able to turn on *Sal4*, which in turn activates four downstream targets, including *Oct4*. Moreover, this study revealed other two four-factors cocktails that can replace OSKM: *Esrrb*, *Sal4* and *Lin28* with either *Nanog* or *Dppa2* (Buganim, 2012). Other strategies to induce pluripotency exploits chromatin-remodelling enzymes, as corepressor *Rco2* (Yang, et al., 2012) or the histone variants (TH2A and TH2B) (Shinigawa et al., 2014) or,

alternatively, the removal of epigenetic barriers to reprogramming as *DOT1L*, a histone H3 lysine 79 (H3K79) methyltransferase.

Of particular interest is *Mbd3*, the scaffold protein of the nucleosome remodelling and deacetylase (NuRD) complex, whose depletion in mouse and human somatic cells significantly enhanced the reprogramming efficiency, up to 100%. Although its role in reprogramming is still under investigation, *Mbd3* physically interacts with the four OSKM factors, maybe altering their biological activity (Rais et al., 2013). Also miRNAs, which promote reprogramming through different mechanisms, can be used to replace transcription factors. ESCC miRNAs are a large family of miRNAs highly expressed in ESCs, which were found to enhance the reprogramming efficiency of MEF to iPSCs. The miR-302/367 cluster, one of the human orthologues of the ESCCs miRNAs, has been identified as a direct target of Oct4 and Sox2. In 2011 it was reported that miR-302/367 cluster, in combination with valproic acid (VPA), which specifically degrades Hdac2 protein, can drive direct reprogramming of both mouse and human fibroblasts in the absence of other pluripotency-associated transcription factors, with an efficiency higher than two-magnitude orders respect to OSKM reprogramming. Many reasons may have contributed to this result, indeed miR302/367 targets hundreds of mRNA targets, including those responsible for chromatin remodelling regulation and cell proliferation (Subramanyam et al., 2011; Anokye-Danso et al., 2011). One of the next-challenges would be to develop robust transgene-free methodologies to induce pluripotency, as those based on chemical reprogramming. Indeed it

has been reported that the DNA-methyltransferase inhibitor 5-azacytidine and the 2i cocktail, a combination of the MEK inhibitor PD0325901 and the GSK3 inhibitor CHIR99021, is able promote the final stage of reprogramming. Other compounds as valproic acid (VPA), a histone deacetylase (HDAC) inhibitor, have been shown to reprogram mouse fibroblasts in the absence of *c-Myc* or human fibroblasts in the absence of *Klf4* and *c-Myc* (Huangfu et al., 2008). Also vitamin C has been identified as a positive modulator of pluripotency. In fact, by promoting the activity of the Jhdmlb demethylase, it reduces the methylation levels at Histone 3 lysine 36 and suppresses cell senescence. Moreover, vitamin c is thought to preserve the imprinted status of the locus *Dlk-Dio3* through histone modifications that prevent Dnmt3a binding. Other compounds have been exploited to enhance reprogramming as the TGF- $\beta$  inhibitor 616452 or forskolin (FSK), an AMP agonist, which has been successfully used as *Oct4* replacer (Hou et al., 2013).

### **1.5.3 Road to the future: iPSCs clinical application**

iPSCs technology holds tremendous potential for regenerative medicine, allowing their differentiation into a wide range of specialized cell types and tissues to be used in the future to replace damaged or diseased tissues in patients with experience of trauma, diabetes, cardiovascular diseases, metabolic deficiencies or degenerative neurological disorders. Regarding the potential clinical applicability of ESCs/iPSCs, Geron has reported the use of ESCs-derived oligodendrocytes for the treatment of spinal cord

injury ([www.clinicaltrials.gov](http://www.clinicaltrials.gov)), while Advanced Cell Technology has received FDA approval to conduct a clinical trial using ESCs-derived retinal cells for the treatment of Stargardt disease and age-related macular degeneration (Wu and Hochedlinger, 2011). Regarding the application of iPSCs for therapeutic purpose, the most compelling study used iPSCs-derive hematopoietic cells for the treatment of a humanized mouse model of sickle cell anemia (Hanna et al., 2007). In this study, iPSCs were derived from a transgenic mouse carrying a mutated sequence of the human haemoglobin gene and then genetically corrected through HDR. Transplantation of iPSCs-derived hematopoietic cells into the transgenic mice resulted in restoration of haemoglobin to normal levels and improved phenotype. Although this remarkable result, the translational potential of this strategy for sickle cell anemia in human beings remains to be determined. These studies, in addition with many pre-clinical studies in animal models, may support the therapeutic potential of iPSCs for the treatment of genetic disorders as well as for regenerative medicine. However, before bringing iPSCs biology into cell-based therapy, several hurdles need to be considered, as the efficiency of cell-lineage specification and of cell-purification, to exclude the risks of teratomas, and the development of new cell delivery methods for relevant organs (Wu and Hochedlinger, 2011). To overcome the potential tumorigenicity of the cells, the use of progenitors or terminally differentiated cells may increase the safety of iPSCs for clinical applications. Also the transdifferentiation of an adult cell type into another one would help to circumvent the tumorigenicity associated to pluripotent

stem cells and would provide an alternative source of clinically relevant cell type. However this approach brings with it the disadvantage that somatic cells, contrary to iPSCs, have a limited lifespan and are not expandable (Wu and Hochedlinger, 2011).

To further limit the risk of insertional mutagenesis and the consequent tumour formation non-integrant systems, as those based on Sendai virus, DNA-based episomal reprogramming, mRNA or protein transduction, are obviously the most favoured reprogramming techniques for safety issues. For their therapeutic usage, iPSCs can be derived directly from patients who require therapy, thus minimizing the risk of transplant rejection. Fibroblasts, keratinocytes or peripheral mononuclear cells are the preferred donor cells, since they are easily obtained from the patients. However, the production of autologous clinical grade cells for each patient would be difficult and financially prohibitive. An alternative would be to create a bank of clinical grade iPSC lines from healthy volunteer donors, which can be expanded and differentiated for use in a large cohort of patients, selected in order to maximize HLA-matching, thus minimizing the risk of allograft rejection (Taylor et al., 2012).

iPSCs can be relevant also for disease modelling, to recapitulate pathological conditions *in vitro* by patient-derived somatic cell conversion to iPSCs, followed by their differentiation into disease-specific cell types. The factors that may influence the amenability of diseases to *in vitro* modelling are: the disease onset in patients, the complexity of the underlying genetic defects and the cell-autonomous nature of the disorder (Wu and Hochedlinger, 2011). If

relevant iPSC and target cells would be globally available for research and drug development purposes, this would result in better standardized conditions and in a substantial improvement in safety, feasibility and accuracy. Assuming that disease features can be reproduced in vitro, one of the major limitation is the lack of established lineage-specific differentiation protocols to derive purified cells for large-scale screenings. Moreover, disease modelling with iPSCs is hampered by the heterogeneity of the maturation stage of the differentiated iPSC, due to donor source and to culture conditions (Diecke et al., 2014).

#### **1.5.4 iPSCs as alternative source of HSCs**

In principle, iPSCs can be exploited to derive HSC or hematopoietic progenitor cells (HPCs) of clinical interest for therapeutic application.

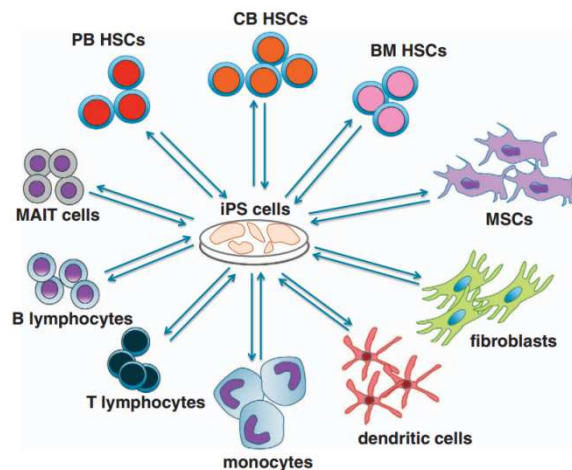
Seminal studies on embryogenesis and ESCs differentiation have provided major insights into key pathways that control ESCs/iPSCs hematopoietic commitment, allowing the identification of the essential role of Wnt, Notch, Hedgehog and TGF $\beta$ -Smad signaling pathways in HSCs development and maintenance (Sluvkin, 2012). In the mouse embryo, the yolk sac is the first to generate hematopoietic cells, including macrophage, megakaryocytes and red blood cells. The first HSCs able to full hematopoietic reconstitution are observed in the aorta-gonado-mesonephros region, vitelline and umbilical arteries. After expansion in the fetal liver, HSCs migrate in the bone marrow, which become the major site of haematopoiesis in post-natal life. Given the high similarities

of hematopoietic differentiation patterns between hESCs and ESCs, the knowledge on both molecular mechanisms and niche factors critical for hematopoietic specification has allowed the establishment of *in vitro* protocols to generate hematopoietic stem cells and progenitors, although further understanding of HSCs biology is a prerequisite toward optimized hematopoietic differentiation protocol.

To date, three different methods to derive hematopoietic cells from iPSCs have been described: co-culture with murine mesenchymal OP9 cells, dissociation of teratomas induced in iPSCs-injected immunodeficient mice (Amabile et al., 2013) or iPSCs transduction with Lhx2, a LIM-homeobox transcription factor, although the latter is eligible only for hematopoietic conversion of murine iPSC (Sluvkin, 2012). Other strategies relies on the dissociation of embryoid bodies formed by iPSCs at 7-10 days; this methods, however, tends to be more variable into iPSC-derived hematopoietic progenitor cells (Focosi et al., 2014). NK lymphocytes, with therapeutic potential both for cancer and infectious disease have been generated, however the major limitation is represented by the two-steps stromal-cells-co-culture and the need for sorting of rare CD45<sup>+</sup> CD34<sup>+</sup> population.

An interesting application of iPSCs is the generation of autologous antigen-specific T-lymphocytes for immunotherapy. Eventually, T-lymphocytes can be engineered with chimeric antigen receptors to confer them new desired antigen specificities (Themeli et al., 2013). The most established system for iPSC-derived T cell differentiation is the co-culture with the OP9-stromal cell line,

deficient in macrophage colony-stimulating factor (MCSF), engineered to ectopically express the Notch ligand Delta 1 (Dll1) or Notch-ligand Delta 4 (Dll4) to support both murine and human T-lineage differentiation. In 2011 Carpenter reported also the generation of iPSC-derived B-lymphocytes, by showing iPSC differentiation when co-cultured with OP9 stromal cells (Carpenter et al., 2011).



**Figure 1.18. Blood cell types generated from iPSC.** Summary of blood cell types successfully used as a source for iPSCs cell generation and of blood cells successfully redifferentiated to blood cells to date (Focosi et al., 2014).

Red-blood cells are the ideal candidate for iPSC-based clinical trials, because of the short half-life and absence of nucleus, thus sparing the patients the risks of oncogenicity. iPSCs have the potential to produce pathogen-free RBCs, however the major limitation for clinical translation are the large amount of RBSs required to generate a unit ( $10^{12}$ ), the poor enucleation efficiency and the switching to adult-type ( $\beta$ ) globin-forms. Also autologous

iPSC-derived platelets have been generated and will be soon tested in clinical trials (Advanced Cell Technologies).

## Scope of the thesis

The aim of this work is to develop an optimized strategy that allows for efficient generation of disease-free iPSCs from primary cells, by exploiting targeted genome editing to overcome the risks of insertional mutagenesis or unregulated gene expression associated to integrating-vector based gene-therapy.

The feasibility of this approach, based on site-specific genome editing of primary somatic cells through Integrase-Defective Lentiviral Vectors (IDLV) and engineered Zinc Finger Nucleases (ZFNs) and on cell-reprogramming of the so-generated gene-corrected cells, will be assessed on SCID-X1, an immunological disorder caused by mutations in the Interleukin 2 Receptor Common Gamma-chain (*IL2RG*) gene.

By inserting a corrective *IL2RG* cDNA downstream its endogenous promoter, this strategy will allow the correction a broad spectrum of SCID-X1 mutations downstream to the insertion site, with the same set of ZFNs and donor vector, resulting in the reconstitution of both physiologic expression and function of the corrected gene.

Overall, the new strategy presented in this work, which couples gene-correction with cell-reprogramming, will allow the generation of disease-free iPSC, with tremendous potential for cell-gene therapy, paving the way for the development of novel and safer therapeutic approaches for SCID-X1, as well as for other monogenic disorder.



## References

Abraham R.T. (2001) Cell cycle checkpoint signaling through the ATM and ATR kinases. *Genes Dev.*15(17): 2177-96.

Aiuti A., Roncarolo M.G. (2009) Ten years of gene therapy for primary immune deficiencies. *Hematology Am Soc Hematol Educ Program*, 2009: 682-9.

Aiuti, A., Biasco L., Scaramuzza S., Ferrua F., Cicalese M.P., Baricordi C., Dionisio F., Calabria A., Giannelli S., Castiello M.C., Bosticardo M., Evangelio C., Assanelli A., Casiraghi M., Di Nunzio S., Callegaro L., Benati C., Rizzardi P., Pellin D, Di Serio C, Schmidt M, Von Kalle C., Gardner J., Mehta N., Neduva V., Dow J.D., Anne Galy A., Miniero R., Finocchi A., Metin A., Banerjee P.P., Orange J.S., Galimberti S., Valsecchi M.G., Biffi A., Montini E., Villa A., Ciceri F., Roncarolo M.G., Luigi Naldini (2013). Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott–Aldrich syndrome. *Science*, 341: 123-151.

Aiuti, A., Cattaneo F., Galimberti, S., Benninghoff, U., Cassani, B., Callegaro, L., Scaramuzza, S., Andolfi G., Mirolo, M., Brigida, I., Tabucchi, A., Carlucci, F., Eibl, M., Aker, M., Slavin, S., Al-Mousa H., Al Ghonaium A., Ferster A., Duppenhaler A., Notarangelo, L., Wintergerst U., Buckley R.H., Bregni, M., Marktel, S., Valsecchi, M.G., Paolo Rossi, Ciceri F., Miniero, R., Bordignon C. and Roncarolo, M.G. (2009) Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *N. Engl. J. Med.* 360, 447–458.

Anliker B., Abel T., Kneissl S., Hlavaty J., Caputi A., Brynza J., Schneider I.C., Münch R.C., Petznek H., Kontermann R.E., Koehl U., Johnston I.C., Keinänen K., Müller U.C., Hohenadl C., Monyer H., Cichutek K., Buchholz C.J. (2010) Specific gene transfer to neurons, endothelial cells and hematopoietic progenitors with lentiviral vectors. *Nat Methods*, (11):929-35.

Anokye-Danso F., Trivedi C.M., Jühr D., Gupta M., Cui Z., Tian Y., Zhang Y., Yang W., Gruber P.J., Epstein J.A., Morrissey E.E. (2011) Highly efficient miRNA-mediated reprogramming of mouse

and human somatic cells to pluripotency. *Cell Stem Cell*, 8: 376-388.

Banasik, M.B., and P.B. McCray, Jr. (2010) Integrase-defective lentiviral vectors: progress and applications. *Gene Ther.*, 17:150-157.

Bauer D.E., Kamran S.C., Lessard S., Xu J., Fujiwara Y., Lin C., Shao Z., Canver M.C., Smith E.C., Pinello L., Sabo P.J., Vierstra J., Voit R.A., Yuan G.C., Porteus M.H., Stamatoyannopoulos J.A., Lettre G., Orkin S.H. (2013). An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* 6155: 253-257.

Bauer T.R., Allen J.M., Hai M., Tuschong L.M., Khan I.F., et al. (2008) Successful treatment of canine leukocyte adhesion deficiency by foamy virus vectors. *Nat Med* 14: 93–97.

Biffi A., Montini E., Lorioli L., Cesani M., Fumagalli F., Plati T., Baldoli C., Martino S., Calabria A., Canale S., Benedicenti F., Vallanti G., Biasco L., Leo S., Kabbara N., Zanetti G., Rizzo W.B., Mehta N.A., Cicalese M.P., Casiraghi M., Boelens J.J., Del Carro U., Dow D.J., Schmidt M., Assanelli A., Neduva V., Di Serio C., Stupka E., Gardner J., von Kalle C., Bordignon C., Ciceri F., Rovelli A., Roncarolo M.G., Aiuti A., Sessa M., Naldini L. (2013) Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. *Science* 341: 123-158.

Biffi, A., Capotondo A., Fasano S., del Carro U., Marchesini S., Azuma H., Malaguti M.C., Amadio S., Brambilla R., Grompe M., Bordignon C., Quattrini A., Naldini L. (2006) Gene therapy of metachromatic leukodystrophy reverses neurological damage and deficits in mice. *J. Clin. Invest.* 116, 3070–3082.

Biffi, A., De Palma M., Quattrini A., Del Carro U., Amadio S., Visigalli I., Sessa M., Fasano S., Brambilla R., Marchesini S., Bordignon C., Naldini L. (2004) Correction of metachromatic leukodystrophy in the mouse model by transplantation of genetically modified hematopoietic stem cells. *J. Clin. Invest.* 113, 1118–1129.

Bitinaite J, Wah D.A., Aggarwal A.K., Schildkraut I. (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A*. 95(18):10570-5.

Blackwell, T.K., and H. Weintraub. (1990) Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science*. 250:1104-1110.

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

Bogdanove, A.J., Schornack S. and Lahaye T. (2010) TAL effectors: finding plant genes for disease and defense. *Current opinion in plant biology*. 13:394-401.

Boztug, K., Schmidt M., Schwarzer A., Banerjee, P.P., Avedillo Díez I., Dewey R.D., Böhm M., Nowrouzi A., Ball C.R., Glimm H., Naundorf S., Köhlcke K., Blasczyk R., Kondratenko I., Maródi L., Orange J.S., von Kalle C. and Klein C. (2010) Stem-cell gene therapy for the Wiskott-Aldrich syndrome. *N. Engl. J. Med.* 363, 1918–1927.

Branda C.S., Dymecki S.M. (2004) Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. *Dev Cell*. (1):7-28.

Branzei D., Foiani M. (2010) Maintaining genome stability at the replication fork. *Nat Rev Mol Cell Biol*. (3):208-19.

Braun C.J. et al. (2014) Gene therapy for Wiskott–Aldrich syndrome—long-term efficacy and genotoxicity. *Sci. Transl. Med.* 6, 227ra233.

Brown B.D., Gentner B., Cantore A., Colleoni S., Amendola M., Zingale A., Baccarini A., Lazzari G., Galli C., Naldini L. (2007) Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state. *Nat Biotechnol*. 25(12):1457-67

Brown B.D., Naldini L. (2009) Exploiting and antagonizing microRNA regulation for therapeutic and experimental applications. *Nat Rev Genet*, 10(8):578-85.

Buchholz D.E, Hikita S.T., Rowland T.J., Friedrich A.M., Hinman C.R., Johnson L.V and Clegg D.O. (2009). Derivation of functional retinal pigmented epithelium from induced pluripotent stem cells. *Stem Cells*, 10:2427-34.

Bushman F., Lewinski M., Ciuffi A., Barr S., Leipzig J., Hannenhalli S., Hoffmann C. (2005) Genome-wide analysis of retroviral DNA integration. *Nat Rev Microbiol*. 3: 848-58.

Candotti, F. et al. (2012) Gene therapy for adenosine deaminase-deficient severe combined immune deficiency: clinical comparison of retroviral vectors and treatment plans. *Blood*, 120, 3635–3646.

Canver M. C., Smith E. C., Sher F., Pinello L., Sanjana N.E., Shalem O., Chen D.D, Schupp P.G., Vinjamur D.S, Garcia S.P., Luc S., Kurita R., Nakamura Y., Fujiwara Y., Maeda T., Yuan G., Zhang F., Orkin S.H. & Bauer D.E. (2015) BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*, 527, 192–197

Capotondo, A. et al. Brain conditioning is instrumental for successful microglia reconstitution following hematopoietic stem cell transplantation. *Proc. Natl Acad. Sci. USA* 109, 15018–15023 (2012).

Carpenter L., Malladi R., Yang C.T., Franch A., Plikington K.J., Forsey R.W. Sloane-Stanley J., Silk K.M., Davies T.J., Fairchild P.J., Enver T., Watt S.M. (2011) Human induced pluripotent stem cells are capable of B-cell lymphopoiesis. *Blood*. 14;117(15): 4008-11.

Cartier, N. et al. (2009) Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* 326, 818–823.

Cartier, N. et al. (2012) Lentiviral hematopoietic cell gene therapy for X-linked adrenoleukodystrophy. *Methods Enzymol.* 507, 187–198.

Cavazzana-Calvo M, Fischer A. (2007) Gene therapy for severe combined immunodeficiency: are we there yet? *J Clin Invest.* 117(6):1456-65.

Cavazzana-Calvo, M. et al. (2010) Transfusion independence and HMGA2 activation after gene therapy of human  $\beta$ -thalassaemia. *Nature* 467, 318–322.

Cho S.W. (2014) Analysis of off-target effects of CRISPR/cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 24: 132-141.

Ciccio A., Elledge S.J. (2010) The DNA damage response: making it safe to play with knives. *Mol Cell*, 40: 179-204.

Collier L.S., Carlson C.M., Ravimohan S., Dupuy A.J., Largaespada D.A. (2005) Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature*, 436: 272-6

Copelan, E. A. (2006) Hematopoietic stem-cell transplantation. *N. Engl. J. Med.* 354, 1813–1826.

Cronin J., Zhang X., and Raiser J. (2005) Altering the Tropism of Lentiviral Vectors through Pseudotyping. *Curr Gene Ther.* 4: 387–398.

de Boer B.A., Ruijter J.M., Voorbraak F.P., Moorman A.F. (2009) More than a decade of developmental gene expression atlases: where are we now? *Nucleic Acids Res.*

de Felipe P, Luke GA, Hughes LE, Gani D, Halpin C, Ryan MD (2006) E unum pluribus: multiple proteins from a self-processing polyprotein. *Trends Biotechnol* 24: 68-75.

de Semir D, Aran J.M. (2006) Targeted gene repair: the ups and downs of a promising gene therapy approach. *Curr Gene Ther.* 6(4):481-504.

Deyle D.R., Russell D.W.(2009) Adeno-associated virus vector integration. *Curr Opin Mol Ther.* 11(4):442-7.

Dorrell C., Gan O.I., Pereira D.S., Hawlet R.G., Dick J.E. (2000). *Blood*, 95: 102-110.

Doyon, Y., T.D. Vo, M.C. Mendel, S.G. Greenberg, J. Wang, D.F. Xia, J.C. Miller, F.D. Urnov, P.D. Gregory, and M.C. Holmes. 2011. Enhancing zinc-finger nuclease activity with improved obligate heterodimeric architectures. *Nature methods.* 8:74-79.

Durai S., Mani M., Kandavelou K., Wu J., Porteus M.H., Chandrasegaran S. (2005) Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res.* 33(18):5978-90.

Ellis J. (2005) Silencing and variegation of gammaretrovirus and lentivirus vectors. *Hum Gene Ther.* 16:1241-6.

Engelman A., Cherepanov P. (2012) The structural biology of HIV-1: mechanistic and therapeutic insights. *Nat Rev Microbiol.* 10(4):279-90.

Ferrua, F., Brigida, I. & Aiuti, A. (2010) Update on gene therapy for adenosine deaminase-deficient severe combined immunodeficiency. *Curr. Opin. Allergy Clin. Immunol.* 10, 551–556.

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

Focosi D., Amabile G., Di Ruscio A., Quaranta P., Tenen D.G. and Pistello M. (2014) Induced pluripotent stem cells in hematology: current and future applications. *Blood cancer Journal.*

Foley, J.E., J.R. Yeh, M.L. Maeder, D. Reyon, J.D. Sander, R.T. Peterson, and J.K. Joung. 2009. Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). *PloS one*. 4:e4348.

Follenzi A., Ailles L.E., Bakovic S., Geuna M. and Naldini L. (2000) Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. *Nature Genetics*, 25: 217 – 222.

Frazer K.A., Murray S. S., Schork N.J. and Topol E.J.(2009) Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics* 10, 241-251.

Frecha C.,Lévy C., Costa C., Nègre D., Amirache F., Buckland R., Russell S.J., Cosset F.L. and Verhoeyen E. (2011) Measles Virus Glycoprotein-Pseudotyped Lentiviral Vector-Mediated Gene Transfer into Quiescent Lymphocytes Requires Binding to both SLAM and CD46 Entry Receptors. *J Virol*. 85(12): 5975–5985.

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

Gabriel, R., A. Lombardo, A. Arens, J.C. Miller, P. Genovese, C. Kaeppl, 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*. 29:816-823.

Gabriel, R., R. Eckenberg, A. Paruzynski, C.C. Bartholomae, A. Nowrouzi, A. Arens, S.J. Howe, A. Recchia, C. Cattoglio, W. Wang, K. Faber, K. Schwarzwaelder, R. Kirsten, A. Deichmann, C.R. Ball, K.S. Balaggan, R.J. Yanez-Munoz, R.R. Ali, H.B. Gaspar, L. Biasco, A. Aiuti, D. Cesana, E. Montini, L. Naldini, O. Cohen-Haguenauer, F. Mavilio, A.J. Thrasher, H. Glimm, C. von Kalle, W. Saurin, and M. Schmidt. 2009. Comprehensive genomic access to vector integration in clinical gene therapy. *Nat Med*. 15:1431-1436.

Gabriel R, von Kalle C. & Schmidt M. (2015) Mapping the precision of genome editing. *Nat. Biotech.* 33: 150-151

Gansbacher B. (2003) Report of a second serious adverse event in a clinical trial of gene therapy for X-linked severe combined immune deficiency (X-SCID). *J. Gene Med.* 3:261-2.

Gaspar H.B., Parsley K.L., Howe S., King D., Gilmour K.C., Sinclair J., Brouns G., Schmidt M., Von Kalle C., Barington T., Jakobsen M.A., Christensen H.O., Al Ghonaium A., White H.N., Smith J.L., Levinsky R.J., Ali R.R., Kinnon C., Thrasher A.J. (2004) Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet*, 364(9452):2181-7.

Gaspar, H. B. et al. (2011). Hematopoietic stem cell gene therapy for adenosine deaminase-deficient severe combined immunodeficiency leads to long-term immunological recovery and metabolic correction. *Sci. Transl. Med.* 3, 97ra80; erratum 5, 168er1.

Gaspar H.B., Cooray S., Gilmour K.C., Parsley K.L., Adams S., Howe S.J., Al Ghonaium A., Bayford J., Brown L., Davies E.G., Kinnon C., Thrasher A.J. (2011) Long-term persistence of a polyclonal T cell repertoire after gene therapy for X-linked severe combined immunodeficiency. *Sci. Transl. Med.* 3, 97ra79.

Gennery A.R., Slatter M.A., Grandin L., Taupin P., Cant A.J., Veys P., Amrolia P.J., Gaspar H.B., Davies E.G., Friedrich W., Hoenig M., Notarangelo L.D., Mazzolari E., Porta F., Bredius R.G.M., Lankester A.C., Wulffraat N.M., Seger R., Güngör T., Fasth A., Sedlacek P., Neven B., Blanche S., Fischer A., Cavazzana-Calvo M., Landais P. (2010) Transplantation of Haematopoietic Stem Cells and Long Term Survival for Primary Immunodeficiencies in Europe: entering a new century, do we do better? *J Allergy Clin Immunol* 126 602-610.

Genovese P., Schirotti G., Escobar G., Di Tomaso T., Firrito C., Calabria A., Moi D., Mazziere R., Bonini C., Holmes M.C., Gregory P.D., van der Burg M., Gentner B., Montini E., Lombardo

A., Naldini L. (2014) Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* 510: 235-240.

Gentner B., Visigalli I., Hiramatsu H., Lechman E., Ungari S., Giustacchini A., Schira G., Amendola M., Quattrini A., Martino S., Orlacchio A., Dick J.E., Biffi A., Naldini L. (2010) Identification of hematopoietic stem cell-specific miRNAs enables gene therapy of globoid cell leukodystrophy. *Sci Transl Med.* 17;2(58):58ra84

Ginn S.L., Curtin J.A., Kramer B., Smyth C.M., Wong M., Kakakios A., McCowage G.B., Watson D., Alexander S.I., Latham M., Cunningham S.C., Zheng M., Hobson L., Rowe P.B., Fischer A., Cavazzana-Calvo M., Hacein-Bey-Abina S., Alexander I.E. (2005) Treatment of an infant with X-linked severe combined immunodeficiency (SCID-X1) by gene therapy in Australia. *Med J Aust.* 182(9):458-63.

Girard L., Hanna Z., Beaulieu N., Hoemann C.D., Simard C., Kozak C.A. and Jolicoeur P. (1996). Frequent provirus insertional mutagenesis of Notch1 in thymomas of MMTVD/myc transgenic mice suggests a collaboration of c-myc and Notch1 for oncogenesis. *Genes Dev.*, 10, 1930–1944.

Goverdhan S, Puntel M., Xiong W., Zirger J.M., Barcia C., Curtin J.F. (2005). Regulatable gene expression systems for gene therapy applications: progress and future challenges. *Mol Ther*, 12:189–211.

Gupta, A., X. Meng, L.J. Zhu, N.D. Lawson, and S.A. Wolfe. 2011. Zinc finger protein independent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. *Nucleic acids research.* 39:381-392.

Hacein-Bey Abina, S. et al. (2015) Outcomes following gene therapy in patients with severe Wiskott–Aldrich syndrome. *J. Am. Med. Assoc.* 313, 1550–1563.

Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, Morillon E, Clappier E, Caccavelli L, Delabesse E, Beldjord K, Asnafi V, MacIntyre E, Dal Cortivo L, Radford I, Brousse N, Sigaux F, Moshous D, Hauer J, Borkhardt A, Belohradsky BH,

Wintergerst U, Velez MC, Leiva L, Sorensen R, Wulffraat N, Blanche S, Bushman FD, Fischer A, Cavazzana-Calvo M (2008) Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* 118: 3132-3142.

Hacein-Bey-Abina S, von Kalle C, Schmidt M, Le Deist F, Wulffraat N, McIntyre E, Radford I, Villeval JL, Fraser CC, Cavazzana-Calvo M, Fischer A (2003a) A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med* 348: 255-256.

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

Hacein-Bey-Abina, S. et al. (2014) A modified  $\gamma$ -retrovirus vector for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* 371, 1407–1417.

Hacein-Bey-Abina, S., et al (2010) Efficacy of Gene Therapy for X-Linked Severe Combined Immunodeficiency. *N Eng J Med* 363:335-364.

Händel E.M., Alwin S., and Cathomen T. (2009) Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. *Mol. Ther.* 17: 104–111.

Hanna J, Wernig M, Markoulaki S, Sun C.W., Meissner A, Cassady J.P., Beard C, Brambrink T, Wu L.C., Townes T.M., Jaenisch R. (2007) Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science*, 318(5858):1920-3.

Ho L., Crabtree G.R. (2010) Chromatin remodelling during development. *Nature*, 463:474-84.

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

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

Holmes R., Zuniga-Pflucker J..C (2009) The OP9-DL1 system: generation of T-lymphocytes from embryonic or hematopoietic stem cells in vitro. *Cold Spring Harb Protoc.*

Holthausena C., Wymana ., Kanaar R. and Thomas J. (2010) Mini-review Regulation of DNA strand exchange in homologous recombination DNA Repair 9: 1264–1272.

Howe S.J., Mansour M.R., Schwarzwaelder K., Bartholomae C., Hubank M., Kempinski H., Brugman M.H., Pike-Overzet K., Chatters S.J, de Ridder D., Gilmour K.C., Adams S., Thornhill S.I., Parsley K.L., Staal F.J., Gale R.E., Linch D.C., Bayford J., Brown L., Quaye M., Kinnon C., Ancliff P., Webb D.K., Schmidt M., von Kalle C., Gaspar H.B., Thrasher A.J. (2008) Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J Clin Invest* 118: 3143-3150.

Hsu P.D., Lander E.S., Zhang F. (2014) Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* 157: 1262-1278.

Hurt, J.A., S.A. Thibodeau, A.S. Hirsh, C.O. Pabo, and J.K. Joung. 2003. Highly specific zinc finger proteins obtained by directed

domain shuffling and cell based selection. *Proc Natl Acad Sci U S A*. 100:12271-12276.

Hyongbum K. and Kim J.S. (2014) A guide to genome engineering with programmable nucleases. *Nat.Rev.Gen.* 15: 321-334.

Joung J.K., Sander J.D. (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* 14: 49-55.

Kalman L., Lindegren M.L., Kobrynski L., Vogt R., Hannon H., Howard J.T., Buckley R. (2004) Mutations in genes required for T-cell development: IL7R, CD45, IL2RG, JAK3, RAG1, RAG2, ARTEMIS, and ADA and severe combined immunodeficiency: *HuGE review. Genet Med* 6: 16-26.

Kang, E.M. et al (2010) Retrovirus gene therapy for X-linked chronic granulomatous disease can achieve stable long-term correction of oxidase activity in peripheral blood neutrophils. *Blood* 115, 783–791.

Kang, H. J. et al. (2011) Retroviral gene therapy for X-linked chronic granulomatous disease: results from phase I/II trial. *Mol. Ther.* 19, 2092–2101.

Kay, M. A. (2011). State-of-the-art gene-based therapies: the road ahead. *Nat Rev Genet* 113(5): 1006-15.

Kelly A. Frazer<sup>1</sup>, Sarah S. Murray<sup>1</sup>, Nicholas J. Schork<sup>1</sup> & Eric J. Topol (2009) Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics* 10: 241-251.

Kim Y.G., Cha J., Chandrasegaran S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A*. 93(3):1156-60.

Klug A. (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu Rev Biochem.* 79:213-31

Knorr D.A., Ni Z., Hermanson D., Hexum M.K., Bendzick L., Cooper L. J., Lee D.A., Kaufman D.S. (2013) Clinical-scale derivation of natural killer cells from human pluripotent stem cells for cancer therapy. *Stem Cells Transl Med.* 4:274-83.

Kochenderfer J.N., Dudley M.E., Kassim S.H., Somerville R.P., Carpenter R.O., Stetler-Stevenson M., Yang J.C., Phan G.Q., Hughes M.S., Sherry R.M., Raffeld M., Feldman S., Lu L., Li Y.F., Ngo L.T., Goy A., Feldman T., Spaner DE., Wang M.L., Chen C.C., Kranick S.M., Nath A., Nathan D.A., Morton K.E., Toomey M.A., Rosenberg S.A. *J Clin Oncol.* 33(6):540-9.

Kohn D.B., Kantoff P.W., Eglitis M.A., McLachlin J.R., Moen R.C., Karson E., Zwiebel J.A., Nienhuis A., Karlsson S., O'Reilly R. (1987) Retroviral-mediated gene transfer into mammalian cells. *Blood Cells* 13(1-2):285-98.

Krägeloh-Mann, Groeschel S., Kehrer C., Opher K., Nägele T., Handgretinger R. and Müller I. (2013) Juvenile metachromatic leukodystrophy 10 years post-transplant compared with a non-transplanted cohort. *Bone Marrow Transplant.* 48, 369–375 (2013).

Lei F., Haque R., Weiler L., Vrana K.E., Song J. (2009) T lineage differentiation from induced pluripotent stem cells. *Cell Immunol.* 260: 1-5.

Li, L., J.M. Olvera, K.E. Yoder, R.S. Mitchell, S.L. Butler, M. Lieber, S.L. Martin, and F.D. Bushman. (2001) Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection. *EMBO J.* 20:3272-3281.

Lieber M.R., Ma Y., Pannicke U., Schwarz K. (2003) Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol.* 4(9):712-20.

Liu X., Zhao T. (2013) Clinical Therapy Using iPSCs: Hopes and Challenges. *Genomics Proteomic Bioinformatics* 11: 294-298.

Liu, R., Paxton W.A, Choe S., Ceradini D., Martin S.R., Horuk R., MacDonald M.E., Stuhlmann H., Koup R.A., Landau N.R. (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of

some multiply-exposed individuals to HIV-1 infection. *Cell* 86, 367–377.

Lombardo A., Cesana D., Genovese P., Di Stefano B., Provasi E., Colombo D.F., Neri M., Magnani Z., Cantore A., Lo Riso P., Damo M., Pello O.M., Holmes M.C., Gregory P.D., Gritti A., Broccoli V., Bonini C., Naldini L. (2011) Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat Methods* 8: 861-869.

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

Lombardo A. and Naldini L. (2014) Genome Editing: A Tool For Research and Therapy: Targeted genome editing hits the clinic. *Nature Medicine*, 20, 1101–1103.

Maeder, M.L., S. Thibodeau-Beganny, A. Osiaik, D.A. Wright, R.M. Anthony, M. Eichinger, T. Jiang, J.E. Foley, R.J. Winfrey, J.A. Townsend, E. Unger- Wallace, J.D. Sander, F. Muller-Lerch, F. Fu, J. Pearlberg, C. Gobel, J.P. Dassie, S.M. Pruett-Miller, M.H. Porteus, D.C. Sgroi, A.J. Iafrate, D. Dobbs, P.B. McCray, Jr., T. Cathomen, D.F. Voytas, and J.K. Joung. (2008) Rapid "opensource" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Molecular cell*. 31:294-301.

Maetzig T., Galla M., Baum C. and Schambach A. (2011) Gammaretroviral Vectors: Biology, Technology and Application. *Viruses* 3:677-713.

Maier P., von Kalle C., Laufs S. (2010) Retroviral vectors for gene therapy. *Future Microbiology* 10: 1507-1523.

Martin J. Evans (2001) The cultural mouse. *Nat. Med.* 7, 1081–1083.

Maruyama T., Dougan S.K., Truttmann M.C., Bilate A.M., Ingram J.R., Ploegh H.L. (2015) Increasing the efficiency of precise

genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol.* 33:538-42.

Maston G.A., Evans S.K., Green M.R.(2006) Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet*, 7:29-59.

Matrai J., Cantore A., Bartholomae C..C, Annoni A., Wang W., Acosta-Sanchez A., Samara-Kuko E., De Waele L., Ma L., Genovese P., Damo M., Arens A., Goudy K., Nichols T.C., von Kalle C., Roncarolo M..G, Schmidt M., Vandendriessche T., Naldini L. (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., M.B. Noyes, L.J. Zhu, N.D. Lawson, and S.A. Wolfe. 2008. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol.* 26:695-701.

Métais J.Y., Dunbar C.E. (2008) The MDS1-EVI1 gene complex as a retrovirus integration site: impact on behavior of hematopoietic cells and implications for gene therapy. *Mol Ther.* (3):439-49.

Miller, J.C., M.C. Holmes, J. Wang, D.Y. Guschin, Y.L. Lee, I. Rupniewski, C.M. Beausejour, A.J. Waite, N.S. Wang, K.A. Kim, P.D. Gregory, C.O. Pabo, and E.J. Rebar. 2007. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol.* 25:778-785.

Miller, J.C., S. Tan, G. Qiao, K.A. Barlow, J. Wang, D.F. Xia, X. Meng, D.E. Paschon, E. Leung, S.J. Hinkley, G.P. Dulay, K.L. Hua, I. Ankoudinova, G.J. Cost, F.D. Urnov, H.S. Zhang, M.C. Holmes, L. Zhang, P.D. Gregory, and E.J. Rebar. 2011. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol.* 29:143-148.

Miller J., McLachlan A.D., Klug A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* (6):1609-14.

Mimitou E.P; Symington L.S. (2009) Nucleases and helicases take center stage in homologous recombination. *Trends Biochem Sci.* (5):264-72.

Mingozzi F. and High K. A. (2011) Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. *Nature Rev. Genet.* 12, 341–355.

Miyoshi H., Smith K.A., Mosier D.E., Verma I.M., Torbett B.E. (1999) Transduction of human CD34+ cells that mediate long-term engraftment of NOD/SCID mice by HIV vectors. *Science*, 283(5402):682-6.

Modlich U., Navarro S., Zychlinski D., Maetzig T., Knoess S., Brugman M. H., Schambach A., Charrier S., Galy A., Thrasher A. J., Bueren J. and Baum C. (2009) Insertional Transformation of Hematopoietic Cells by Self-inactivating Lentiviral and Gammaretroviral Vectors. *Molecular Therapy*, 11, 1919–1928.

Moore, M., A. Klug, and Y. Choo. 2001. Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc Natl Acad Sci U S A.* 98:1437-1441.

Montini, E., D. Cesana, M. Schmidt, F. Sanvito, C.C. Bartholomae, M. Ranzani, F. Benedicenti, L.S. Sergi, A. Ambrosi, M. Ponzoni, C. Doglioni, C. Di Serio, C. von Kalle, and L. Naldini. 2009. The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. *J Clin Invest.* 119:964-975.

Montini, E., D. Cesana, M. Schmidt, F. Sanvito, M. Ponzoni, C. Bartholomae, L. Sergi Sergi, F. Benedicenti, A. Ambrosi, C. Di Serio, C. Doglioni, C. von Kalle, and Naldini L. (2006) Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. *Nat Biotechnol.* 24:687-696.

Moynahan M.E. and Jasin M. (2010) Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat. Rev. Mol. Cell Biol.* 11: 196–207.

Mussolino C. and Cathomen T. (2012) TALE nucleases: tailored genome engineering made easy. *Current opinion in biotechnology*. 23:644-650.

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

Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S (2008) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* 26: 101-106.

Naldini, L. (2011). Ex vivo gene transfer and correction for cell-based therapies. *Nature Rev. Genet.* 12, 301–315.

Naldini L. (2015) Gene therapy returns to centre stage. *Nat. Rev.* 526: 351-360.

Nayak, S. and Herzog, R. W. (2010) Progress and prospects: immune responses to viral vectors. *Gene Ther.* 17, 295–304.

Ni Z., Knorr D.A., Kaufman D.S. (2013) Hematopoietic and nature killer cell development from human pluripotent stem cells. *Methods Mol Biol.* 1029:33-41.

Nolta J.A., Kohn D.B. (1992) Comparison of the effects of growth factors on retroviral vector-mediated gene transfer and the proliferative status of human hematopoietic progenitor cells. *J Clin Invest.* 90(2): 342–348.

Notarangelo L.D., Fischer A., Geha R.S., Casanova J.L., Chapel H., Conley M.E., Cunningham-Rundles C., Etzioni A., Hammartröm L., Nonoyama S., Ochs H.D., Puck J., Roifman C., Seger R., Wedgwood J. (2009) Primary immunodeficiencies: 2009 update. *J. Allergy Clin. Immunol.* 124, 1161–1178.

Ott M.G., Schmidt M., Schwarzwaelder K., Stein S., Siler U., Koehl U., Glimm H., Kühlcke K., Schilz A., Kunkel H., Naundorf S., Brinkmann A., Deichmann A., Fischer M., Ball

C., Pilz I., Dunbar C., Du Y., Jenkins N.A., Copeland N.G., Lüthi U., Hassan M., Thrasher A.J., Hoelzer D., von Kalle C., Seger R., Grez M. (2006) Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat Med.* (4):401-9.

Pabo C.O., Peisach E., Grant R.A. (2001) Design and selection of novel Cys2His2 zinc finger proteins. *Annu Rev Biochem.* 70:313-40.

Pasi C.E., Dereli-Oz A., Negrini S., Friedli M., Fragola G., Lombardo A., Van Houwe G., Naldini L., Casola S., Testa G., Trono D., Pelicci P.G., Halazonetis T.D. (2011) Genomic instability in induced stem cells. *Cell Death Differ* 18: 745-753.

Paruzynski, A., A. Arens, R. Gabriel, C.C. Bartholomae, S. Scholz, W. Wang, S. Wolf, H. Glimm, M. Schmidt, and C. von Kalle. 2010. Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing. *Nature protocols.* 5:1379-1395.

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

Phillips, C.M., X. Meng, L. Zhang, J.H. Chretien, Urnov F.D. and A.F. Dernburg. (2009) Identification of chromosome sequence motifs that mediate meiotic pairing and synapsis in *C. elegans*. *Nature cell biology.* 11:934-942.

Pierce A.J., Johnson R.D., Thompson L.H., Jasin M. (1999) XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev*;13(20):2633-8.

Porteus M.H. and Carroll D. (2005) Gene targeting using zinc finger nucleases. *Nat Biotechnol.* (8):967-73

Porteus M. (2011) Translating the lessons from gene therapy to the development of regenerative medicine. *Mol. Ther.* 19: 439-441

Provasi E., Genovese P., Lombardo A., Magnani Z., Liu P.Q., Reik A., Chu V., Paschon D.E., Zhang L., Kuball J., Camisa B., Bondanza A., Casorati G., Ponzoni M., Ciceri F., Bordignon C., Greenberg P.D., Holmes M.C., Gregory P.D., Naldini L and Bonini C. (2012) Editing T cell specificity towards leukemia by zinc finger nucleases and lentiviral gene transfer. *Nat Med* 18: 807-815.

Rais Y., Zviran A., Geula S., Gafi O., Chomsky E., Viukov S., Mansour A.A., Caspi I., Krupalnik V., Zerbib M. (2013) Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* 502: 65-70.

Ramalingam S., Kandavelou K., Rajenderan R., and Chandrasegaran S. (2011) Creating designed zinc-finger nucleases with minimal cytotoxicity. *J Mol Biol.* 405:630- 641.

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

Reynolds A.E., Ryckman B.J., Baines J.D., Zhou Y., Liang L., Roller R.J.( 2001) U(L)31 and U(L)34 proteins of herpes simplex virus type 1 form a complex that accumulates at the nuclear rim and is required for envelopment of nucleocapsids., 75:8803-17.

Sadelain M, Papapetrou E.P. and Bushman F.D. (2012) Safe harbours for the integration of new DNA in the human genome. *Nature Reviews Cancer*, 12, 51-58.

Santiago, Y., E. Chan, P.Q. Liu, S. Orlando, L. Zhang, F.D. Urnov, M.C. Holmes, D.Guschin, A. Waite, J.C. Miller, E.J. Rebar, P.D. Gregory, A. Klug, and T.N. Collingwood. 2008. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci U S A.* 105:5809-5814.

Schmidt, M., K. Schwarzwaelder, C. Bartholomae, K. Zaoui, C. Ball, I. Pilz, S. Braun, H. Glimm, and C. von Kalle. (2007) High-

resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nature methods*. 4:1051-1057.

Scobie L, Hector R.D., Grant L., Bell M., Nielsen A.A., Meikle S., Philbey A., Thrasher A.J., Cameron E.R., Blyth K., Neil J.C. (2009) A novel model of SCID-X1 reconstitution reveals predisposition to retrovirus-induced lymphoma but no evidence of gamma C gene oncogenicity. *Mol Ther*. 17(6):1031-8.

Segura M. M., Alba R., Bosch A., (2008). Advances in helper-dependent adenoviral vector research. *Curr. Gene Ther*. 8, 222–235  
Shen B., Zhang W., Zhang J., Zhou J., Wang J., Chen L., Wang L., Hodgkins A., Iyer V., Huang X. and Skarnes W. (2014) Efficient genome modification by CrisPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* 11: 399-402.

Sluvkin I.I. (2013) Hematopoietic specification from human pluripotent stem cells: current advances and challenges toward de novo generation of hematopoietic stem cells. *Blood*, 122: 4035-4046.

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

Smithies, O., R.G. Gregg, S.S. Boggs, M.A. Koralewski, and Kucherlapati R.S. (1985) Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature*. 317:230-234.

Sollu, C., K. Pars, T.I. Cornu, S. Thibodeau-Beganny, M.L. Maeder, J.K. Joung, R. Heilbronn, and Cathomen T. (2010) Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic acids research*. 38:8269-8276.

Stein, S., Ott M.G., Schultze-Strasser S., Jauch A., Burwinkel B, Kinner A, Schmidt M, Krämer A, Schwäble J, Glimm H, Koehl U, Preiss C, Ball C, Martin H, Göhring G, Schwarzwaelder K, Hofmann W.K., Karakaya K, Tchatchou S, Yang R, Reinecke P, Kühlcke K, Schlegelberger B, Thrasher AJ, Hoelzer D, Seger R, von Kalle C, Grez M. (2010) Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation

after gene therapy for chronic granulomatous disease. *Nature Med.* 16, 198–204.

Sternecker J.L., Reinhardt P., Scholer H.R. (2014) Investigating human disease using stem cell models. *Nat Rev Genet* 15: 625-639.

Subramanyam D., Lamouille S., Judson R.L., Liu J.Y., Bucay N., Derynck R. and Blomch R. (2011) Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nat. Biotech.* 29: 443-448

Suda T. and Liu D. (2007) Hydrodynamic Gene Delivery: Its Principles and Applications. *Molecular Therapy* (2007) 15 12, 2063–2069.

Suzuki Y., Craigie R. (2007) The road to chromatin - nuclear entry of retroviruses. *Nat Rev Microbiol.* 3:187-96.

Szczepiek, M., V. Brondani, J. Buchel, L. Serrano, D.J. Segal, and Cathomen T.(2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol.* 25:786-793.

Szostak J.W., Orr-Weaver T.L., Rothstein R.J., Stahl F.W. (1983) The double-strand-break repair model for recombination. *Cell*, 33(1):25-35.

Takahashi K., Tanabe K., Ohnuki M., Narita M., Ichisaka T., Tomoda K., Yamanaka S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131(5):861-72.

Takahashi K., Yamanaka S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126(4):663-76

Taylor B.S., Sobieszczyk M.E., McCutchan F.E. and Hammer S.M. (2008). The Challenge of HIV-1 Subtype Diversity. *The new england Journal of medicine* 35: 1590-1602.

Taylor C.J., Peacock S., Chaudhry A.N., Bradley J.A., Bolton E.M. (2012) Generating an iPSC Bank for HLA-Matched Tissue

Transplantation Based on Known Donor and Recipient HLA Types. *Cell Stem Cell* 11: 147-152.

Themeli M., Kloss C.C., Ciriello G., Fedorov V.D., Perna F., Gonen M., Sadelain M. (2013) Generation of tumor-targeted human T lymphocytes from induced pluripotent stem cells for cancer therapy. *Nat Biotechnol.* 31:928-33.

Theunissen T.W., Powell B.E., Wang H., Mitalipova M., Faddah D.A., Reddy J., Fan Z.P., Maetzel D., Ganz K., Shi L., Lungjangwa T., Imsoonthornruksa S., Stelzer Y., Rangarajan S., D'Alessio A., Zhang J., Gao Q., Dawlaty M.M., Young R.A., Gray N.S., Jaenisch R. (2014) Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell.* 15: 471-87.

Thomas C.E., Ehrhardt A. and Kay M.A (2003) Progress and problems with the use of viral vectors for gene therapy. *Nat. Rev. Gen.* 4: 356-356.

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

Thomas, K. R., Folger, K. R. and Capecchi, M. R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. *Cell* 44:419–428.

Touzot F., Hacein-Bey-Abina S., Fischer A., Cavazzana M. (2014) Gene therapy for inherited immunodeficiency. *Expert Opin Biol Ther* 14: 789-798.

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

Tuerk, C., and Gold L.. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*: 505-510.

Turitz Cox D.B, Platt R.J., Zhang F (2015) Therapeutic genome editing: prospects and challenges. *Nat Med* 21: 121-131.

Uren A.G., Kool J., Berns A. and van Lohuizen M. (2005) Retroviral insertional mutagenesis: past, present and future. *Oncogene*, 24: 7656–7672.

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

van Lohuizen M., Verbeek S., Krimpenfort P., Domen J., Saris C., Radaszkiewicz T. and Berns A. (1989). Predisposition to lymphomagenesis in pim-1 transgenic mice: cooperation with c-myc and N-myc in murine leukemia virus-induced tumors. *Cell*, 56, 673–682.

Wang, X. *et al.* *Nat. Biotechnol.* 33, 187–197 (2015).

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

Woods N.B., Bottero V., Schmidt M., von Kalle C., Verma I.M. (2006) Gene therapy: therapeutic gene causing lymphoma. *Nature* 440: 1123.

Wu S.M. and Hochedlinger K (2011) Harnessing the potential of induced pluripotent stem cell for regenerative medicine. *Nature Cell Biol* 13: 497-505.

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

Yu J., Vodyanik M. A., Smuga-Otto K., Antosiewicz-Bourget J., L. Frane J., Tian S., Nie J., Jonsdottir G. A., Ruotti V., Stewart R., Slukvin I. I., Thomson J.A. (2007) Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science* 318: 1917-1920.

Zufferey R., Dull T., Mandel R.J., Bukovsky A., Quiroz D., Naldini L., Trono D. (1998) Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J Virol.* 72(12):9873-80.

Zychlinski D., Schambach A., Modlich U., Maetzig T., Meyer J., Grassman E., Mishra A., Baum C. (2008) Physiological promoters reduce the genotoxic risk of integrating gene vectors. *Mol Ther.* 4):718-25. 2008.

## **CHAPTER 2**

# **Targeted Correction and Reprogramming of SCID-X1 Fibroblasts Rescues IL2RG in iPSC-derived T-lymphoid Cells**

Claudia Firrito<sup>1,2</sup>, Pietro Genovese<sup>1</sup>, Giulia Schioli<sup>1,3</sup>, Mauro Biffi<sup>1</sup>, Tiziano Di Tomaso<sup>1</sup>, Oscar Quintana-Bustamante<sup>4</sup>, Andrew R. Gennery<sup>5</sup>, Philip D. Gregory<sup>6</sup>, José Carlos Segovia<sup>4</sup>, Juan A. Bueren<sup>4</sup>, Angelo Lombardo<sup>1,3,7</sup>, Luigi Naldini<sup>1,3,7</sup>

<sup>1</sup> San Raffaele Telethon Institute for Gene Therapy, San Raffaele Scientific Institute, Milan, Italy.

<sup>2</sup> Università degli Studi di Milano-Bicocca; Milan; 20126; Italy.

<sup>3</sup> Vita-Salute San Raffaele University, Milan, Italy.

<sup>4</sup> Division of Hematopoietic Innovative Therapies, CIEMAT/CIBERER, Madrid, Spain.

<sup>5</sup> Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, United Kingdom.

<sup>6</sup> Sangamo BioSciences Inc., Richmond, CA, United States.

<sup>7</sup> shared senior authorship.

Correspondence to: naldini.luigi@hsr.it

*Under consideration for publication in EMBO Molecular Medicine*

## Summary

Gene replacement by integrating vectors has been successfully used to treat X-linked Severe Combined Immunodeficiency (SCID-X1), an invariably lethal immunological disorder. These studies, however, highlighted the risks of insertional mutagenesis and unregulated transgene expression, calling for the development of safer gene therapy approaches based on targeted genome editing. Here, by optimizing donor design and Zinc Finger Nucleases activity we relieved several blocks limiting gene targeting and achieved efficient knock-in of a corrective *IL2RG* transgene downstream its endogenous promoter in two SCID-X1 fibroblasts cell lines. Because this cell type does not physiologically express *IL2RG*, we coupled correction with exogenous selection from the same genomic locus to enrich for gene-corrected fibroblasts, and reverted them to pluripotency using a Myc-less reprogramming lentiviral vector. Co-excision of this vector together with the selector cassette allowed obtaining several gene-corrected, reprogramming factor-free iPSCs with a normal karyotype. Finally, by showing selective growth advantage of T-lymphoid cells generated from the corrected iPSCs, we provided evidence of the efficacy of functional correction of the *IL2RG* mutant allele, paving the way to the development of novel therapeutic approaches for the treatment of SCID-X1.

## Introduction

X-linked Severe Combined Immunodeficiency (SCID-X1) is a rare immunological disease caused by mutations occurring throughout the Interleukin 2 Receptor common  $\gamma$ -chain (*IL2RG*) gene, with a mutational hot-spot in exon 5. Because  $\gamma$ -chain provides the signalling subunit of several receptors for essential lymphopoietic cytokines, its absence results in the lack of circulating T and NK cells and in profound B cell abnormalities (Kalman et al, 2004). SCID-X1 is invariably fatal unless treated by bone marrow transplantation. In the presence of an HLA-matched family donor, patients have high chances to be cured. For other types of donors, success rates are lower and full restoration of immunity may not be achieved. Pioneering clinical trials have demonstrated the long term therapeutic benefit of autologous Hematopoietic Stem and Progenitor Cells (HSPC)-based gene therapy using gamma-Retroviral Vectors ( $\gamma$  -RV) (Touzot et al, 2014). The success of these trials can be ascribed to the strong selective growth advantage conferred by gene correction to the T-cell lineage, while functional restoration of the NK and B cell compartments was less evident. Because of the limited stem cell transduction by  $\gamma$  -RV and the absence of patient preconditioning, most of the corrected T cells likely have originated from long-lived, transduced T-lymphoid progenitors rather than HSC. Along with the clear proof of therapeutic benefit, these studies also highlighted the potential risks associated with the use of semi-randomly integrating vectors, as shown by the occurrence of vector-driven leukemia in a fraction of

treated patients (Hacein-Bey-Abina et al, 2008; Hacein-Bey-Abina et al, 2003a; Hacein-Bey-Abina et al, 2003b; Howe et al, 2008). Ectopic  $\gamma$  -chain expression was also proposed as a possible cofactor for leukemia development (Woods et al, 2006), although this aspect remains to be clarified. Thus, there is a need for developing safer gene transfer approaches aiming at correcting the gene defect while sparing to the patients the risk of random insertion and unregulated transgene expression.

Whereas  $\gamma$  -RV and Lentiviral Vectors (LV), both carrying self-inactivating Long Terminal Repeats (LTRs) and a moderately active internal promoter to drive IL2RG expression, have now entered clinical testing and may provide a safer gene replacement strategy for this disease, new technologies with the potential to further improve safety and efficacy of gene transfer are emerging as promising alternatives to the use of semi-randomly integrating vectors. To this regard, artificial endonucleases, such as Zinc Finger Nucleases (ZFNs) (Urnov et al, 2010), Transcription Activator-Like Effector Nucleases (Joung & Sander, 2013) and, more recently, RNA-guided Nucleases (Hsu et al, 2014), have been used to induce a DNA double-strand break into a pre-selected site of the genome, and locally activate the Homology-Driven Repair (HDR) pathway (Ciccia & Elledge, 2010). This pathway can be exploited to custom edit the genomic site of interest by delivering into the cells a donor template DNA containing the desired sequence flanked by homology arms to the nucleases target site. By this approach it is possible to replace a specific mutant nucleotide with its wild-type version or insert a corrective template in the

locus in order to treat a broad spectrum of mutations without the need of tailoring nucleases to each individual cluster of mutations (Lombardo et al, 2007a). Importantly, both strategies have the potential to restore the function and physiological expression of the affected gene.

SCID-X1 represents a paradigmatic disease in which to assess the feasibility of targeted gene correction in HSPCs, as phenotypic rescue of the disease can be achieved even with few corrected lymphoid progenitors. To this regard, we have recently developed a protocol that allows targeted correction of the *IL2RG* gene in long-term repopulating human HSPCs (Genovese et al, 2014). While this achievement may pave the way for the development of safer HSC-based gene therapy approaches, its clinical application still faces the relatively low efficiency by which targeted gene editing occurs in the more primitive HSC compartment, thus possibly relying on *ex vivo* or *in vivo* expansion of the corrected cells in order to achieve a therapeutic benefit in the patients. Conversely, the discovery that somatic cells can be reverted to an Embryonic Stem Cell (ESC)-like state by forced expression of pluripotency related transcription factors has opened new perspectives in regenerative medicine (Takahashi et al, 2007). The availability of autologous induced Pluripotent Stem Cells (iPSC) amenable to unlimited *ex vivo* expansion, selection and targeted genetic correction may provide a source of immunologically and ethically compliant cells that can be differentiated into therapeutically relevant cell types, including T and NK cell progenitors (Sterneckert et al, 2014)

## Results

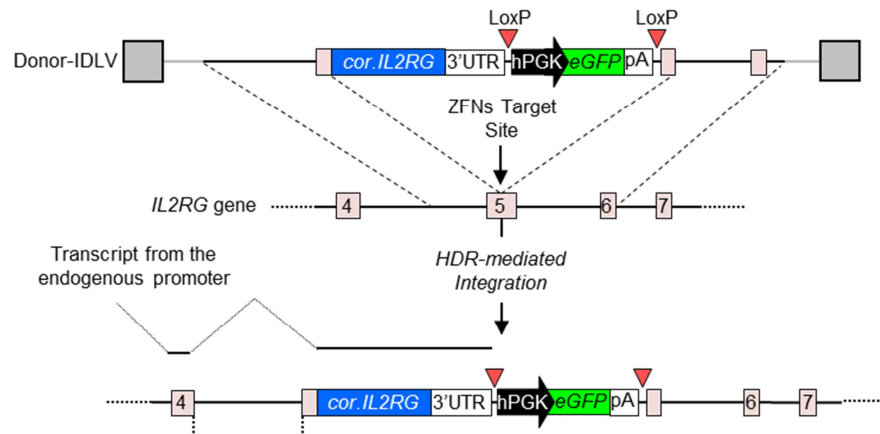
### **The efficiency of targeted integration decreases with increasing ZFN exposure when aiming to correct the *IL2RG* gene.**

We previously showed that ZFNs and Integrase-Defective Lentiviral Vector (IDLV) delivery can be used to knock-in a corrective *IL2RG* cDNA downstream of its endogenous promoter in a human cell line and in lymphoid cells derived *in vivo* from CD34<sup>+</sup> cells, and showed that transcription of the integrated construct was driven by the endogenous *IL2RG* promoter (Genovese et al, 2014; Lombardo et al, 2007a). Here, we extended these studies and identified critical features of construct design that constrain proficiency of targeted gene correction. We first modelled targeted gene-correction in male B-lymphoblastoid cells from a healthy donor. This line was selected because it constitutively expresses the *IL2RG* gene from its single copy X-chromosome, and is not counter-selected in culture when this gene is hypo-functional or inactive. To identify the B-lymphoblastoid cells that carry insertion of the corrective cDNA we inserted downstream of the corrective *IL2RG* cDNA of the donor-IDLV an eGFP-expression cassette (cor.*IL2RG*-IDLV; **Fig. 1A**). We also redesigned the DNA binding domains of the *IL2RG*-ZFNs to increase avidity while maintaining unaltered their genomic recognition site in exon 5 of the *IL2RG* gene. Indeed, when compared to the previously reported *IL2RG*-ZFNs (Set\_07), the newly designed ZFNs set (Set\_11) outperformed the previous one by more than one log in terms of gene targeting efficiency in K-562

cells (**Fig. 1B**). We then transduced male B-lymphoblastoid cells with the cor.IL2RG-IDLV with or without increasing doses of IDLVs encoding for the new set of *IL2RG*-ZFNs (ZFNs-IDLV), and assessed efficiency and specificity of integration. As a positive control, we included in these experiments a donor-IDLV containing only the eGFP-expression cassette (eGFP-IDLV) and found that the percentage of eGFP<sup>+</sup> cells increased proportionally to the dose of the ZFNs-IDLV used (**Fig. 1C**). Conversely, in the cor.IL2RG-IDLV treated cells, we measured a significant drop in the frequency of eGFP<sup>+</sup> cells when increasing the dose of ZFNs-IDLV. We then analyzed by flow cytometry the relative distribution of the eGFP<sup>+</sup> cells between the  $\gamma$ -chain positive and negative cells. As expected from insertion of the sole eGFP cassette (without the corrective cDNA) into the coding sequence of the *IL2RG* gene, 91.7% of the eGFP<sup>+</sup> cells lacked  $\gamma$ -chain expression (**Fig. 1D**; upper right plot), and this phenotype was independent on the dose of the ZFNs-IDLV used. These results were confirmed by Southern blot analysis, which showed that  $\geq 90\%$  of the *IL2RG* alleles contained a single copy of the eGFP-cassette (**Fig. 1E**). On the other hand, editing of the *IL2RG* gene with a functional corrective cDNA (followed by the eGFP expression cassette) should give rise to a population of cells double positive for  $\gamma$ -chain and eGFP. Accordingly, in the samples treated with the lowest dose of ZFNs-IDLVs, 87.5% of the eGFP<sup>+</sup> cells also expressed  $\gamma$ -chain (**Fig. 1D**; bottom left plot), and 86% of the *IL2RG* alleles of the eGFP<sup>+</sup> cells contained integration of the corrective cDNA, either as single cassette or as a concatemer (**Fig. 1E**). Unexpectedly,

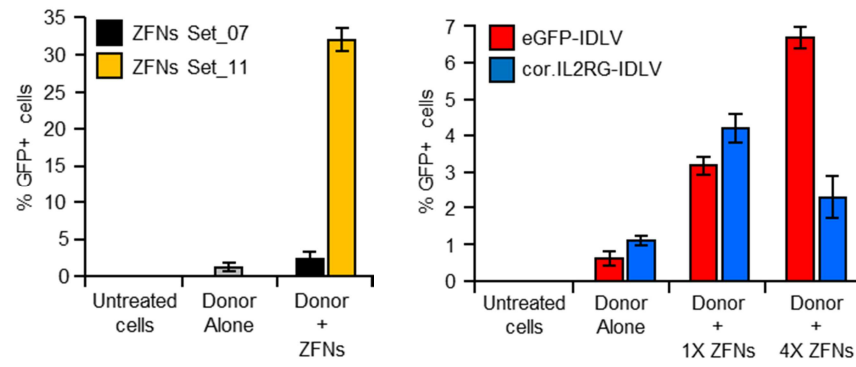
however, when we analysed the samples treated with the highest dose of the ZFNs-IDLVs, we found that only 37% of the eGFP+ cells were  $\gamma$ -chain positive (**Fig. 1D**; bottom right plot). Strikingly, little -if any- detectable integration into the *IL2RG* gene of these cells was found by Southern blot analysis (**Fig. 1E**). These data indicate that, contrary to findings obtained with the eGFP IDLV donor, an increase in the *IL2RG*-ZFNs dose was associated to a decrease in the efficiency of targeted integration of the corrective *IL2RG* cDNA.

A

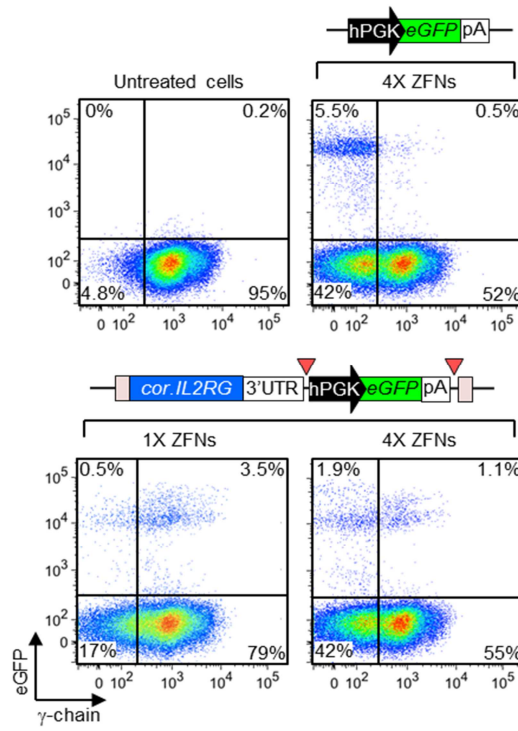


B

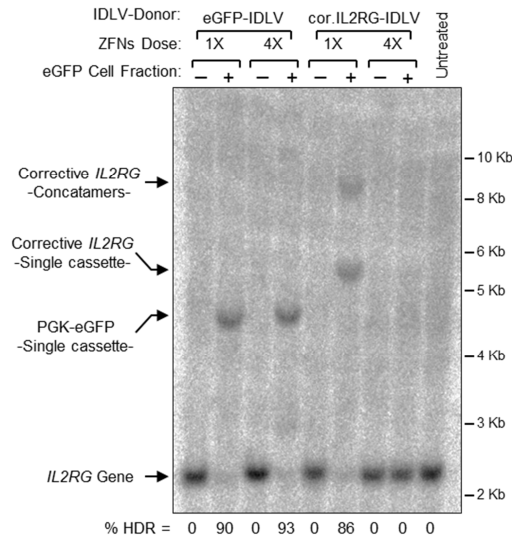
C



D



E



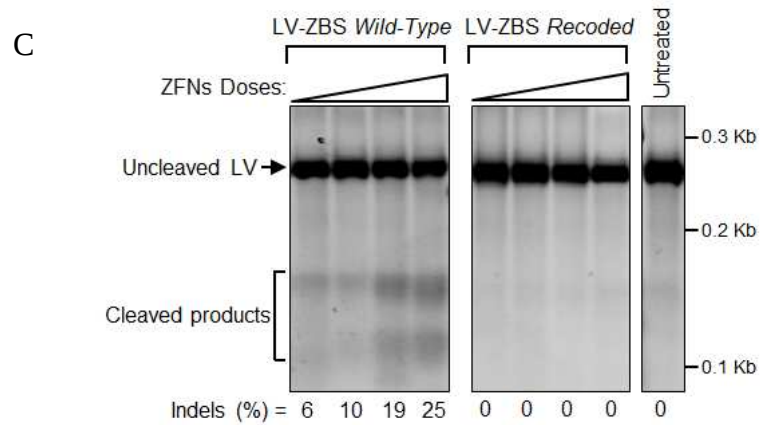
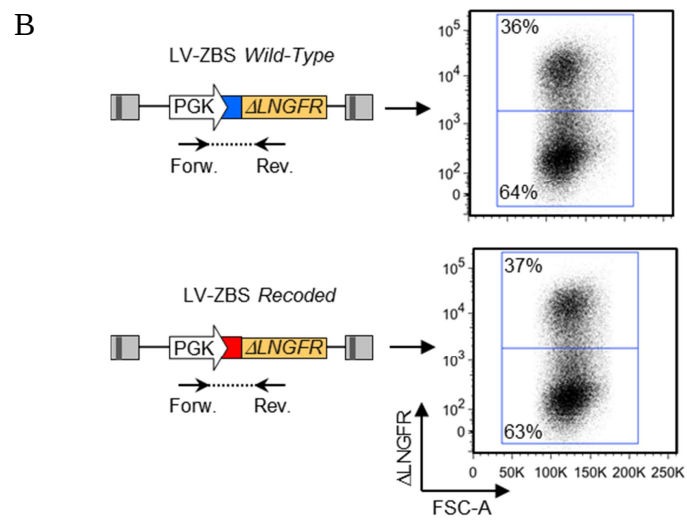
**Figure 1. Targeted integration into *IL2RG* in B-lymphoblastoid cells.** **A**, Schematic representation of targeted insertion of the corrective *IL2RG* cDNA plus the LoxP-flanked eGFP-cassette into exon 5 of the *IL2RG* gene. Upon HDR-mediated repair of the *IL2RG* gene with the donor-IDLV, expression of the corrective *IL2RG* cDNA is regulated by the endogenous promoter. IDLV: Integrase Defective Lentiviral Vector; cor.*IL2RG*: the promoter-less *IL2RG* cDNA, comprising exon 5 to 8 of the *IL2RG* gene; 3' UTR: 3' untranslated region of the *IL2RG* gene; PGK-GFP-pA: the eGFP-expression cassette flanked by LoxP sites; SA: splice acceptor; SD: splice donor. **B**, Histogram showing the percentage of GFP+ cells upon treatment with the old (Set\_07) or the new (Set\_11) *IL2RG*-ZFNs, with or without a PGK-GFP Donor IDLV. Data are represented as Mean  $\pm$  s.e.m. of 3 independent

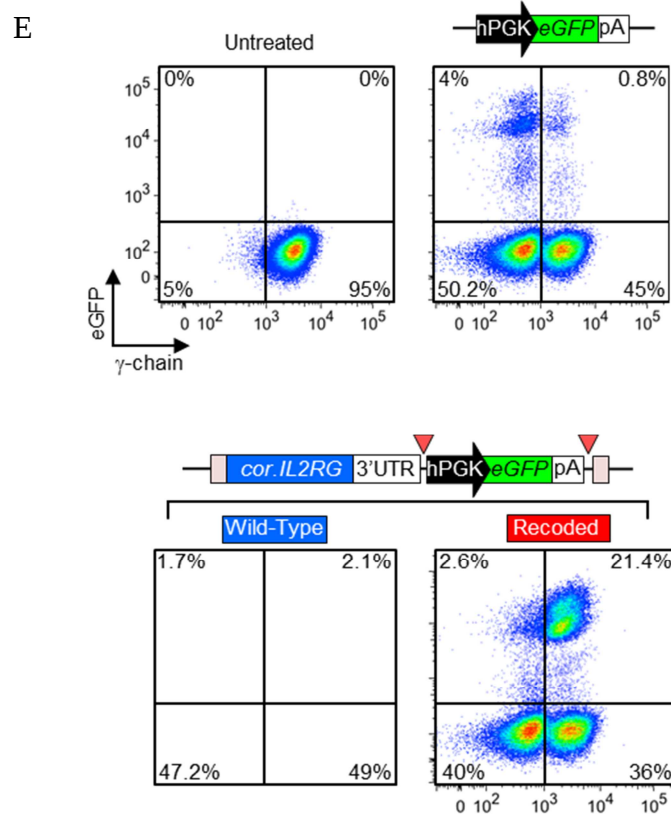
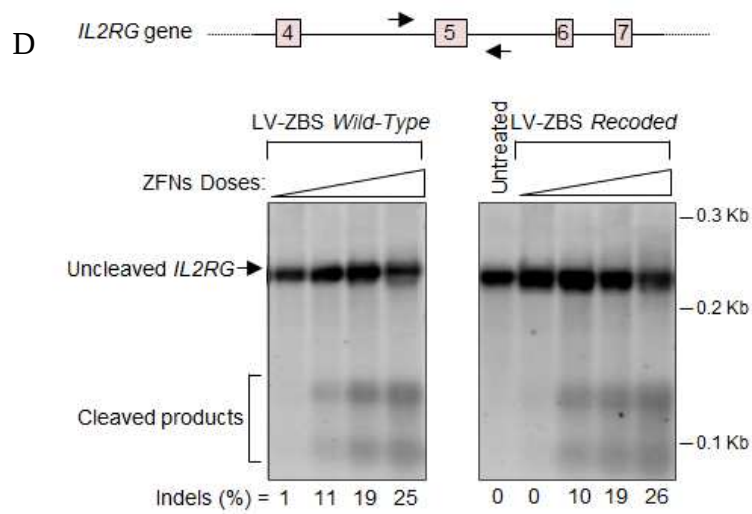
experiments. **C**, Percentage of GFP<sup>+</sup> cells upon treatment with either the eGFP-IDLV or the cor.IL2RG-IDLV donor and exposed to low (1X ZFNs) or high (4X ZFNs) doses of the *IL2RG*-ZFNs. **D**, Representative flow cytometry analysis of B-lymphoblastoid cells from (C) treated with the indicated donor-IDLVs (schematics on the top of the plots). The percentage of positive cells for each quadrant is indicated. **E**, Cells from (C) were sorted according to eGFP expression and then analyzed by Southern blot using a probe that recognizes the *IL2RG* locus outside of the homology arms of the donor IDLV. On the left of the blot are indicated the different configurations of the *IL2RG* locus. At the bottom of the blot are reported the percentages of HDR for each condition, as measured by densitometric analysis.

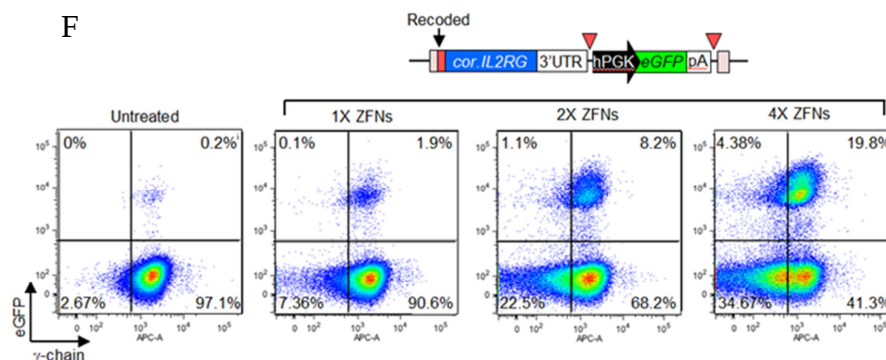
### **Improved donor design rescues targeted correction of the *IL2RG* gene.**

This inconsistency might be explained by the undesired activity of the *IL2RG*-ZFNs on the sequence of the corrective cDNA, which contains a fully matching *IL2RG*-ZFNs target site. To investigate this hypothesis, we introduced silent mutations into the *IL2RG*-ZFNs target site, in order to abrogate ZFNs binding while maintaining the coding frame of the *IL2RG* cDNA (**Fig. 2A**). To assess resistance of the recoded sequence to ZFNs activity, we first introduced the wild-type *IL2RG*-ZFNs target site or its recoded counterpart in K-562 cells by integrase competent Lentiviral Vector (LV) transduction (**Fig. 2B**). By this approach, we generated two cell populations in which the two ZFNs target sites were semi-

randomly distributed throughout the genome. We then exposed these cells to a wide range of ZFNs doses and measured the rate of insertions/deletions (indels) introduced by the error-prone Non-Homologous End Joining (NHEJ) repair pathway (Ciccia & Elledge, 2010) into the LV-bearing *IL2RG* ZFNs target sites. By using this assay we found that recoding of the right ZFN target site was sufficient to fully abrogate ZFNs activity at all ZFNs doses tested (**Fig. 2C**). This effect was not due to a different delivery rate of the *IL2RG*-ZFNs, as the extent of indels measured at the endogenous *IL2RG* gene was superimposable between these two cell populations (**Fig. 2D**). Based on this data, we then recoded the right ZFN target site present in the corrective *IL2RG* cDNA of the donor-IDLV (rec.*IL2RG*-IDLV), and compared the gene targeting efficiency between the cor.*IL2RG*-IDLV and the rec.*IL2RG*-IDLV in male B-lymphoblastoid cells treated with a high dose of the improved *IL2RG*-ZFNs. Remarkably, the rec.*IL2RG*-IDLV consistently outperformed (more than 6-fold) its parental counterpart in terms of eGFP<sup>+</sup> cells at all ZFNs doses tested (**Fig. 2E**), including higher doses than those previously tested (reaching up to 64% *IL2RG* editing, calculated as the sum of the  $\gamma$ -chain negative plus the  $\gamma$ -chain positive/eGFP<sup>+</sup> cells). By using the recoded construct, the percentage of eGFP<sup>+</sup> cells was proportional to the dose of ZFNs used, and nearly every eGFP<sup>+</sup> cell was also  $\gamma$ -chain positive (**Fig. 2F**). In summary, by recoding the ZFNs binding site of the donor vector we were able to achieve efficient targeted insertion of the corrective *IL2RG* cDNA while maintaining expression of the edited gene.







**Figure 2. Optimized design of the IDLV-donor for efficient targeted integration into *IL2RG*.**

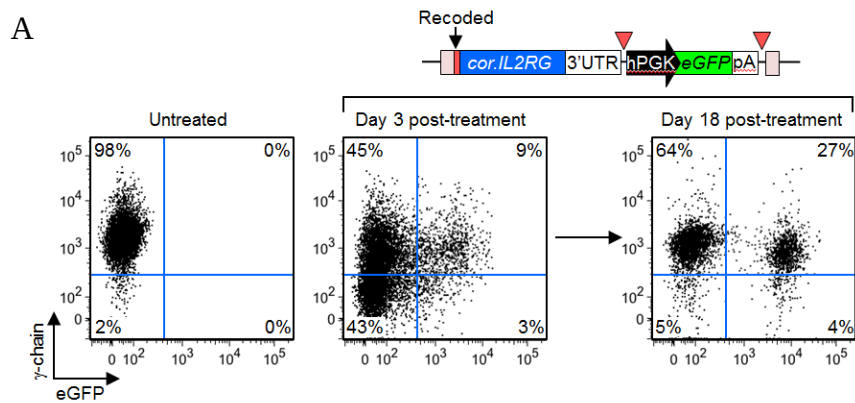
**A**, Sequence of the *IL2RG*-ZFNs binding site before and after its partial recoding. **B**, Representative flow cytometry analysis of K-562 cells transduced with the lentiviral vectors expressing the  $\Delta$ LNGFR marker and containing the wild-type (top) or the recoded (bottom) ZFNs-Binding Site (ZBS). Arrows below each vector indicate the location of the PCR primers used to analyze the cells for activity of the *IL2RG*-ZFNs as described in (C). FSC-A: Forward Scatter-Area.  $\Delta$ LNGFR: truncated Nerve Growth Factor Receptor. **C**, Mismatched selective endonuclease assay performed on genomic DNA from cells in (B) upon their exposure to increasing doses of the *IL2RG*-ZFNs. The indels rate for each sample is indicated below the gels. **D**, Top: schematic of the *IL2RG* locus depicting the PCR primers (arrows) used to assess the frequency of indels by the mismatched selective endonuclease assay. Bottom: indels analysis of the *IL2RG* locus on cells from Figure 2C. The indels rate for each sample is indicated below the

gels **E**, Representative flow cytometry analysis of B-lymphoblastoid cells treated with the IL2RG-ZFNs and the indicated donor-IDLVs. **F**, Representative flow cytometry dot plots of B-lymphoblastoid cells treated with rec.IL2RG-IDLV donor and increasing doses of IL2RG-ZFNs.

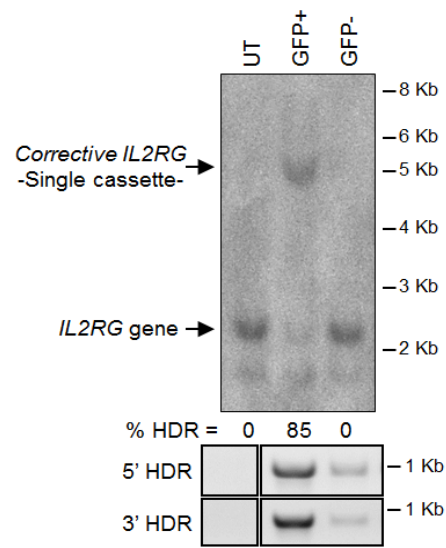
### **Functional expression of $\gamma$ -chain in primary T-lymphocytes upon insertion of the corrective cDNA.**

We then assessed functionality of the corrective cDNA by targeting its integration in human primary T-lymphocytes, a cell type that depends on physiological expression of the IL2RG gene for survival and proliferation (Kalman et al, 2004). Because of this dependency, T-lymphocytes are best suited to assess if correction of the *IL2RG* gene would impart a selective growth advantage to these cells over those carrying inactivating SCID-X1 mutations. Thus, according to a previously optimized protocol for gene-targeting in hematopoietic stem cells (Genovese et al, 2014), we used mRNA electroporation to transiently express the new set of *IL2RG*-ZFNs into T-cells from healthy male donors after their transduction with the rec.IL2RG-IDLV. At short-term post-treatment, nearly half of the treated T-lymphocytes lost surface expression of  $\gamma$ -chain (**Fig. 3A**; middle plot), likely because of disruption of the open reading frame of the gene by the error-prone NHEJ repair pathway. While these cells rapidly disappeared from the culture due to a selective growth disadvantage, at later time-point up to 27% of the treated T-lymphocytes were double positive

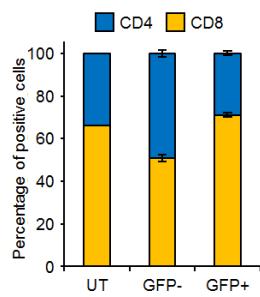
for  $\gamma$ -chain and eGFP (**Fig. 3A**; right plot). Southern blot and PCR analyses performed on the sorted eGFP<sup>+</sup> T-cells confirmed targeted integration in up to 85% of the *IL2RG* gene and the expected integration junctions between the corrective cassette and the *IL2RG* gene (**Fig. 3B**). When analysed by flow cytometry for expression of the CD4 and CD8 markers, the double positive T-lymphocytes were phenotypically indistinguishable from their eGFP-negative counterparts and from untreated controls (**Fig. 3C**). Within these subpopulations, the eGFP-positive and -negative cells displayed a similar distribution of T-cell subsets (**Fig. 3D**). Furthermore, eGFP-positive, -negative and untreated controls expressed comparable levels of proinflammatory cytokines, IL-2 and the activation marker CD107a upon stimulation with PMA and ionomycin (**Fig. 3E**). Overall, these data show that editing of the *IL2RG* gene with the corrective cDNA supports  $\gamma$ -chain expression and its normal function in primary T-lymphocytes.



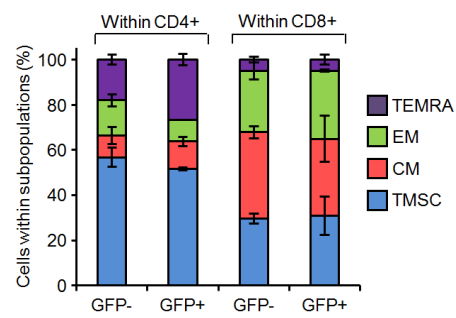
B



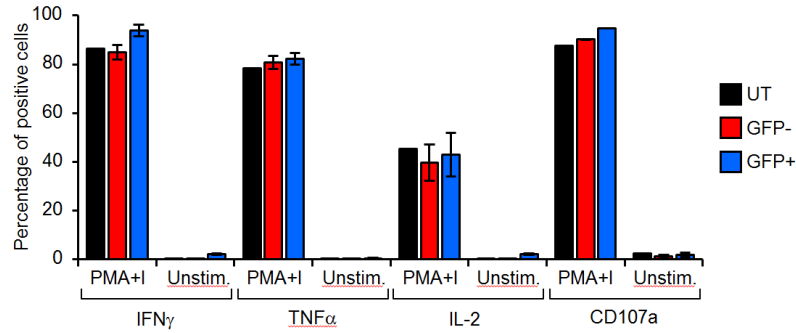
C



D



E



**Figure 3. Targeted integration into IL2RG of primary T-lymphocytes.**

**A**, Representative flow cytometry dot plots of T-lymphocytes from a male healthy-donor treated or not with the IL2RG-ZFNs and the donor-IDLV in which the right ZFN binding site of the corrective cDNA was recoded. These analyses were performed 3 or 18 days post-treatment. **B**, Southern blot analysis of eGFP-positive and -negative T-cells from (A) performed and analyzed as in Figure 1E. **C**, **D**, Distribution of CD4<sup>+</sup> and CD8<sup>+</sup> T-cells (**C**) and, within these populations, of the indicated functional T-cell subsets (**D**) at 20 days after treatment as measured by multiparametric flow cytometry analysis. Data are represented as Mean  $\pm$  s.e.m. of 3 independent experiments. TEMRA: Terminally Effector Memory T-cells, defined as CD45RA<sup>+</sup> CD62L<sup>-</sup>; EM: Effector Memory T-cells, defined as CD45RA<sup>-</sup> CD62L<sup>-</sup>; CM: Central Memory T-cells, defined as CD45RA<sup>-</sup> CD62L<sup>+</sup>; TMSC: T Memory Stem Cells, defined as CD45RA<sup>+</sup> CD62L<sup>+</sup>. **E**, Histogram showing the

percentage of T-cells expressing the indicated pro-inflammatory cytokines or the activation marker CD107a upon stimulation or not with PMA plus Ionomycin (PMA+I). Data are represented as Mean  $\pm$  s.e.m. of 3 independent experiments.

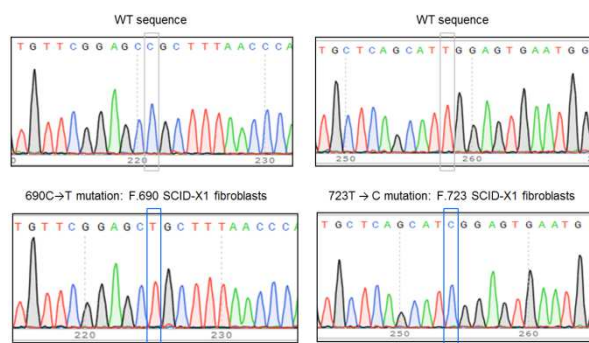
### **Generation of disease- and vector-free iPSC from SCID-X1 fibroblasts.**

Having optimized the design of the donor vector and shown its functionality in a relevant cell type, we then exploited our strategy to correct mutations in the exon 5 of *IL2RG* of primary fibroblasts from SCID-X1 patients. One patient had a 690C→T mutation (namely F.690 cells) and the other had a 723T→C mutation (namely F.723 cells; **Fig. 4A**). To enrich for gene-corrected fibroblasts, which do not express *IL2RG* and thus cannot be selected for its correction, we included downstream of the corrective cDNA an excisable eGFP- or a Puromycin Resistance (PuroR)-expression cassette (**Fig. 4B**). We delivered these donor DNAs by IDLV transduction and, 12hrs later, we transfected the fibroblasts with *in vitro* transcribed mRNAs encoding for the optimized *IL2RG*-ZFNs. This treatment yielded up to 5.9 or 4% eGFP-positive F.690 or F.723 cells, respectively (3.7 $\pm$ 0.3% for F.690 or 3.5 $\pm$ 0.1% for F.723; Mean $\pm$ sem; n=5 or 2 independent experiments, respectively; **Fig. 4C** left and **Fig. 4C** right. Similar results were obtained by targeting an eGFP-expression cassette within the *AAVS1* locus of these cells using IDLV with homology regions to the *AAVS1* locus. Notably, these levels were

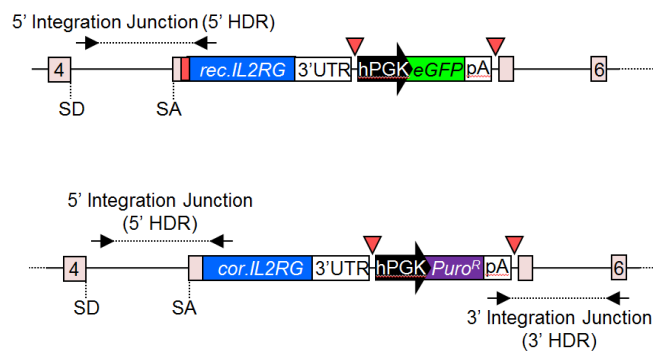
significantly higher than those observed when the *IL2RG* or the *AAVS1* donor IDLVs were either delivered alone or co-delivered with unrelated ZFNs. These data would indicate that, in cells co-exposed to the ZFNs and their cognate donor IDLVs, insertion of the eGFP cassettes likely occurred through HDR. We then developed a novel Cre-excisable LV (Reprogramming Recoded EXcisable LV: R<sup>2</sup>EX LV; **Fig. 4D**) to reprogram the PuroR- or the eGFP-sorted F.690 cells, and obtained 8 or 23 ES-like clones, respectively, containing the expected integration junctions of the corrective cassette into *IL2RG* (**Fig. 4E**). Sanger sequencing performed on the genomic DNA of a subset of these clones confirmed correction of the mutant *IL2RG* allele (**Suppl. Fig. 2A**). Further analyses performed on randomly selected clones showed that these cells homogeneously expressed *bona fide* pluripotency markers (**Fig. 4F**), were stably maintained in culture for long periods of time (up to 2-years) with an ES-like morphology, and contained on average 1.2 (for PuroR) or 2 (for eGFP+) copies per diploid genome of the R<sup>2</sup>EX LV. Interestingly, however, the iPSC clones generated from reprogramming the eGFP-positive fibroblasts did not express eGFP (**Suppl. Fig. 2B**). Yet, these cells displayed a normal karyotype (**Suppl. Fig. 2C**), had reactivated expression of the endogenous pluripotency genes (**Fig. 4G-H**), and were competent for *in vivo* differentiation into tissues of the three germ layers (**Fig. 4I**). Similarly, we also generated iPSCs from the parental non-corrected F.690 cells and from healthy donors-derived fibroblasts. Finally, we exploited the IDLV technology to transiently express the Cre recombinase and obtained several

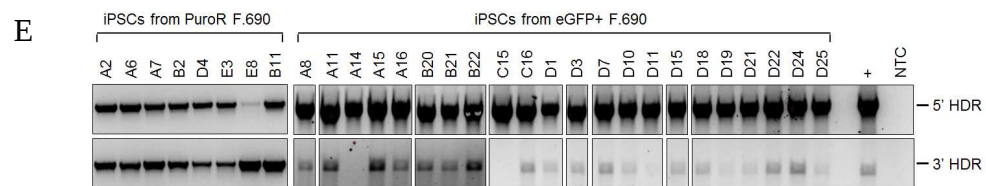
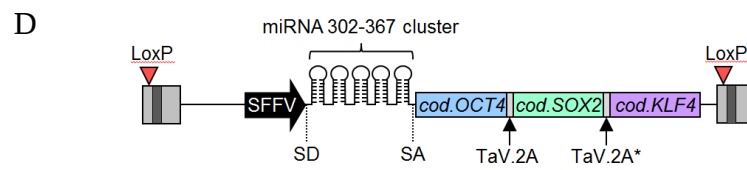
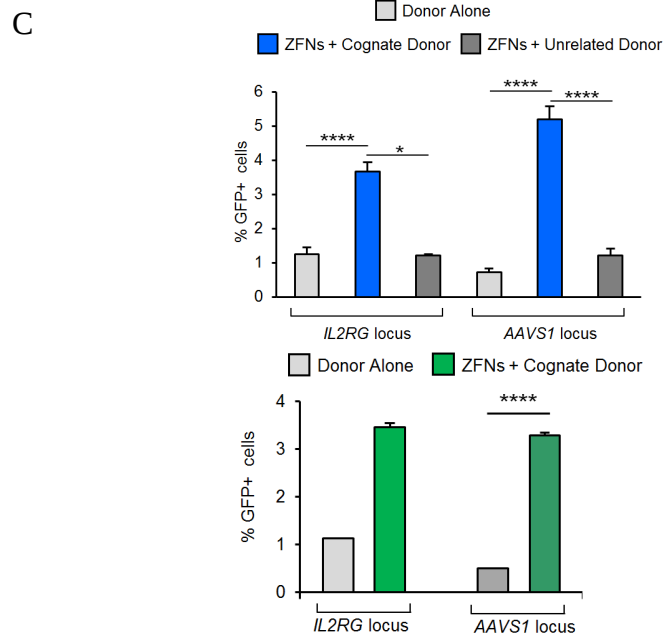
reprogramming factors-free, gene-corrected F.690 iPSCs still displaying a normal karyotype (**Suppl. Fig. 2 and Supp. Fig. 3**). Concomitantly with the excision of the R<sup>2</sup>EX LV, also the loxP-flanked reporter cassette downstream of the corrective cDNA was excised (**Suppl. Fig. 3**).

A

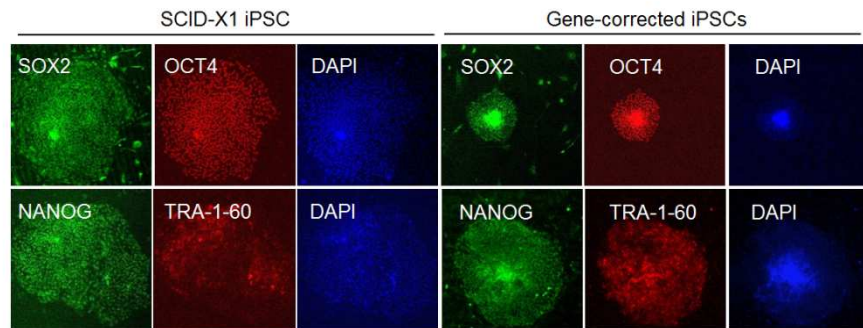


B

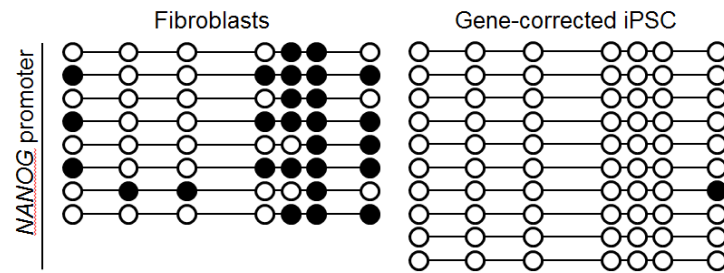




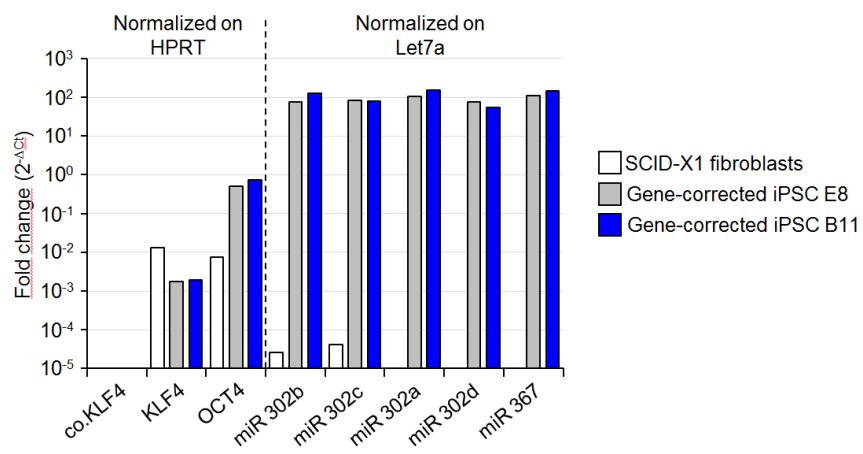
F



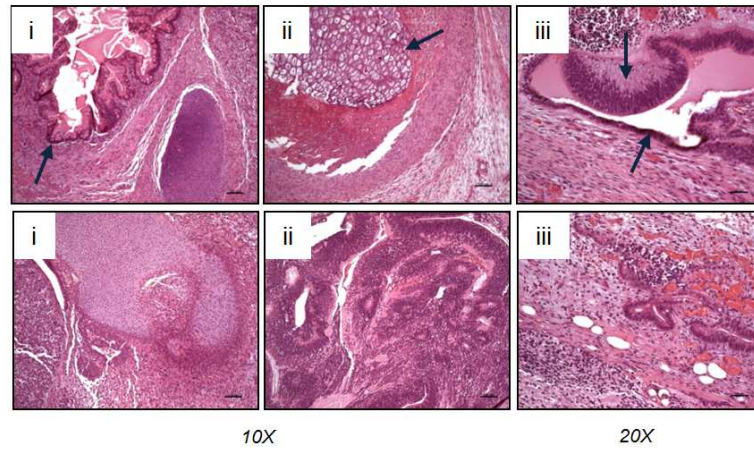
G



H



I



**Figure 4. Gene-correction and reprogramming of SCID-X1 fibroblasts.**

**A**, Sanger sequencing of the region encompassing the *IL2RG* exon 5 of wild-type, F.690 or F.723 fibroblasts showed the presence of the expected mutations in the latter two samples. **B**, Schematic of the *IL2RG* locus upon targeted integration of the corrective cDNA and its accompanying PuroR or eGFP--expression cassette. Arrows indicate the location of the PCR primers used to analyze the iPSC clones for targeted integration. **C**, left, Percentage of eGFP-positive F.690 cells measured by flow cytometry at 14 days from the indicated treatments. Data are represented as Mean+s.e.m. of 5 independent experiments (for the *IL2RG* locus: Donor Alone, ZFNs+cognate Donor, ZFNs+unrelated Donor, n=17, 36, 3 replicates, respectively; for the *AAVS1* locus: Donor Alone, ZFNs+cognate Donor, ZFNs+unrelated Donor, n=6, 6, 3 replicates, respectively). **C**, right, Percentage of eGFP-positive F.723 cells measured by flow cytometry at 15 days from the indicated treatments. Data are represented as Mean  $\pm$  s.e.m. (for the *IL2RG*

locus: Donor Alone, ZFNs+cognate Donor, n=2 or 3 technical replicates, respectively; for the *AAVS1* locus: Donor Alone, ZFNs+cognate Donor, n=3 or 3 technical replicates, respectively).

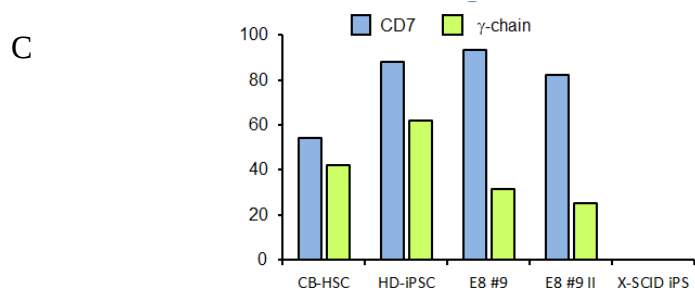
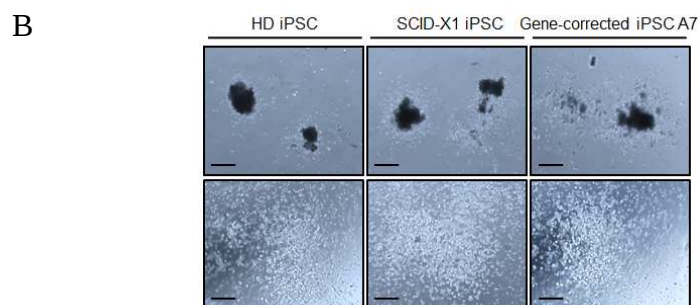
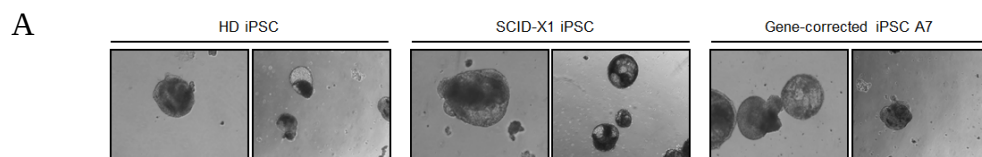
**D**, Schematic of the Reprogramming Recoded EXcisable LV ( $R^2EX$  LV), which co-expresses the miRNA 302-367 cluster (Anokye-Danso et al, 2011) together with codon optimized versions (*cod.*) of the transcription factors OCT4, SOX2 and KLF4. SFFV: Spleen Focus Forming Virus promoter; TaV.2A: wild-type or codon optimized (\*) sequence of the *Thosea asigna* Virus 2A self-cleaving peptide (de Felipe et al, 2006) to functionally link the reprogramming genes. **E**, Analysis of the indicated iPSC clones from the PuroR or the eGFP-positive F.690 cells using the PCR strategy depicted in (B). +: positive control; NTC: No Template Control. **F**, Immunofluorescence analysis of the indicated iPSC clones for expression of the pluripotency markers TRA1-60, NANOG, OCT4 and SOX2. DAPI: nuclear staining. **G**, Methylation analysis of the *NANOG* promoter of SCID-X1 fibroblasts and a gene-corrected iPSC clone. Methylation and non-methylation represented by filled and open circles, respectively **H**, Gene expression profile of pluripotency-associated genes and of the LV-encoded *co.KLF4* in SCID-X1 fibroblasts and gene-corrected iPSCs clones. Data are represented as fold change relative to HPRT (for *co.KLF4*, *KLF4* and *OCT4*) or to *Let-7a* (for *miR302b*, *miR302c*, *miR302a*, *miR302d*, *miR367*). **I**, Representative hematoxylin and eosin staining of teratomas from two gene-corrected iPSC clones. Top: arrows indicate (i) smooth muscle cells, cartilage and glandular epithelium from mesodermal and

endodermal differentiation; (ii) cartilage and immature bone from mesodermal differentiation; (iii) retinal cells from ectodermal differentiation; Bottom: arrow indicate the presence of (i) cartilage, smooth muscle cells and epithelial cells from mesodermal differentiation; (ii) immature neuroglial cells and rosette from neuroectodermal differentiation; (iii) smooth muscle cells, respiratory epithelium and immature neuroglial cells from mesodermal, ectodermal and endodermal differentiation.

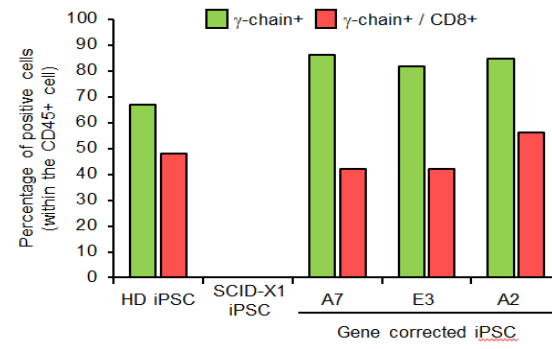
### **IL2RG correction rescues $\gamma$ -chain expression in hematopoietic cells derived from SCID-X1 iPSC.**

To assess if genetic correction of *IL2RG* leads to functional rescue of  $\gamma$ -chain expression in iPSC-derived T-cell progenitors, we exploited a previously described hematopoietic differentiation protocol based on co-culture of embryoid bodies (EB)-derived hematopoietic progenitors with OP9-DL1 stromal cells (Holmes & Zuniga-Pflucker, 2009). EB-derived hematopoietic progenitors from healthy-donor, SCID-X1 and gene-corrected iPSCs gave rise to myeloid and erythroid colonies in Colony Forming Cell (CFC) assay (**Fig. 5A** and **Fig. 5B**). Cells dissociated from the EBs were also cultured onto OP9-DL1 feeder cells in the presence of T-lymphoid promoting cytokines for 30 days, when the cells were analysed by flow cytometry for expression of T-cell markers and  $\gamma$ -chain. These analyses showed that gene-corrected and HD-derived iPSCs were capable of differentiation in hematopoietic CD45<sup>+</sup> cells at comparable frequencies and, importantly, expressed similarly  $\gamma$ -chain on the cell surface (**Fig. 5C, 5D** and **5E**). A

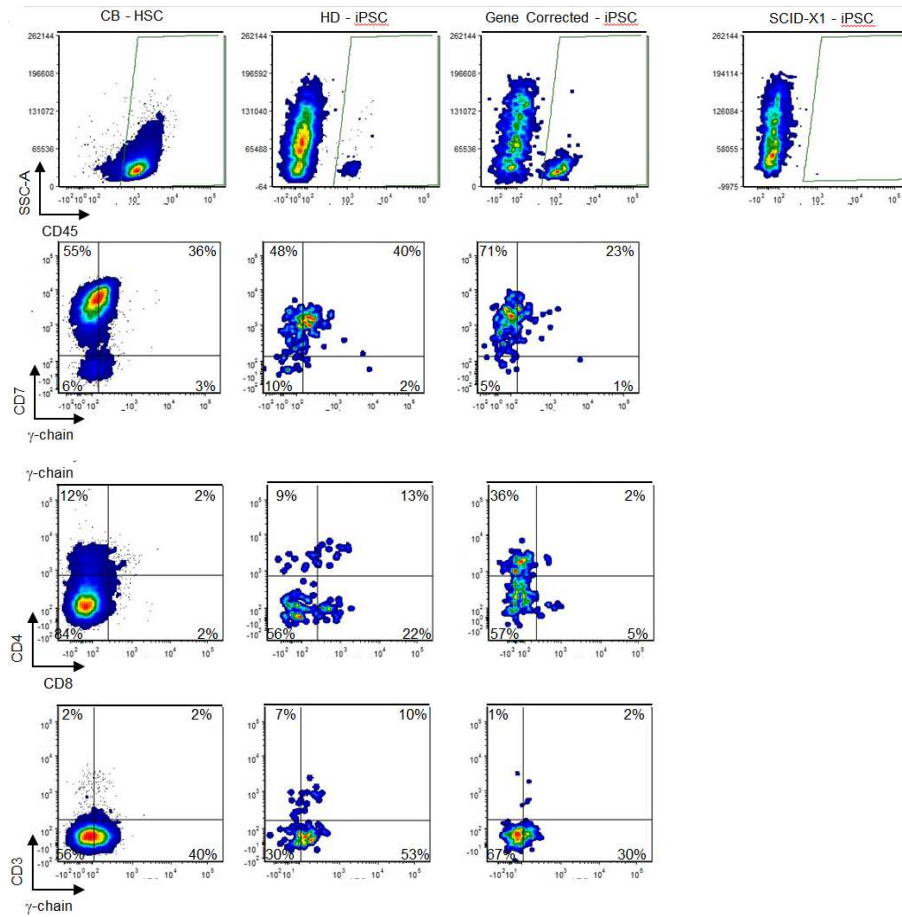
detectable fraction of the  $\gamma$ -chain positive cells also expressed T-lymphoid markers, including CD8, CD4, CD7 and CD3. Intriguingly, SCID-X1 iPSCs failed to generate CD45+ cells in the latter culture conditions. Overall, these data show that gene correction rescues  $\gamma$ -chain expression in hematopoietic cells derived from SCID-X1 iPSCs, and suggest an *in vitro* selective growth advantage of the corrected progenitors over their isogenic, not-corrected counterparts.



D



E



**Figure 5. Hematopoietic differentiation of the iPSCs.**

**A**, Representative bright field pictures of EBs derived from the indicated iPSC lines. **B**, Representative bright field pictures of erythroid (top) and myeloid (bottom) colonies from CFU assays performed with mechanically dissociated EBs from the indicated iPSC clones. HD iPSC: Healthy-Donor derived iPSC. **C**, Histogram showing the percentage of CD7 or  $\gamma$ -chain positive cells upon *in vitro* differentiation of Cord Blood (CB)-derived HSPCs or iPSC clones. For the excised iPSC clone E8 #9, a duplicate of the differentiation is shown. **D**, Histogram showing the percentage of cells positive for the indicated markers upon *in vitro* differentiation of the indicated iPSC clones. **E**, Representative flow cytometry analysis of the differentiated iPSC clones for expression of the indicated hematopoietic markers.



## Discussion

Here we report targeted correction of SCID-X1 mutations in primary patient's fibroblasts and their subsequent conversion to a pluripotent state by an excisable low-copy reprogramming LV. At variance with SCID-X1 iPSC, the cells carrying knock-in of the corrective cDNA were competent for *in vitro* differentiation into T-cell progenitors expressing  $\gamma$ -chain, a cell population of potential therapeutic relevance for the treatment of SCID-X1. This achievement was made possible by identifying and overcoming several limiting steps affecting the efficiency and fidelity of targeted gene correction. These studies showed that retention of the nucleases target site within the corrective cDNA was a primary cause of ineffective gene correction. Two independent but not mutually exclusive mechanisms might explain this phenomenon. First, the nucleases may induce a DNA DSB within the incoming reverse transcribed donor-IDLV, which is then processed for degradation by the endogenous cell machinery. Because of this, the pool of donor DNA template available for HDR is reduced, thus leading to a decrease in the overall gene targeting efficiency. Second, the nucleases may act on the already corrected gene, which is then repaired by the error-prone NHEJ repair pathway. These newly introduced mutations might abrogate expression of the corrected gene or reduce stability of its encoded protein, as suggested by the presence of a significant fraction of  $\gamma$ -chain negative, eGFP positive B-lymphoblastoid cells (**Fig. 1E**). While being underestimated in cells treated with nucleases pair with low-binding activity, these issues were clearly evident in cells treated

with highly-active nucleases, thus affecting the targeting efficiency in a dose-dependent manner. Importantly, both issues were solved by abrogating activity of the nucleases on the corrective cDNA *via* recoding of the ZFNs binding sites. Here, we were forced to recode only the “right” ZFN target site, as nucleotide changes within the “left” ZFN target site would have impaired efficiency of gene targeting. Indeed, the left ZFN target site present within the template DNA is used by the HDR pathway as annealing site for the invading strand originating from the endogenous locus. Recoding of the “right” ZFN binding site in the corrective cDNA resulted in a 10-fold increase in gene-targeting efficiency, with nearly all eGFP-positive cells now expressing  $\gamma$ -chain at comparable levels to those measured in wild-type B-lymphoblastoid cells (**Fig. 2D**) or primary T-lymphocytes (**Fig. 3A**). By combining delivery of the improved donor construct with ZFN-encoding mRNAs in two SCID-X1 fibroblast cell lines, we were able to achieve targeted gene correction in up to 5% of the treated cells (**Fig. 4**).

Because *IL2RG* is not expressed in primary fibroblasts and iPSCs, reprogramming of these bulk-treated populations would have come at the cost of performing labour-intensive and time-consuming molecular screenings in order to identify the few corrected clones. To bypass this problem, we enriched the gene-corrected fibroblasts by eGFP expression or puromycin selection prior to reprogramming. This strategy proved to be extremely valuable, not only because it robustly increased the yield of gene-corrected iPSC *per* reprogramming experiment, but also because we found that the

transgene expression cassette was unexpectedly silenced during the reprogramming process, thus preventing its usage to enrich for gene corrected iPSCs. While being instrumental to our aims, transcriptional repression or eventual reactivation of the reporter cassette in the progeny of the differentiating iPSCs might alter physiologic expression of the proximal corrected gene, thus limiting therapeutic application of this strategy. Because of this, the reporter cassettes were programmed for Cre-mediated excision, a procedure that would have been anyway required to eliminate the reprogramming vector from the genome of the iPSC. To this end, among the several non-integrating technologies currently available to introduce the Cre recombinase into the iPSCs, we chose the IDLV given its high gene delivery rate and tolerability in this cell type. Indeed the IDLV-Cre proved to be very effective in co-excising up to 2 copies of the randomly integrated reprogramming LV and the targeted selector cassette in a single round of infection, thus leading to several gene-corrected and reprogramming factor-free iPSC clones (**Suppl. Fig. 3**). Importantly, all clones analysed displayed a normal karyotype (**Suppl. Fig. 2**), indicating that the gene-targeting, reprogramming and excision procedures had at least no major macroscopic effects on the genome integrity of these cells. The adoption of a Myc-less reprogramming strategy might have contributed to the achievement of this result (Pasi et al, 2011). Finally, by promoting differentiation of the gene-corrected iPSC into T-lymphoid progenitors, we confirmed that the edited *IL2RG* locus, which underwent through several steps of genetic and epigenetic modifications, was able to express functional levels of  $\gamma$

-chain in response to proper differentiation stimuli (**Fig. 5**). These experiments, together with those conducted in primary T-lymphocytes (**Fig. 3**), validate the design of the corrective cDNA in light of its therapeutic usage for the treatment of SCID-X1 caused by mutations within or downstream exon 5 of *IL2RG*. Yet, further biological understanding of T-cell development and technical improvements in the *in vitro* T-cell differentiation protocols will eventually enable generating higher yield of gene-corrected T-lymphoid progenitors capable of long-term thymic engraftment and disease correction in SCID-X1 patients, and potentially recapitulating the clinical achievements of the initial HSPC-based clinical trial with a strategy based on targeted genome editing and cell reprogramming.

## Experimental procedure

### Vectors

Donor transfer constructs were generated from the HIV-derived self-inactivating transfer construct pCCLsin.cPPT.hPGK.eGFP.Wpre. The integrase-detective 3<sup>rd</sup> generation packaging plasmid pMD.Lg/pRRE.D64VInt was generated by replacing the *Bcl*I-*Afl*II fragment from plasmid pCMVDR9-D64V (Naldini et al., 1996). IDLV stocks were prepared as previously described (Follenzi and Naldini, 2002). Briefly, 293T cells were co-transfected by calcium phosphate precipitation with the required transfer vector plasmid, the pMD.Lg/pRRE.D64VInt packaging plasmid, the pMD2.VSV-G envelope encoding plasmid, and pRSV-Rev in the following amounts: 36/12.5/9/6.25 µg DNA per 15 cm dish, respectively. 1mM sodium butyrate was added to the collection medium. Vector particles were concentrated 500-fold by ultracentrifugation and measured by HIV-1 Gag p24 immunocapture (Perkin Elmer). Yield ranged from 20 to 200µg p24/ml, depending on the vector type. IDLV stocks were titered by a qPCR designed to discriminate the reverse transcribed vector genome from plasmid carried over from transient transfection (Matrai et al., 2011). Sequence and maps of *AAVS1- PGK.GFP* were previously reported (Lombardo et al., 2011). Cloning strategies, sequence and maps of the donor-IDLVs, the IDLV-Cre and the reprogramming LVs were previously reported (Genovese et al, 2014).

### Zinc finger nucleases

Zinc fingers nucleases were designed and assembled by the Biotech Company “Sangamo BioSciences”. ZFPs for targeting the *AAVS1* site and exon 5 of *IL2RG* gene were assembled from an archive of

in vitro selected modules (Moore et al., 2001) and optimized in the binding  $\alpha$ -helix (Hockemeyer et al., 2009; Lombardo et al., 2007).

| ZFN            | Recognition sequence | Finger 1    | Finger 2 | Finger 3    | Finger 4 |
|----------------|----------------------|-------------|----------|-------------|----------|
| AAVS1<br>-ZFP1 | CCACTGTGGGG<br>T     | YNWHLQ<br>R | RSDHLTT  | HNYARD<br>C | QNSTRIG  |
| AAVS1<br>-ZFP2 | TAGGGACAGGA<br>T     | QSSNLAR     | RTDYLVD  | YNTHLTR     | QGYNLAG  |
| IL2RG<br>-ZFP1 | ACTCTGTGGAA<br>G     | RKSTLTD     | ARSTRIT  | RSDLSK      | QRSNLKV  |
| IL2RG<br>-ZFP2 | AAAGCGGCTCC<br>G     | RSDNLSV     | RNAHRIN  | TSYGRNE     | ARSTRTN  |

Assembled ZFPs were cloned in-frame as NH<sub>2</sub>-terminal fusions to the catalytic domain of FokI into pcDNA 3.1 (Invitrogen). The obligate heterodimer FokI domains (opti- FokI) generate an active nuclease only by heterodimerization and when incorporated into ZFNs induce DSB with higher specificity (Miller et al., 2007).

Amino acid sequence of the obligate heterodimeric *FokI* domains (Miller et al., 2007) (the optimized aminoacids are underlined):

- 'plus' variant:

QLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFM  
KVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQA  
DEMQRVVKENQTRNKHINPNEWKVPSSVTEFKFLFVSGHFKGNYKA  
QLTRLNHKTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRKFNNGEINF

- 'minus' variant:

QLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFMK  
VYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQAD  
EMERYVEENQTRNKHLPNEWKVPSSVTEFKFLFVSGHFKGNYKAQ  
LTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRKFNNGEINF

Both pairs of ZFNs were transiently expressed as mRNAs. Plasmid templates for ZFNs mRNA production were linearized and purified by phenol/chloroform extraction followed by ethanol DNA precipitation. 2 µg/reaction of linearized plasmid template was in vitro transcribed at 37 °C for 2 hr using T7 RNA polymerase and 7.5 mM nucleotide triphosphates (MEGAscript Kit; Ambion). Triphosphate-derivatives of pseudouridine and 5-methylcytidine (TriLink) were used to generate modified nucleoside-containing RNAs. Cap0 mRNAs was generated by supplementing the reactions with 6 mM m<sup>7</sup>(3'-O-methyl)-G(5')ppp(5')G, a non-reversible cap analog (ARCA, New England Biolabs) and lowering the concentration of GTP to 1.5 mM.

After TURBO DNase treatment (4U/reaction, 1 hr at 37°C), mRNAs were poly(A) tailed with *E. Coli* Poly(A) Polymerase (8U/reaction) for 1 hr at 37°C (PolyA tailing kit; Ambion), yielding ≥ 150 nt polyA. Transcripts were purified by RNeasy Plus Mini Kit (Qiagen). All RNA samples were quantified by spectrophotometry and analysed by denaturing agarose gel electrophoresis for quality assurance.

### **Cell culture and transduction**

Human B-lymphoblastoid cells, HEK293T, K-562 and Mouse Embryonic Fibroblasts (MEF) were maintained as described (Lombardo et al., 2007a). For gene targeting in B-lymphoblastoid cells, 1x10<sup>6</sup> cells were incubated overnight with the indicated donor-IDLVs at Multiplicity of Infection (MOI) of 50, either alone

or together with cognate IDLVs expressing the indicated ZFNs (1µg p24/ml for each ZFN-expressing IDLV). Alternatively, donor-IDLV transduced cells were electroporated with *in vitro* transcribed mRNAs encoding for the indicated ZFNs (from 10 to 50µg of each ZFN mRNA/ml; P3 Primary Cell 4D-Nucleofector X Kit, program EW113; Lonza). K-562 cells were transduced with the indicated LVs at MOI of 0.2. Human fibroblasts from healthy donors and SCID-X1 patients were obtained under informed consent and upon approval of the San Raffaele Hospital Bioethical Committee and of the Great North Childrens' Hospital, Newcastle upon Tyne, respectively. For targeted integration, 10<sup>5</sup> SCID-X1 fibroblasts were transduced at MOI of 50 with the indicated donor-IDLV and then transfected with mRNAs encoding for the indicated ZFNs (0.5µg of each ZFN mRNA/ml; TransIT®-mRNA; Mirus). T-lymphocytes from healthy donors' peripheral blood mononuclear cells were isolated and activated as described (Provasi et al, 2012). After 48 hours of stimulation, T-cells were infected with the indicated IDLVs at MOIs ranging from 50 to 150. The following day, 5x10<sup>5</sup> cells were electroporated with mRNAs encoding for the ZFNs (from 56 to 225µg/ml; P3 Primary Cell 4D-Nucleofector X Kit, program EO 115; Lonza) and then expanded (Lombardo et al, 2011) to perform flow cytometry, DNA analyses and sorting (MoFlo™ XDP Cell Sorter; Beckman Coulter, Inc.). For cytokines release assay, T cells were stimulated at 37°C for 5 hours with PMA (50ng/ml) plus Ionomycin (1µg/ml) in the presence of 2µl per ml of culture of BD Golgi Plug (BD Pharmingen). Antibody anti human CD107a was added to the cells from the beginning of the

stimulation. Cells were then fixed and permeabilized using BD Cytotfix/Cytoperm Kit (BD Pharmingen) and prepared for flow cytometry.

### **Cell reprogramming and iPSC differentiation**

Human fibroblasts were reprogrammed to pluripotency as previously described (Lombardo et al, 2011). Briefly,  $10^5$  cells were transduced with the indicated reprogramming LV at MOI of 1 or 3, plated onto mitotically inactivated MEF and then cultured in human ESC medium for the following 25-30 days, when ESC-like colonies were picked for expansion and analyzed for expression of pluripotency markers (Takahashi et al, 2007). For Cre-mediated excision, single cell-derived iPSC treated with the ROCK inhibitor Y27632 (Watanabe et al, 2007) (Sigma) were incubated overnight with the IDLV expressing the Cre-recombinase from the EF1A promoter and used at 150ng p24/ml. Single cell-derived clones were then expanded and their genomic DNA analyzed by quantitative PCR to measure the rate of excision of the reprogramming LV and the reporter cassette using the assays listed in **Table 1**. For T-cell differentiation, we slightly modified a previously described protocol (Holmes & Zuniga-Pflucker, 2009). Briefly, iPSCs were harvested by treatment with 1mg/ml collagenase IV for one hour. The cell clumps were then cultured in ultra-low attachment plates (Costar) for 2 days in DMEM-KO (Gibco), 20% FBS (Euroclone), 1% NEAA (Invitrogen), 1mM L-Glutamine, 0.1mM  $\beta$ -mercaptoethanol (Gibco) and 0.5ng/ml hrBMP-4 (Peprotech). Then, medium was replaced with the

following: Stempro 34 (Invitrogen) supplemented with 2mM L-Glutamine (Lonza), Ascorbic Acid (50µg/ml, Invitrogen), 0.1mM β-mercaptoethanol (Gibco), 100ng/ml of hrSCF, hrFlt3 ligand and TPO (Peprotech), 10ng/ml hrIL3 and hrIL6 (Peprotech), 50ng/ml hrBMP4 (Peprotech), 200ng/ml Wnt11 (R&D) and 5ng/ml hrVEGF (Peprotech). Medium was changed every 3-4 days and at day 7 was replaced by fresh medium containing rhWnt3a instead of rhWnt11 (R&D). After 14 days in this culture condition, EB containing hematopoietic progenitor cells were enzymatically dissociated and single cells were seeded onto OP9-DL1 feeder cells for T-cell specification in: α-MEM supplemented with 20% FBS, 2mM L-glutamine, 1% Pen/Strep, 0.1mM β-mercaptoethanol, 10ng/ml hrFlt3L (Peprotech), 20ng/ml hrIL-7 (Peprotech). Colony-Forming Unit assay was also performed by growing single-cells in semi-solid medium (Methocult H4434 Classic, StemCell Technologies). After 14 days, immunophenotypic analysis on the differentiated cells was performed by flow cytometry analysis (Genovese et al, 2014). Pictures were taken using a digital inverted microscope (EVOS, AMG). The use of human primary fibroblasts and T-lymphocytes was approved by the San Raffaele Hospital Bioethical Committee (protocols TIGET-HPCT and TIGET-PERIBLOOD, respectively). For the teratoma assay, 10<sup>6</sup> iPSCs were injected subcutaneously into 8- to 11-week-old NOD-SCID\_IL2Rg<sup>-/-</sup> mice (Jackson Laboratory). Mice were sacrificed 6 to 8 weeks after cells injection, when teratomas were isolated, fixed in 4% buffered paraformaldehyde and then stained with hematoxylin-eosin for histopathological analyses. The experimental

protocol was approved by the Institutional Animal Care and Use Committee of the San Raffaele Scientific Institute (IACUC 528).

### **Karyotyping**

Chromosome analysis was done on slide preparations of cell suspensions. Monolayer cell cultures were treated with colcemid at a final concentration of 0.1 µg/ml for 2 hours at 37°C, and mitoses were mechanically removed. After hypotonic treatment with 0.075 M KCl and fixation in methanol:acetic acid (3:1 vol/vol), the cell suspension was dropped onto a slide and air-dried. Cells grown on coverslips were treated the same way except that the colcemid concentration was 0.3 µg/ml. Chromosome counts and karyotype analyses were done on metaphases stained with a standard Q banding.

### **Immunofluorescence and flow cytometry analyses**

For immunofluorescence analysis, cells were fixed in 0.5% paraformaldehyde for 1 hour at room temperature (RT) and washed 3 times in PBS. Cells were then permeabilized with Triton X 0.1% (SIGMA) and blocked with 5% FBS (Euroclone) at RT for 1 hour. Cells were then incubated overnight at 4° C with anti-NANOG, anti-OCT4, anti-Tra-1-60 and anti-SOX2 primary antibodies, washed 3 times in PBS and incubated with the appropriate secondary antibodies at RT for 1 hour. Afterwards cells were washed again in PBS and incubated with TO-PRO for nuclei staining. Pictures were taken by confocal microscopy, using an Axioskop 2 plus direct microscope (Zeiss) equipped with Radiance 2100 three-laser

confocal device (Bio-Rad). Fluorescent signals from the individual fluorophores were sequentially acquired from single optical sections and processed with ImageJ software. Flow cytometry analyses were performed using FACSCantoII or LSRFortessa (BD Pharmingen) with antibodies listed in **Table 2**, according to the manufacturer's instructions. Fluorochrome- and dose-matched isotypes were used as controls. 7-Aminoactinomycin (7-AAD) was used to exclude positive, non-viable cells from the analysis. Analysis was performed on  $1-5 \times 10^5$  cells with FCS express v4.0 (DeNovo software). Cell-Sorting was performed using MOFlo XPD Cell Sorter (Beckman Coulter).

| <b>Antibodies</b> | <b>Conjugated</b> | <b>Company</b>   |
|-------------------|-------------------|------------------|
| anti-human CD132  | APC               | BioLegend        |
| anti-human CD45   | APC-H7            | BD Pharmingen    |
|                   | PeCy7             | Beckmann Coulter |
| anti-human CD3    | PE                | BD Pharmingen    |
|                   | PeCy7             | Beckmann Coulter |
| anti-human CD4    | PB                | BD Pharmingen    |
| anti-human CD8    | APC-H7            | Beckmann Coulter |
| anti-human LNGFR  | APC               | BD Pharmingen    |
| anti-human CD33   | PeCy7             | Beckmann Coulter |
| anti-human CD13   | APC               | BD Pharmingen    |
| anti-human CD14   | PE                | BD Pharmingen    |
| anti-human SSEA-4 | APC               | BioLegend        |
| anti-human CD7    | PE                | Immunotech       |
| anti-human CD34   | PeCy7             | BD Pharmingen    |
| anti-human CD19   | PE                | BD Pharmingen    |
| anti-human CD235a | APC               | BD Pharmingen    |

**Table 2.** Antibodies used for immunophenotypic analyses.

## **Molecular analyses**

### **Mismatch selective endonuclease assay**

Cell assay was used to measure the extent of mutations consequent to NHEJ at the ZFN target sites (Guschin et al., 2010; Miller et al., 2007). Briefly, PCR was performed using primers flanking the

ZFN-recognition site. The PCR product was denaturated, allowed to re-anneal and digested with Surveyor nuclease assay (Transgenomic). Because this enzyme cuts DNA at sites of duplex distortions, the re-annealing products between wild type and mutant alleles, carrying mutations or deletions consequent to ZFN activity, are specifically digested. The reaction products were separated on a Spreadex EL1200 Wide Mini gel (Elchrom Scientific), stained by GelRed (Biotium) and the bands quantified by ImageQuant TL software. The ratio of the uncleaved parental fragment to the two lower migrating cleaved products was calculated using the formula  $(1 - ((\text{sum of the cleaved products}) / (\text{sum of cleaved products and parental fragment})))^{1/2} \times 100$ . The list of primers used for Cel1 analysis is given below.

Mismatch selective endonuclease assay (LV-genome wide)

Forward Cel1- hPGK sense primer: 5'-GTGTGGGGCGGTAGTGTGG-3'

Reverse Cel1- LNGFR antisense primer: 5' AGAAGCAGCAACAGCAGCAGG-3'

Mismatch selective endonuclease assay (*IL2RG*)

Forward Cel1-*IL2RG* primer: 5'- TTCTCCCTTCTCTCATAGACACCC -3'

Reverse Cel1-*IL2RG* primer: 5'- CTCATGGATTGGGTCATGTGG -3'.

### **Targeted integration analysis**

Genomic DNA was isolated with Blood & Cell Culture DNA Midi Kit, DNeasy® Tissue Kit or QIAamp DNA Micro Kit (QIAGEN) according to the starting number of cells.

Extraction of genomic DNA from colonies in CFC assays was performed with Lysis Buffer as previously described (Biffi et al.,

2013). Briefly, cells are resuspended in lysis buffer supplemented with proteinase K and kept 4h at 37°C. After a step of freezing-thawing to disrupt the cells, proteinase K is inactivated by heating 10' 95°C. Finally, cell lisates are centrifuged 5' 8000rpm and the supernatant is used to perform PCR analysis.

To detect targeted integration in the *IL2RG* gene, 40-200 ng genomic DNA was subjected to PCR with a combination of AmpliTaq Gold® (Applied Biosystems) and TaqExtender™ PCR Additive (Stratagene) using primers indicated below. PCR amplicons were resolved on agarose gel and visualized by ethidium bromide staining. For Southern Blot analyses, genomic DNA was extracted with Blood & Cell Culture DNA Midi Kit (QIAGEN) and digested with BspHI for *IL2RG*. Matched DNA amounts were separated on 1% agarose, transferred to a nylon membrane and probed with 32P radiolabeled sequences. Membranes were exposed in a Storage Phosphor Screen. For Q-PCR to establish vector integrations, 200 ng genomic DNA were analyzed using primers and probes complementary to the vector backbone sequence (PBS), the GFP sequence and the human *TERT* gene, the latter amplification used as normalizer, as previously described (Brown et al., 2006; Santoni de Sio et al., 2006). Standard curves for RRE and GFP were generated by serial dilutions of DNA from human cell lines (CEM) containing a known number of vector integrations.

Targeted integration into *IL2RG* by HDR

5' integration junction

Forward *IL2RG* primer: 5'- GCTAAGGCCAAGAAAGTAGGGCTAAAG -3'

Reverse *IL2RG* cDNA exon 6 recoded primer: 5'  
AGCCAGAAGTACACGCACAGC -3'

3' integration junction

Forward SV40pA primer: 5'-ACCTCTACAAATGTGGTATGGCTG -3'

Reverse *IL2RG* primer: 5'- TTCCTTCCATCACCAAACCCTCTTG -3'.

Southern blot analysis to detect targeted integration in *IL2RG*

Forward *IL2RG* probe primer: 5'- AGGGATACTGTGGGACATTGGAG -3'

Reverse *IL2RG* probe primer: 5'- AGGTCCTTCTATCTGTCTGGTTG -3'

### Gene expression analyses

For gene expression analyses, mRNA was extracted using RNeasy Micro Kit (QIAGEN) and cDNA was synthesized using SuperScript VILO cDNA Synthesis Kit (Invitrogen). The resulting cDNA was amplified before quantitative PCR by Taqman PreAmp Master Mix Kit (Applied Biosystems) according to the manufacturer's instructions. Gene expression was performed in triplicate with the TaqMan Gene Expression assays listed in **Table 3** in a Viia7 Real-time PCR thermal cycler. The relative expression level of the indicated genes was calculated by the  $\Delta\Delta C_t$  method and normalized to the indicated genes.

| Gene Symbol | Assay ID      |
|-------------|---------------|
| HPRT        | Hs99999909_m1 |
| Let7a       | hsa-miR Let7a |
| miR-367     | hsa-miR 367   |
| miR-302a    | hsa-miR 302a  |
| miR-302b    | hsa-miR 302b  |
| miR-302c    | hsa-miR 302c  |
| miR-302d    | hsa-miR 302d  |
| KLF4        | Hs00358836_m1 |
| OCT4        |               |

**Table 3. List of the TaqMan Gene Expression assays.**

### **Bisulfite sequencing**

For bisulfite sequencing, genomic DNA from the indicated samples was treated with EpiTect Bisulfite kit (Qiagen) according to manufacturer's instruction and then used to PCR amplify the *Nanog*-promoter region using primers listed below. PCR fragments were purified and cloned into pCRII-TOPO TA and ten clones for each sample were verified by sequencing by the M13 universal primer.

Nanog Forward: 5'-TGGTTAGGTTGGTTTTAAATTTTGG -3'

Nanog Reverse: 5'- AACCCACCCTTATAAATTCTCAATTA -3'

### **Western Blot analyses**

For Western Blot analysis, cells were trypsinized, harvested and lysed and protein content was quantified by BCA (bicinchoninic acid) Protein Assay Kit (Thermo Scientific). A total protein amount of 40-80 ug per lane was loaded and electrophoresed on Nupage precast polyacrylamide 4-12% BIS-TRIS gels. The following antibodies were used for Oct3/4, Santa Cruz, 1:1000; for Sox2, 1:1000, for Klf4, 1:1000; for GFP, 1:2000; for mouse 1:5000; for rabbit 1:5000.

### **Statistical analyses**

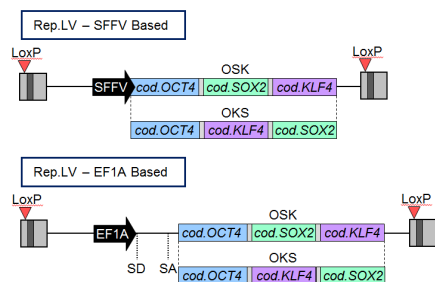
Statistical analyses were performed by unpaired Student's t test for pairwise comparison or one way analysis of variance (ANOVA) with Bonferroni's multiple comparison post-test for three or more

groups, as indicated. Values are expressed as Mean  $\pm$  standard error of the mean (SEM).

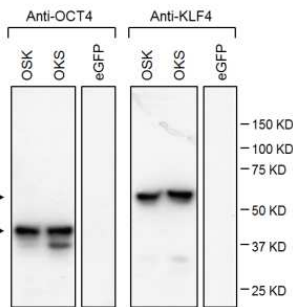


Supplementary Figures

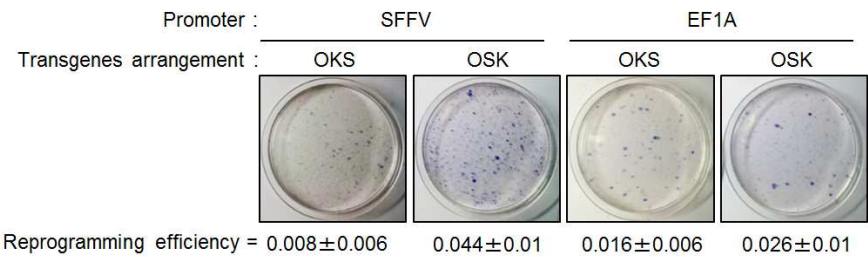
A



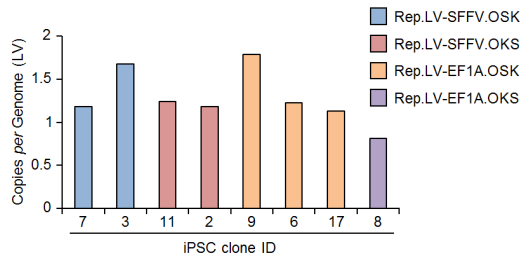
B



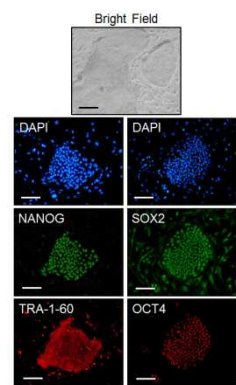
C



D



E



**Supplementary Figure 1. Development of an optimized reprogramming LV.**

To generate pluripotent stem cells from the gene-corrected fibroblasts we developed a Cre-excisable LV platform optimized to reprogram somatic cells from low-copy integrants. First, we constructed a panel of LVs (schematics in panel **A**) containing LoxP sites in the sin-LTRs and expressing a polycistronic mRNA encoding for the human OCT4, SOX2 and KLF4 from the SFFV or the EF1A promoter. The reprogramming genes were arranged in two different permutations, namely OSK or OKS, and were codon optimized to improve translation. Stoichiometric expression of these factors was achieved through the use of the *Thosea asigna* Virus 2A self-cleaving peptide (de Felipe et al, 2006), as shown by Western blot analysis of HEK293T cells transduced with the SFFV-based LVs (panel **B**). By transducing human primary fibroblasts from SCID-X1 patients or healthy donors at multiplicity of infection of 1 with the different reprogramming LVs and then counting the number of alkaline phosphatase-positive colonies, we found that the LV-SFFV.OSK outperformed all the other vectors in terms of reprogramming efficiency ( $0.044 \pm 0.01\%$ , mean  $\pm$  S.D of 3 independent experiments; panel **C**). Notably, the ES-like colonies that emerged 2-3 months after LV-SFFV.OSK transduction expressed *bona fide* markers of pluripotency as gauged by immunofluorescence analysis for TRA1-60, NANOG, OCT4 and SOX2 (DAPI: nuclear staining; representative images in panel **D**) and harbored a single-copy integrant of the reprogramming vector (panel **E**). Finally, by including the miR-302/367 cluster (Anokye-



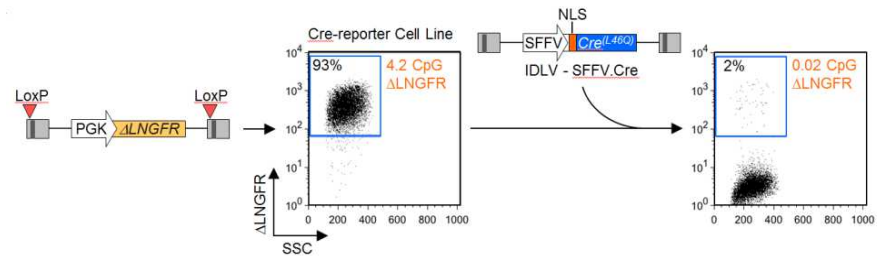
C

|                       | Chromosome |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |   |   |
|-----------------------|------------|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|---|---|
| Clone ID<br>Karyotype | 1          | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | X | Y |
| B11<br>2n: 46,XY      |            |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |   |   |
| A7<br>2n: 46,XY       |            |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |   |   |
| E8 #5<br>2n: 46,XY    |            |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |   |   |
| E8 #EL14<br>2n: 46,XY |            |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |   |   |
| D22<br>2n: 46,XY      |            |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |   |   |
| SCID-X1<br>2n: 46,XY  |            |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |   |   |

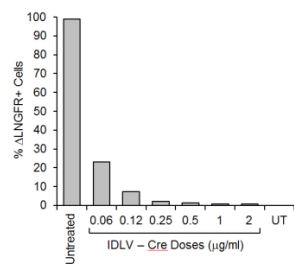
**Supplementary Figure 2. Characterization of the gene-corrected iPSCs.**

**A**, Sequence of the junction from two iPSC clones indicating the expected nucleotide sequence for targeted gene-correction. **B**, Pictures of eGFP-enriched fibroblasts (left) and of a representative iPSC clone derived from the reprogramming of eGFP fibroblasts. **C**, Karyotype analysis of the indicated iPSC clones.

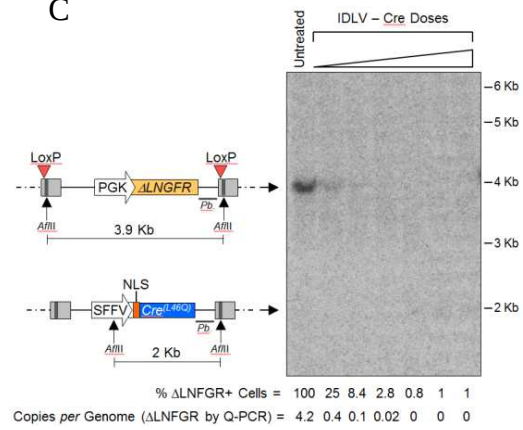
A



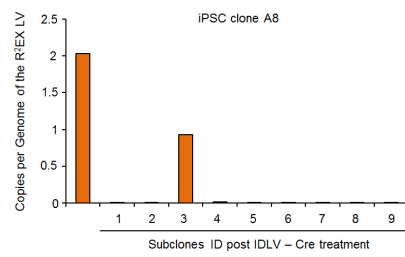
B



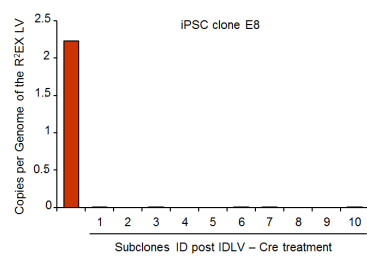
C



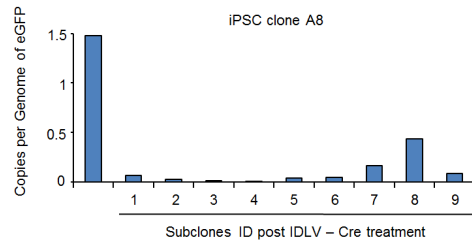
D



E



F



**Supplementary Figure 3. Co-excision of the reprogramming vector and the reporter cassette from the gene-corrected iPSC by IDLV-Cre.**

Since it was previously shown that transcriptional reactivation of the reprogramming genes can decrease the *in vitro* differentiation potential of the iPSC towards (Ramos-Mejia et al, 2012), or that it can eventually lead to the development of hematological malignancies when these cells are transplanted into mice (Nakagawa et al, 2008), we designed the R<sup>2</sup>EX LV for excision from the genome of the iPSCs by transient expression of the Cre recombinase. To test the feasibility of this approach, we first transduced with IDLV-Cre an engineered K-562 cell line engineered to contains 4.2 copies *per* genome of a LV with LoxP sites in the SIN LTRs and expressing then  $\Delta$ LNGFR marker (panel A, showing on the left schematic of the excisable LV and flow cytometry and quantitative PCR analysis of the transduced cells upon their enrichment to near purity by magnetic microbeads selection). A single administration of the IDLV-Cre was sufficient to reduce the vector load by more than 2 log without any sign of obvert cell toxicity, resulting in almost complete abrogation of cell surface expression of the  $\Delta$ LNGFR marker (panel A, showing on

the right schematic of the IDLV-Cre and flow cytometry and quantitative PCR analyses of the transduced cells). The excision efficiency was directly proportional to the dose of IDLV-Cre used (panel **B**, showing the percentage of  $\Delta$ LNGFR<sup>+</sup> cells as measured by flow cytometry analysis 20 days post-treatment with the indicated doses of the IDLV-Cre), approaching 100% in most of the samples, and reaching up to 80% even in those treated with the lowest vector dose. Southern blot and quantitative PCR analyses confirmed excision of the  $\Delta$ LNGFR expression cassettes (panel **C**, showing at the left of the blot schematics of the vectors and the digestion and probing strategy, and reporting at the bottom of the blot the percentage of  $\Delta$ LNGFR<sup>+</sup> cells and the copies per cells of  $\Delta$ LNGFR as measured by flow cytometry or by quantitative PCR, respectively). Notably, no detectable background integration of the IDLV-Cre was observed in these experiments. We then tested the IDLV-Cre platform on two gene-corrected iPSC clones, one derived from the PuroR and the other from the eGFP gene-corrected fibroblasts. In accordance with the results obtained in the Cre-reporter cell line, a single administration of IDLV-Cre resulted in the generation of several reprogramming factors-free, gene-corrected iPSCs (panel **D**, showing the copies *per* genome of the R<sup>2</sup>EX LV as measured by quantitative PCR for the *cod.KLF4* sequence on the clones before or after treatment with the IDLV-Cre). Together with excision of the R<sup>2</sup>EX LV, we also documented excision of the eGFP-reporter cassette (panel **E**, showing the copies *per* genome of the eGFP as measured by quantitative PCR on the A8 clone from panel D).



## References

- Anokye-Danso F, Trivedi CM, Juhr D, Gupta M, Cui Z, Tian Y, Zhang Y, Yang W, Gruber PJ, Epstein JA, Morrissey EE (2011) Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell* **8**: 376-388
- Ciccio A, Elledge SJ (2010) The DNA damage response: making it safe to play with knives. *Mol Cell* **40**: 179-204
- de Felipe P, Luke GA, Hughes LE, Gani D, Halpin C, Ryan MD (2006) E unum pluribus: multiple proteins from a self-processing polyprotein. *Trends Biotechnol* **24**: 68-75
- Genovese P, Schirotti G, Escobar G, Di Tomaso T, Firrito C, Calabria A, Moi D, Mazzieri R, Bonini C, Holmes MC, Gregory PD, van der Burg M, Gentner B, Montini E, Lombardo A, Naldini L (2014) Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* **510**: 235-240
- Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, Morillon E, Clappier E, Caccavelli L, Delabesse E, Beldjord K, Asnafi V, MacIntyre E, Dal Cortivo L, Radford I, Brousse N, Sigaux F, Moshous D, Hauer J, Borkhardt A, Belohradsky BH, Wintergerst U, Velez MC, Leiva L, Sorensen R, Wulffraat N, Blanche S, Bushman FD, Fischer A, Cavazzana-Calvo M (2008) Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* **118**: 3132-3142
- Hacein-Bey-Abina S, von Kalle C, Schmidt M, Le Deist F, Wulffraat N, MacIntyre E, Radford I, Villeval JL, Fraser CC, Cavazzana-Calvo M, Fischer A (2003a) A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med* **348**: 255-256
- Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E, Sorensen R, Forster A, Fraser P, Cohen JJ, de Saint Basile G, Alexander I, Wintergerst U, Frebourg T, Aurias A, Stoppa-Lyonnet D, Romana S, Radford-Weiss I, Gross F, Valensi F, Delabesse E, Macintyre E, Sigaux F, Soulier J, Leiva LE,

Wissler M, Prinz C, Rabbitts TH, Le Deist F, Fischer A, Cavazzana-Calvo M (2003b) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**: 415-419

Holmes R, Zuniga-Pflucker JC (2009) The OP9-DL1 system: generation of T-lymphocytes from embryonic or hematopoietic stem cells in vitro. *Cold Spring Harb Protoc* **2009**: pdb prot5156

Howe SJ, Mansour MR, Schwarzwaelder K, Bartholomae C, Hubank M, Kempinski H, Brugman MH, Pike-Overzet K, Chatters SJ, de Ridder D, Gilmour KC, Adams S, Thornhill SI, Parsley KL, Staal FJ, Gale RE, Linch DC, Bayford J, Brown L, Quaye M, Kinnon C, Ancliff P, Webb DK, Schmidt M, von Kalle C, Gaspar HB, Thrasher AJ (2008) Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J Clin Invest* **118**: 3143-3150

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

Joung JK, Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* **14**: 49-55

Kalman L, Lindegren ML, Kobrynski L, Vogt R, Hannon H, Howard JT, Buckley R (2004) Mutations in genes required for T-cell development: IL7R, CD45, IL2RG, JAK3, RAG1, RAG2, ARTEMIS, and ADA and severe combined immunodeficiency: HuGE review. *Genet Med* **6**: 16-26

Lombardo A, Cesana D, Genovese P, Di Stefano B, Provasi E, Colombo DF, Neri M, Magnani Z, Cantore A, Lo Riso P, Damo M, Pello OM, Holmes MC, Gregory PD, Gritti A, Broccoli V, Bonini C, Naldini L (2011) Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat Methods* **8**: 861-869

Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L (2007a) Gene editing in human stem cells using zinc

finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol* **25**: 1298-1306

Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L (2007b) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol* **25**: 1298-1306

Matrai J, Cantore A, Bartholomae CC, Annoni A, Wang W, Acosta-Sanchez A, Samara-Kuko E, De Waele L, Ma L, Genovese P, Damo M, Arens A, Goudy K, Nichols TC, von Kalle C, MK LC, Roncarolo MG, Schmidt M, Vandendriessche T, Naldini L (2011) Hepatocyte-targeted expression by integrase-defective lentiviral vectors induces antigen-specific tolerance in mice with low genotoxic risk. *Hepatology* **53**: 1696-1707

Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S (2008) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* **26**: 101-106

Pasi CE, Dereli-Oz A, Negrini S, Friedli M, Fragola G, Lombardo A, Van Houwe G, Naldini L, Casola S, Testa G, Trono D, Pelicci PG, Halazonetis TD (2011) Genomic instability in induced stem cells. *Cell Death Differ* **18**: 745-753

Provati E, Genovese P, Lombardo A, Magnani Z, Liu PQ, Reik A, Chu V, Paschon DE, Zhang L, Kuball J, Camisa B, Bondanza A, Casorati G, Ponzoni M, Ciceri F, Bordignon C, Greenberg PD, Holmes MC, Gregory PD, Naldini L, Bonini C (2012) Editing T cell specificity towards leukemia by zinc finger nucleases and lentiviral gene transfer. *Nat Med* **18**: 807-815

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

Sternecker JL, Reinhardt P, Scholer HR (2014) Investigating human disease using stem cell models. *Nat Rev Genet* **15**: 625-639

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

Touzot F, Hacein-Bey-Abina S, Fischer A, Cavazzana M (2014) Gene therapy for inherited immunodeficiency. *Expert Opin Biol Ther* **14**: 789-798

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* **11**: 636-646

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

Woods NB, Bottero V, Schmidt M, von Kalle C, Verma IM (2006) Gene therapy: therapeutic gene causing lymphoma. *Nature* **440**: 1123

## **CHAPTER 3**



## Final Discussion

### Summary

Gene replacement by integrating vectors has been successfully used to treat several inherited disorders, including lysosomal storage disorders (LSD), thalassemia and primary immunodeficiencies (PIDs). For X-linked severe combined immunodeficiency (SCID-X1), a fatal monogenic disorder caused by mutation of the *IL2RG* gene, the early clinical studies have clearly demonstrated the efficacy of integrating vector-based gene replacement therapy, which achieved efficient lymphoid reconstitution, thanks also to the selective growth advantage of the genetically modified cells. However, these studies also highlighted the potential risk of insertional mutagenesis associated to random vector-insertion and to unregulated transgene expression, thus calling for the development of safer gene therapy approaches. In contrast, gene-correction strategies, aimed at site-specific genome-editing, would restore both function and physiologic expression of the mutated gene. In a proof-of-concept study we already reported targeted genome editing in HSCs (Genovese et al., 2014), opening new perspectives for cell-based gene therapy. Even though very fascinating, this approach is currently hampered by limited gene-targeting efficiency, when compared to integrating vector-based gene transfer, and by their *ex-vivo* expansion, which is still elusive and somehow challenging. To this aim, the development of new strategies to select or to expand the few edited cells will be required

in light of future clinical translation. These findings, together with Yamanaka's discovery that somatic cells can be reverted to a pluripotent state by simply overexpressing few reprogramming factors, have opened new perspective for the treatment of human genetic diseases.

Here we reported gene correction of SCID-X1 primary fibroblasts followed by reprogramming to pluripotency, to provide a source of genetically corrected pluripotent cells, amenable to indefinitely expansion and differentiation into clinically relevant cell types.

## **Aim**

In order to generate gene-corrected iPSCs from SCID-X1 patient cells, we first validated our strategy by knocking-in a corrective *IL2RG* cDNA transgene downstream of its endogenous promoter in B-lymphoblastoid cells, which constitutively express *IL2RG*, and in primary T-lymphocytes, which requires *IL2RG* for their survival and growth, and provided evidence of physiologic activity of the gene-edited *IL2RG* gene. We then coupled correction with exogenous selection of the gene-corrected fibroblasts, which do not physiologically express *IL2RG*, by including an excisable GFP- or a Puromycin Resistance (PuroR) expression cassette downstream of the corrective cDNA. The enriched gene-corrected cells so generated were then reverted to pluripotency to obtain a potentially unlimited source of gene-corrected induced pluripotent stem cells (iPSC), by means of a novel reprogramming vector that expresses OCT4, SOX2, KLF4 and microRNA cluster 302-367. This

approach resulted in the generation of corrected bona-fide iPSCs, as confirmed by molecular analyses for targeted integration, which were characterized for their pluripotent state. IDLV-mediated transient delivery of the Cre-recombinase allowed co-excision of both the reprogramming vector and the selector cassette, resulting in the generation of several gene-corrected, reprogramming-factor free iPSCs with normal karyotype. Finally, by differentiating genetically corrected iPSC to T-lymphoid progenitor cells, which are lacking in SCID-X1 patients, and showing a selective growth advantage of those derived from corrected iPSCs, we provided evidence of the functional correction of the *IL2RG* mutant allele. Overall these data demonstrate the feasibility of our targeted gene editing approach, which couples gene correction with cell reprogramming to generate disease-free iPSC, thus paving the way for the development of novel and safer therapeutic option for SCID-X1.

## **Conclusions and future perspectives**

Here we reported the generation of disease-free iPSCs from SCID-X1 primary fibroblasts and their differentiation into T-lineage hematopoietic cells. From a clinical point of view, this work might support the future therapeutic application of iPSCs to derive both HSCs for autologous transplantation or mature immune cells for adoptive immunotherapies. On the other hand, these experiments underline also the requirement to develop more robust and efficient protocols to derive T-lymphoid cells from iPSC, allowing for

future clinical translation. In fact, despite the intensive efforts, the generation of hematopoietic stem cells (HSCs) from iPSC has remained an elusive goal. To this aim, several groups have focused their attention to a deeper understanding of the biology of HSCs development, thus defining criteria, including the use of cell surface markers, to identify definitive hematopoietic progenitors, possibly allowing selection and expansion of these cells.

Although clinical translation still requires further improvement and scaling up of the protocol, our strategy will potentially provide an endless supply of therapeutically relevant cells, including T and NK-cells, to be used for supportive or curative treatment. Indeed gene-corrected iPSCs will be also helpful to model and predict clinical efficacy of a gene-editing therapeutic approach.

Of note, these immune cells will also represent a valuable cell-source to treat other relevant diseases, including immunodeficiencies, as well as pathogen infections or tumours.

## Publications

Firrito C., Genovese P., Schioli G., Biffi M., Di Tomaso T., Quintana-Bustamante O., Gennery A.R., Gregory P.D., Segovia J.C., Bueren J.A., Lombardo A., Naldini L. Targeted correction and Reprogramming of SCID-X1 Fibroblasts Rescues IL2RG in iPSC-derived T-lymphoid Cells. Under Revision for Embo. Mol. Med.

Genovese P., Schioli G., Escobar G., Di Tomaso T., Firrito C., Calabria A., Moi D., Mazzieri R., Bonini C., Holmes MC., Gregory PD., van der Burg M., Gentner B., Montini E., Lombardo A., Naldini L. (2014) Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* 510: 235-40.



## Acknowledgments

I would like to thank Prof. Luigi Naldini for giving me the opportunity to work in his excellent laboratory, where I have been involved in a very challenging project, surrounded by very interesting and challenging projects.

I thank Angelo Lombardo for his supervision, stimulating ideas and fruitful discussion.

I thank all the persons that were involved in this project: Pietro Lo Riso, Mauro Biffi for his technical help, Tiziano Di Tomaso for gene-expression analyses and mRNA production, Pietro Genovese and Giulia Schirotti for their work in primary T-lymphocytes and all the lab members, for scientific discussion and for sharing both good time and bad times in our daily research activity.

I would like to thank also Prof. Massimo Pizzato for the long-lasting support he gave me from the beginning, for believing in me and for conveying the passion for Science.

I express my gratitude to my friends, especially Angelo, Carolina, Donata, Francesca, Luisa for all the pleasant moments spent together and their sincere friendship; thank also to Qi, for his support and encouragement during these months.

My sincere gratitude goes to my family, for the patience and the loving and invaluable support. A special thanks goes to Nicolò, for the joy you brought in our lives!