

GENERATION OF XANTHOMONAS DERIVED TALE PROTEINS  
THAT INHIBIT GENE TRANSCRIPTION

by  
ŞEYDA ŞAZİYE TEMİZ

Submitted to the Graduate School of Engineering and Natural Sciences  
in partial fulfillment of  
the requirements for the degree of  
Master of Science

Sabancı University

July 2013

GENERATION OF *XANTHOMONAS* DERIVED TALE PROTEINS  
THAT INHIBIT GENE TRANSCRIPTION

APPROVED BY:

Assoc. Prof.Dr. Batu Erman



(Thesis Supervisor)

Prof.Dr. Canan Atılgan



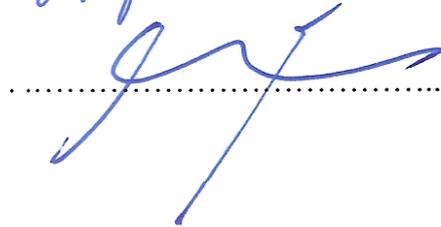
Prof.Dr. Selim Çetiner



Prof.Dr. Uğur Sezerman



Assist.Prof.Dr. Erdal Toprak



DATE OF APPROVAL: .....30.07.2013.....

© Şeyda Şaziye Temiz, 2013

ALL RIGHTS RESERVED

## **ABSTRACT**

### **GENERATION OF XANTHOMONAS DERIVED TALE PROTEINS THAT INHIBIT GENE TRANSCRIPTION**

Şeyda Şaziye Temiz

Biological Sciences and Bioengineering, MSc. Thesis, 2013

Thesis supervisor: Batu Erman

**Keywords:** Golden Gate cloning, IL-7R alpha, NF-kappa B,  
Transcription activator-like effector, TALEN

In the first part of this study, we aimed to mutate several transcription factor binding sites in the IL-7R alpha gene locus by generating transcription activator like effector (TALE) nucleases (TALEN) targeting these sites. We designed, constructed and expressed 3 pairs of TALENs targeting the NF-kappaB, Notch, and glucocorticoid receptor (GR) binding sites in the IL7R gene enhancer. We generated cell lines with insertion and deletion (INDEL) mutations induced by these TALENs at these target sites and determined the effects of these mutations on IL-7R alpha gene expression. We assessed TALEN induced mutations in murine Neuro-2a and RLM11 cell lines by a modified restriction fragment length polymorphism (RFLP) assay and by DNA sequencing. We demonstrate that mutations induced by TALEN pairs targeting the IL7R enhancer NF-kappaB site reduce IL-7R alpha gene expression, while mutations in the Notch binding site did not change IL-7R expression. In the second part of this study, we aimed to inhibit the transcription activation function of the NF-kB protein by competitive binding of target sites with TALE proteins. We generated plasmids encoding TALE-dsRed fusion proteins that were designed to bind NF-kB binding sites in a reporter cell line. TNF-alpha treatment of this cell line results in NF-kB nuclear translocation and a resultant increase in GFP fluorescence. TALE-dsRed fusion proteins with increasing numbers of DNA binding repeats competed for NF-kB binding to this reporter and resulted in reduced GFP expression upon TNF treatment. Our experiments demonstrate that TALENs and TALEs can efficiently inhibit gene transcription.

## ÖZET

### GEN TRANSKRİPSİYONUNU BASKILAYAN XANTHOMONAS KÖKENLİ TALE PROTEİNLERİNİN OLUŞTURULMASI

Şeyda Şaziye Temiz

Biyoloji Bilimleri ve Biyomühendislik, Master Tezi, 2013

Tez Danışmanı: Batu Erman

Anahtar Kelimeler: Golden Gate klonlama, IL-7R alfa, NF-kappa B,  
Transcription activator-like effector, TALEN

Bu çalışmanın ilk bölümünde IL7R alfa gen bölgesindeki çeşitli transkripsiyon faktör bağlanma bölgelerinde oluşturulan transcription activator-like effector (TALE) nükleaz (TALEN) plazmidleri ile mütasyonlar oluşturmayı amaçladık. IL-7R geninin enhancer bölgesindeki NF-kappaB, Notch ve GR bağlanma bölgelerini hedefleyen 3 çift TALEN plazmidini oluşturduk. TALEN plazmidlerini hücrede ifade ederek mütasyonlu hücre hatları oluşturduk ve bu mütasyonların IL7R ifadesine etkisini araştırdık. Neuro-2a ve RLM11 hücre hatlarında oluşturduğumuz mütasyonları modifiye edilmiş RFLP yöntemi ile DNA sekanslaması ile belirledik. NF-kB bağlanma bölgesindeki mütasyonların IL-7R ifadesinde azalmaya neden olduğunu, Notch bağlanma bölgesindeki mütasyonların ise IL-7R ifadesini değiştirmedeğini gözlemledik. Bu çalışmanın ikinci bölümünde, NF-kB proteinlerinin transkripsiyon aktivasyonundaki rolünü NF-kB bağlanma bölgelerine kompetitif olarak bağlanan TALE proteinleri ile inhibe etmeyi amaçladık. TNF-alfa uygulanmasına NF-kB translokasyonu sonucunda GFP ifadesi ile cevap veren reporter hücre hattındaki NF-kB bağlanma bölgesine bağlanacak TALE-dsRed füzyon proteinleri ifade eden plazmidler oluşturduk. TALE-dsRed füzyon proteinlerinin NF-kB proteininin bağlanmasını kompetitif olarak inhibe ederek reporter hücrelerde GFP ifadesini azalttığını gözlemledik. Çalışmalarımız, TALE ve TALEN proteinlerini gen transkripsiyonunu baskılayabileceğimizi göstermektedir.

*To my family...*

## ACKNOWLEDGEMENT

First, I would like to thank my supervisor Assoc. Prof. Dr. Batu Erman for the guidance, support and patience during my master's project. I am so grateful to him for the opportunity of working on such a unique project in his lab.

I would like to thank to the members of my thesis committee, Prof. Dr. Selim Çetiner, Prof. Dr. Uğur Sezerman, Assist. Prof. Dr Erdal Toprak, and Prof. Dr. Canan Atılgan for their support and helpful criticism for my thesis evaluation.

I would also like to thank to my colleagues Nazlı Keskin, Canan Sayitođlu, Emre Deniz, Bahar Shamloo and Gülperi Yalçın for sharing all cheerful moments for these two years in the lab. I am happy for being a member of such an understanding and friendly group.

I would like to express my appreciation to my precious friends Tuğçe Öz and Tuğçe Altınuşak for their moral support and patience. Their presence is a gift of Sabancı University for me which made these two years incredibly beautiful.

I owe special thanks to my family for their patience, motivation, and continuous and unconditional love. Their endless confidence in me is the most important factor for all the things I have achieved.

Finally, I would like to thank the Scientific and Technological Research Council of Turkey, TÜBİTAK BİDEB for the support during my thesis project. This project is supported by TÜBİTAK 109T315.

## TABLE OF CONTENTS

1. INTRODUCTION.....	1
1.1. Transcription Activator Like Effectors.....	1
1.1.1. Special Structural Features of TAL Effector Proteins.....	1
1.1.2. Crystal Structure of TAL Effector Protein.....	4
1.1.3. Designing Custom TAL Effector Proteins.....	6
1.1.4. TALE Assembly Platforms.....	7
1.1.5. Targeted Genome Modification Using TALENs.....	9
1.1.6. Types of Genome Modification.....	11
1.1.7. Scaffold Optimization.....	13
1.1.8. Applications of Genome Editing Using TALENs.....	15
1.2. Interleukin-7 Signaling.....	16
1.2.1. Interleukin-7 and Interleukin-7 Receptor.....	16
1.2.2. IL-7R Signaling Pathways.....	17
1.2.3. Importance of IL-7R Signaling for Lymphopoiesis.....	18
1.2.4. Regulation of IL7R alpha Gene.....	20
1.2.4.1. Notch.....	22
1.2.4.2. NF- $\kappa$ B.....	23
1.2.4.3. Glucocorticoid Receptor (GR).....	24
2. AIM OF THE STUDY.....	26
3. MATERIALS AND METHODS.....	27
3.1. Materials.....	27
3.1.1. Chemicals.....	27
3.1.2. Equipment.....	27
3.1.3. Buffers and Solutions.....	27
3.1.4. Growth Media.....	28
3.1.4.1. Bacterial growth media.....	28
3.1.4.2. Mammalian cell culture growth media.....	28
3.1.5. Cell Types.....	29
3.1.6. Commercial Molecular Biology Kits.....	29
3.1.7. Enzymes.....	29

3.1.8. Vectors and Primers.....	29
3.1.9. DNA Molecular Weight Marker.....	32
3.1.10. DNA Sequencing.....	32
3.1.11. Software and Computer Based Programs.....	32
3.2. Methods.....	33
3.2.1. Bacterial Cell Culture.....	33
3.2.1.1. Bacterial culture growth.....	33
3.2.1.2. Competent cell preparation and transformation.....	33
3.2.1.3. Plasmid DNA isolation.....	34
3.2.2. Vector Construction.....	34
3.2.3. Construction of TALE Expression Plasmids.....	35
3.2.3.1. Identification of TALE and TALEN target sites.....	35
3.2.3.2. Assembly of custom TAL Effector and TALEN constructs using Golden Gate TALEN kit.....	35
3.2.4. Mammalian Cell Culture.....	42
3.2.4.1. Maintenance of mammalian cell lines.....	42
3.2.4.2. Transient transfection of adherent cells with PEI.....	43
3.2.4.3. Transient transfection of suspension cells.....	43
3.2.4.4. Infection.....	44
3.2.4.5. Flow cytometric analysis.....	44
3.2.5. TALEN Induced Mutation Screening.....	44
3.2.5.1. Genomic DNA extraction.....	46
3.2.5.2. Restriction Fragment Length Polymorphism (RFLP) Analysis.....	46
4. RESULTS.....	48
4.1. Use of TALENs to Mutate Transcription Factor Binding Sites of the IL-7R Gene.....	48
4.1.1. Commercially Designed GR Binding Site TALEN Pair.....	49
4.1.1.1. Cloning of the left ECR3 GR binding site TALEN upstream of the eGFP in retroviral plasmid.....	51
4.1.1.2. Cloning of the right ECR3 GR binding site TALEN upstream of the dsRed in retroviral plasmid.....	53

4.1.1.3.	Stable expression of ECR3 GR binding site TALEN pair constructs in the NIH3T3 cell line.....	55
4.1.1.4.	Cloning of the left ECR3 GR binding site TALEN into a CMV IRESeGFP mammalian expression plasmid.....	56
4.1.1.5.	Cloning of the right ECR3 GR binding site TALEN-IRES-dsRed cassette in viral plasmid downstream of CMV promoter.....	58
4.1.1.6.	Ectopic expression of the ECR3 GR binding site TALEN pair in Neuro-2a cells.....	60
4.1.2.	Assembly of TALENs Targeting Notch Binding Site of the IL7R Using Golden Gate TALEN Kit.....	62
4.1.2.1.	Construction of a TALEN pair targeting the IL7R ECR2-ECR3 Notch binding site in the pCAGT7 backbone.....	62
4.1.2.2.	Cloning of the designed Notch binding site TALEN monomers into mutant FokI destination vectors.....	65
4.1.2.3.	Construction of the TALEN pair targeting Notch binding site in the pC-Goldy backbone.....	68
4.1.2.4.	Expression of the designed Notch binding site TALEN pair in Neuro-2a cells and detection of site-specific mutations.....	69
4.1.2.5.	Expression of the designed Notch binding site TALEN pair in RLM11 cells and detection of site-specific mutations.....	72
4.1.2.6.	Detection of Mutation at the Notch Binding Site TALEN Target Site for Different Backbones.....	74
4.1.2.7.	Expression of IL7R on RLM11 cells transfected with Notch binding site TALEN pairs.....	75
4.1.3.	Assembly of TALENs targeting the NF- $\kappa$ B binding site of the IL7R using Golden Gate TALEN kit.....	76
4.1.3.1.	Construction of a TALEN Pair targeting the IL7R ECR3 NF- $\kappa$ B binding site in the pCAGT7 backbone.....	76
4.1.3.2.	Construction of a TALEN pair targeting the IL7R ECR3 NF- $\kappa$ B binding site in the pC-Goldy backbone.....	79

4.1.3.3. Expression of the designed NF- $\kappa$ B binding site TALEN pair in RLM11 cells and detection of site-specific mutations...	80
4.1.3.4. Expression of IL7R on RLM11 cells transfected with NF- $\kappa$ B binding site TALEN pair.....	83
4.2. Use of TALE as Competitive Inhibitors.....	83
4.2.1. Construction of pCAGdsRed Plasmid.....	84
4.2.2. Golden Gate TALE Assembly of the Competitive Inhibitors of NF- $\kappa$ B Binding Using the pCAGdsRed Plasmid as Backbone Vector.....	85
4.2.3. Expression of TALE-dsRed Constructs in HEK293 6.1.1 Cells and the Effect on GFP Expression.....	91
5. DISCUSSION.....	93
6. CONCLUSION.....	102
REFERENCES.....	104
APPENDIX.....	114
APPENDIX A: Chemicals Used In the Study.....	114
APPENDIX B: Equipment Used In the Study.....	116
APPENDIX C: DNA Molecular Weight Marker.....	118
APPENDIX D: FACS Analysis of GFP and dsRed Expression Levels in NIH3T3 Cell Line Infected with ECR3 GR Binding Site TALEN Plasmids.....	119
APPENDIX E: FACS Analysis of GFP and dsRed Expression Levels in Neuro-2a Cell Line Transfected with CMV ECR3 GR Binding Site TALEN Plasmids.....	120
APPENDIX F: FACS Analysis of IL7R $\alpha$ Expression Levels on RLM11 Cell Line Transfected with Constructed TALEN Plasmids.....	121
APPENDIX G: Representative Sequence Analysis of Mutations Induced at Notch Binding Site TALEN Target Sites of Neuro-2a Cells.....	122
APPENDIX H: FACS Analysis of GFP Expression Levels in HEK293 6.1.1 Reporter Cell Line Transfected with TALEdsRed Expression Plasmids.....	123

## LIST OF FIGURES

Figure 1.1 Transcription activator-like effector (TALE) protein structure and DNA recognition code.....	2
Figure 1.2 Tandem repeats in the DNA binding domain of TALE protein from <i>Xanthomonas axonopodispv. citri</i> str. 306.....	3
Figure 1.3 Crystal structure of TALE.....	5
Figure 1.4 TALE based custom proteins can be used to target DNA.....	6
Figure 1.5 TALEN structure for genome editing.....	10
Figure 1.6 TALEN induced genome editing.....	12
Figure 1.7 IL-7 receptor signaling pathway.....	18
Figure 1.8 IL-7R expression by lymphocytes.....	19
Figure 1.9 IL7R gene loci with different transcription factor binding sites.....	21
Figure 1.10 Notch signaling.....	23
Figure 1.11 NF- $\kappa$ B signaling pathways.....	24
Figure 1.12 Glucocorticoid receptor signaling.....	25
Figure 3.1 Golden Gate assembly of custom TALE and TALEN constructs.....	36
Figure 3.2 Timeline for TAL effector and TALEN construction using TALEN Golden Gate kit.....	37
Figure 3.3 Strategies for construction of TALENs in the mammalian expression plasmid.....	42
Figure 3.4 General strategy for detection of TALEN induced mutation at target site.....	45
Figure 3.5 Modified RFLP assay to increase mutation detection efficiency.....	45
Figure 4.1 Schematic representation of the mouse IL7R $\alpha$ gene locus.....	49
Figure 4.2 Binding sites of the commercially designed TALEN pair targeting the GR binding site of the IL7R enhancer region.....	50
Figure 4.3 Construction of the pMIGII left-GR TALEN IRESeGFP plasmid.....	52
Figure 4.4 Construction of the pSP72-right ECR3 GR binding site TALEN plasmid.....	54

Figure 4.5 Strategy for cloning the right ECR3 GR binding site TALEN from the pSP72 plasmid upstream of the IRES-dsRed cassette in the pMIGII plasmid backbone.....	55
Figure 4.6 Infection of NIH3T3 cells with virus produced using GR binding site TALEN fluorescence reporter plasmids.....	56
Figure 4.7 Strategy for cloning of the left ECR3 GR binding site TALEN downstream of CMV promoter and upstream of the IRES-eGFP cassette.....	58
Figure 4.8 Strategy for cloning of the right GR binding site TALEN-IRES-dsRed cassette downstream of CMV promoter.....	60
Figure 4.9 Transfection of murine Neuro-2a cells with CMV GR binding site TALEN fluorescence reporter plasmids.....	61
Figure 4.10 Golden Gate reaction #1 for the left Notch binding site TALEN.....	63
Figure 4.11 Golden Gate reaction #1 for the right Notch binding site TALEN.....	64
Figure 4.12 Golden Gate reaction #2 for the Notch binding site TALEN constructs in the pCAGT7.....	65
Figure 4.13 Cloning of the left Notch binding site TALEN into the pCAGT7 FokI ELD backbone.....	66
Figure 4.14 Cloning of the right Notch binding site TALEN construct into the pCAGT7 FokI KKR plasmid.....	68
Figure 4.15 Golden Gate reaction #2 for Notch binding site TALEN pair in the pC-Goldy backbone.....	69
Figure 4.16. Binding sites for the assembled TALEN pair targeting the Notch binding site of the IL7R enhancer region.....	70
Figure 4.17 Mutation detection at the Notch TALEN target site of Neuro-2a cells using a modified RFLP assay.....	71
Figure 4.18 Site directed mutagenesis in Neuro-2a cells using the Notch binding site TALEN pair.....	72
Figure 4.19 Mutation detection at the Notch TALEN target site of RLM11 cells using the modified RFLP assay.....	73
Figure 4.20 Site directed mutagenesis in RLM11 cells using the Notch binding site TALEN pair.....	74
Figure 4.21 Mutation detection at the Notch binding site of RLM11 cells using a Notch binding site TALEN pair in different backbones.....	75

Figure 4.22 IL7R expression levels of untransfected and Notch TALEN transfected RLM11 cells.....	76
Figure 4.23 Golden Gate reaction#1 for the left NF- $\kappa$ B binding site TALEN.....	77
Figure 4.24 Golden Gate reaction#1 for the right NF- $\kappa$ B binding site TALEN.....	78
Figure 4.25 Golden Gate reaction#2 for the left and right NF- $\kappa$ B binding site TALENs in the pCAGT7 backbone.....	79
Figure 4.26 Golden Gate reaction#2 for the left and right NF- $\kappa$ B binding site TALENs in the pC-Goldy backbone.....	80
Figure 4.27 Binding site of the assembled TALEN pair targeting the NF- $\kappa$ B binding site of the IL7R enhancer region.....	81
Figure 4.28 Mutation detection at NF- $\kappa$ B TALEN target site of RLM11 cells using the modified RFLP assay.....	82
Figure 4.29 Site directed mutagenesis in RLM11 cells transfected with TALEN pair targeting NF- $\kappa$ B binding site.....	82
Figure 4.30 IL7R expression levels of untransfected and NF- $\kappa$ B TALEN transfected RLM11 cells.....	83
Figure 4.31 Strategy of designing TALEdsRed constructs as competitive inhibitors.....	84
Figure 4.32 Cloning strategy for construction of the backbone plasmid with a fluorescent reporter, pCAGdsRed.....	85
Figure 4.33 Golden Gate reaction#1 for the NF- $\kappa$ B reporter A, B6, B5 and B4 plasmids.....	87
Figure 4.34 Golden Gate reaction#1 for the NF- $\kappa$ B reporter B1, B2 and B3 plasmids.....	88
Figure 4.35 Golden Gate reaction #2 for the TALEdsRed12 and TALEdsRed13 plasmid construction.....	89
Figure 4.36 Golden Gate reaction #2 for the TALEdsRed14 and TALEdsRed15 construction.....	90
Figure 4.37 Golden Gate reaction #2 for the TALEdsRed16 and TALEdsRed17 construction.....	91
Figure 4.38 GFP expressions of TALEdsRed transfected HEK293 6.1.1 cells after TNF- $\alpha$ treatment.....	92

Figure D.1 FACS analysis of GFP and dsRed expression in ECR3 GR binding site TALEN infected NIH3T3 cell line.....	119
Figure E.1 FACS analysis of GFP and dsRed expression in CMV ECR3 GR binding site TALEN transfected Neuro-2a cell line.....	120
Figure F.1 FACS analysis of IL7R $\alpha$ expression on RLM11 cell line transfected with TALEN expression plasmids targeting NF- $\kappa$ B binding site and Notch binding site of IL-7R enhancer region.....	121
Figure G.1 Sequencing analysis of the Notch binding site TALEN transfected Neuro-2a cells.....	122
Figure H.1 FACS analysis of GFP expression in transfected HEK293 6.1.1 reporter cell line.....	123

## LIST OF TABLES

Table 3.1 List of vectors used in this project.....	30
Table 3.2 List of primers used in this project.....	31
Table 3.3 List of software and computer based programs used in this study.....	32
Table 3.4 Binding sites of TALEN pair and spacer sequences.....	35
Table 3.5 Components and amounts for Golden Gate reaction #1.....	38
Table 3.6 Optimized colony PCR conditions.....	39
Table 3.7 Components for the first part of Golden Gate reaction #2.....	40
Table 3.8 Components for second part of Golden Gate reaction #2.....	40
Table 3.9 Optimized PCR conditions for TALEN target site amplification.....	46

## LIST OF ABBREVIATIONS

$\alpha$	Alpha
$\beta$	Beta
$\gamma$	Gamma
K	Kappa
CD	Cluster of differentiation
cDNA	Complementary DNA
CLP	Common lymphoid progenitor
CMV	Cytomegalovirus
CSL	CBF1/ RBPjk/ Su(H)/ Lag1
DN	Double Negative
DP	Double Positive
DSB	Double stranded break
ECR	Evolutionarily Conserved Region
ETP	Early Thymic Precursors
FLASH	Fast Ligation-based Automatable Solid-phase High-throughput
FTOC	Fetal Thymic Organ Culture
GABP	GGAA binding protein
GR	Glucocorticoid Receptor
GRE	Glucocorticoid Response Element
HR	Homologous Recombination
Hrp	Hypersensitive response and Pathogenicity
HSC	Hematopoietic stem cell
Hsp90	Heat Shock Protein 90
I $\kappa$ B	Inhibitory of kappa B

ICA	Iterative Capped Assembly
IKK	Inhibitory Kappa B Kinase
IL	Interleukin
IL-7R	Interleukin-7 Receptor
INDEL	Insertion and Deletion
IRES	Internal ribosome entry site
JAK	Janus Kinase
LDL	Low-Density Lipoprotein
LIC	Ligation Independent Cloning
LPS	Lipopolysaccharide
LTR	Long terminal repeat
MAML	Mastermind-like
NICD	Notch Intracellular Domain
NF- $\kappa$ B	Nuclear factor - kappa light chain enhancer of activated B cells
nGRE	Negative Glucocorticoid Response Element
NHEJ	Non-homologous End Joining
NK cell	Natural Killer Cell
NLS	Nuclear Localization Signal
PEI	Polyethylenimine
REAL	Restriction Enzyme and Ligation
Runx1	Runt-related transcription factor 1
RFLP	Restriction Fragment Length Polymorphism
RVD	Repeat Variable Di-residue
SCID	Severe Combined Immunodeficiency
SP	Single Positive
STAT	Signal transducer and activator of transcription
T3S	Type III Secretion System

TAL	Transcription Activator-like
TALE	Transcription Activator-like Effector
TALEN	Transcription Activator-like Effector Nuclease
TALER	Transcription Activator-like Effector Recombinases
TBE	Tris-borate-EDTA
TCR	T-cell Receptor
TD	Translocation Domain
TLR-4	Toll-like Receptor 4
TNF- $\alpha$	Tumor Necrosis Factor Alpha
ZFN	Zinc Finger Nuclease

## **1. INTRODUCTION**

### **1.1 Transcription Activator like Effectors**

Transcriptional activator-like (TAL) effector proteins are produced by Gram-negative bacterial plant pathogens that belong to the genus *Xanthomonas* which cause various diseases in different plant species. These pathogens secrete TAL effector proteins through a Hrp (hypersensitive response and pathogenicity)-type III secretion system (T3S) into the cytoplasm of host plant cells using a bacterial translocon complex. Once translocated to the eukaryotic plant cell, bacterial TAL effector proteins interfere with different plant pathways to contribute to infection. Once inside the plant cell cytoplasm, TALE proteins translocate to nucleus with the help of a NLS and target various elements in the plant genome. By binding to host plant cell gene promoters, TAL effector proteins lead to the transcriptional activation of the host genes[1].

#### **1.1.1 Special Structural Features of TAL Effector Proteins**

TAL effector proteins are composed of an N-terminal translocation domain, a central domain with array of repeat units for DNA binding, and a C-terminal region containing a nuclear localization signal (NLS) and an acidic transcriptional activation domain (Figure1.1).

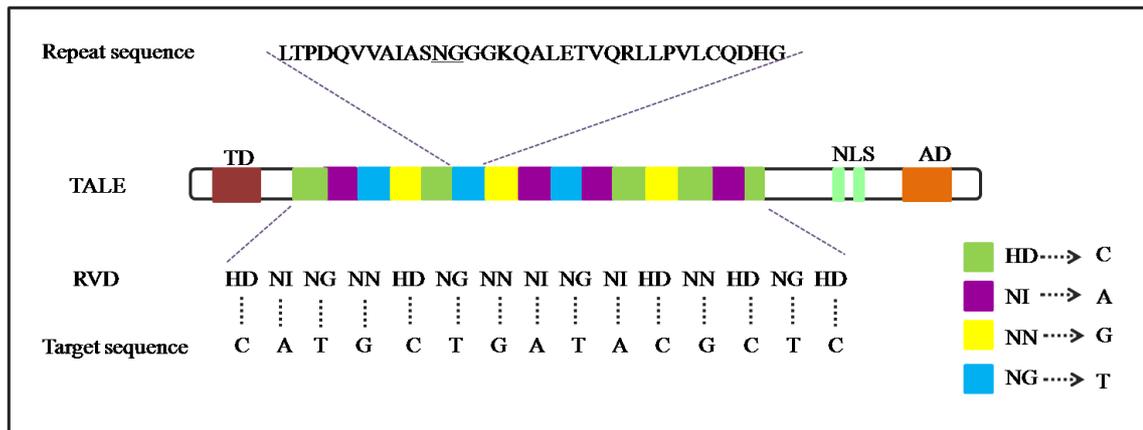


Figure 1.1 Transcription activator-like effector (TALE) protein structure and DNA recognition code. TALE proteins from *Xanthomonas* species consist of an N-terminal translocation domain (TD), a central repeat array for DNA binding, a C-terminal domain with two nuclear localization signals (NLS) and a transcriptional activation domain (AD). Each DNA binding repeat is composed of 34 identical amino acids with the exception of the 12<sup>th</sup> and 13<sup>th</sup> residues, RVDs that determine DNA binding specificity. The consensus repeat sequence is shown in single letter amino acid code above the protein schematic, with the RVD underlined. The DNA binding base preferences of four common RVDs (coded by colored boxes) are shown[2].

The characteristic central DNA binding domain of TALE proteins consists of tandem repeat units with 34 amino acids residues followed by a single half repeat of 20 amino acids. In each repeat unit, only two adjacent amino acid residues at position 12 and 13 are polymorphic and named ‘repeat-variable di-residues’ (RVDs) (Figure 1.2). The DNA binding specificity of a TAL effector protein is determined by the number and order of the different RVD containing repeats. Each RVD in a repeat recognizes a single nucleotide mediated by a code (summarized in Fig.1) that results in specific DNA binding. The correlation between the number of repeat units of TALE binding domains and the length of its target DNA sequence indicated the presence of a code determining RVDs specificity[3]. In 2010, binding specificities of RVDs were validated using computational analysis[4]. In naturally occurring TALE proteins, certain RVDs bind to their corresponding repeat with high specificity such that HD binds to C, NG binds to T and NI binds to A. On the other hand, some RVDs show degeneracy of recognizing two different bases or being nonselective towards bases. Repeat units containing NN RVDs recognize both A and G bases; whereas NS repeats recognizes all four base pairs.

```

LTPEQVVAIASNI GGKQALETVQALLPVLCQAHG
LTPEQVVAIASHD GGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNI GGKQALETVQRLLPVLCQAHG
LTPEQVVAIASHD GGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNI GGKQALETVQRLLPVLCQAHG
LTPDQVVAIASHD GGKQALETVQRLLPVLCQAHG
LTPDQVVAIASHD GGKQALETVQRLLPVLCQAHG
LTPQQVVAIASNG GGKQALETVQRLLPVLCQAHG
LTPEQVVAIASHD GGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNG GGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNG GGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNG GGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNG GGKQALETVQRLLPVLCQAHG
LTLDQVVAIASNG GGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNI GGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNI GGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNG GGRPALESIVAQLSRPDPALA

LTPEQVVAIASN~GGKQALETVQRLLPVLCQAHG

```

Figure 1.2 Tandem repeats in the DNA binding domain of TALE protein from *Xanthomonas axonopodispv. citri* str. 306 (gb|AAM39243.1). Each repeat consists of 34 amino acid residues, where the 12<sup>th</sup> and 13<sup>th</sup> are polymorphic, repeat variable di-residues (RVDs) (highlighted according to the code given in Fig 1.1 for amino acid recognition). Amino acid sequence at the bottom is the consensus.

Comparison of naturally occurring TALE protein RVD sequences and the corresponding DNA binding sites at the promoters of host genes indicates that at a gross level, the code of ‘one RVD to one base’ is not context dependent; in other words, base preference of one RVD is not affected by the preference of adjacent RVDs. However, a big unknown in TALE protein DNA binding is whether all RVDs must contact their corresponding bases or whether ‘mismatches’ can be allowed for efficient binding. Most recognition sites of naturally occurring TALEs are preceded by a thymidine base at position -1 (the base before the TALE binding site). Although no sequence conservation exists between repeat units comprising the DNA binding domain and the amino acid sequence preceding the first repeat, secondary structure prediction studies indicate a degree of conservation in this -1 repeat[3]. TAL effector proteins also make contacts with the -1 T residue by a so called ‘0 repeat’ or ‘cryptic repeat’, present at the N-terminus of the central repeat domain. This interaction was found to be necessary for DNA binding and activation of target genes. The direct relationship between the identity of hypervariable residues (RVD) of repeat units and the sequence of TAL effector protein binding sites in host gene promoters enables the design of artificial TAL effector

proteins targeted to specific binding sites [3, 4]. The ability to assemble custom repeat arrays of TAL effector proteins that can bind desired DNA sequences has recently allowed for the design of artificial transcription factors and DNA binding domains with various functions[5-7].

### **1.1.2 Crystal Structure of TAL Effector Proteins**

The final verification of the code governing TALE protein DNA binding came from crystallization studies in 2012 where two groups determined the atomic scale structure of two TALE proteins. The DNA binding domain of naturally occurring TAL effector protein PthXo1 from the rice pathogen *X.oryzae* was crystallized as bound to its DNA target (PDB:3UGM)[8]. In addition, crystal structure of an artificially engineered TAL effector protein, dHax3 was reported as both DNA-free and DNA-bound states (PDB: 3V6P and 3V6T, for DNA-free and DNA-bound structures, respectively)[9].

The structures in these studies, consistently show that repeat units of TAL effector proteins form a right handed, superhelical structure around a relatively unperturbed B-form DNA helix such that RVDs make contacts with residues in the DNA major groove (Figure 1.3). The external diameter for superhelical wrapping of the TAL effector protein around the DNA duplex is approximately 60 Å. Each repeat unit corresponding to 34 amino acids in the primary sequence, consists of a left-handed helix bundle, in which a short and a long  $\alpha$  helix are connected with a loop. In this structure, residues 3-11 of each repeat unit form the short  $\alpha$  helix, whereas residues 14-33 form the long  $\alpha$  helix, placing the 12<sup>th</sup> and 13<sup>th</sup> residues (the RVD) in the loop inserted into the DNA major groove. These structures identify a proline residue at 27<sup>th</sup> position of the repeats which generates a kink in the long helix, which is claimed to be critical for the sequential packaging of repeat units and the association of the tandem array of repeats with the DNA structure [8, 9]. The 13<sup>th</sup> residue in each RVD makes sequence specific contacts with target DNA; whereas, 12<sup>th</sup> residue interacts with a backbone carbonyl oxygen atom of a conserved alanine residue located at the C-terminus of each repeat. In other words, the first position of the RVD stabilizes the confirmation of the RVD loops rather than recognizing DNA and it is the second position (the 13<sup>th</sup> residue) of every repeat that contributes to the DNA binding specificity[8, 9].

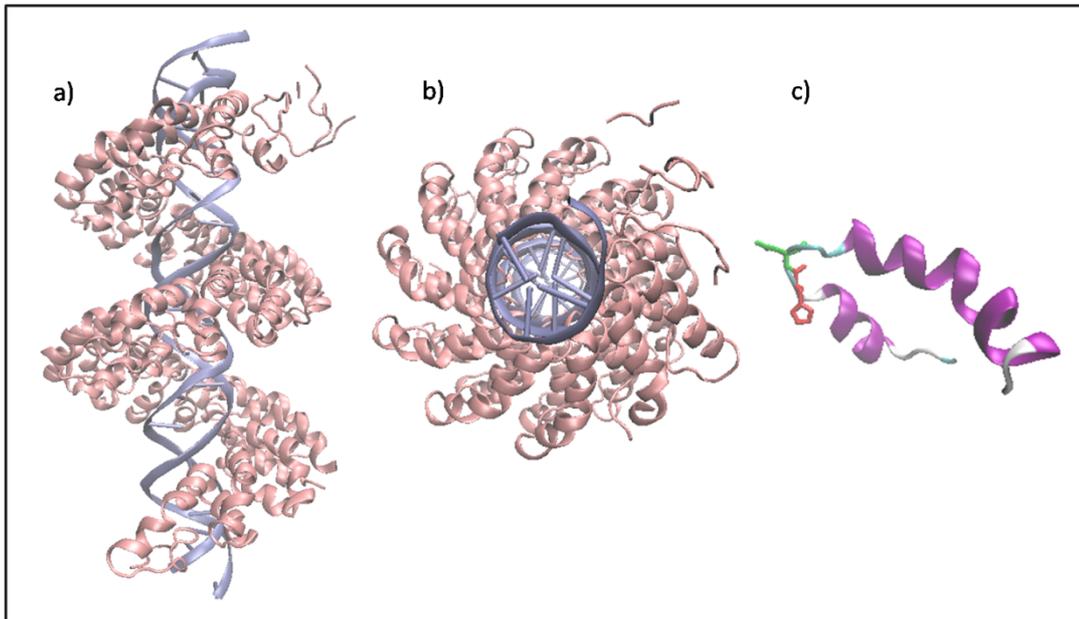


Figure 1.3 Crystal structure of the natural TAL Effector protein, PthXo1. a) Side view of PthXo1 in its DNA bound state and b) top view, the protein backbone is indicated in pink and the DNA double helix is shown in grey. c) Structure of a single repeat unit containing an HD RVD, the H residue is shown in red, the D residue in green and alpha helices in purple (PDB: 3UGM, [8]).

The biochemical basis behind the sequence specific interaction of RVDs with DNA is clearly demonstrated by these two structural studies. In these structures, an HD RVD recognizes a cytosine base utilizing van der Waals interactions between an aspartate residue at the second position of the RVD with the cytosine and hydrogen bonds between a carboxylate oxygen atom of the aspartate and the N4 atom of cytosine. In these structures, an NG RVD interacts with a thymine base such that the smallest amino acid, glycine, at second position provides sufficient space for the 5-methyl group of the thymine base and forms van der Waals interactions with this methyl group. In these structures, an NN RVDs interacts with less specificity, binding both adenosine and guanosine by forming a hydrogen bond between the second position asparagine and the N7 nitrogen of the adenosine and guanosine purine rings. The NS RVD is also nonselective because the second position serine makes hydrogen bonds with the N7 atom of adenosine and guanosine purine rings. Curiously, these structural studies do not yield clues about contacts for interaction of the NS RVD with pyrimidines. Isoleucine, the second position residue of the NI RVD, forms non-polar van der Waals contacts between its aliphatic side chain and the C8 of adenine purine ring or the C5 of a cytosine pyrimidine ring [8, 9].



repeat arrays to targets with guanine bases was found to be less than that of NN containing arrays[14]. In a recent study, the NH RVD was found to be more specific for recognizing guanine over adenine when compared to the other RVDs targeting guanine; NN and NK. Although NH may be a more specific guanine binder, the activity of TALEs with NH containing repeats was less than those with NN containing repeats [14, 15]. Therefore, the current practice of designing artificial TALE proteins must take into consideration the affinity and specificity of individual repeats and often relies on empirically determined rules for binding.

#### **1.1.4 TALE Assembly Platforms**

The presence of multiple repeat sequences different only in two amino acid residues makes the assembly of custom TALEs using common molecular biology techniques, difficult. Although there are commercial DNA synthesis companies such as Collectis Bioresearch and Life Technologies providing custom synthesized genes encoding TALEs[16], synthesis of highly repetitive sequences is complicated and currently too expensive for high-throughput genome editing experiments[2].

An understanding of the features required for TAL effector protein activity has recently enabled the engineering of TAL effector protein coding genes using different assembly platforms, generated in three different laboratories: a) standard cloning-based methods, b) Golden Gate assembly methods and c) solid-phase assembly methods[17].

Standard cloning based methods assemble TALE repeat arrays through sequential restriction digestion and ligation of plasmids encoding units of single or multiple TALE repeat domains. Unit assembly [18], REAL (Restriction Enzyme and Ligation) [19] and REAL-Fast Assembly[20] are three reported methods using standard cloning assembly methods. Although the use of basic molecular cloning techniques seems like an advantage for these methods, it is not possible to perform high-throughput assembly.

The Golden Gate assembly method uses type IIS restriction endonuclease enzymes, which generate multiple sticky ends of fragments that can be assembled in groups of up to 10 repeat unit fragments in the specified order in one single ligation reaction. Golden Gate assembly protocol takes approximately 5 consecutive days. This

assembly entails a two-step ligation reaction, where repeat units are first assembled in intermediary array plasmids and then joined in a final expression plasmid. Sequencing is performed for identification of the clone with correct number of repeat units. The Golden Gate assembly method is advantageous because of its simplicity, speed and low cost. As a result, this is the most popular TALE assembly method in published work [5, 7, 10]. However, assembling large numbers of TAL effector repeat arrays is difficult using the Golden Gate method, making high throughput assembly not feasible.

Currently, there are four different high-throughput TALE assembly methods based on solid phase assembly[2]. First, the FLASH (Fast Ligation-based Automatable Solid-phase High-throughput) system uses an archive of 376 plasmids encoding one, two, three and four TAL effector repeats with variously ordered RVDs that are assembled in an iterative fashion on solid phase magnetic beads. After assembly, the final TAL effector repeat array is released from magnetic beads by restriction enzyme digestion and cloned directly into an expression vector. Using this technique, 96 different DNA fragments encoding the final full-length repeat array with the desired number of repeats can be assembled in less than one day [21]. The second protocol, iterative capped assembly (ICA) involves the addition of monomer units to growing chains of TALE repeats while blocking incomplete extension of chains using hairpin ‘capping’ oligonucleotides. This method allows the synthesis of up to 21 repeat arrays in 3 hours [22]. The third technique, ligation independent cloning (LIC), is based on the use of a library of plasmids encoding repeat unit combinations containing long, unique single stranded DNA overhangs that anneal with overhangs of other fragments without any need for ligation. It is possible to construct plasmids encoding more than 600 TALEs in a single day using the LIC strategy[12]. Finally a magnetic bead based TALE assembly method described by Wang et al. (2012), enables synthesis of over one hundred TALE repeat arrays with 16-20 units in three days[23].

Reagent kits for these different assembly platforms for TAL effector protein construction are currently provided by a non-profit plasmid distribution service Addgene (<http://www.addgene.org/TALEN/>). Software for designing arrays of TAL effector proteins, detailed protocols for plasmid construction, and reference collections are available on websites such as TALE-NT[24] and TALengineering.org[17]. An active and open newsgroup was established by the Joung research group at Harvard

Medical School, USA, for discussion of projects and problems related to TAL effector proteins (<https://groups.google.com/group/talengineering>).

### 1.1.5 Targeted Genome Modification Using TALENs

Designer nucleases are important tools for site directed mutagenesis at the genomic level. Genome editing by nucleases is not only a very useful tool for studying the function of targeted genes, but also has found spectacular success in the clinic for treating patients suffering from diseases caused by monogenic mutations. For treatment of HIV-1 infection, zinc finger nucleases (ZFNs) were designed to disrupt the CCR5 (chemokine receptor 5) gene, which is a co-receptor required to infect T cells. Engraftment of *ex vivo* expanded HIV-1 resistant autologous CD4<sup>+</sup> T cells resulted in lower viral count and higher CD4<sup>+</sup> T-cell count in mice compared to wild-type CD4<sup>+</sup> T cell engrafted mice[25]. The approach of using ZFNs in HIV treatment have entered Phase 2 clinical trials[17].

ZFNs have traditionally been used as artificial (designed) nucleases. ZFNs contain a DNA binding domain composed of 3-4 synthetic zinc finger motifs fused to the non-sequence-specific DNA cleavage domain of the type II restriction enzyme FokI. The crystal structure of zinc finger transcription factors indicate that ZFNs bind DNA whereby each zinc finger motif recognizes a specific DNA sequence by inserting an  $\alpha$  helix into the major groove of the DNA double helix[26]. In this structure, amino acids within each zinc finger motif make contacts with 4 bases of the DNA helix (3 on one strand and one on the opposite strand). Thus, a zinc finger DNA binding protein with 4 motifs can contact up to 12 bases of DNA and zinc finger motifs can be modularly assembled to recognize long DNA sequences.

The FokI restriction enzyme is a type IIS restriction endonuclease and dimerization of its endonuclease domain is required for its activity for creating double stranded DNA breaks. For this reason, ZFNs are designed to have two subunits resulting in the formation of a heterodimer on two closely oriented 'inverted' half sites. ZFN monomers bind to these two half-sites separated by a spacer region on which the FokI domain from each heterodimer assembles and generates a double stranded break (DSB)[27]. ZFNs have been used for genome modification of various model organisms.

However, generation of sequence-specific ZFNs is complicated, due to two main reasons. First, there exists a crosstalk between individual zinc finger motifs such that the motif in the second position affects the binding specificity of the motif in the first position, etc. This limits the modular use of ZFNs for assembling designed DNA binding domains. In other words, the DNA binding of zinc finger nucleases is context-dependent. Secondly, some zinc finger motifs are not specific to the targeted site, such that they can bind and cleave alternative sites, resulting in off target specificity. Because ZFNs are used to introduce DSB in genomic DNA that results in the generation of mutations, off target specificity may lead to unwanted mutations throughout the genome [28].

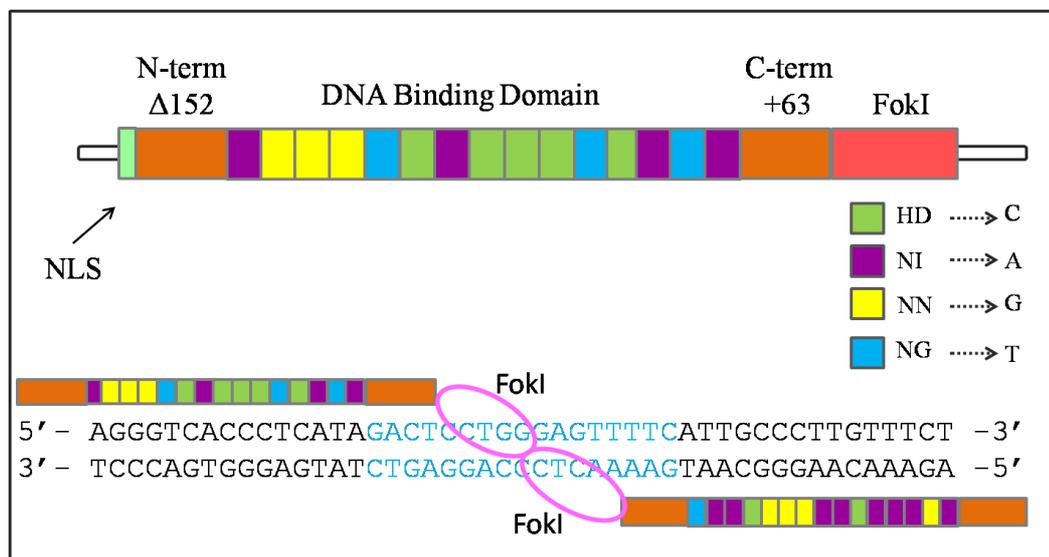


Figure 1.5 TALEN structure for genome editing. For targeted genome modification, a pair of TALEs, each fused to a FokI DNA cleavage domain is designed to bind a target DNA sequence (black bases). The FokI enzyme requires dimerization for its DNA cleavage activity and assembles on the intervening spacer sequence (blue bases) to cleave in this region. TALEN enzymes have a modified structure compared to naturally occurring TALE proteins. The domain structure of TALEN proteins is as follows: the NLS (light green) is located at the N-terminus; N-terminal and C-terminal segments (orange) flank the DNA binding domain; the FokI domain (pink) is fused to the C-terminus. Each repeat unit in the DBD is color coded (as in Figure 1.1) to indicate the RVD-DNA binding code [5, 29].

Transcription activator-like effector nucleases (TALENs) for targeted genome engineering (Figure 1.5) have generated much interest since the discovery of the one RVD to one base code [3, 4]. As in the case of zinc finger nucleases, TALENs consist of a DNA binding domain fused to a FokI restriction enzyme DNA cleavage domain.

Because FokI only cleaves DNA as a dimer, TALENs are designed as heterodimers, such that two monomers bind to individual target sites separated by a short spacer region. The length of the spacer region is important for FokI dimerization and DNA cleavage [30]. Several groups have used TALENs to modify endogenous genes in yeast[31], fruit flies[32], zebrafish[33-36], frogs[37], plants[7], livestock[38], mice[39] and human somatic and pluripotent stem cells[40]. The simple one RVD to one base code makes the construction of TALE repeat arrays targeting any DNA sequence easy and routine. In a recent study, TALENs were found to be significantly more mutagenic than ZFNs [34]. In another study, side-by-side analysis of ZFNs and TALENs with overlapping binding sites for endogenous targets has shown that TALENs were less cytotoxic than ZFNs with similar gene disruption activities[41]. Lower toxicity is likely a result of lower rates of off target cleavage by TALENs when compared to ZFNs, which may result in unwanted mutation of alternative gene loci. These parameters make TALENs superior over ZFNs for targeted gene modification.

#### **1.1.6 Types of Genome Modification**

Genome editing using site-specific nucleases depends on the generation of DNA double stranded breaks (DSB). Cellular repair of DNA DSBs induced at spacer regions occur either by non-homologous end joining (NHEJ) or if a homologous piece of DNA is present, by homologous recombination (HR) (Figure 1.6).

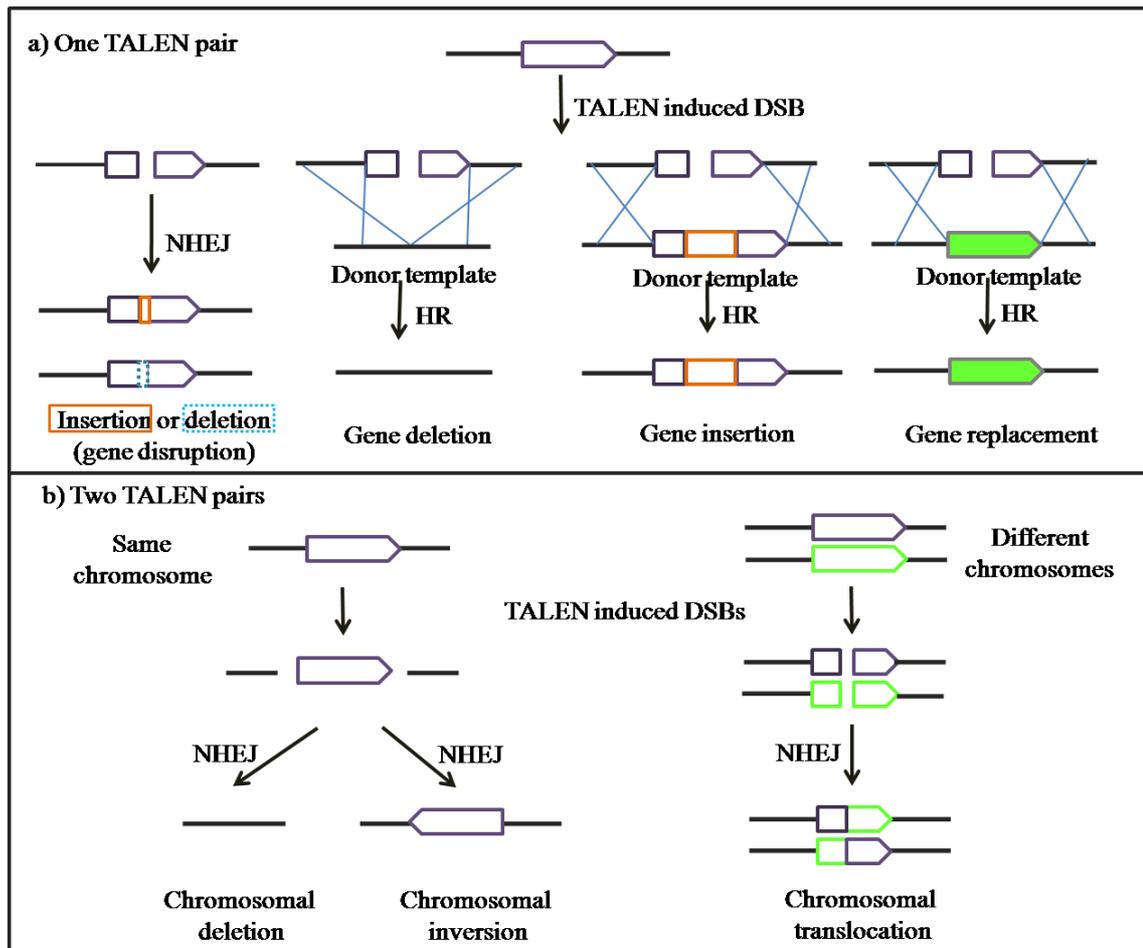


Figure 1.6 TALEN induced genome editing. Genome editing after DSB creation occurs either by non-homologous end joining (NHEJ) or by homologous recombination (HR). a) In the case of targeted genome editing using one TALEN pair, NHEJ results in small insertions and deletions (INDELs) at the site of the DSBs. HR can be used for gene deletion, gene insertion (for example an epitope tag) or gene replacement (for example a fluorescent reporter gene such as GFP) depending on the donor template used. b) If two TALEN pairs create DSBs on the same chromosome, NHEJ mediated repair may result in chromosomal deletion or inversion. If DSBs are generated on different chromosomes, translocations may occur. This mode of DNA repair may be problematic if off-target specificity is not minimized [2, 29].

NHEJ is an error-prone mechanism in which broken DNA ends are simply re-joined leading to small insertions or deletions (INDELs) at the site of the double stranded break. INDELs induced in the protein coding sequences of genes will often yield frame-shift mutations leading to a knock out of gene function. Recently, it was reported that NHEJ mediated reading frame correction can be used to restore protein function in Duchenne muscular dystrophy, a genetic disease caused by mutations in the coding region for the dystrophin gene[42]. Homologous recombination repairs double stranded breaks using a homologous sequence as a template. In this case, a DNA

template with sequences homologous to those flanking the site of the double stranded break is introduced to the cell, together with TALEN encoding plasmids. Depending on the sequences within the donor template, homology directed repair can result in gene deletion, gene addition or gene replacement. Gene addition can be used to integrate specific genes under the control of specific promoter elements or to insert an epitope tag for labeling proteins encoded by endogenous genes. Gene replacement involves exchange of genetic information between an endogenous genomic region and an exogenous DNA template[2]. Use of single stranded homologous oligonucleotides as donors rather than a template plasmids was recently shown to be effective for homology directed repair of DNA double stranded breaks [35]. Introduction of two pairs of TALENs into cells at the same time may lead to more complex genome alterations. If two TALEN pairs target the same chromosome, this results in either large chromosomal deletions or inversions. On the other hand, targeting different chromosomes may lead to translocations[17]. In a recent study, large chromosomal deletions and inversions were obtained in livestock by targeting the same chromosome with two TALEN pairs [38].

### **1.1.7 Scaffold Optimization**

Recent work has identified specific structural features of TAL effector proteins that are important in the construction of proteins with the desired specificity and activity. The main difference between various commonly available TALEN architectures is the length and sequence of the N-terminal and C-terminal amino acid sequences flanking the TALE DBD. In naturally occurring TALE proteins, the N-terminal region contains sequences necessary for secretion into host plant cells. On the other hand, the C-terminal region contains both the nuclear localization signals and a transcriptional activation domain. In the earliest report of targeting DNA double stranded breaks with TAL effector-nuclease fusions, the DNA binding repeat domain was flanked by 287 amino acid N-terminal region and a 231 amino acid C-terminal region. This active TALEN protein pair recognized two 12 bp-long target sites on DNA separated by a spacer of 12-30 bp [30]. Even though this was the first of its kind to generate DSBs in genomic DNA, it was not known if the cutting efficiency or specificity was optimal and whether they could be improved. It is possible that amino

acid sequences flanking the DBD, necessary for TAL effector protein function in plant cells may interfere with the catalytic activity of TALENs.

For this reason, several groups generated truncations in the N and C terminal regions of TALENs to optimize DNA cleavage activity. Miller et al. (2011) tested the activity of TALENs with different C-terminal linker lengths separating the DBD from the FokI catalytic domain. They found that TALENs with the highest activity contained a truncated 136 residue N-terminal and a 63 amino acid C-terminal domain. These truncated TALEN proteins resulted in a genome mutation rate between 5-20% across a spacer size range of 12-20 bp [6]. Another study determined that the minimal DNA binding domain of TALEN proteins must have at least 47 amino acids in the C-terminal linker between the TALE DBD and the FokI catalytic domain in addition to a truncated 153 residue N-terminal domain. This study showed that these truncated TALEN proteins could cleave DNA with a spacer length of 12-21 bp between two target sites. An even shorter C-terminal linker with only 17 amino acids was also shown to be active when used for targeting 12 bp spacers[41]. These studies indicate that there is a correlation between the spacer length of the DNA sequence within which the FokI enzyme cleaves and the length of the C-terminal linker region separating the DBD from the FokI domain. This constraint likely affects the positioning of the two FokI enzymatic domains in a heterodimeric structure that is necessary for cleavage.

A second generation TALEN scaffold named Goldy TALEN was recently reported to have improved genome editing efficiency in zebrafish [35]. Although the Goldy scaffold uses 136 residue N-terminal domain and a 63 amino acid C-terminal linker domain, like the previously described scaffold [6], there are nine different amino acid substitutions at the N-terminal and 5 different amino acids substitutions at C-terminal linker domains. Efficient gene knockout was obtained in livestock using TALEN pairs assembled in a Goldy scaffold[38]. Recently, it was reported that DNA binding domains of 15 RVDs in the Goldy TALEN scaffold with spacers ranging 13-19bp resulted in highly efficient genome editing in zebrafish[43].

Various TALEN protein scaffolds optimize FokI domain dimerization to generate active TALEN heterodimers at target sites. However, optimizing the cleavage activity at a target site may also increase the probability of homodimeric TALEN proteins composed of identical subunits that bind to and cleave unwanted, 'off-target'

sites. Off target cleavage is a critical parameter for the efficacy and safety of designed TALEN pairs. In the case of ZFNs, off-target cleavage and the associated cytotoxicity were reduced using mutant FokI cleavage domains. Specific residues on the dimer interface of the FokI cleavage domain were mutated such that homodimerization of TALEN monomers were prevented by electrostatic and hydrophobic interactions [44]. The idea for mutant FokI cleavage domains to prevent homodimerization was successfully applied to TALEN proteins. In fact, obligate heterodimeric TALEN pairs induced similar or higher mutation frequencies in zebrafish genes when compared to TALENs with the same DNA binding domain with wild type FokI cleavage domains. Moreover, the frequency of abnormal embryos that developed after obligate heterodimeric TALEN pair encoding mRNA microinjection was less than that generated by mRNAs encoding homodimeric TALEN pairs, with wild type FokI domains[45]. Obligate heterodimer TALEN scaffolds were also used in studies that generated gene knockouts in zebrafish[36] and *Xenopus* embryos[37].

### **1.1.8 Applications of Genome Editing Using TALENs**

TALENs have been used in various model organisms for targeted genome modification. In most of the studies, a single TALEN pair was used to induce NHEJ to create small insertions and deletions (INDELs) for generation of gene knockouts [21, 32, 39, 42, 46]. Use of two TALEN pairs to create double stranded breaks on the same chromosome generates large chromosomal deletions and inversions [38]. Introduction of TALEN pairs together with a donor template, even a short single-stranded DNA allows insertion of a desired sequence into the target site [7, 35, 40]. In addition, homology directed repair of DNA double stranded breaks can be used for fusion of endogenous genes to sequences encoding epitope tags or fluorescent reporter proteins such as GFP to track protein expression, distribution and interaction with other proteins [17] (Figure 1.6).

New animal models of human diseases can be rapidly created using TALENs to induce mutations without any need for embryonic stem cell cultures and targeting vectors. Recently, an animal model for familial hypercholesterolemia was created using TALENs targeting a gene that encodes low-density lipoprotein (LDL) receptor in livestock[38]. In a recent study, a phenotypic model of Hermansky-Pudlak syndrome,

which results in decreased pigmentation and bleeding problems, was created by injecting mRNAs encoding a TALEN pair together with synthetic oligodeoxynucleotides into one-cell stage mouse embryos to generate chocolate missense mutations in the RAB38 gene encoding a small GTPase for the regulation of intracellular vesicle trafficking. In this study, germline mutations created through homology directed repair of TALEN induced DSBs were corrected using a donor template with wild type sequence [47]. Thus, TALEN mediated gene modification has a great potential to be used in gene therapy to correct or disrupt genes or gene products, especially in the case of diseases with genetic components.

Another use for TALEN technology is the generation of mutants for conducting structure-function studies probing the function of protein coding genes and genomic regulatory regions. In this study we generated TALENs to make mutations in a putative transcription factor binding sites in the enhancer of the IL7R gene and also we generated TALE proteins that competitively inhibit the important transcription factor NF- $\kappa$ B. The significance of these two TALEN targets is described in the section below.

## **1.2 Interleukin-7 signaling**

### **1.2.1 Interleukin-7 and Interleukin-7 Receptor**

Interleukin-7 (IL-7) is an essential and non-redundant cytokine necessary for the development, differentiation and survival of lymphocytes. The human IL-7 gene is 72kb long and is located on chromosome 8 encoding a protein of 20 kD, whereas the murine IL-7 gene is 41 kb in length and is located on chromosome 3, encoding a protein of about 18 kD. The active form of human IL-7 has a protein size of 25 kD due to post translational glycosylation. It is a single chain protein consisting of four  $\alpha$  helices with a hydrophobic core. Human IL-7 is produced by nonhematopoietic cells, such as bone marrow stromal cells and epithelial cells of the thymus, skin and intestine[48, 49].

IL7 was discovered in 1988 as a result of its proliferative activity on immature murine B-cells *in vivo*. Later studies on IL-7<sup>-/-</sup> and IL7 receptor (IL7R)<sup>-/-</sup> knockout mice displayed a significant decrease in the number of T lymphocytes, indicating a role of the IL-7 cytokine in development. IL7 also has a role in maintaining stable numbers of naive and memory T-cells in the peripheral immune system. The proliferative effect

of IL7 on lymphocytes makes it a potent therapeutic for lymphoid regeneration in lymphopenic states such as after chemotherapy or radiotherapy (reviewed in [50]).

IL7 signals lymphocytes by binding to its specific receptor IL-7R composed of a heterodimer of two transmembrane proteins: the specific  $\alpha$  chain (IL7R $\alpha$ , also known as CD127) and a common cytokine receptor  $\gamma$  chain ( $\gamma_c$ ), which is shared by the receptors of IL-2, IL-4, IL-9 and IL-15. Both of these subunits are necessary for high affinity binding of IL-7. The human IL-7R $\alpha$  gene is localized to chromosome 5 with a size of about 20kb whereas the murine IL-7R $\alpha$  gene is on chromosome 15 with approximate size of 22 kb. Both human and murine genes contain eight exons and seven introns. The mature form of IL7R is composed of 439 amino acid residues with a molecular weight of 49.5 kD. IL7R is expressed mainly by lymphoid lineage cells, namely T-lymphocytes, progenitor B-lymphocytes, and NK cells. IL7R is also expressed by cells of innate immune system such as certain dendritic cells, macrophages derived from bone-marrow and lymphoid tissue inducer cells (LTi). In addition, it was demonstrated that IL7R $\alpha$  was present on non-hematopoietic cells such as human intestinal cells, human endothelial cells and several non-lymphoid cancer cells such as lung, melanoma, renal, colon and breast cancer cells [49, 51]. The role of this receptor on these non-lymphoid lineages is currently not known.

### **1.2.2 IL-7R Signaling Pathways**

Extracellular IL-7 binding induces dimerization of the IL7R $\alpha$  and  $\gamma$  chains. As a result, JAK kinases bound to the intracellular domains of the IL7R $\alpha$  chain and  $\gamma$  chain are activated such that JAK3 phosphorylates JAK1 and the  $\alpha$ -chain. Phosphorylated residues promote the recruitment of PI3K and STAT proteins. PI3K phosphorylates Akt, which promotes cell survival through degradation of pro-apoptotic proteins such as Bad and Bax. Phosphorylated STAT proteins dimerize and translocate to the nucleus and function as transcription factors that induce the expression of target genes such as Bcl-2, Cyclin D1, SOCS-1 and c-myc (Figure 1.7)[52].

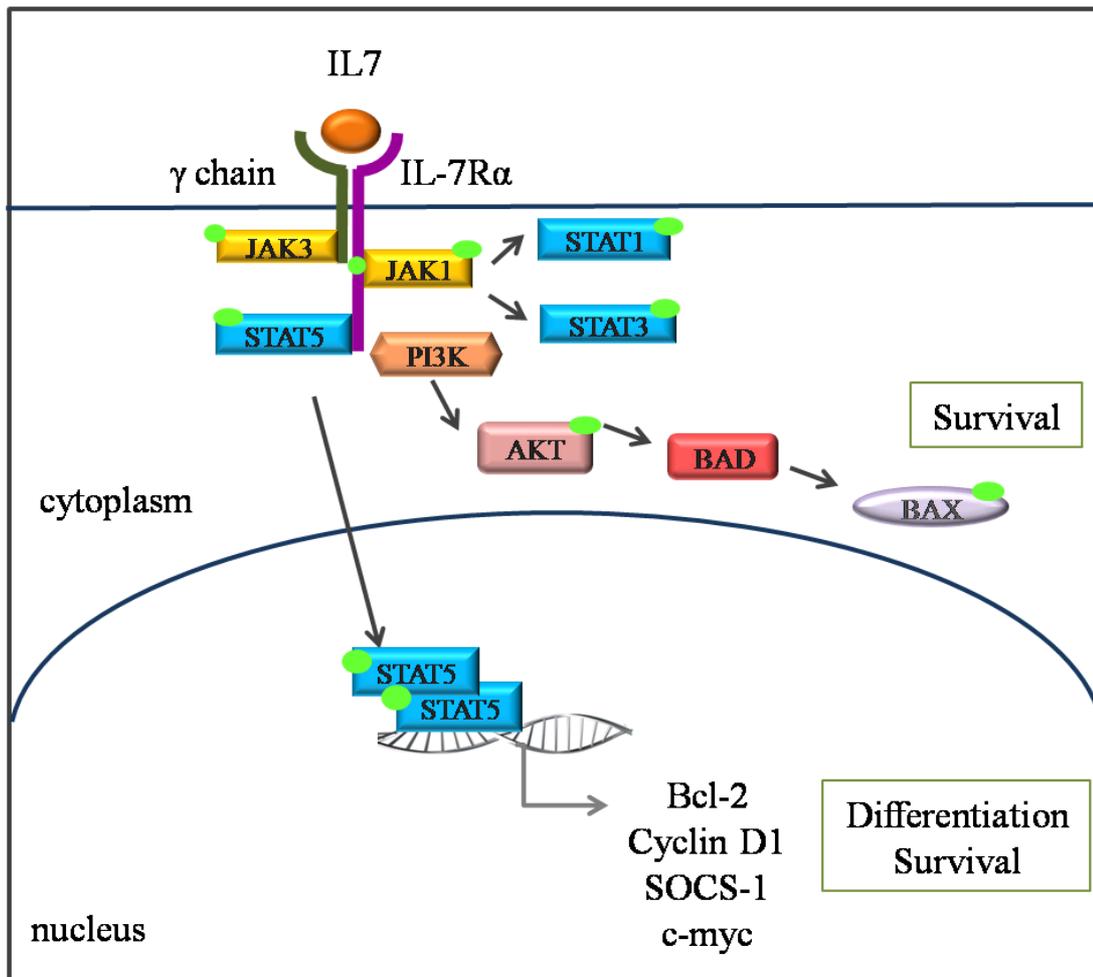


Figure 1.7 The IL-7 receptor signaling pathway[52].

### 1.2.3 Importance of IL-7R Signaling for Lymphopoiesis

B cell development occurs mainly in the bone marrow and can be divided into different stages according to the expression of intra cellular and surface markers, rearrangement of status of the antibody encoding immunoglobulin heavy and light chains, and their cell cycle status[53]. Figure 1.8 shows the different stages of B cell development in the bone marrow and the expression of IL-7R in these stages.

The importance of the IL-7 response for mouse B cell development was demonstrated by a block in the transition from the pro-B cell to the pre-B cell stage in IL-7R deficient mice[54]. IL-7R signaling has a role in regulating the accessibility of chromosomes containing the immunoglobulin heavy chain genes to the gene recombination machinery during B lymphocyte development. Immunoglobulin gene recombination is important for the generation of the primary antibody repertoire diversity [55]. Attenuation of IL-7R signaling by the transcription factor IRF-4,

upregulated by pre-BCR signals, affects activation of light chain rearrangement in pre-B lymphocytes [56]. IL-7R signaling is necessary for expression of transcription factors such as EBF, important for transition from the pro-B stages to the more mature stages [57].

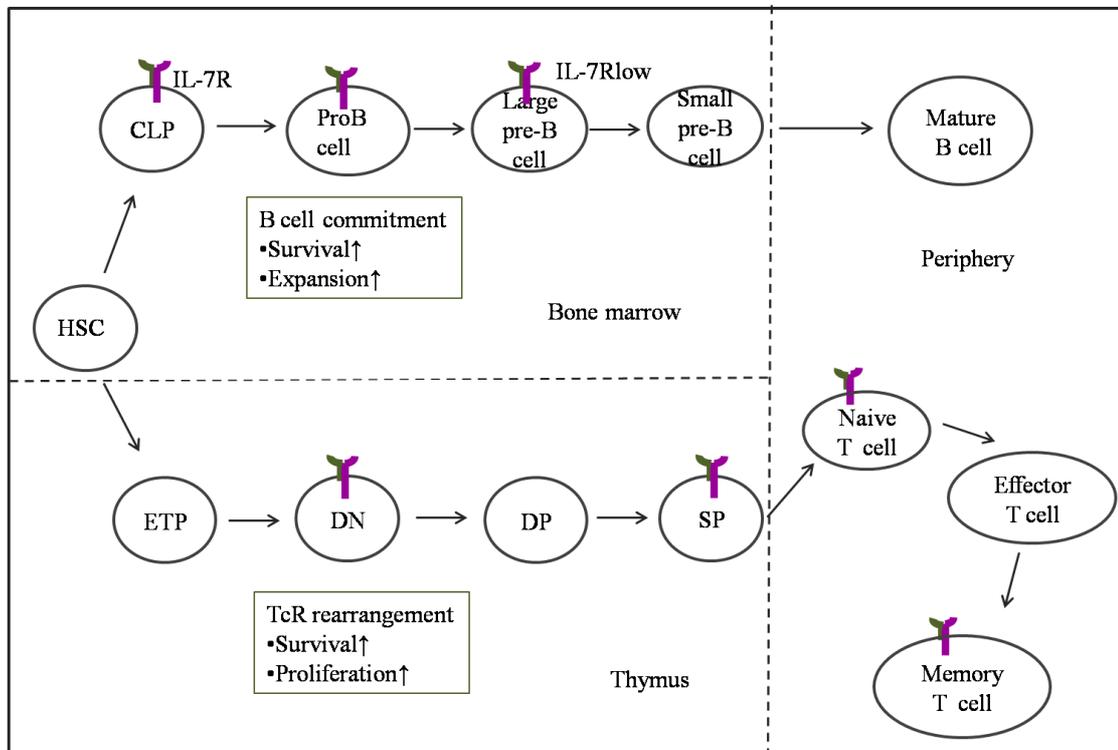


Figure 1.8 IL-7R expression by lymphocytes[58].B lymphocytes of the bone marrow and T lymphocytes of the thymus express IL7R on the cell surface at different stages of development. The expression of IL7R is dynamically regulated during development.

IL-7R expression is tightly regulated in T cell development. It is expressed on double-negative thymocytes, absent on double-positive thymocytes and re-expressed by single positive cells (Figure 1.8). The double negative stage of T cells can be divided into four sub-populations according to the surface expression of CD44 and CD25, known as DN1 through DN4.  $\beta$ -selection of thymocytes occurs at the DN3 stage. Developmental arrest of IL-7R deficient cells at the DN3 stage indicates that IL-7 signaling is essential for survival and proliferation of  $\beta$ -selected cells. In addition, absence of IL-7 signaling can be compensated by overexpression of anti-apoptotic proteins such as Bcl-2 or by loss of pro-apoptotic factors such as Bim or Bax (as reviewed in [59]). IL-7R signaling blocks the differentiation of DN cells to the DP stage by inhibiting expression of the transcription factors TCF-1, LEF-1 and ROR $\gamma$ t [60].

Thus, IL-7R $\alpha$  expression is terminated in thymocytes that reach the double positive (DP) stage. IL-7R $\alpha$  is re-expressed in post DP intermediate cells (CD4<sup>+</sup>CD8<sup>low</sup>) on which CD8 coreceptor transcription is selectively downregulated. Sustained TCR signaling results in differentiation of intermediate cells into CD4 single positive cells. Intermediate cells that no longer receive TCR signals differentiate into CD8 single positive cells as a result of IL-7 signaling (as reviewed in [61]). Thus, IL7R signaling plays critical roles at different stages of T lymphocyte development and lack of signaling or misregulated signaling can cause diseases such as SCID or lymphoma [62, 63].

In the peripheral immune system, IL-7R $\alpha$  is expressed on all naive CD4 and CD8 T cells. Upon antigen stimulation of effector T cells, IL-7R expression is decreased whereas, paradoxically, receptor expression for other cytokines such as IL-2, IL-4 and IL-15 is increased. IL-7R expression is also upregulated in memory cells. IL-7R expression in naive and memory cells is important not only for their survival but also for maintaining a long term homeostatic balance between these peripheral cells [64].

#### **1.2.4 Regulation of IL7R alpha Gene**

The expression profile of IL7R changes during the different developmental stages of both B- and T-lymphocytes. Regulation of IL7R expression at these stages is controlled by different transcription factors that are tightly regulated during development. Various transcription factors that control IL7R $\alpha$  expression at the transcriptional level have been identified. Figure 1.9 shows the IL7R $\alpha$  gene locus and the bioinformatically identified transcription factor binding sites in this locus.

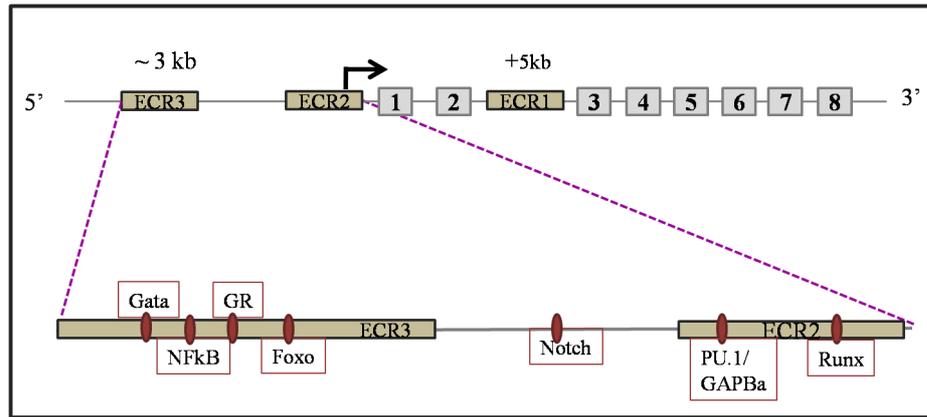


Figure 1.9 IL7R gene loci with different transcription factor binding sites

In the promoter region of the IL-7R gene, a GGAA motif serves as a binding site for PU.1 which is an ETS family transcription factor. It was demonstrated that PU.1 is required for IL-7R expression in developing B cells[65]. Although T cells do not express PU.1, this GGAA motif is occupied by another ETS family transcription factor, GGAA binding protein (GABP) that regulates IL-7R expression. GABP binding to the GGAA motif in the absence of PU.1 can promote IL7R expression in committed B cells, but not in early B cell progenitors[66].

Runx1 is a transcription factor that regulates IL-7R expression and it has a binding site in promoter region of IL-7R gene. Studies on Runx1 deficient mice showed that this transcription factor was necessary for the positive selection and maturation of CD4 single positive thymocytes. It was suggested that the loss of survival signals due to an absence of IL7R $\alpha$  expression in Runx1 deficient mice was the reason behind the reduction in the number of CD4 single positive T lymphocytes[67].

The IL-7R $\alpha$  gene locus has an evolutionarily conserved region (ECR) about 3 kb upstream of the transcription initiation site. This ECR contains binding sites for the GATA, Foxo, glucocorticoid receptor (GR) and NF- $\kappa$ B transcription factors. GATA-3 is a zinc-finger transcription factor important in lymphocyte development. In addition to its role in T lymphocytes, GATA-3 and CD127 were found to be molecular markers for mouse thymic NK-cell development. Loss of CD127 expression on early thymocytes precursors in Gata3 deficient mice suggested that generation of CD127 positive NK cells is GATA-3 dependent[68].

Foxo transcription factors, a subgroup of the Fork head family, have roles in the regulation of apoptosis, cell cycle progression, glucose metabolism and stress resistance. A Foxo binding site in the IL-7R $\alpha$  gene is located about 3.5 kb upstream of

transcription initiation site according to detailed bioinformatics analysis. IL-7R $\alpha$  expression in CD44<sup>lo</sup> CD4<sup>+</sup> and CD8<sup>+</sup> T cells was severely impaired in Foxo1 knockout mice, indicating a direct regulatory effect of Foxo1 on IL-7R $\alpha$  transcription[69]. Other transcription factors binding to enhancer region of IL-7R gene locus, Notch, NF- $\kappa$ B and GR whose binding sites were mutated in this study using TALEN technology, will be explained in detail below.

#### **1.2.4.1 Notch**

The Notch signaling is highly conserved in all metazoans with its roles in the regulation of cell proliferation, differentiation and cell death. The Notch receptor is a transmembrane protein and interacts with transmembrane ligands Delta and Serrate (Jagged in mammals) on neighboring cells. Binding of Notch to its ligand induces two proteolytic cleavages. The first cleavage by ADAM-family metalloproteases separates the extracellular domain of the receptor and the second cleavage driven by a  $\gamma$ -secretase enzyme complex releases the Notch intracellular domain (NICD) from the plasma membrane. NICD translocates to nucleus and interacts with the DNA binding protein CBF1/ RBPjk/ Su(H)/ Lag1 (CSL) and its co-activator Mastermind-like (MAML) to upregulate expression of target genes (Figure 1.10)[70, 71].

Notch signaling is important for T-lineage specification as demonstrated by studies in which induced deletion of Notch1 in hematopoietic progenitors resulted in a reduction in thymus size and a decrease in the number of thymocytes. In fact, the absence of Notch signaling in the thymus drives differentiation of lymphoid cells to the B cell lineage[72].

Overexpression of the intracellular active form of Notch1 in human early thymic precursors (ETPs) in a fetal thymic organ culture (FTOC), upregulated IL7R $\alpha$  expression whereas deficiency of Notch1 signaling resulted in the down-regulation of IL7R $\alpha$  expression and a developmental arrest at the  $\beta$ -selection check point. In addition, a putative RBP-Jk-binding site was identified about 1000bp upstream of transcription initiation site of IL-7R $\alpha$  after chromatin immunoprecipitation and luciferase assays showing IL7R $\alpha$  is a direct transcriptional target of Notch1 [73].

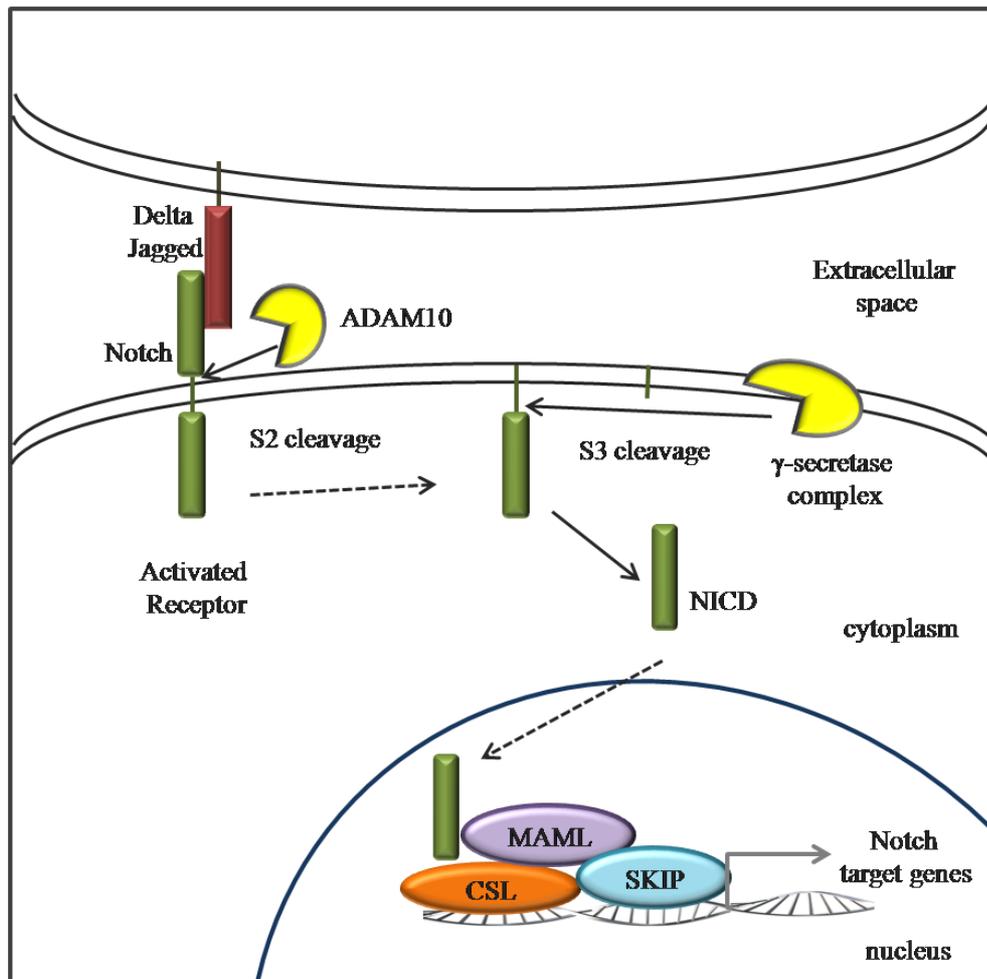


Figure 1.10 Notch signaling [70]. The Notch intracellular domain (NICD) is released from the membrane upon ligand binding induced cleavage of the Notch receptor on the plasma membrane. Cleaved NICD translocates into the nucleus and binds a preexisting CSL (RBP-J $\kappa$ ) transcription factor complex; helps recruit the adaptor protein Mastermind-like (MAML) and results in transcriptional activation.

#### 1.2.4.2 NF- $\kappa$ B

NF- $\kappa$ B is an important regulator in the immune system controlling the expression of numerous genes that are necessary in processes like cell survival, differentiation and proliferation. Tight regulation of the NF- $\kappa$ B pathway is important because its inappropriate activation is associated with different diseases such as cancer, autoimmunity and chronic inflammation. In the resting state, an NF- $\kappa$ B transcription factor composed of a heterodimer of the p50 and p65 proteins is bound to its inhibitor, I $\kappa$ B. The NF $\kappa$ B-I $\kappa$ B complex resides in the cytoplasm because of the shielding of the nuclear localization signal of NF- $\kappa$ B. Various external stimuli can activate NF- $\kappa$ B; three of them are summarized in Figure 1.11. Although various numbers of proteins are involved in each pathway initiated with different stimuli, all of them intersect in

activation of an I $\kappa$ B kinase, which phosphorylates I $\kappa$ B, leading to its ubiquitinylation and subsequent proteosomal degradation. Removal of I $\kappa$ B results in the activation of the NF- $\kappa$ B dimer which translocates to the nucleus and binds to its target sites for gene activation (as reviewed in [74, 75]).

An NF- $\kappa$ B binding site is present in the promoter region of the IL-7R $\alpha$  gene. Whether this site has a functional significance for IL7R gene transcription has not been addressed. A microarray study linking NF- $\kappa$ B signaling to IL7R gene transcription identified IL7R as a TNF-inducible gene[76].

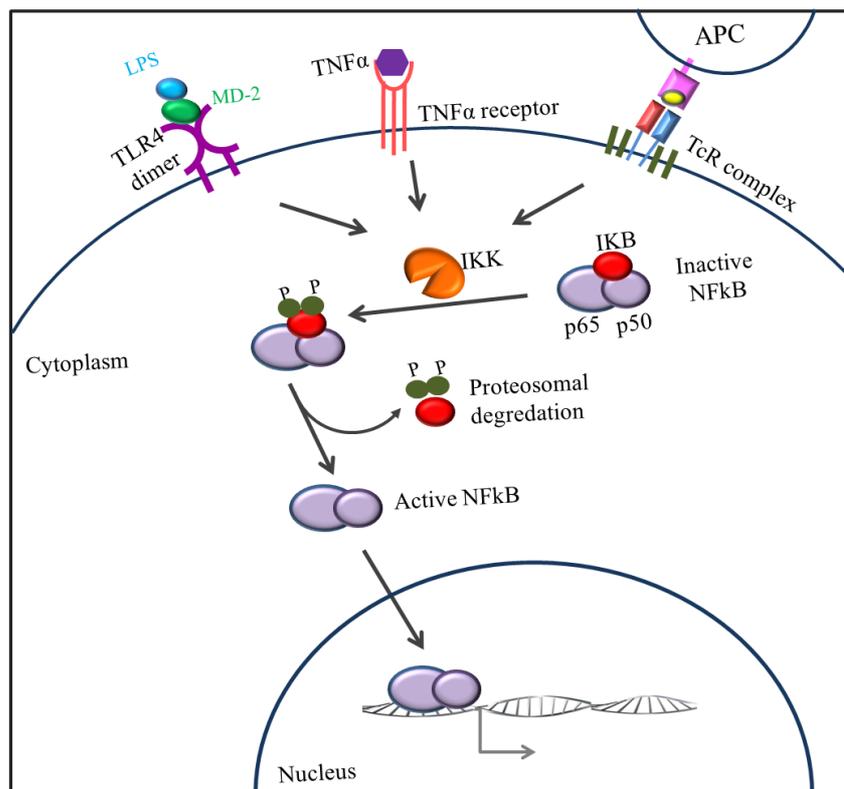


Figure 1.11 NF- $\kappa$ B signaling pathways.

### 1.2.4.3 Glucocorticoid Receptor (GR)

Glucocorticoids are secreted by cells of adrenal cortex as a response to the effects of cytokines released during inflammation such as TNF- $\alpha$  and IL-1 $\beta$ . Glucocorticoids act as anti-inflammatory factors by inhibiting cytokine mediated signaling pathways and inducing apoptosis in certain cells of the immune system [77].

The glucocorticoid receptor, an inactive transcription factor resident in the cytoplasm of unstimulated cells is released from chaperones after binding its ligand, in glucocorticoid signaled cells and translocates to the nucleus. GR bound to its ligand can function by activating or inhibiting the transcription of various genes by binding to

glucocorticoid response elements (GRE) in their promoters (Figure 1.12). Another mechanism for transcription regulation of GR is protein-protein interaction. For instance, GR inhibits NF- $\kappa$ B regulated gene expression by binding to the p65 subunit of NF- $\kappa$ B to block the mechanistic interaction of p65 with the transcription machinery[78].

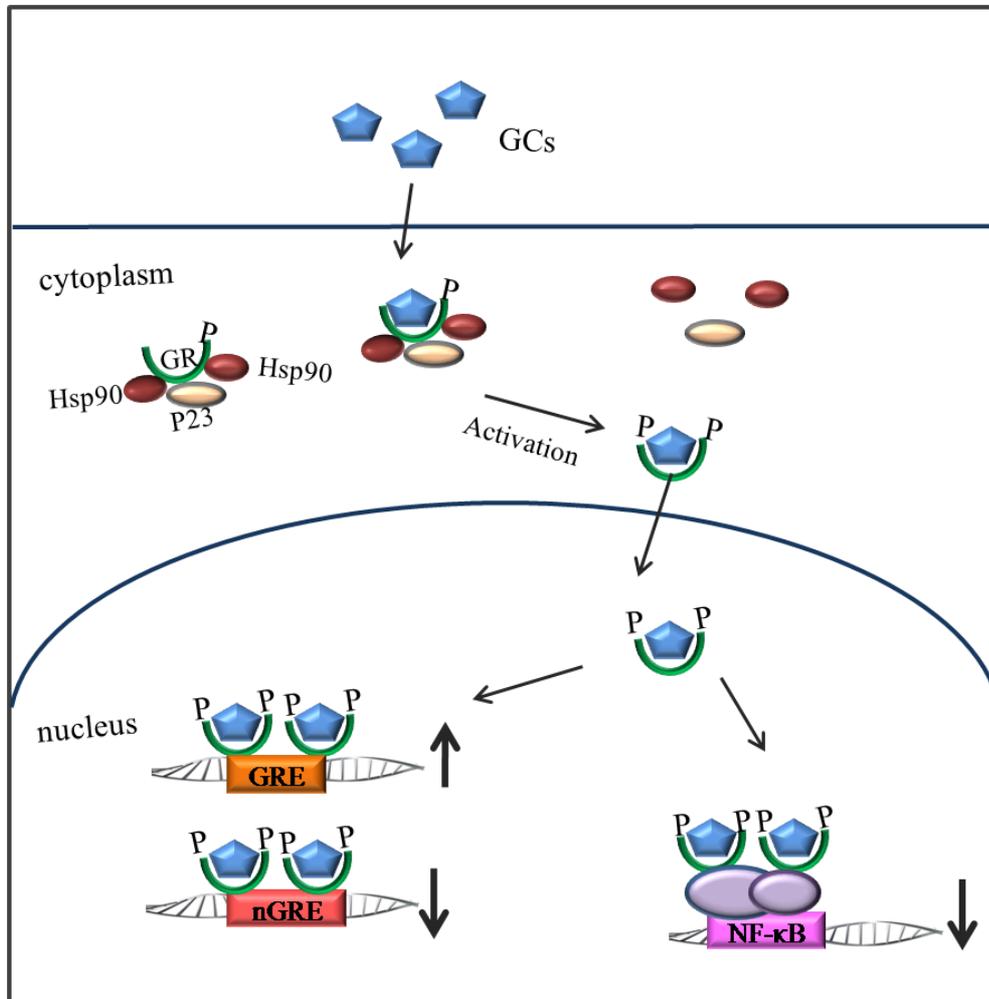


Figure 1.12 Glucocorticoid receptor signaling. In the absence of signaling, cytoplasmic GR (blue pentagon) is bound by the chaperones Hsp90 and p23. In the presence of the hydrophobic glucocorticoid ligand, which freely diffuses into the cell, GR dissociates from its chaperones, is phosphorylated and translocates into the nucleus. GR can activate genes with GRE elements, inhibit genes with nGRE elements and also inhibit NF- $\kappa$ B responsive genes by binding to the p65 subunit[77].

A GRE was identified 3000bp upstream of the IL-7R $\alpha$  transcription initiation site. Recruitment of GR to this GRE activates transcription [79]. In a recent study, it was shown that both IL-7R mRNA and protein expression were upregulated in T cells treated with dexamethasone, a synthetic glucocorticoid [80].

## 2. AIM OF THE STUDY

Transcription activator-like effector (TALE) proteins from the plant pathogen *Xanthomonas* have a central DNA binding domain (DBD) composed of highly conserved repeat units. These repeats are polymorphic at their 12<sup>th</sup> and 13<sup>th</sup> amino acid residues, named as repeat variable di-residues (RVDs) which determine the binding specificity of the DNA binding domain. A simple one RVD to one base code and the modular structure of the DBD allows for the engineering of TALE proteins with novel DNA binding specificity and functionality. TALE nucleases (TALENs) that consist of a non-sequence-specific FokI cleavage domain fused to a target site specific DNA binding domain are used for genome editing. Cellular repair of TALEN induced double stranded breaks (DSBs) occurs either by non-homologous end joining (NHEJ) or by homologous recombination (HR) that results in mutations at target sites.

In the first part of this study, we aimed to mutate transcription factor binding sites of the IL-7R $\alpha$  gene locus using TALEN technology to determine the effects of these mutations on IL-7R expression. We designed and constructed three TALEN pairs targeting binding sites of glucocorticoid receptor (GR), Notch and NF- $\kappa$ B in the IL-7R gene enhancer. We performed a modified restriction fragment length polymorphism (RFLP) assay and DNA sequencing analysis to detect mutations induced at target sites of Neuro-2a and RLM11 cell lines transfected with TALEN constructs. We further analyzed the expression of IL-7R of mutated cell lines. In the second part of the study, we designed, constructed and expressed TALEdsRed fusion proteins binding to an NF- $\kappa$ B binding site in a reporter cell line in order to inhibit the transcription activation function of NF- $\kappa$ B upon TNF- $\alpha$  treatment by competitive binding. The aim of this project in the long term is to generate novel cell lines with specific mutations.

### 3. MATERIALS AND METHODS

#### 3.1 Materials

##### 3.1.1 Chemicals

All the chemicals used in this project are listed in the Appendix A.

##### 3.1.2 Equipment

All the equipment used in this project are listed in the Appendix B.

##### 3.1.3 Buffers and Solutions

Standard buffers and solutions used in this project were prepared according to the protocols in Sambrook et al., 2001.

Calcium Chloride (CaCl<sub>2</sub>) solution: 60 mM CaCl<sub>2</sub>, 15% glycerol and 10mM PIPES at pH 7.00 were mixed and the solution was filter-sterilized and stored at 4°C for competent cell preparation.

5X Tris-Borate-EDTA (TBE) Buffer: 54 g Tris base, 27.5 g Boric acid and 20mL of 0.5 M EDTA at pH 8.00 were dissolved in 1L of dH<sub>2</sub>O and stored at RT.

1% (w/v) Agarose gel: 1 g of agarose was dissolved in 100 mL of 0.5X TBE buffer by heating in a microwave oven. 0.001 % (v/v) of ethidium bromide was added to solution for visualization of nucleic acids.

Phosphate-buffered saline (PBS) : 1 tablet of PBS was dissolved in 200 mL of dH<sub>2</sub>O. The solution was filter-sterilized for use in mammalian cell culture and stored at 4°C.

Polyethylenimine (PEI) (1 $\mu$ g/ $\mu$ L): 50 mg PEI was dissolved in 50 mL dH<sub>2</sub>O that has been heated to ~80°C and cooled to room temperature. After neutralizing to pH 7.00, the solution was filter-sterilized, aliquoted and stored at -20°C.

FACS buffer: 0.5 g Bovine serum albumin (BSA) and 0.5 g sodium azide were dissolved in 500 mL 1X HBSS and stored at 4°C.

### **3.1.4 Growth Media**

#### **3.1.4.1 Bacterial growth media**

Liquid media: 20 g Luria-Broth (LB) was dissolved in 1 L of dH<sub>2</sub>O and autoclaved at 121°C for 15 min. For selection, ampicillin with final concentration of 100  $\mu$ g/mL, kanamycin with final concentration of 50 $\mu$ g/mL, spectinomycin with final concentration of 50 $\mu$ g/mL and tetracycline with final concentration of 10 $\mu$ g/mL were added to liquid medium after autoclave.

Solid media: 35 g LB agar was dissolved in 1 L of dH<sub>2</sub>O and autoclaved at 121°C for 15 min. For selection, antibiotics with previously indicated concentrations were added to autoclaved medium after cooling down to 50°C. Autoclaved and antibiotic added medium was poured onto sterile Petri dishes. Solid agar plates were stored at 4°C.

#### **3.1.4.2 Mammalian cell culture growth media**

Adherent cell lines Phoenix, NIH3T3, Neuro-2A, HEK293 6.1.1 were grown in DMEM cell culture medium that is supplemented with 10% heat inactivated fetal bovine serum (FBS), 2 mM L-glutamine, 100 unit/mL penicillin and 100 unit/mL streptomycin.

Suspension cell line RLM11 were grown in RPMI 1640 cell culture medium that is supplemented with 10% heat inactivated fetal bovine serum (FBS), 2 mM L-Glutamine, 100 unit/mL, 100 unit/mL streptomycin, non-essential amino acids, vitamin and 50  $\mu$ M 2-mercaptoethanol.

Both adherent and suspension cell lines were frozen in fetal bovine serum (FBS) containing DMSO at a final concentration of 10% (v/v). Freezing medium is stored at 4°C.

### **3.1.5 Cell Types**

*E. coli* DH-5 $\alpha$  competent cells were used for bacterial transformation of plasmids.

Neuro-2A mouse neuroblastoma cell line (ATCC: CCL-131) was used for transfection experiments. Phoenix cell line was derived from HEK293T cell line such that constructs for production of gag, pol and envelope protein were placed in that cells and used for virus production. NIH 3T3 is mouse embryonic fibroblast cell line (ATCC: CCL-1658) and used for infection experiments. HEK293 6.1.1 is derived form of human embryonic kidney cell line such that 4 NF- $\kappa$ B binding sites fused with fos promoter and GFP were integrated to the genome.

RLM11, a radiation-induced BALB/c murine CD4 single positive thymoma T cell line, was used for transfection and analysis of IL7R expression level with FACS[81].

### **3.1.6 Commercial Molecular Biology Kits**

- QIAGEN Plasmid Midi Kit, 12145, QIAGEN, Germany
- QIAquick Gel Extraction Kit, 28704, QIAGEN, Germany
- GenElute Mammalian Genomic DNA Miniprep Kit, G1N350, SIGMA, Germany
- GenElute PCR Clean-Up Kit, NA1020, SIGMA, Germany
- CloneJET<sup>TM</sup> PCR Cloning Kit, K1232, Thermo Fisher Scientific,

### **3.1.7 Enzymes**

All enzymes and their corresponding buffers used in this project are from NEB and Fermentas.

### **3.1.8 Vectors and Primers**

Vectors and primers used in this project are listed in Table 3.1 and Table 3.2.

Vector Name	Purpose	Bacterial Resistance
pcDNA-GFP	Transfection efficiency control	Ampicillin
pIRES2eGFP	Cloning	Kanamycin
pMIGII <sub>dsRED</sub>	Cloning	Ampicillin
pcDNA3.1(+)	Cloning	Ampicillin
pHD1-pHD10 pNG1-pNG10 pNN1-pNN10 pNI1-pNI10	Module plasmids for TALE / TALEN construction	Tetracycline
pFUS_A pFUS_B1- pFUSB10	Array plasmids for TALE / TALEN construction	Spectinomycin
pLR-HD pLR-NG pLR-NN pLR-NI	Last repeat plasmids for TALE / TALEN construction	Tetracycline
pCAG-T7-TALEN(Sangamo)	Backbone plasmid for TALEN construction	Ampicillin
pCAG-T7-TALEN(Sangamo)-FokI-KKR-Destination	Backbone plasmid for TALEN construction with mutant FokI	Ampicillin
pCAG-T7-TALEN(Sangamo)-FokI-ELD-Destination	Backbone plasmid for TALEN construction with mutant FokI	Ampicillin
pC-Goldy TALEN	Backbone plasmid for TALEN construction	Ampicillin
pJET1.2/blunt	Cloning of PCR products	Ampicillin

Table 3.1 List of vectors used in this project

Primer Name	Sequence	Purpose
Notch for BamHI	ATAGGATCCATTGAAACCATAACCACCCTC	Notch TALEN target site amplification
Notch rev Bgl2	GCGAGATCTCCCTTCTCTCTAATTCTGTT	Notch TALEN target site amplification
Kpl11 For	CCAAGGAATAAACCCAAGGA	NFKB TALEN target site amplification
Kpl11 Rev	TCTCCAAGCAACAAAAGAA	NFKB TALEN target site amplification
pCR8_F1	TTGATGCCTGGCAGTTCCT	Colony PCR of Golden GATE reaction #1
pCR8_R1	CGAACCGAACAGGCTTATGT	Colony PCR of Golden GATE reaction #1
TAL_F1	TTGGCGTCGGCAAACAGTGG	Colony PCR of Golden GATE reaction #2
TAL_R2	GGCGACGAGGTGGTCGTTGG	Colony PCR of Golden GATE reaction #2
SeqTALEN_ 5-1	CATCGCGCAATGCACTGAC	Sequencing of final TALEN construct
pJET1.2 forward sequencing primer	CRACTCACTATAGGGAGAGCGCC	Colony PCR and sequencing of cloned PCR products
pJET1.2 reverse sequencing primer	TTCTTGTAGCTAAAAGGTACCGTC	Colony PCR and sequencing of cloned PCR products

Table 3.2 List of primers used in this project

### 3.1.9 DNA Molecular Weight Marker

DNA molecular weight marker used in this project is given in Appendix C.

### 3.1.10 DNA sequencing

DNA sequencing was commercially performed by McLab, CA, USA.  
(<http://www.mclab.com/home.php>)

### 3.1.11 Software and Computer Based Programs

The software and computer based programs used in this project

Program Name	Website/ Company	Purpose
CLC Main Workbench 6.1.1	<a href="http://www.clcbio.com/">http://www.clcbio.com/</a>	Primer design, molecular cloning, sequence data management
FlowJo 7.6.5	<a href="http://www.flowjo.com/">http://www.flowjo.com/</a>	FACS data analysis
TAL Effector Nucleotide Targeter 2.0	<a href="https://tale-nt.cac.cornell.edu/">https://tale-nt.cac.cornell.edu/</a>	TALE / TALEN design tool
Quantity One	Bio – Rad	Gel image analysis
Visual Molecular Dynamics (VMD)	<a href="http://www.ks.uiuc.edu/Research/vmd/">http://www.ks.uiuc.edu/Research/vmd/</a>	Crystal structure display and analysis

Table 3.3 List of software and computer based programs used in this study

## 3.2 Methods

### 3.2.1 Bacterial Cell Culture

#### 3.2.1.1 Bacterial culture growth

*E. coli* DH5 $\alpha$  bacterial cells were grown overnight (~16 h) at 37°C shaking at 250 rpm in Luria Broth (LB). Bacterial cells were either spread or streaked on LB Agar plates to obtain single colonies and grown overnight (~16 h) at 37°C. Antibiotics were added to growth media depending on the application.

For long-term storage of bacterial cells, glycerol was added to the overnight grown culture to a final concentration of 15% in 1 mL. Bacterial glycerol stocks were stored at -80°C.

#### 3.2.1.2 Competent cell preparation and transformation

*E. coli* DH5 $\alpha$  competent cells were prepared using stock of previously prepared competent cells. 50 $\mu$ L from previously prepared competent cells were grown in 50 mL LB without selective antibiotic overnight at 37°C shaking at 250 rpm. Next day, 4 mL from the overnight culture was diluted within 400 mL LB and incubated under same growth conditions until the OD<sub>590</sub> reaches to 0.375. Then, previously prepared ice-cold CaCl<sub>2</sub> solution was used for resuspension of bacterial cell pellet after successive centrifugation steps and for final preparation. 200 $\mu$ L aliquots of competent cells prepared were frozen immediately in liquid nitrogen and then stored at -80°C. Competency of prepared cells was tested by transforming varying concentrations of pUC19 plasmid.

For transformation of competent cells, CaCl<sub>2</sub> treated chemically competent bacterial cells were taken from -80°C and ~100 pg of plasmid DNA was added before cells were completely thawed. After 30 min of incubation on ice, the cells were heat-shocked at 42°C for 90 seconds and transferred back to ice rapidly to chill for 60 seconds. 800 $\mu$ L of sterile LB without antibiotics added and cultures were incubated for 45 minutes at 37°C for recovery of cells and expression of antibiotic resistance gene encoded by the plasmid. Transformed cells were spread onto LB agar plates containing appropriate antibiotic for selection using sterile glass beads. Then, the plates were incubated overnight at 37°C.

### **3.2.1.3 Plasmid DNA isolation**

Plasmid DNA isolation was performed either by the alkaline lysis protocol or by QIAGEN Plasmid Midi Kits. For plasmid isolation, either single colony of *E.coli* from LB agar plates or glycerol stock was grown overnight at 37°C shaking at 250 rpm in liquid medium containing selective antibiotics with appropriate concentrations. The concentration and purity of kit-isolated plasmid DNA were determined using Nanodrop.

### **3.2.2 Vector Construction**

Restriction Enzyme Digestion: Digestion reactions containing template DNA, enzyme and its compatible buffer were incubated at the optimum temperature of the enzyme used for 2 hours. ~ 300 ng of template DNA was used for diagnostic digestions whereas the amount of template DNA used for gel extraction and cloning purposes was at least 1 µg.

Agarose Gel Electrophoresis and Gel Extraction: Agarose gels to observe DNA samples and digestion products were prepared in varying concentrations from 0.8% to 2% depending on the size of DNA fragments to be separated. Agarose gel was prepared by dissolving appropriate amount of agarose powder in 0.5X TBE, heating for 3-5 min in a microwave. After cooling-down of the solution to room temperature, ethidium bromide was added at a final concentration of 0.001% (v/v) and the gel was poured onto the gel apparatus for solidification. 0.5X TBE was also used as running buffer. DNA samples were mixed with 6X DNA loading dye before loading to the gel. Electrophoresis was performed at 100-135 V for 45-75 minutes and the bands were observed under UV light. Gel extraction of DNA samples was performed by QIAGEN Gel Extraction Kit.

Dephosphorylation of Vector Ends: 5' phosphate groups of linearized vector DNA were dephosphorylated using Calf Intestinal Alkaline Phosphatase (CIAP) prior to insert ligation, to prevent vector re-ligation.

Ligation: Ligation was performed using T4 DNA Ligase (Fermentas), in 1:3 and 1:6 molar vector to insert ratio using 100ng vector. In addition, ligation reaction mixture without insert was prepared as negative control for each ligation. The ligation reaction was incubated at 16°C for 16 hours in a final volume of 20µL. Then, half of the ligation mixture was transformed into chemically competent bacteria.

### 3.2.3 Construction of TALE Expression Plasmids

#### 3.2.3.1 Identification of TALE and TALEN target sites

The software used for design of transcription activator like effectors (TALE) and transcription activator like effector nucleases (TALEN) is available for use as an online tool (TAL Effector Targeter and TALEN Targeter (old version with design guidelines), TALE-NT; <https://tale-nt.cac.cornell.edu/>). DNA sequence entered is scanned for potential TALEN recognition sites based on either preset design guidelines defined by four different articles or user-provided spacer and RVD lengths. The software gives coordinates and sequences of recognition sites for right and left TALEN monomers and the spacer sequence. In addition, RVD sequences necessary for construction of custom TALENs were also provided as software output. Binding sites of TALEN pair and spacer sequences are given in Table 3.4

	Left TALEN binding sequence	Spacer	Right TALEN binding sequence
GR TALEN	AGCACATGCTGTACCAAA AT	ATTATGTCTTAACTTAA CTTTGTTCTTTTAC	ATCTTCACAACCTAAA GG
Notch TALEN	AGGGTCACCCTCATA	GACTCCTGGGAGTTTTTC	ATTGCCCTTGTTTCT
NFKB TALEN	TGCCCCACCCAAAAGGGG TAA	GCACACCAGTGGAAAT CCCCTGAG	CAAACCTAGCACATGC TGTA

Table 3.4 Binding sites of TALEN pair and spacer sequences

#### 3.2.3.2 Assembly of custom TAL Effector and TALEN constructs using Golden Gate TALEN kit

TAL effector DNA binding domain is composed of tandem repeat modules. 12<sup>th</sup> and 13<sup>th</sup> amino acids within each repeat module, called repeat-variable di-residues (RVDs), are responsible for nucleotide recognition. NI, NN, NG and HD are the four most common RVDs, each preferentially bind to nucleotides A, G, T, and C, respectively. Design of custom TALE and TALENs were performed using TALEN Golden Gate Kit, which was obtained from Addgene. The Golden Gate TALEN kit was reported by Cermak et al (2011) and contains a set of module plasmids with each

individual RVDs, array plasmids for intermediate cloning and backbone expression plasmids to make final TALEN expression constructs[5].

The custom TALEN or TAL effector construct is assembled by using successive rounds of Golden Gate cloning, in which digestion by Type IIS restriction endonucleases such as BsaI and Esp3I is performed to create unique 4 bp overhangs on DNA fragments. These unique overhangs flanking each RVD were designed such that up to 10 RVD-encoding repeat module plasmids can be ligated in a single reaction. Assembly of repeat modules into array plasmids is followed by assembly of array plasmids into final expression vectors (Figure 3.1). Construction of TAL effector or TALEN construct was achieved in 5 days (Figure 3.2)

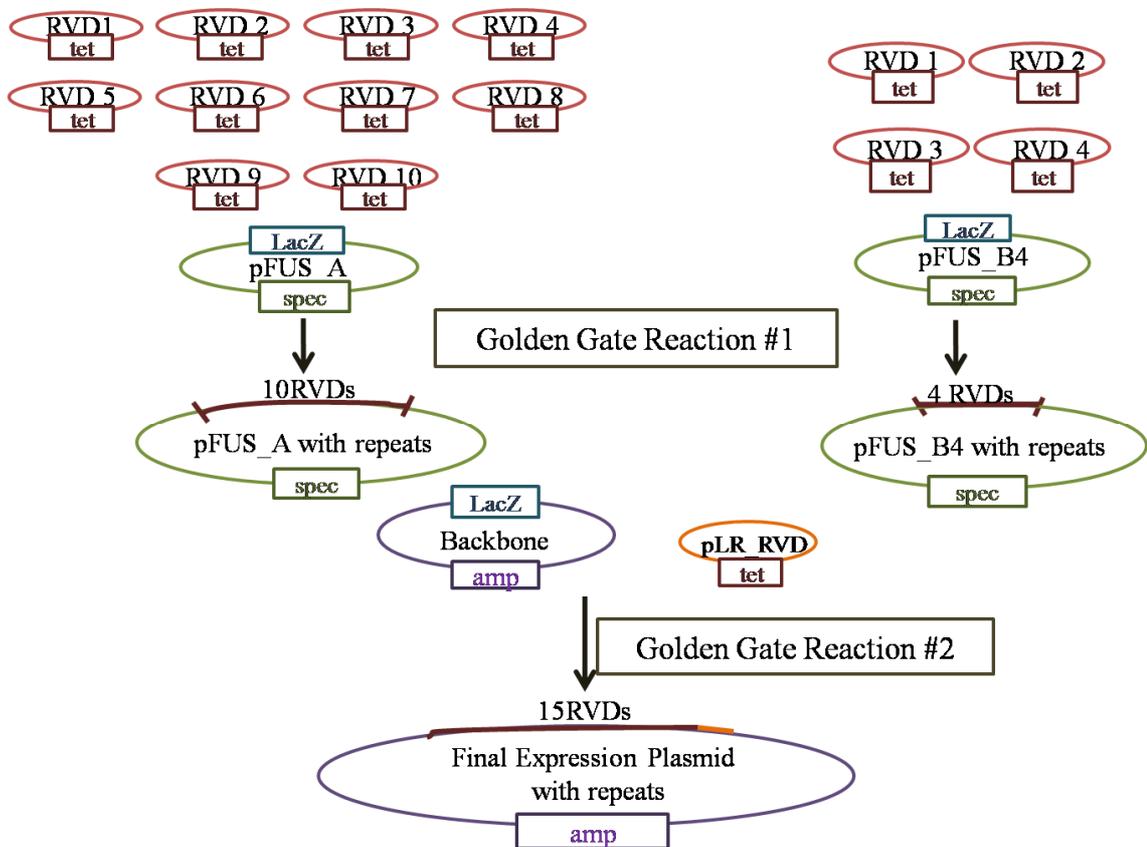


Figure 3.1 Golden Gate assembly of custom TALE and TALEN constructs

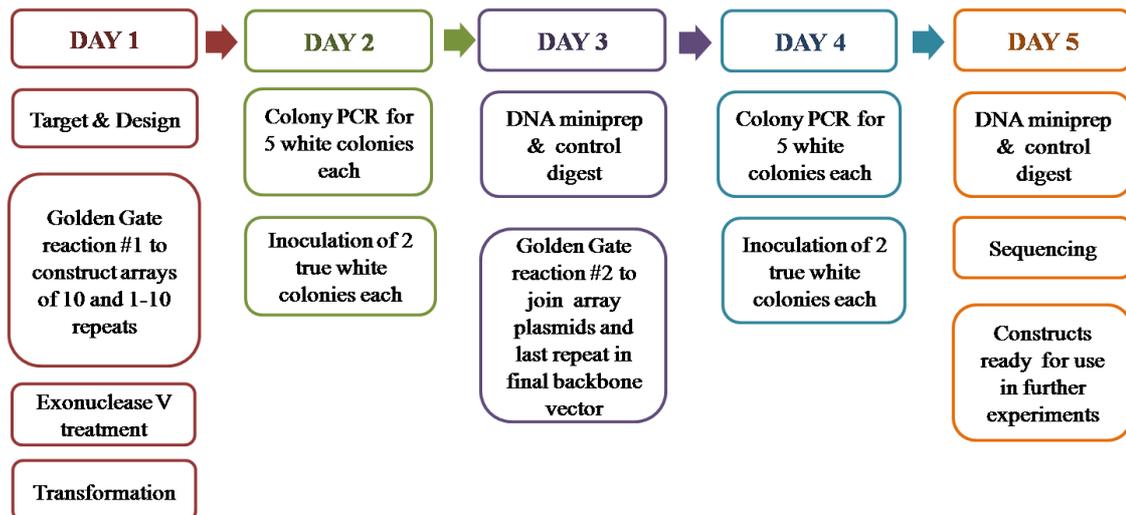


Figure 3.2 Timeline for TAL effector and TALEN construction using TALEN Golden Gate kit [5].

Day 1: After identification of possible TALEN target sites and determination of the RVD sequence, two separate array plasmids were assembled for “n” RVD repeat module containing TALEN expression plasmids. First 10 module plasmids selected according to the order of RVD sequence were cloned into array plasmid pFUS\_A. Then, modules were selected for remaining RVDs, 11 – (n-1), and cloned into the array plasmid pFUS\_B#n-1. RVD encoding modules for second array were selected starting with plasmid #1. Last RVD (#n) was not included in this reaction as it was provided by a different, “last repeat” plasmid and included in the second step of Golden Gate cloning.

Golden gate reaction #1 was set according to Table 3.5 for each intermediary array plasmid, called as reaction A for first array plasmid and reaction B for the second one. For example a TALEN encoding plasmid with 17 repeats was generated by cloning 10 repeats into the pFUS-A plasmid and 6 repeats into the pFUS-B6 plasmid. The contents of these plasmids were transferred in later days of the procedure into the pC-Goldy TALEN destination expression plasmid along with the contents of LR plasmids, in a four plasmid reaction.

Components	Used amount
Each of module vectors	150 ng
pFUS vector	150 ng
BsaI (NEB)	1 $\mu$ L
BSA (2 mg/ ml)	1 $\mu$ L
T4 DNA ligase (NEB)	1 $\mu$ L
10X DNA ligase buffer	2 $\mu$ L
dH <sub>2</sub> O	Up to 20 $\mu$ L
Total	20 $\mu$ L

Table 3.5 Components and amounts for Golden Gate reaction #1

Reactions were incubated in a thermo cycler for following cycle:

10 X (37°C/5 min + 16°C/10 min) + 50°C/5 min + 80°C/5min

In order to degrade unligated linear dsDNA fragments of incomplete ligation products, and linearized vectors, 1 $\mu$ L of exonuclease V (RecBCD) (NEB) and 1 $\mu$ L of 10mM ATP were added to reaction and incubated at 37°C for 1 hour.

Chemically competent DH5 $\alpha$  *E.coli* cells were transformed with 2  $\mu$ L of the reaction and plated on LB agar containing 50 $\mu$ g/mL spectinomycin, with X-gal and IPTG for blue/white screening of colonies.

Day 2: Correct assembly of TALEN RVD repeat modules into intermediary arrays was controlled first by performing colony PCR with 5 white colonies picked from each plate. A PCR master mix was prepared according to colony PCR conditions shown in Table 3.6 using pCR8\_F1 and pCR8\_R1 as forward and reverse primers, respectively, individual colonies were resuspended in this solution.

Component	Volume
10X standard Taq buffer(Mg free)	2.5 $\mu$ L
25 mM MgCl <sub>2</sub>	2 $\mu$ L
10mM dNTP each	0.5 $\mu$ L
Forward primer	0.2 $\mu$ L
Reverse primer	0.2 $\mu$ L
Taq polymerase (5U/ $\mu$ l)	0.125 $\mu$ L
dH <sub>2</sub> O	19.475 $\mu$ L
Total	25 $\mu$ L

Table 3.6 Optimized colony PCR conditions

PCR was performed according to following cycle;

95°C/4 min + 30X (95°C/30 s + 55°C/30 s + 72°C/135 s) + 72°C/10 min

Depending on the colony PCR results, two correct clones were inoculated into 3mL LB containing 50 $\mu$ g/mL spectinomycin and incubated overnight at 37°C shaking at 200 rpm.

Day 3: Plasmid DNA was isolated from overnight cultures of pFUS\_A and pFUS\_B plasmids containing repeats. Correct assembly of array was controlled by restriction enzyme digestion with AflIII and XbaI and agarose gel electrophoresis. Double digestion with these enzymes releases the repeat arrays and size of fragments produced was 1048 bp for pFUS\_A containing 10 RVDs whereas size of fragments varied for pFUS\_B plasmids.

Correctly assembled intermediary arrays and sequence encoding the n<sup>th</sup> repeat were assembled into the final expression backbone vector. TALEN Golden Gate Kit contains 4 different backbone plasmids, pTAL1 and pTAL2 for TAL effector design whereas pTAL3 and pTAL4 for TALEN construction and expression of designed TALEN in yeast. For expression of TALENs in mammalian cells after assembly, different final expression plasmids obtained from Addgene which are pCAG-T7-TALEN (Sangamo), pCAG-T7-TALEN (Sangamo)-FokI-KKR-Destination, pCAG-T7-

TALEN (Sangamo)-FokI-ELD-Destination and pC-Goldy TALEN. For Golden Gate reaction# 2, digestion and ligation were performed in 2 steps due to BsmBI restriction enzyme working at 55°C which inhibited the activity of T4 DNA ligase. The first part of the reaction was set according to Table 3.7.

Components	Used amount
Reaction A	150 ng
Reaction B	150 ng
pLR vector	150ng
Expression backbone vector	75ng
NEB Buffer 4 (10X)	1.5 µL
BsmBI (NEB)	0.5 µL
dH <sub>2</sub> O	Up to 15µL
Total	15µL

Table 3.7 Components for the first part of Golden Gate reaction #2

After incubation of the first part of the reaction at 55°C for 10 minutes, the second part of the reaction was set according to Table 3.8

Components	Used amount
ATP (10 mM)	2 µL
NEB Buffer 4 (10X)	0.5 µL
T4 DNA ligase	1 µL
DTT (0.2M)	1µL
Water	0.5 µL
Total	20 µL

Table 3.8 Components for second part of Golden Gate reaction #2.

Reactions were incubated in a thermo-cycler using the following cycle:  
16°C /15min + 55°C / 15 min + 80°C / 5 min

Chemically competent DH5 $\alpha$  *E.coli* cells were transformed with 2  $\mu$ L of the reaction and plated on LB agar containing 100  $\mu$ g/mL ampicillin, with X-gal and IPTG for blue/white screening of colonies.

Day 4: Colony PCR was performed to check the assembly of intermediary arrays into the final expression plasmid and 5 white colonies were picked from the plate. The colony PCR mix was prepared according to Table 3, using TAL\_F1 and TAL\_R2 as forward and reverse primers. After resuspending individual colonies in a reaction mixture, colony PCR was performed according to the following cycle;

95°C/4 min + 30X (95°C/30 s + 55°C/30 s + 72°C/3 min) + 72°C/10 min

Depending on colony PCR results, two correct clones were inoculated into 3mL LB containing 100 $\mu$ g/mL ampicillin and incubated overnight at 37°C shaking at 200 rpm.

Day 5: Plasmid DNA was isolated from overnight cultures and correct assembly of the final full-length repeat array was verified by restriction enzyme digestion with AatII and StuI and agarose gel electrophoresis. In addition, BspEI control digest, which cut only in HD modules of 2-10, was performed to determine final array integrity. DNA midpreps were prepared from correctly assembled TALEN plasmids, and sequenced using the SeqTALEN 5-1 and TAL\_R2 primers.

To express TALEN pairs in mammalian cells, either TALEN constructs will be cloned to a plasmid containing mammalian expression promoter (Strategy 1) or the destination plasmid selected in Golden Gate reaction #2 would contain promoter for mammalian expression (Strategy 2). GR TALEN monomers that were commercially synthesized were cloned first to pMIGII backbone plasmids containing fluorescent reporter both to track expression of TALEN pair and to produce virus (Strategy 1a). GR TALEN pair was cloned to CMV promoter containing plasmids together with fluorescent reporters to express these plasmids ectopically in mammalian cells (Strategy 1b). Two different TALEN pairs were designed using Golden Gate TALEN kit, TALEN pairs targeting Notch binding site and NF- $\kappa$ B binding site. After assembling repeat monomers in array plasmids of pFUS\_A and pFUS\_B, as destination vector either pCAGT7 (Strategy 2a) or pC-Goldy backbone were selected (Strategy 2c). In

addition, TALEN repeats constructed in pCAGT7 were cloned to PCAGT7-FokI-ELD or pCAGT7-FokI-KKR mutants (Strategy 2b). Figure 3.3 summarizes strategies of constructing TALEN pairs in different mammalian expression plasmid.

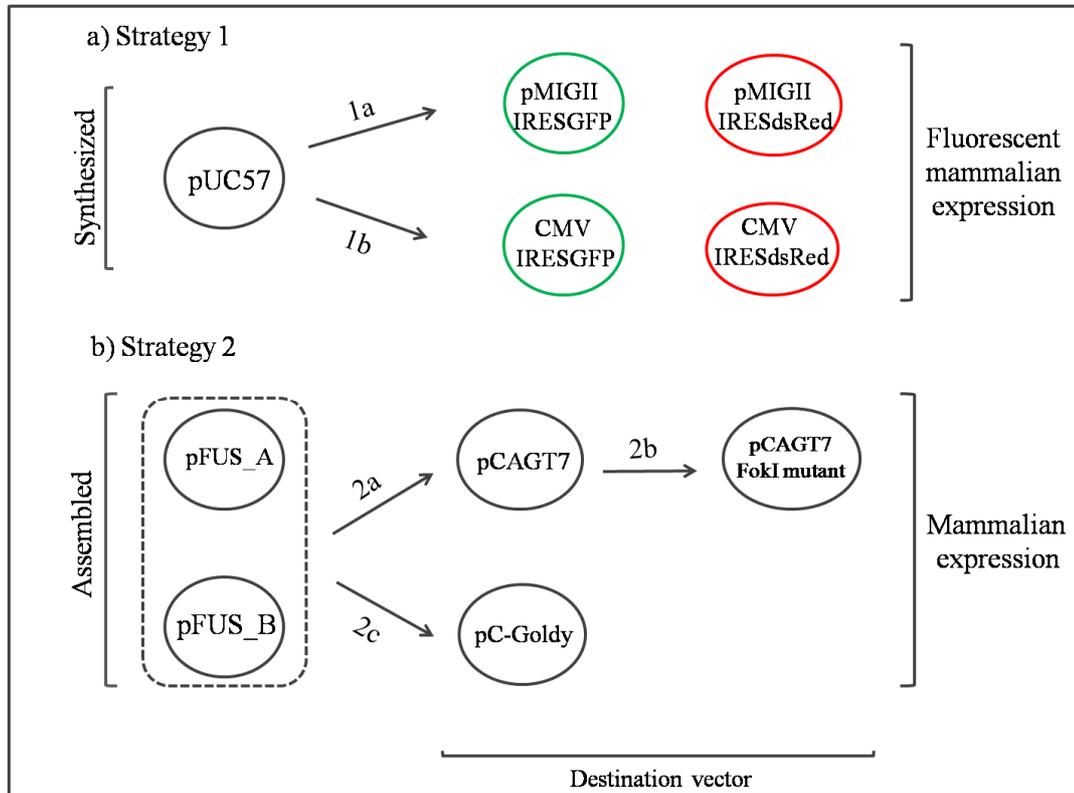


Figure 3.3 Strategies for construction of TALENs in mammalian expression plasmids

### 3.2.4 Mammalian Cell Culture

#### 3.2.4.1 Maintenance of mammalian cell lines

Phoenix, NIH-3T3, Neuro-2A and HEK293 6.1.1 were adherent cell lines used in this project and were grown in DMEM cell culture medium that is supplemented with 10% heat inactivated fetal bovine serum (FBS), 2 mM L-glutamine, 100 unit/mL penicillin and 100 unit/mL streptomycin in 10 cm tissue culture dishes. RLM11 was the only suspension cell line used in this project and was grown in RPMI 1640 cell culture medium that is supplemented with 10% heat inactivated fetal bovine serum (FBS), 2 mM L-Glutamine, 100 unit/mL, 100 unit/mL streptomycin, non-essential amino acids, vitamin and 50  $\mu$ M 2-mercaptoethanol in tissue culture flasks. All cultures were maintained in a humidified incubator supplied with 5% CO<sub>2</sub> at 37°C and split into fresh medium when they reach to ~80% confluency. TALEN transfected cells were incubated in a humidified incubator supplied with 5% CO<sub>2</sub> at 32°C after transfection for 72 hours.

For preparation of frozen stocks of both adherent and suspension cell lines, cells at exponential growth phase were resuspended in ice-cold freezing medium. They were stored at -80°C for 24-48 hours and then transferred to liquid nitrogen tank for long-term storage. After thawing, cells were immediately washed with growth medium to remove any residual DMSO.

#### **3.2.4.2 Transient transfection of adherent cells with PEI**

Transient transfection of adherent cell lines was achieved using polyethylenimine (PEI). PEI is a cationic polymer, which forms complex with negatively charged DNA and bind to cell surface. DNA is taken into the cell via endosomal vesicles and osmotic swelling release plasmid DNA to the cytoplasm [ref. Pei]. One day before transfection,  $4.0 \times 10^6$  adherent cells were split onto 10 cm tissue culture dishes. On the day of transfection, 7 µg of total DNA was diluted in 1 mL serum-free DMEM without phenol red in a sterile tube. PEI (1µg/µL) was added to diluted DNA based on 3:1 ratio of PEI (µg) to total plasmid DNA (µg) and mixed immediately by vortexing. After 15 minutes of incubation at room temperature, DNA/PEI mixture was added drop by drop on cells in tissue culture dishes. Viral supernatant from Phoenix cells was harvested twice as 24 hour and 48 hour after transfection. Viral supernatants were mixed, passed through 45-µm filters, aliquoted in 1 mL and stored at -80°C. Neuro-2A cells transfected with TALENs harvested 72 hours after transfection for FACS analysis. HEK293 6.1.1 cells transfected with TALE-dsRed constructs were treated with TNF-α approximately 38 hours after transfection and harvested 9 hours after that.

#### **3.2.4.3 Transient transfection of suspension cells**

Transient transfection of suspension cell line, RLM11, was done by using Neon electroporation system (Invitrogen). One day before transfection, cells were split 1:10 ratio.  $10^7$  cells were washed twice with filter sterilized 1X PBS. After removal of supernatant, 10 µg DNA was added onto pellet and cells were resuspended in 100µL of HBS. Mixture was taken into 100µL Neon golden tips and placed in electroporation cuvette. Optimum transfection condition for delivery of DNA into cells is 1500 V with a single pulse in 20 milliseconds. Then, cells were transferred to tissue culture flasks containing pre-warmed RPMI supplemented with 10% FBS, 2 mM L-Glutamine, 100 unit/mL, 100 unit/mL streptomycin, non-essential amino acids, vitamin and 50 µM 2-

mercaptoethanol. TALEN transfected RLM11 cells were incubated at 32°C-incubator supplied with 5% CO<sub>2</sub> for 72 hours and harvested for further TALEN genotyping assays.

#### **3.2.4.4 Infection**

5 x 10<sup>5</sup> NIH3T3 cells were split for each well of 6-well plate one day before infection. On the day of infection, viral supernatant was thawed and polybrene was added to a final concentration of 6 µg/ mL. After removing the supernatant of NIH3T3 cells, viral supernatant containing polybrene was added onto cells and incubated for 2 hours at 37°C. Then, medium was replaced with DMEM supplemented with 10% FBS, 2 mM L-glutamine, 100 unit/mL penicillin and 100 unit/mL streptomycin. Virus treatment protocol was repeated in the next day. Cells were harvested for FACS analysis 48 hour after first virus treatment.

#### **3.2.4.5 Flow cytometric analysis**

10<sup>6</sup> cells were used for each flow cytometric analysis. Flow cytometric analysis of cells was performed using BD FACSCanto. For analysis of cells expressing fluorescent proteins GFP and dsRed, cells were washed twice with FACS buffer and resuspended in 500µL of FACS buffer for analysis. GFP and dsRed expression levels were detected with FITC and PE channels, respectively. For analysis of IL7R expression, cells were washed twice with FACS buffer and incubated with CD127-Biotin antibody against mouse IL7Rα –Biotin at 4°C for 30 min in dark. After washing twice with FACS buffer to remove unbound antibody, cells were incubated with SA-Alexa 647 at 4°C for 30 min in dark. Cells were washed twice with FACS buffer and resuspended 500µL of FACS buffer for analysis. IL7R expression levels were detected with Alexa-647 channel.

#### **3.2.5 TALEN Induced Mutation Screening**

General strategy for detection of mutation at TALEN target site is given in Figure 3.4. In this study, mutation in TALEN target site was detected by loss of restriction enzyme cut site, restriction fragment length polymorphism (RFLP) assay.

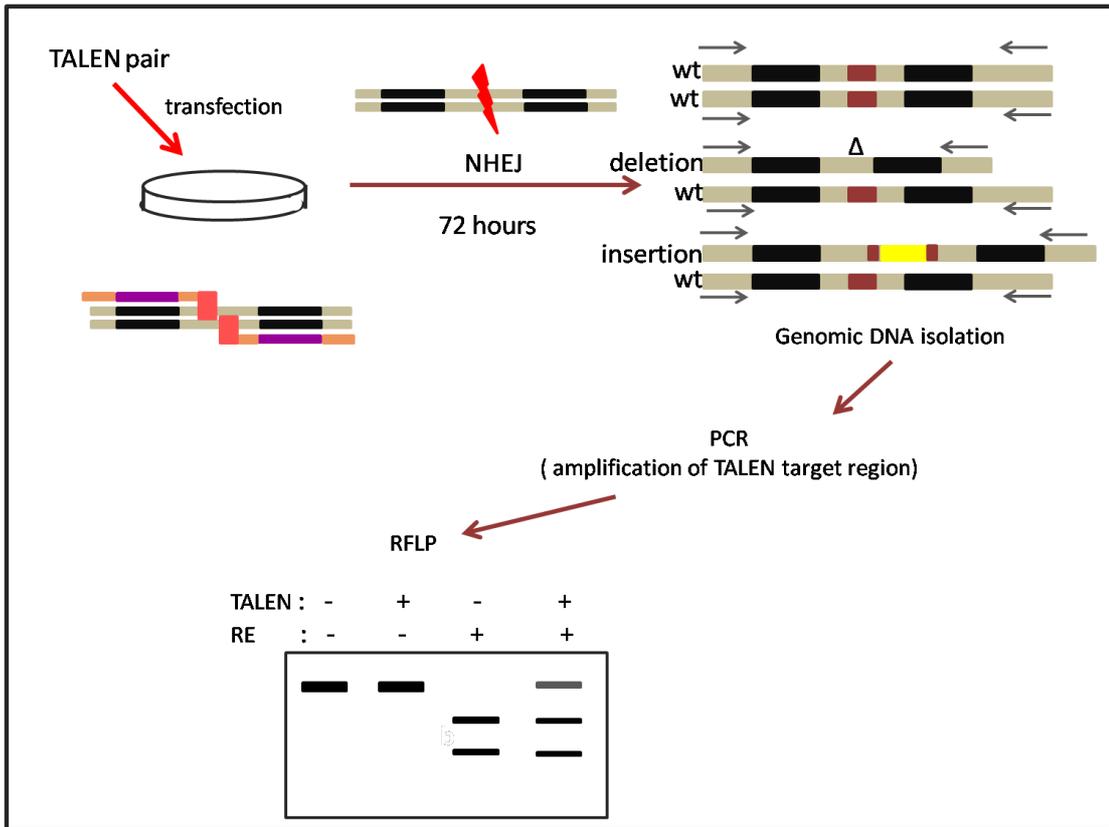


Figure 3.4 General strategy for detection of TALEN induced mutation at target site.

In order to improve detection efficiency, RFLP assay was modified, which is summarized in Figure 3.5.

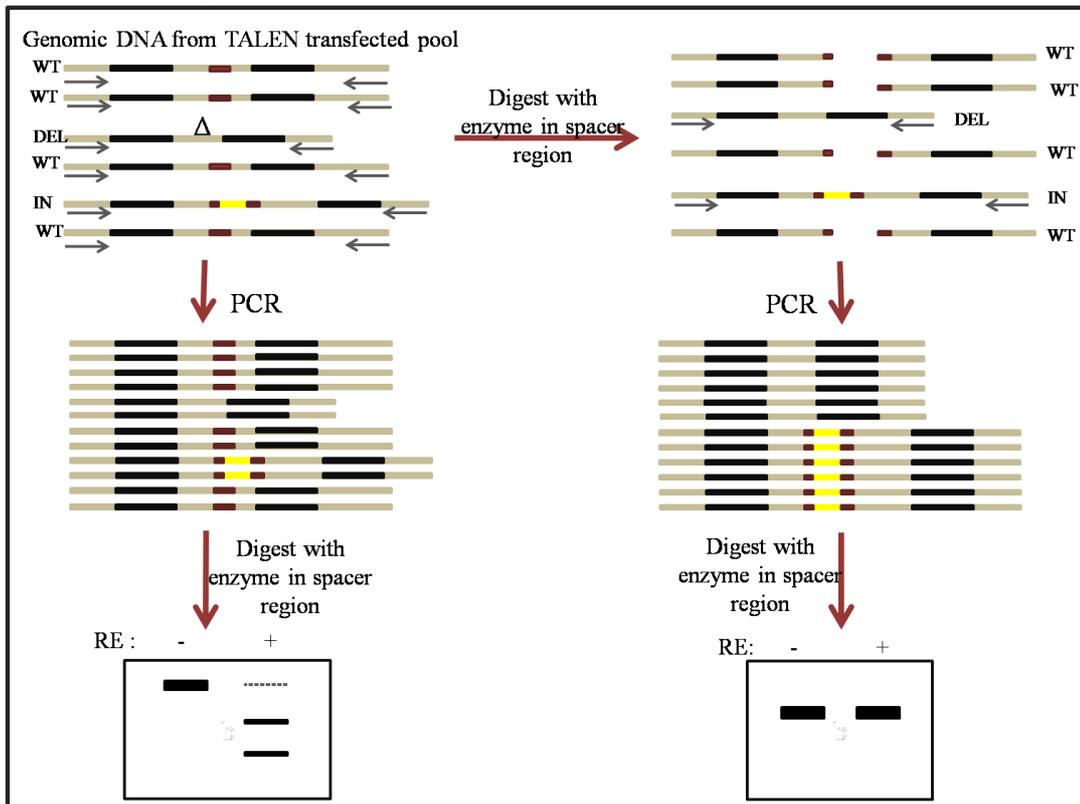


Figure 3.5 Modified RFLP assay to increase mutation detection efficiency

### 3.2.5.1 Genomic DNA extraction

72 hours after transfection with TALENs, genomic DNA of the cells was isolated by using GenElute Mammalian Genomic DNA Miniprep Kit (SIGMA).

### 3.2.5.2 Restriction Fragment Length Polymorphism (RFLP) analysis

Isolated genomic DNA was digested with enzyme in the spacer region, BsaJI and BstNI for TALENs targeting Notch binding site, and BsrI for TALENs targeting NFkB binding site.

PCR reaction was performed using digested genomic DNA according to optimized PCR conditions given in Table 3.9 using primers for TALEN target site amplification.

Component	Volume
Template genomic DNA	1.0 $\mu$ L
5X Phusion high fidelity buffer	4.0 $\mu$ L
10mM dNTP each	0.4 $\mu$ L
Forward primer (10mM)	1.0 $\mu$ L
Reverse primer (10mM)	1.0 $\mu$ L
Phusion Hot Start II DNA polymerase (2U/ $\mu$ l)	0.2 $\mu$ L
dH <sub>2</sub> O	12.4 $\mu$ L
Total	20 $\mu$ L

Table 3.9 Optimized PCR conditions for TALEN target site amplification

PCR was performed according to following cycle;

98°C/4 min + 30X (98°C/30 s + 64°C/30 s + 72°C/60 s) + 72°C/10 min

PCR products were digested once again with the enzyme in spacer region and run on agarose gel to detect whether any undigested band is left indicating presence of mutation at TALEN target site.

Gel extraction was performed for the band in size of the undigested PCR product, and cloned using CloneJET™ PCR Cloning Kit (Thermo Scientific). Uncut bands were cloned to pJET1.2/blunt vector and colony PCR was performed to ensure presence of insert in selected colonies according to conditions provided by the kit. Plasmid DNA was isolated from 3-ml overnight cultures of true colonies and sequenced to evaluate mutations generated at cleavage site.

## **4. RESULTS**

### **4.1. Use of TALENs to Mutate Transcription Factor Binding Sites of the IL7R Gene**

In this study, we generated three pairs of TALENs to mutate transcription factor binding sites of the IL7R enhancer region (Figure 4.1). The first of these TALEN pairs targeted the GR binding site in the 3<sup>rd</sup> ECR of the IL7R gene were commercially designed and synthesized. The second and third TALEN pairs, targeting the Notch binding site and the NF- $\kappa$ B binding sites were assembled using the Golden Gate TALEN construction procedure explained in the Methods section. We hypothesized that IL7R expression would change because of a mutation in the binding sites of the transcription factors critical for its regulation. Thus, if GR, Notch or NF- $\kappa$ B plays a positive role in IL7R expression by binding to their respective sites in ECR3, cells mutated in these binding sites would decrease their levels of IL7R expression.

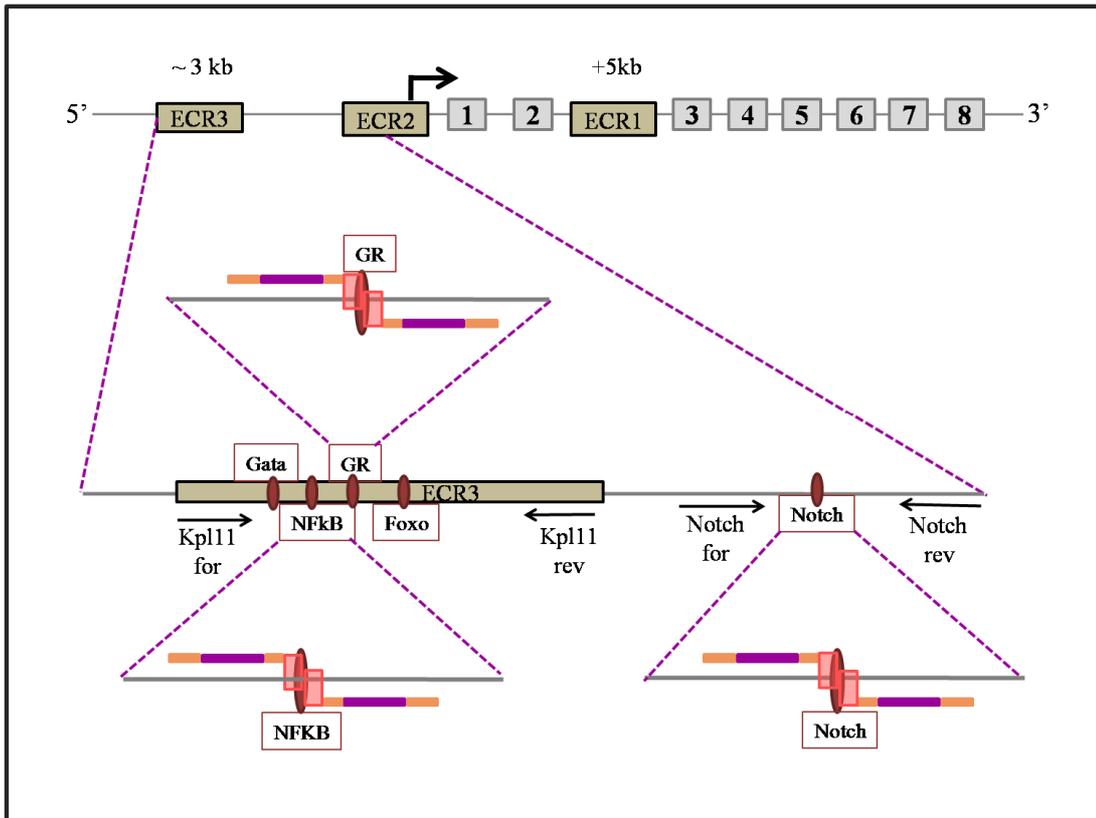


Figure 4.1 Schematic representation of the mouse *IL7R $\alpha$*  gene locus. *IL7R $\alpha$*  gene contains eight exons and three enhancer regions (ECR). ECR3 has binding sites for Gata, NF $\kappa$ B, GR and Foxo. Notch binds upstream region of ECR2. Three different TALEN pairs targeting the binding sites of the transcription factors GR, NF $\kappa$ B and Notch were designed in this study. Arrows indicate binding sites for primer pairs (Kpl11for-Kpl11rev and Notchfor-Notchrev) used for target site amplification.

#### 4.1.1 Commercially Designed GR TALEN Pair

A TALEN pair targeting the glucocorticoid receptor (GR) binding site of the *IL7R* enhancer region was commercially designed, synthesized and cloned in the pUC57 plasmid by Genescript Inc [<http://www.genscript.com/>]. This TALEN pair was named GR TALEN, meaning TALEN targeting GR binding site. Figure 4.2 shows the binding sites of the commercially designed TALENs in the *IL7R* enhancer region together with the binding sites of various different transcription factors and primer binding sites used to analyze the region.

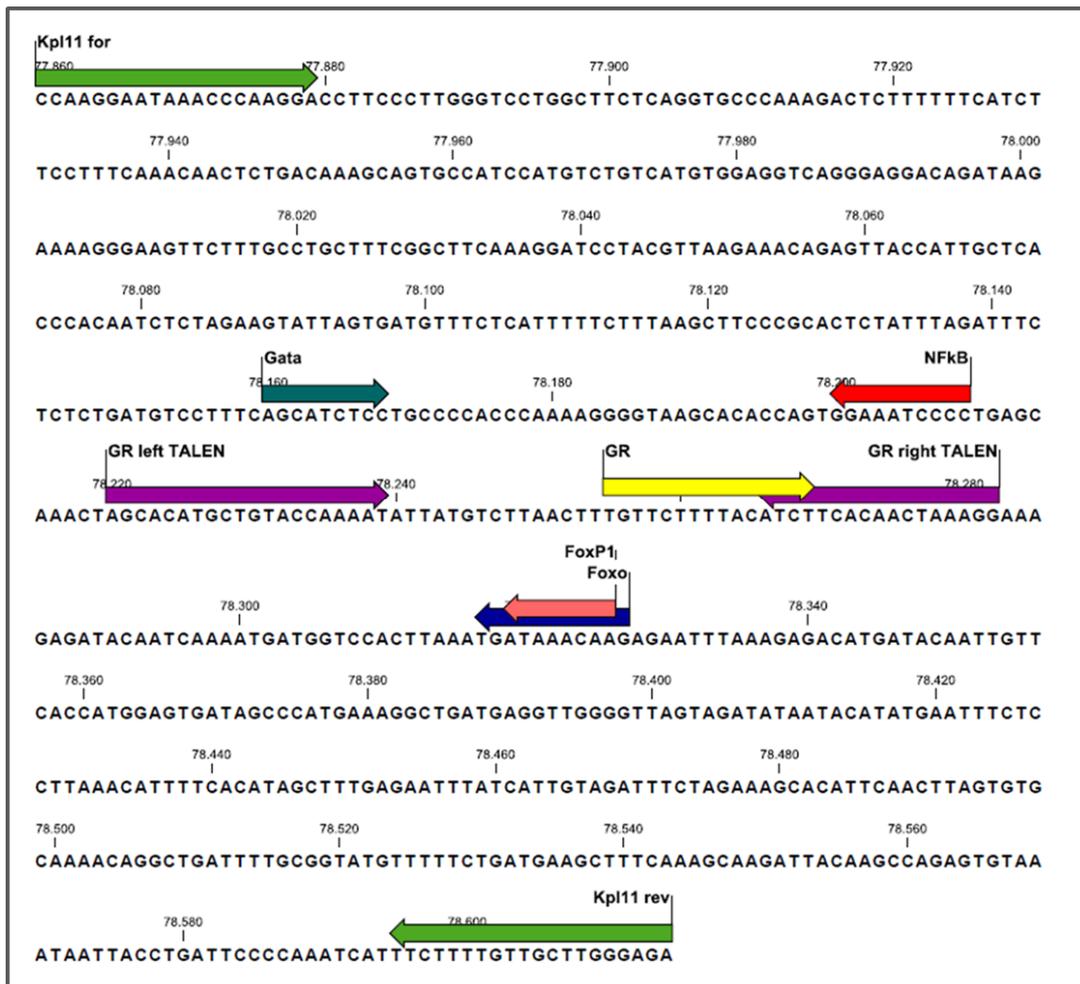


Figure 4.2 Binding sites of the commercially designed TALEN pair targeting the GR binding site of the IL7R enhancer region. Purple arrows indicate the left and right TALEN binding sites. Yellow arrow indicates the GR transcription factor binding site. Green arrows indicate the primer binding sites.

Because the pUC57 plasmid was not a eukaryotic expression plasmid, we first transferred the left and right TALEN coding sequences from this plasmid to a site upstream of the IRES-EGFP and the IRES-dsRed sequences respectively in the pMIGII retroviral mammalian expression plasmid backbone. Plasmid constructs were named pMIGII-GR left TALEN-IRES-eGFP and pMIGII-GR right TALEN-IRES-dsRed. Our reasoning for cloning TALENs into retroviral eukaryotic expression plasmids with fluorescent reporters were two fold. First, we wanted to track the expression of TALEN proteins with the expression of fluorescent reporter proteins. Secondly, we reasoned that upon retroviral infection, our expression cassette should be incorporated into the host cell genome and result in stable expression of TALEN mRNA. We surmised that high levels of TALEN protein expression would result inefficient cleavage of the target

sequence. It is well known that there are size limits for efficiently packaging large RNA molecules into retroviruses. The size of each constructed plasmid was around 10000 bp, which could limit production of virus from our expression plasmids. In addition, stable expression of TALEN proteins may result in over-expression which may increase the probability of generating off-target mutations. Therefore, we decided to transfer the TALEN sequences from the retrovirus plasmids to transient mammalian expression plasmids. The genes encoding TALENs targeting the GR site in the IL7R ECR3 were cloned downstream of a strong mammalian expression promoter, CMV, together with their fluorescent reporter proteins. These experiments are explained in detail below.

#### **4.1.1.1 Cloning of the left ECR3 GR binding site TALEN upstream of the eGFP in retroviral plasmid**

The strategy for cloning the left ECR3GR binding site TALEN upstream of the IRES-eGFP cassette in the pMIGII retroviral plasmid is shown in Figure 4.3. A DNA fragment containing TALE repeats was used to replace TcR  $\alpha\beta$  sequences in the pMIGII-TcR  $\alpha\beta$  eGFP plasmid (this plasmid was previously constructed in the Erman lab for an unrelated project). The plasmid constructed was named as pMIGII-GR left TALEN-IRES-eGFP. The use of an IRES-eGFP cassette resulted in TALEN mRNA expression from a bicistronic mRNA also encoding the eGFP fluorescence reporter.

We digested the pUC57GR left TALEN plasmid with the BglIII-EcoRI restriction enzymes, generating two bands of 4292 bp and 2641 bp. The 4292 bp-band, corresponding to the TALEN encoding sequence was extracted from an agarose gel. We digested the destination plasmid, pMIGIITcR  $\alpha\beta$  eGFP with the BglIII-MfeI restriction enzymes, generating two bands of 6316 bp and 1797 bp. The 6316 bp-band containing the retroviral vector backbone was gel extracted. The restriction enzymes MfeI and EcoRI generate sticky ends that are compatible with each other. Therefore, ligation of the 4292 bp-TALEN encoding fragment with sticky ends of EcoRI and BglIII is possible.

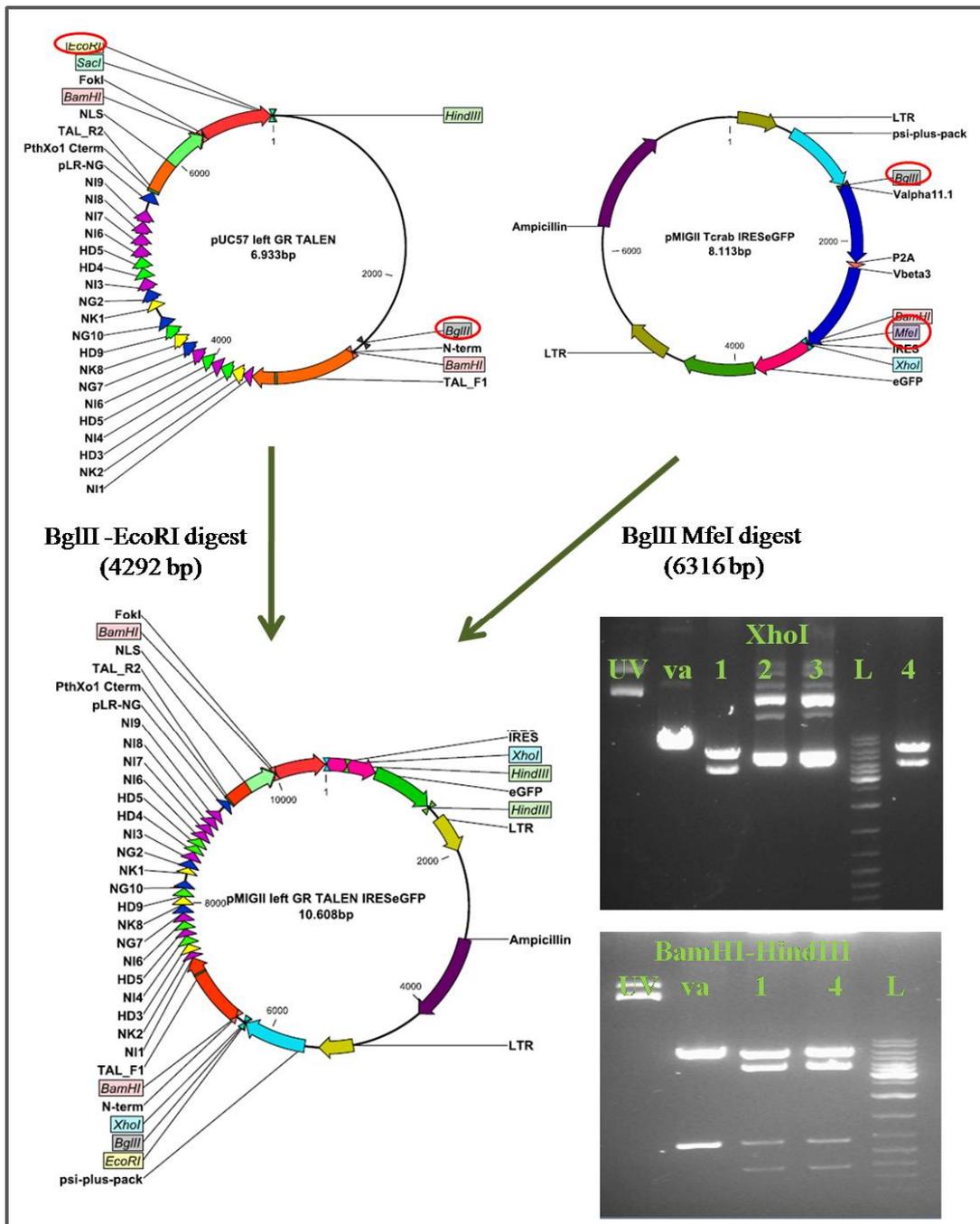


Figure 4.3 Construction of the pMIGII left-GR TALEN IRESeGFP plasmid (Strategy 1a). A flowchart for the cloning procedure is shown on the left, and gel electrophoresis images showing a representative diagnostic digest (using XhoI and BamHI-HindIII restriction enzymes).

After determining the concentrations of gel extracted left GR TALEN encoding fragment and pMIGII vector backbone, we performed a ligation reaction using a 1:3 and 1:6 vector:insert ratio, transformed these reactions into *E. coli* and isolated plasmid DNA from the formed colonies. We then performed a confirmation digest using the XhoI enzyme, which should generate two bands of 6322 bp and 4294 bp in the case of a

correct ligation. Next, we performed a BamHI- HindIII control digest for two of the colonies producing the expected bands in the previous XhoI digestion. The BamHI-HindIII control digest should generate three bands of 5090 bp, 3550 bp and 1113 bp. The results of the confirmation digestions and a map of the constructed plasmid are shown in Figure 4.3.

#### **4.1.1.2 Cloning of the right ECR3 GR binding site TALEN upstream of the dsRed in retroviral plasmid**

The strategy for cloning the right ECR3 GR binding site TALEN upstream of the IRES-dsRed cassette in the pMIGII retroviral plasmid was given in Figure 4.4 and Figure 4.5. First, a DNA fragment containing TALE repeats was cloned into the pSP72 plasmid, an intermediate vector to integrate the necessary restriction enzyme cut sites to the ends of this fragment. Then, the DNA fragment containing the TALEN repeats was cloned upstream of the dsRed sequence in the pMIGII-IRES-dsRed plasmid (This plasmid was previously constructed in the Erman lab for an unrelated project). The plasmid constructed was named pMIGII-GR right TALEN-IRES-dsRed.

We digested the pUC57 GR right TALEN plasmid with BglII-HindIII restriction enzymes generating two bands of 3992 bp and 2635 bp. The 3992 bp-band, corresponding to right TALEN encoding sequence was extracted from an agarose gel. We digested pSP72 with BamHI-HindIII restriction enzymes which should generate fragments of 2432 bp and 30bp. The restriction enzymes BamHI and BglII generate sticky ends that are compatible with each other. The 2432 bp-fragment corresponding to vector backbone was extracted from an agarose gel. After ligation and plasmid DNA isolation, we performed a SacI digestion to control for correct integration, which should generate fragments of 3999 bp and 2433bp. As a second control, we digested this construct with the XhoI-BamHI enzymes, producing 4 bands of 3244 bp, 2432 bp, 641 bp and 123 bp. An agarose gel showing these confirmation digestions of SacI and XhoI-BamHI is shown in Figure 4.4.

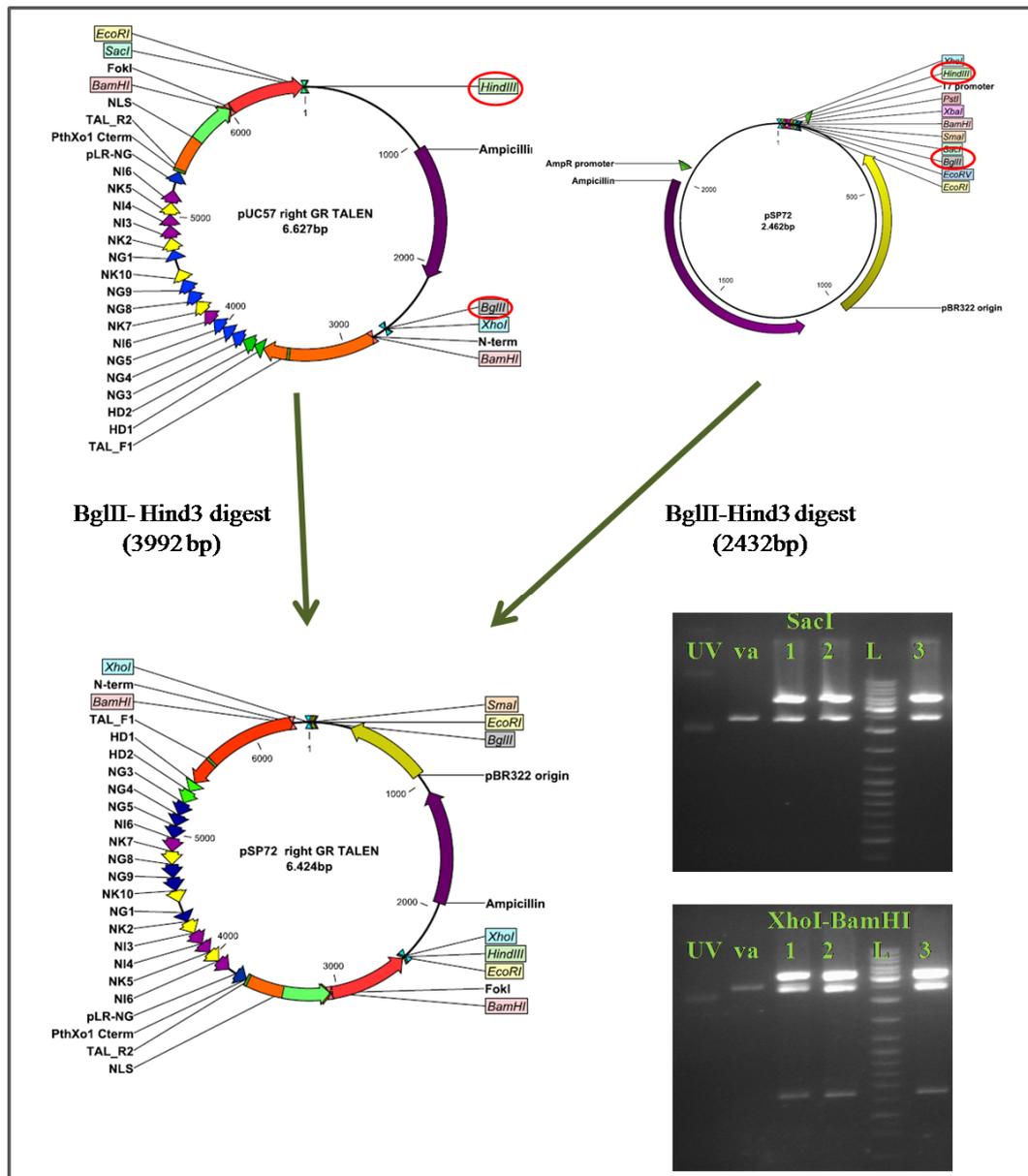


Figure 4.4 Construction of the pSP72-right ECR3 GR binding site TALEN plasmid.

We digested the intermediate pSP72-right ECR3 GR binding site TALEN plasmid construct with *EcoRI*, which should generate two bands of 4007 bp and 2417 bp. The 4007 bp-band, containing right GR TALEN encoding sequence, was gel extracted. We linearized the pMIGII-IRES-dsRed plasmid with *EcoRI* and gel extracted the 6278 bp band. CIAP (Calf Intestine Alkaline Phosphatase) treatment was performed for the band to prevent vector re-ligation.

We determined the concentrations of the gel extracted samples by agarose gel electrophoresis and accordingly, ligation reactions were set using 1:3 or 1:6 vector:insert ratios. We digested the plasmid with *SmaI* to confirm the identity of this

plasmid, which should generate fragments with sizes of 5840 bp, 3422 bp and 1023 bp. An agarose gel showing the *Sma*I confirmation digest and plasmid maps are shown in Figure 4.5.

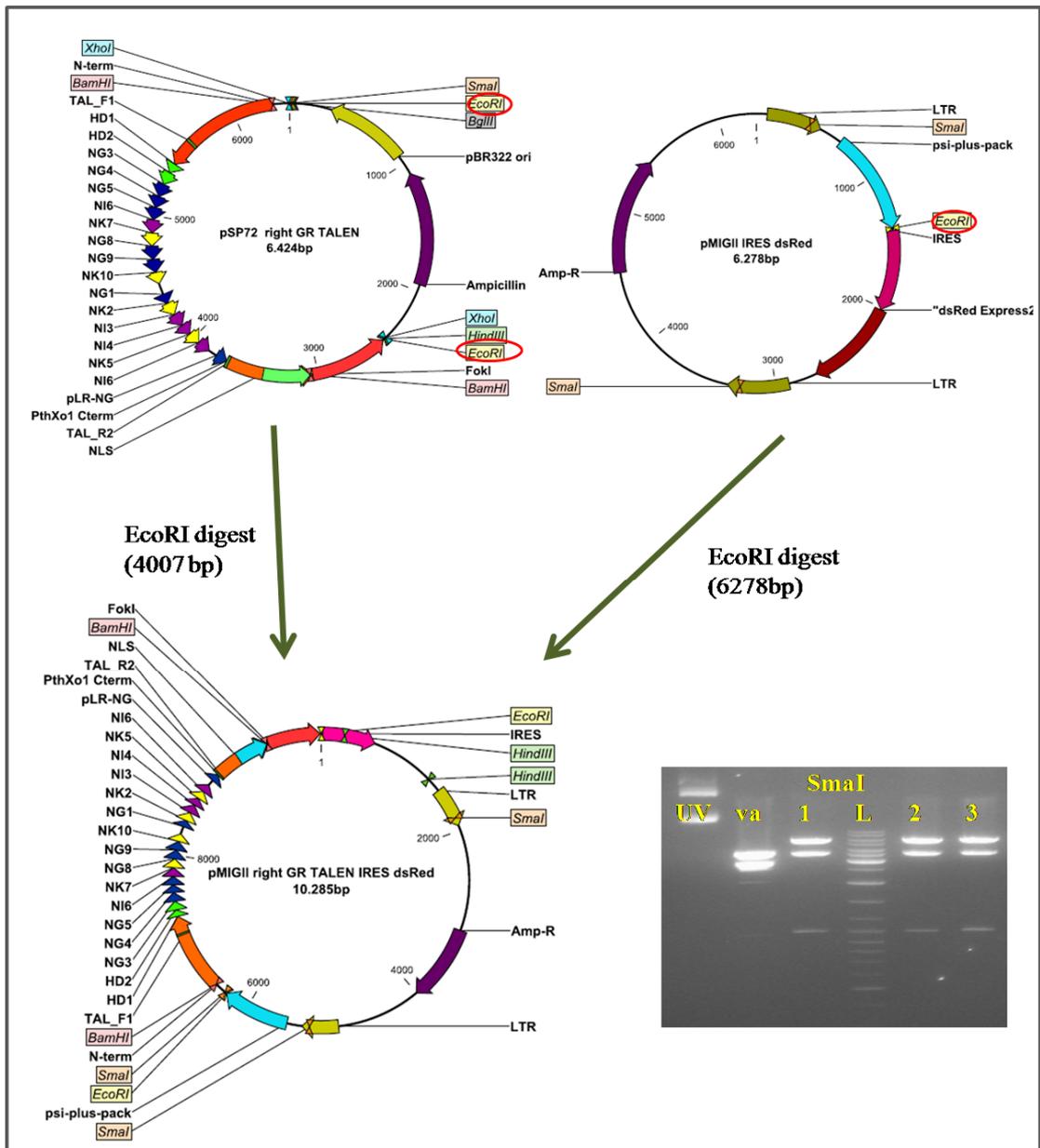


Figure 4.5 Strategy for cloning the right ECR3 GR binding site TALEN from the pSP72 plasmid upstream of the IRES-dsRed cassette in the pMIGII plasmid backbone (Strategy 1a).

#### 4.1.1.3 Stable expression of ECR3 GR binding site TALEN pair constructs in the NIH3T3 cell line

We generated retrovirus from HEK293 Phoenix cells transfected with the [pMIGII-GR left TALEN-IRES-eGFP] and [pMIGII-GR right TALEN-IRES-dsRed] plasmids. Supernatants of transfected cells containing virus were used to infect mouse

NIH3T3 cells. In addition to TALEN constructs, we also generated virus from the pMIGIITcR $\alpha\beta$ -IRESeGFP and pMIGII-IRESdsRed plasmids that were known to produce high titers of infectious virus as a positive control. Infection efficiency was determined by tracking eGFP and dsRed expression on a flow cytometer. Indeed we could not efficiently produce virus from these two plasmids to infect this cell line (data not shown). Although infection was successful for virus generated using the pMIGII backbone plasmid positive controls, we could not detect TALEN eGFP or TALENdsRed expression, indicating the infection was unsuccessful (Figure 4.6). The size of the TALEN plasmids could be the reason behind the low infection efficiency. For this reason, we decided to transfer these GR binding site TALEN reporter constructs into CMV promoter containing mammalian expression plasmids for transfection.

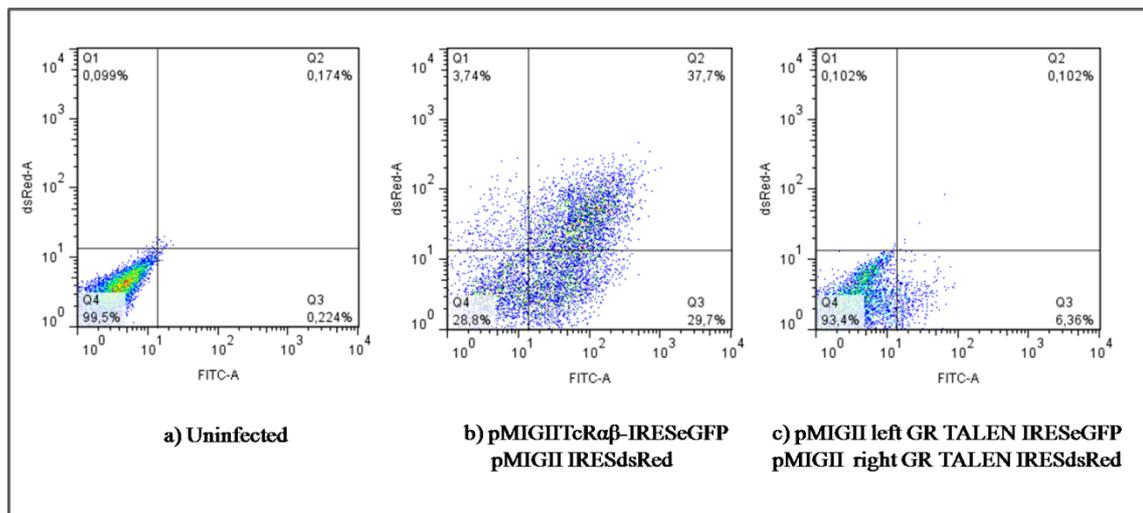


Figure 4.6 Infection of NIH3T3 cells with virus produced using GR binding site TALEN fluorescence reporter plasmids. Dot-plots show relative eGFP fluorescence (labeled FITC on the x-axis) and dsRed fluorescence on the y-axis. From the top left, the four quadrants show GFP negative dsRed positive cells (Q1), GFP positive dsRed positive cells (Q2), GFP positive dsRed negative cells (Q3) and GFP negative dsRed negative cells (Q4).

#### 4.1.1.4 Cloning of the left ECR3 GR binding site TALEN into a CMV IRES eGFP mammalian expression plasmid

The cloning strategy of the left ECR3 GR binding site TALEN into a CMV IRES eGFP mammalian expression plasmid is shown in Figure 4.7. The DNA fragment containing the left ECR3 GR TALEN in pUC57 was cloned into the pIRES2eGFP plasmid. The final plasmid construct was named CMV- left GR TALEN- IRES- eGFP.

The pUC57-left GR TALEN plasmid was digested with EcoRI-BglII, generating bands of 4296 bp and 2645 bp. The 4296 bp-band, which corresponds to the left

TALEN encoding sequence was gel extracted. IRES2eGFP is a mammalian expression plasmid containing a CMV promoter before a multiple cloning site (MCS) and an IRES-eGFP fluorescence reporter (Figure 4.7). It was digested with EcoRI-BglII which produces two bands of 5292 bp and 24 bp. The 5292-bp band corresponding to the vector backbone was extracted from an agarose gel.

We isolated plasmid DNA from colonies formed after ligation and transformation. First, we performed a PstI control digestion which should generate bands of 5491 bp and 4097 bp. Second, we performed an XhoI-HindIII confirmation digest, producing bands of 5038 bp and 4550bp. Plasmid maps and agarose gels for the confirmation digestions of PstI and XhoI-HindIII are shown in Figure 4.7.

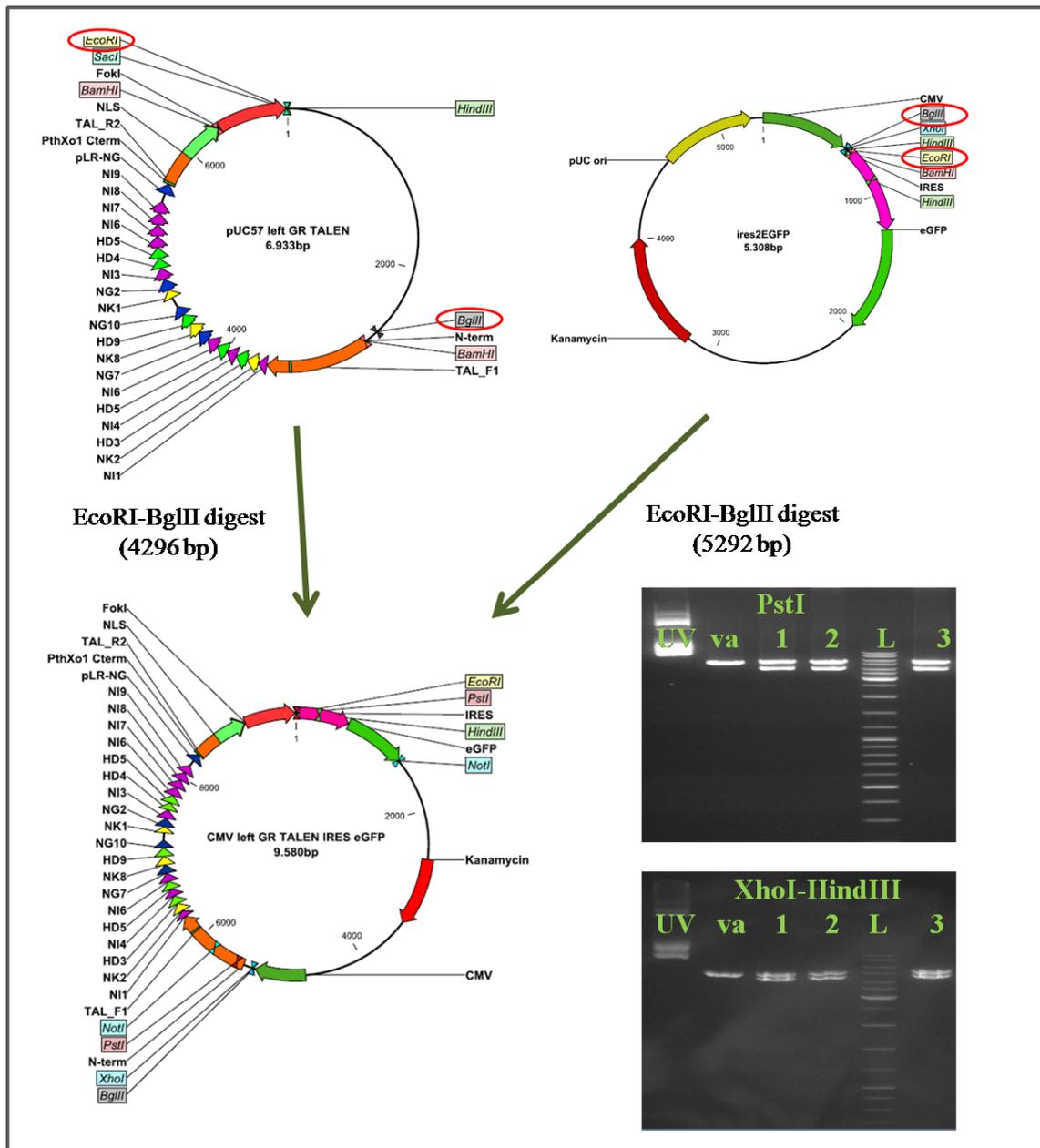


Figure 4.7 Strategy for cloning of the left ECR3 GR binding site TALEN downstream of CMV promoter and upstream of the IRES-eGFP cassette (Strategy 1b).

#### 4.1.1.5 Cloning of the right ECR3 GR binding site TALEN-IRES-dsRed cassette into a CMV promoter containing plasmid

The strategy for cloning of the right ECR3 GR binding site TALEN IRES dsRed cassette from the pMIGII backbone to the pcDNA3.1 plasmid is shown in Figure 4.8. The final plasmid construct was named CMV-right ECR3 GR binding site TALEN IRES dsRed. We digested the pMIGII right GR binding site TALEN-IRES-dsRed plasmid with the SmaI-SalI restriction enzyme pair, which should generate four bands of 5300bp, 3422bp, 1023bp and 544bp. The 5300bp-band, corresponding to the right

GR binding site TALEN-IRES-dsRed cassette was extracted from an agarose gel. pcDNA3.1+ is a mammalian expression plasmid containing a CMV promoter upstream of a multiple cloning site. The destination plasmid pcDNA3.1+ was digested with the EcoRV-XhoI restriction enzyme pair, generating two bands of 5407 bp and 25 bp. The 5407 bp vector backbone was gel extracted and ligated with the 5300 bp band corresponding to the right GR binding site TALEN-IRES-dsRed cassette. The size of final construct was 10703 bp.

We transformed this ligation reaction into E.coli, isolated plasmid DNA from colonies and to confirm the identity of the plasmids, digested them with the EcoRI-BglIII restriction enzyme pair, producing three bands of 5766 bp, 4005 bp and 944 bp. Agarose gel images showing the EcoRI-BglIII confirmation digest and plasmid maps are shown in Figure 4.8.

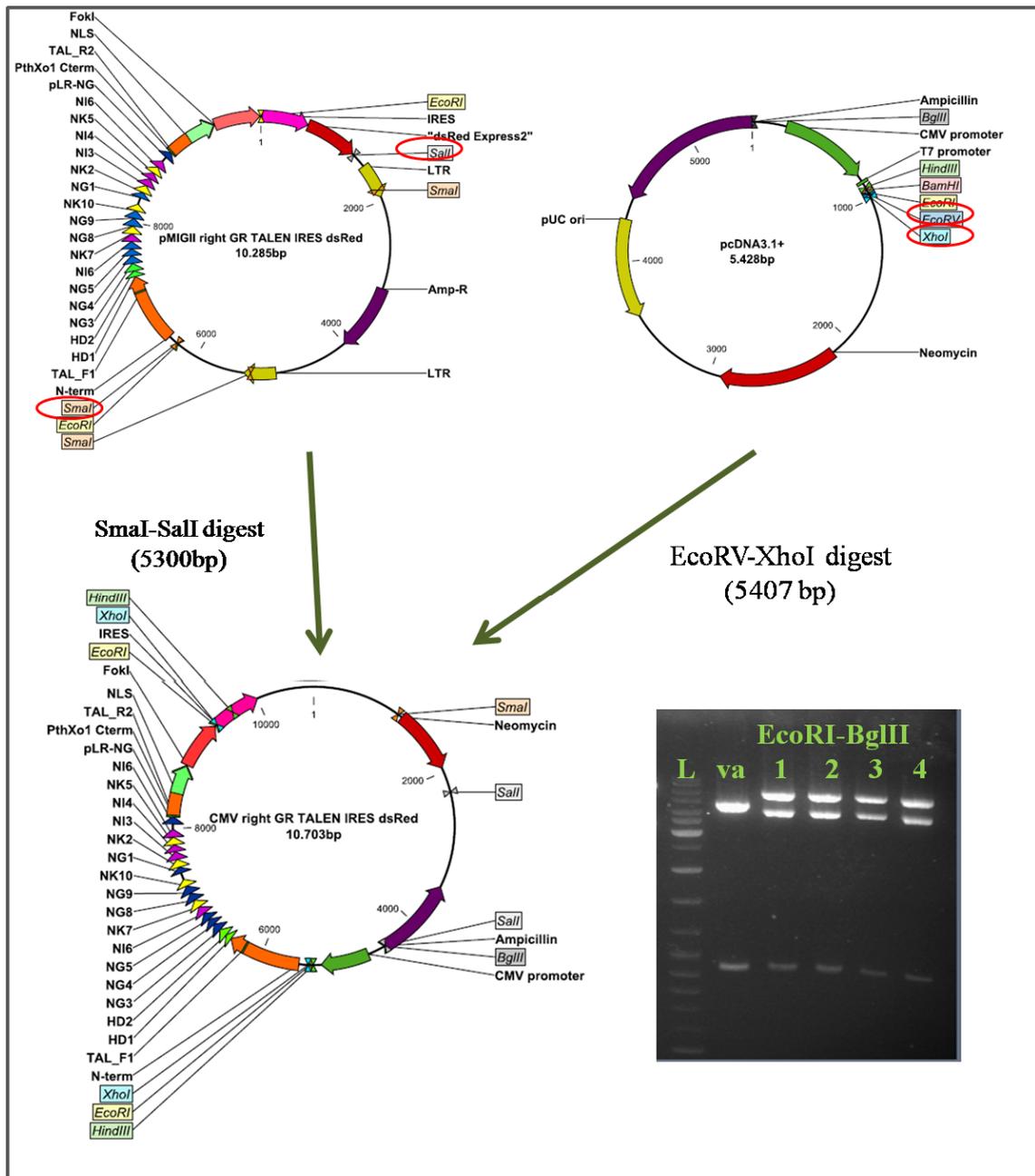


Figure 4.8 Strategy for cloning of the right GR binding site TALEN-IRES-dsRed cassette downstream of CMV promoter (Strategy 1b).

#### 4.1.1.6 Ectopic expression of the ECR3 GR binding site TALEN pair in Neuro-2a cells

To express the TALEN pair targeting GR binding site, we co-transfected the [CMV left ECR3 GR binding site TALEN-IRES eGFP] and [CMV right ECR3 GR binding site TALEN-IRES dsRed] plasmids into Neuro-2a cells. 29.2% of cells were found to be both GFP and dsRed positive, indicating that both TALEN proteins could be expressed in these cells (Figure 4.9). Note that because both plasmids link the

TALEN sequences with the fluorescent reporter genes (GFP and dsRed) use an IRES element, the transcripts encoding the TALENs and the fluorescent proteins are bi-cistronic. Thus the presence of the fluorescent protein in cells indicates the presence of the TALEN proteins.

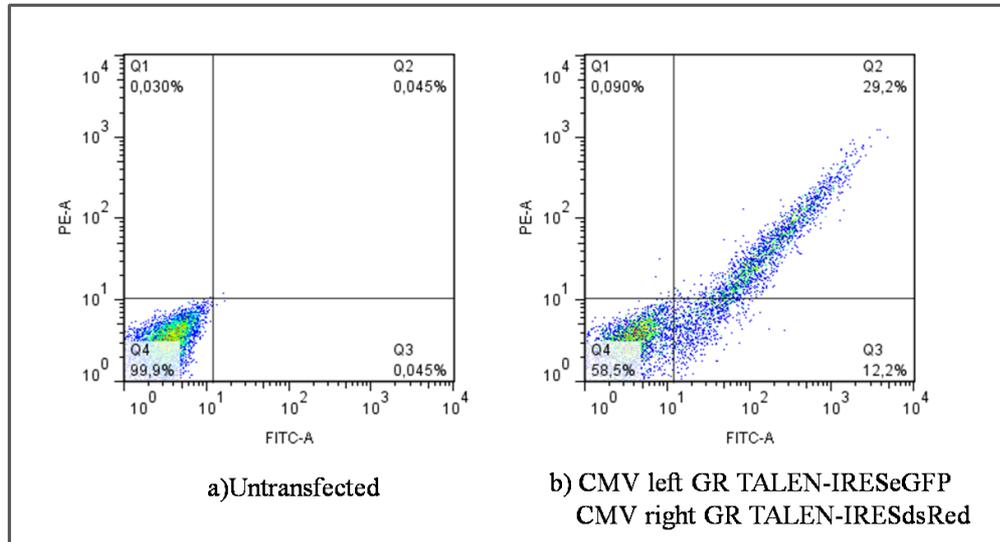


Figure 4.9 Transfection of murine Neuro-2a cells with CMV GR binding site TALEN fluorescence reporter plasmids. Dot-plots show relative eGFP fluorescence (labeled FITC on the x-axis) and dsRed fluorescence on the y-axis. From the top left, the four quadrants show GFP negative dsRed positive cells (Q1), GFP positive dsRed positive cells (Q2), GFP positive dsRed negative cells (Q3) and GFP negative dsRed negative cells (Q4).

We designed the TALEN plasmids targeting the GR binding site using the NK RVD to associate with G. During this study, Christian et al. (2012) found that NK RVD containing TALENs have lower affinity for G when compared with NN RVD containing TALENs[14]. In addition, the GR TALEN pair was designed to have a long C-terminal segment composed of 278 residues. However, a recent study conducted by Miller et al. (2011) indicates that efficient DNA cleavage is correlated with shortened C-terminal domains[6]. Also these TALENs were designed targeting a region that does not have appropriate restriction enzyme recognition sites in the genomic DNA. Thus, even if these TALENs were to cut genomic DNA at the target site and introduce mutations, we did not have an easy way to detect these mutations, such as the RFLP method described in detail in the following sections. For these reasons, we discontinued our work with the TALEN pair targeting the GR binding site in the IL7R gene.

## **4.1.2 Assembly of TALENs Targeting Notch Binding Site of IL7R Gene Using Golden Gate TALEN Kit**

### **4.1.2.1 Construction of a TALEN pair targeting the IL7R ECR2-ECR3 Notch binding site in the pCAGT7 backbone**

The TALEN design program TAL Effector Nucleotide Targeter 2.0 (<https://tale-nt.cac.cornell.edu/>) was used to design a pair of TALENs targeting the Notch binding site in the region between ECR2 and ECR3 of the IL7R gene. The TALEN pair was selected from the output of the software such that both left and right TALENs contain 15RVDs and the spacer region between the TALEN binding sites was 17 bp in length. Left and right TALEN binding sites were preceded by a 5'T base, in accordance with the TALEN design guidelines provided by the TAL Effector Nucleotide Targeter 2.0 program.

A custom TALEN pair targeting the IL7R Notch binding site was designed using the Golden Gate TALEN kit. As explained in detail in the Methods section, TALEN assembly using the Golden Gate TALEN kit takes 5 days and involves two main reactions for the assembly of individual repeat modules into array plasmids and assembly of array plasmids into a final mammalian expression plasmid.

For the construction of the two 15 repeat long Notch binding site TALENs, we cloned the first 10 repeats into the pFUS\_A plasmid and all remaining repeats except the last, into the pFUS\_B4 plasmid (the Golden Gate reaction #1). Thus, we set up two Golden Gate reactions (reaction A and B) for each monomer of the TALEN pair. We transformed Golden Gate reaction #1 into *E. coli*, plated on IPTG/X-gal plates for blue-white screening, performed colony PCR on white colonies using the primers pCR8\_F1 and pCR8\_R1. Note that the number of colonies obtained after transformation was higher for the array vector containing less repeats than the ones with more repeats. For repeats in the pFUS\_A plasmids, we detected bands around 1200 bp, whereas for the pFUS\_B4 plasmids, we obtained around 600 bp bands. In addition to the band of expected size, a smear and ladder of bands were also detected which results from the presence of repeats in clones. After isolation of plasmid DNA from two correct colonies, we performed AflIII-XbaI diagnostic digests which should generate bands reflecting the number of repeats such that a 1048bp-band was obtained for pFUS\_A plasmid with 10 repeats whereas the band size was 430bp for the pFUS\_B4 plasmids

with four repeats. Figure 4.10 and Figure 4.11 show agarose gel images after colony PCR with pCR8\_F1 and pCR8\_R1, confirmation with AflIII-XbaI double digests and the plasmid maps related to reactions A and B of the left Notch binding site TALEN(Notch left TALEN) and the right Notch binding site TALEN (Notch right TALEN), respectively.

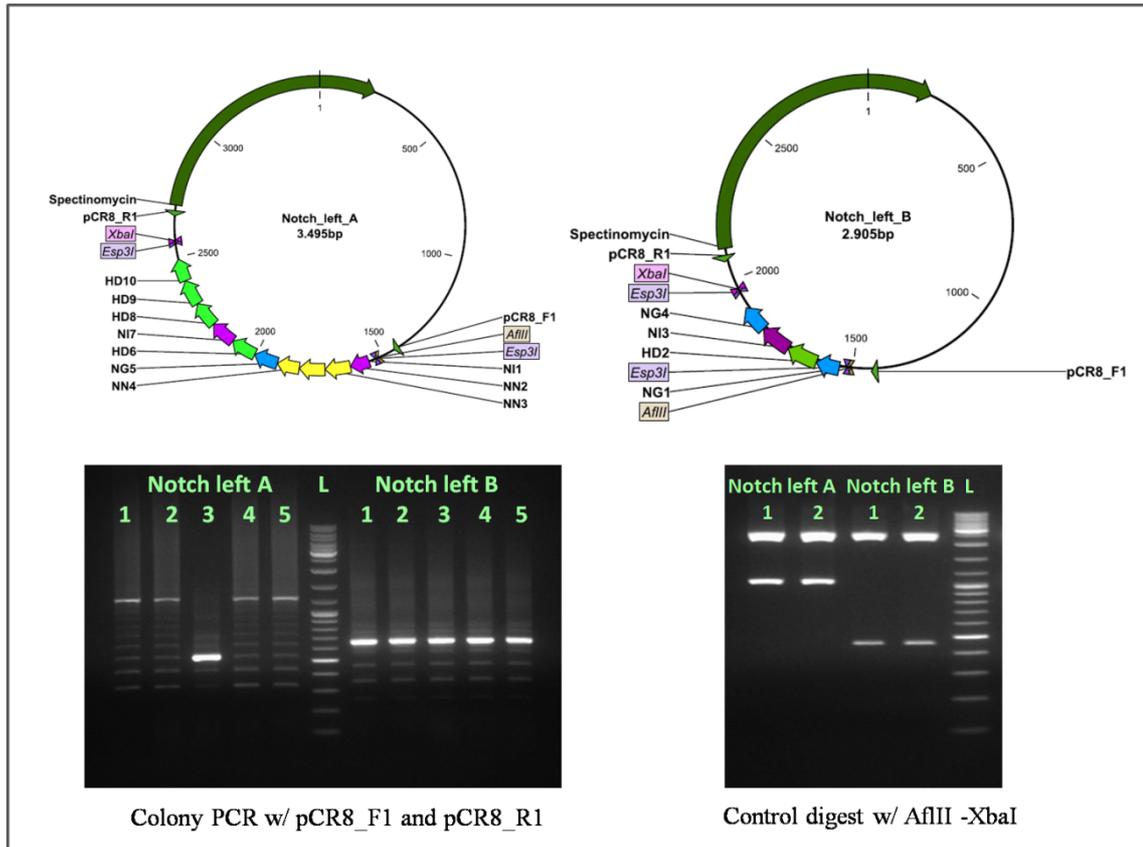


Figure 4.10 Golden Gate reaction #1 for the left Notch binding site TALEN

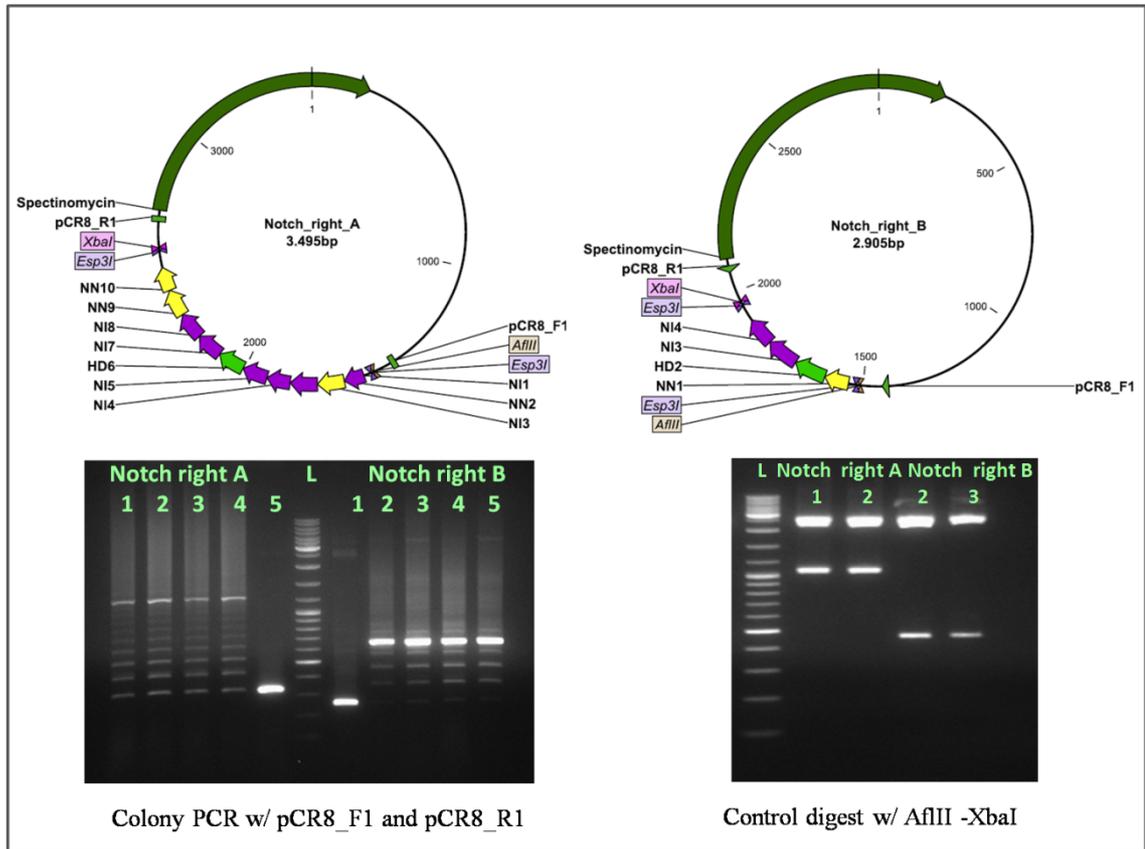


Figure 4.11 Golden Gate reaction #1 for the right Notch binding site TALEN

In the Golden Gate reaction #2, we constructed an intermediary array of repeats coming from reaction A and reaction B, and the last repeat plasmid. We joined these three fragments in the final backbone plasmid, pCAGT7. We performed colony PCR using the TAL\_F1 and TAL\_R2 primers, which gave bands of 1744bp for the 15-repeats and a smear and ladder effect. A confirmation digest with AatII and StuI should result in the generation of a 1724bp-band, indicating the presence of 15 repeats. Figure 4.12 shows agarose gel images after colony PCR with TAL\_F1 and TAL\_R2 primers, control digest with AatII-StuI and plasmid maps of both pCAG left Notch binding site TALEN and pCAG right Notch binding site TALEN.

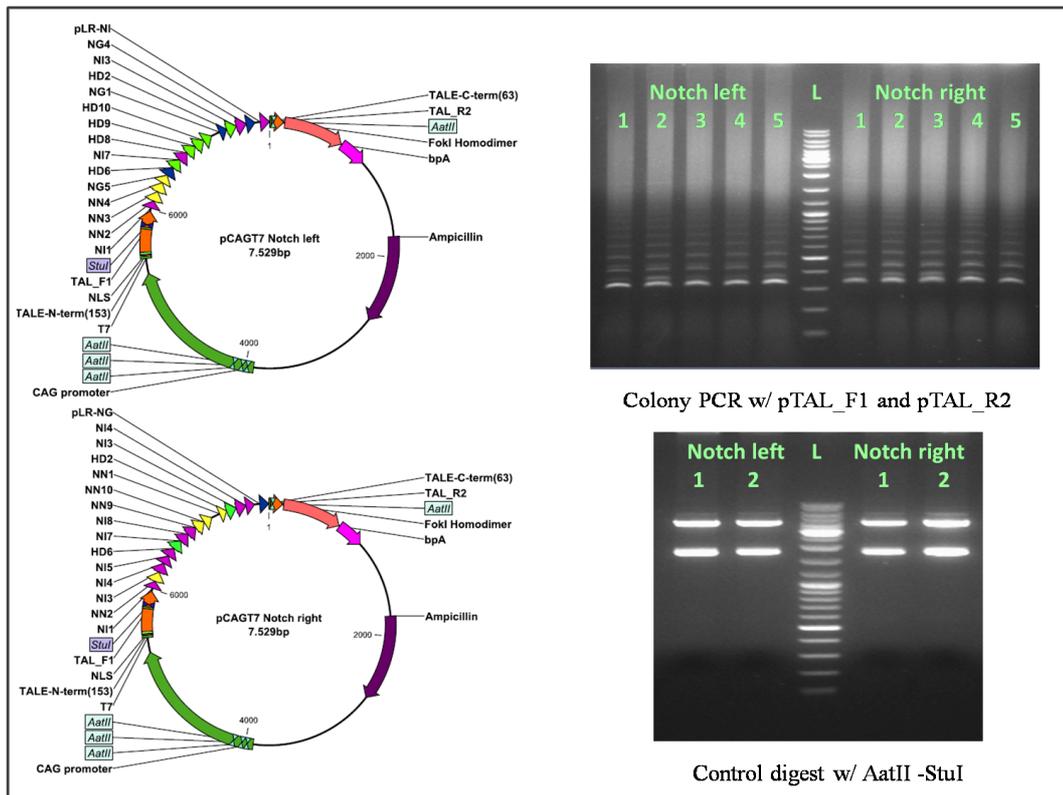


Figure 4.12 Golden Gate reaction #2 for the Notch binding site TALEN constructs in pCAGT7 (Strategy 2a).

#### 4.1.2.2 Cloning of the designed Notch binding site TALEN monomers into mutant FokI destination vectors

Miller et al. (2007) claimed that the safety and efficacy of zinc finger nucleases (ZFNs), which are precursors of the TALENs which also use the dimerized FokI domain to cut DNA, was limited due to homodimerization rather than heterodimerization. This homodimerization resulted in off-target cleavage. Therefore, they engineered an improved ZFN architecture by mutating specific residues of the FokI enzyme dimerization interface such that they cleave target DNA only when monomers paired as heterodimers[44]. In the same manner, two mutant versions of pCAGT7 were developed by the group of Pawel Pelczar to improve TALEN efficiency. pCAGT7-FokI ELD includes the Q486E, I499L and N496D mutations of FokI whereas pCAGT7-FokI KKR includes the I538K, E490K, and H537R mutations. Therefore, to achieve efficient genome modification, we cloned the Notch left TALEN and Notch right TALEN constructs into the pCAGT7-FokI ELD and pCAGT7-FokI KKR plasmids respectively.

To generate these plasmids, we digested the pCAGT7 Notch left TALEN plasmid with the BamHI-BglIII restriction enzymes, generating two bands of 5418bp and 2119bp. The latter band was gel extracted. We digested the pCAGT7-FokI ELD plasmid with BamHI-BglIII which gave two bands of 5414 bp and 1031 bp. The 5414 bp band was extracted from an agarose gel and ligated into the gel extracted pCAGT7 Notch left TALEN fragment. Confirmation digests with ClaI-HincII restriction enzymes generated two bands of 7151 bp and 380 bp. Figure 4.13 shows the cloning strategy and the confirmation digest with ClaI-HincII for the pCAGT7Notch left FokI ELD plasmid.

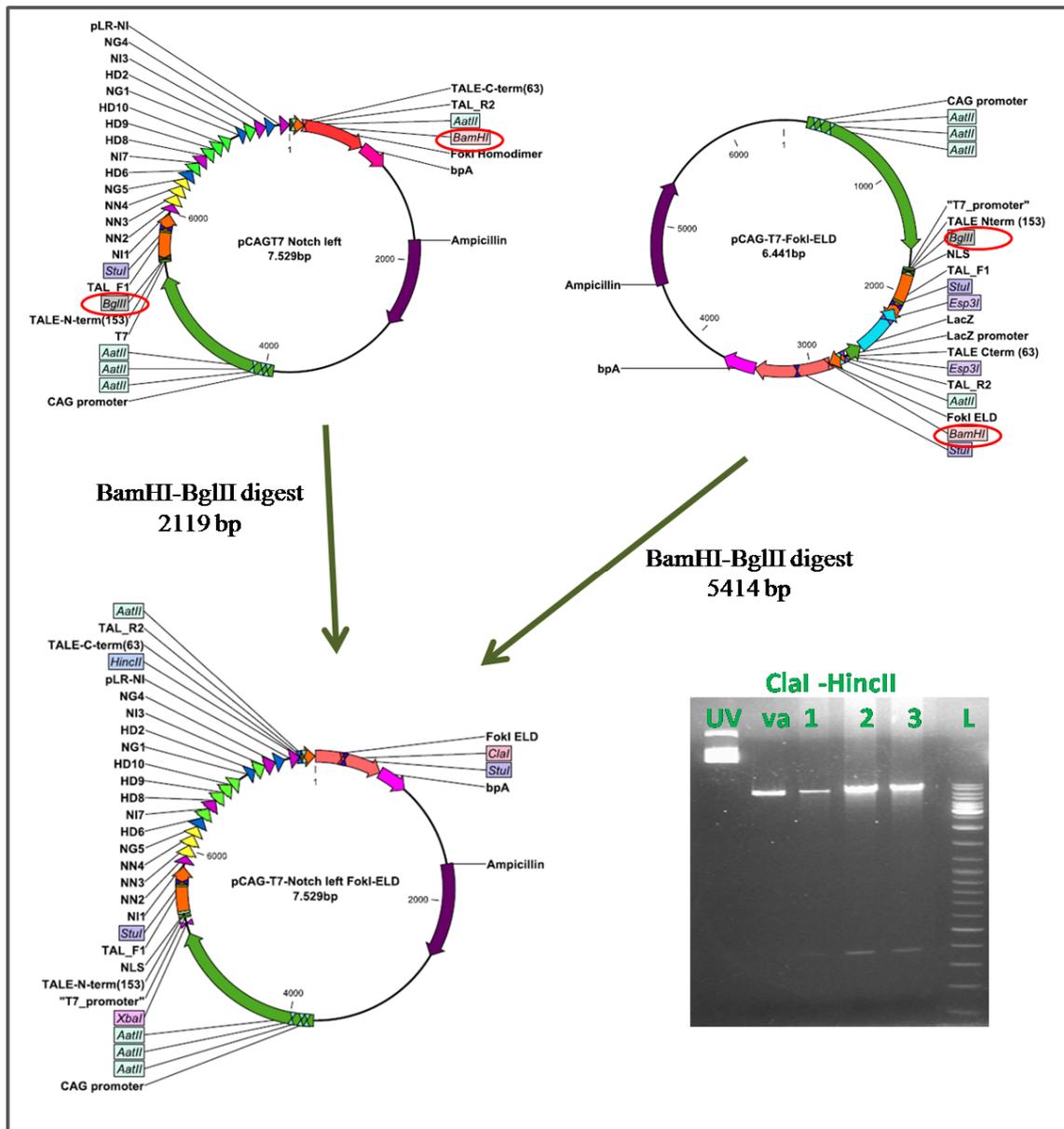


Figure 4.13 Cloning of the left Notch binding site TALEN into the pCAGT7 FokI ELD backbone (Strategy 2b).

To clone the right Notch binding site TALEN into the pCAGT7-FokI KKR plasmid, both pCAG Notch right TALEN and pCAGT7-FokI-KKR were digested with BamHI-BglIII. Digestion of the pCAG Notch right TALEN plasmid with the BamHI-BglIII restriction enzyme pair generates two bands of 5418 bp and 2119 bp. The 2119 bp-band corresponding to the Notch right TALEN repeats was gel extracted. The BamHI-BglIII digestion of the pCAGT7 FokI KKR plasmid generated two bands of 5414bp and 1031bp and the former one was gel extracted. Gel extracted bands were ligated and transformed. After isolation of plasmid DNA, we performed an XbaI-HincII confirmation digest, which generated bands of 5469 bp and 2064 bp. The cloning strategy and agarose gel image of the confirmation digest with XbaI-HincII is shown in Figure 4.14.

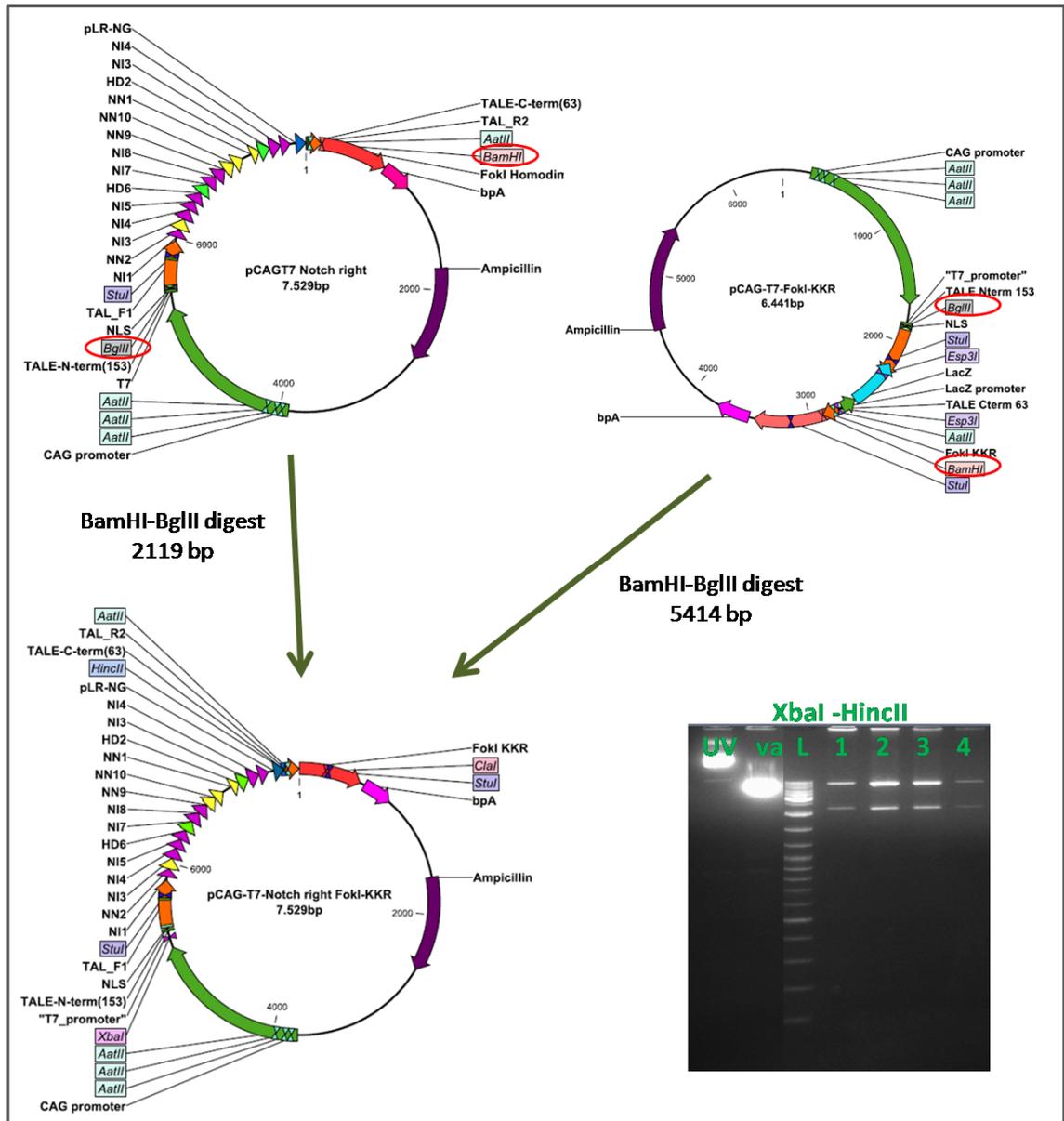


Figure 4.14 Cloning of the right Notch binding site TALEN construct to the pCAGT7 FokI KKR plasmid (Strategy 2b).

#### 4.1.2.3 Construction of the TALEN pair targeting Notch binding site in the pC-Goldy backbone

In a study conducted by Bedell et al (2012), it was shown that the DNA binding domain of a TALEN pair targeting the *ponzr1* locus in the Goldy TALEN scaffold results in a six-fold increased somatic gene mutation rate when compared to the same DNA binding domain in the pTAL backbone[35]. Therefore, we decided from this point onwards to only use the pC-Goldy architecture to express TALENs. To transfer our TALENs into the pC-Goldy plasmid backbone, we repeated Golden Gate reaction #2 for each monomer of the Notch binding site TALEN pair using the previously

constructed pFUS\_A and pFUS\_B array plasmids. We performed colony PCR with the TAL\_F1 and TAL\_R2 primers and observed the expected band size of 1745 bp along with a smear and ladder of smaller sized bands. An AatII -StuI confirmation digest of plasmid DNA extracted from these colonies generated five bands of 3863 bp, 1724 bp, 1074 bp, 87 bp and 57 bp, of which the 1724 bp band reflects the number of repeats assembled, the other bands coming from the plasmid backbone. Agarose gel images showing the colony PCR and confirmation digest with the AatII-StuI enzymes and the maps of the plasmids with assembled repeats is shown in Figure 4.15.

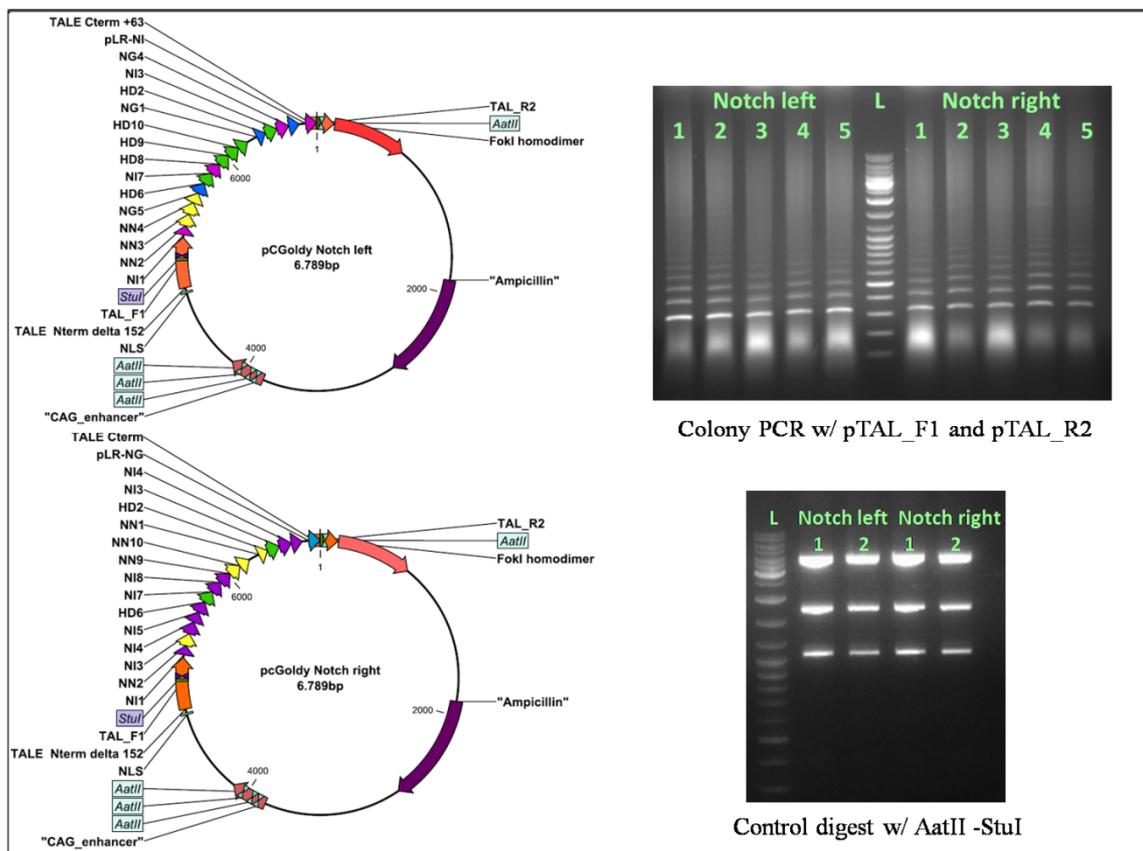


Figure 4.15 Golden Gate reaction #2 for the Notch binding site TALEN pair in the pC-Goldy backbone (Strategy 2c)

#### 4.1.2.4 Expression of the designed Notch binding site TALEN pair in Neuro-2a cells and detection of site-specific mutations

We transfected the TALEN pair targeting Notch binding site in the pC-Goldy backbone into Neuro-2a cells. Although IL7R is only expressed on cells of the lymphoid lineage and these Neuro-2a cells are a cultured neuroblastoma line, we hypothesized that this would be an easy system to observe TALEN mediated mutations because of the high transfection efficiency of this cell line. The location of the binding

sites for the Notch TALEN pair, the forward and reverse primers used in amplifying the TALEN target site and the enzyme cut sites present in the spacer region are shown in Figure 4.16.

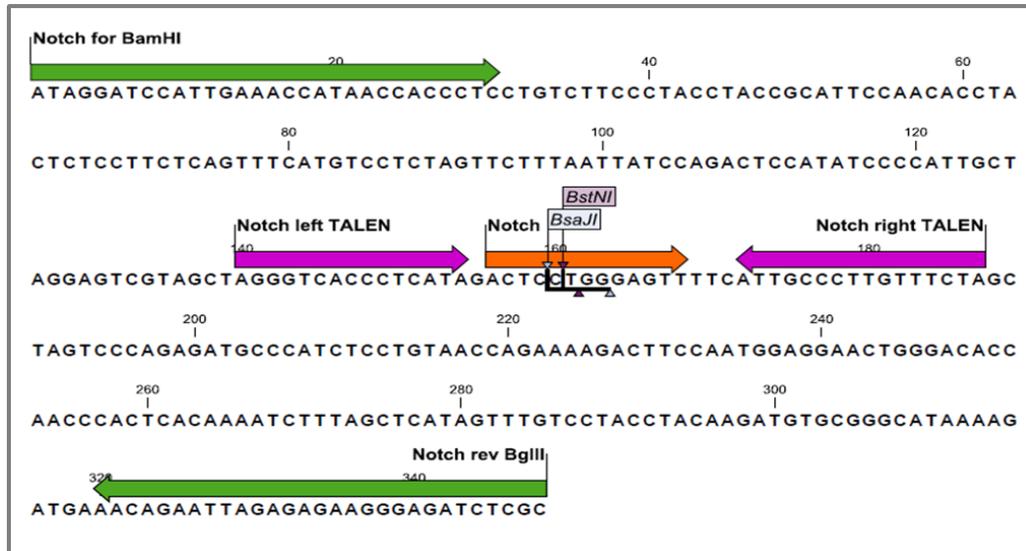


Figure 4.16 Binding sites for the assembled TALEN pair targeting the Notch binding site of the IL7R enhancer region. Purple arrows indicate the left and right TALEN binding sites. Orange arrow indicates the Notch binding site. Green arrows indicate the primer binding sites.

Our strategy for detecting site directed mutations was briefly summarized in the previous chapter (Figure 3.4 and 3.5). We attempted to detect mutations due to TALEN action by a restriction fragment length polymorphism (RFLP) assay with the BsaJI enzyme on PCR amplified target region DNA. In preliminary experiments from TALEN transfected cells, we could not detect an uncut band which is an indication of the presence of a mutated allele in this RFLP assay. For this reason, we modified our RFLP assay by first digesting genomic DNA with the BsaJI enzyme. We expected that BsaJI would only digest wild type alleles and that mutant alleles would remain uncut. We supposed that a PCR reaction with pre-cut genomic DNA would enrich mutant alleles allowing for better detection.

The size of the PCR product from this region was 348 bp and after BsaJI digest, two fragments of 159 bp and 189 bp were generated from the wild type alleles. When the PCR product of the digested sample was digested with BsaJI, the presence of a BsaJI-resistant band indicated the presence of mutated alleles in the genomic DNA pool. The agarose gel image showing PCR products before and after BsaJI digestion for digested and undigested genomic DNA isolated from Neuro-2a cells is shown in Figure

4.17. We extracted the uncut band from these agarose gels and cloned this fragment into the pJET1.2/blunt vector for sequencing. We performed colony PCR for the colonies produced after transformation to control the presence of an insert in this plasmid. We screened more than 100 colonies by colony PCR and sequenced purified plasmid DNA from positive clones.

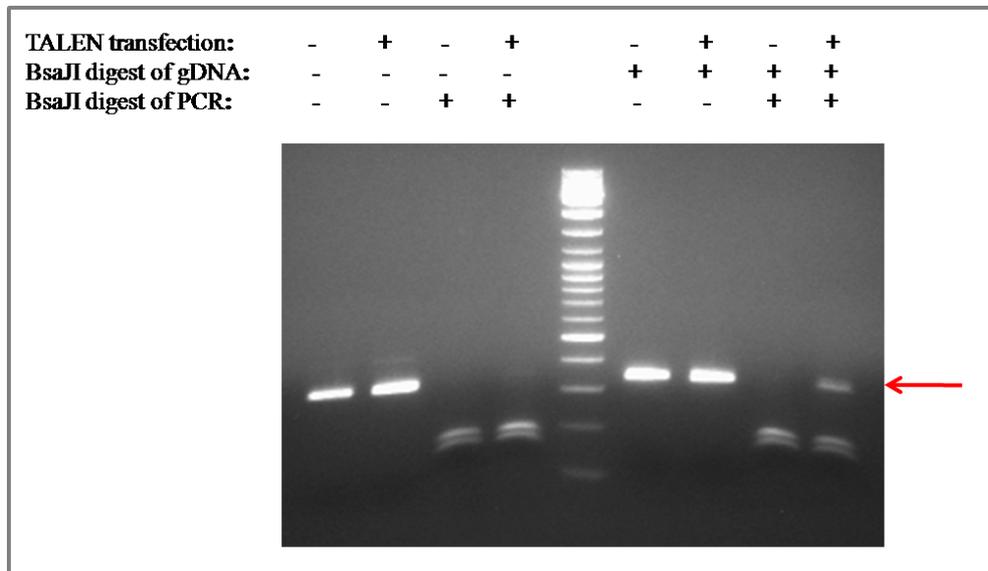


Figure 4.17 Mutation detection at the Notch TALEN target site of Neuro-2a cells using a modified RFLP assay. Uncut PCR product indicating the presence of TALEN induced mutations is shown by a red arrow.

Sequencing analysis of mutations at the TALEN target site for 80 different colonies is shown in Figure 4.18. This result indicates that the Notch TALEN pair generated mutations at the target site of Neuro-2a cell genomic DNA, mediated through imprecise repair of chromosomal double strand breaks (DSB) by non-homologous end joining (NHEJ). Although it is not possible to detect the mutation frequency due to an enrichment step in our modified RFLP technique, we detected 14 different mutations including deletions ranging from 1-11 bp, insertions of 1-5 bp and one single base substitution. All of these mutations were centered on the spacer region between binding sites of TALEN monomers, containing and destroying the Notch binding site.

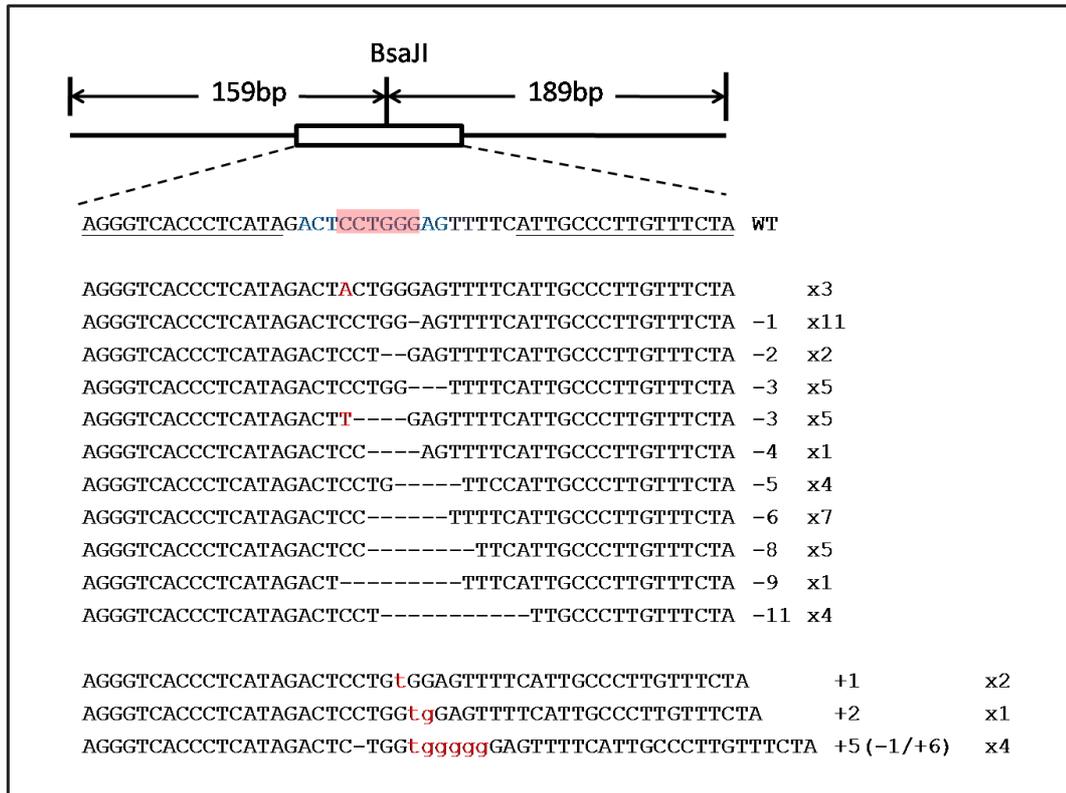


Figure 4.18 Site directed mutagenesis in Neuro-2a cells using the Notch binding site TALEN pair. Underlined regions of wild type sequence indicate the binding sites of TALEN monomers. Blue colored letters in the spacer region represent the Notch binding site, whereas the region highlighted in pink corresponds to the BsaJI cut site. Dashes indicate deletions. Red colored letters are SNPs and small case letters in red are insertions. Numbers on the right hand side indicates number of bases modified and frequency of occurrence of this mutation

#### 4.1.2.5 Expression of the designed Notch binding site TALEN pair in RLM11 cells and detection of site-specific mutations

We had designed the Notch TALEN pair to mutate the Notch binding site of the IL7R gene locus in order to study the functional significance of this site on IL7R expression. After demonstrating that our designed TALEN pair could indeed generate mutations in Neuro2a cells, we transfected the IL7R expressing RLM11 cells with our Notch TALEN pair in the pC-Goldy backbone.

Again, to detect site-specific mutations, we performed our modified RFLP assay (explained previously in Figure 4.16). An agarose gel image for BsaJI digests of PCR products from digested and undigested genomic DNA from TALEN transfected RLM11 cells is shown in Figure 4.19. Again we gel extracted and cloned the uncut BsaJI-resistant band into the pJET1.2/blunt vector for sequencing.

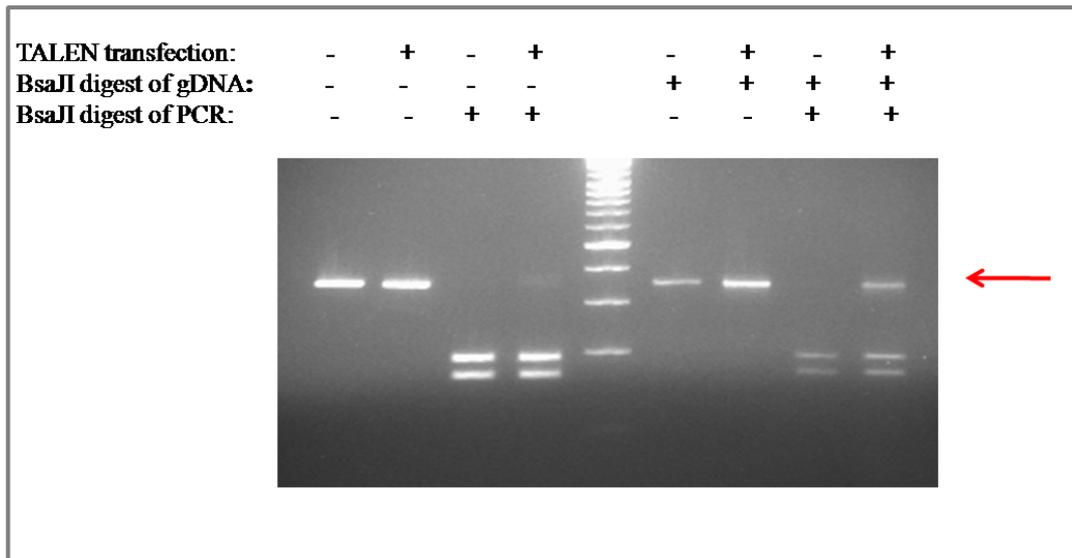


Figure 4.19 Mutation detection at the Notch TALEN target site of RLM11 cells using the modified RFLP assay.

The result of our DNA sequencing analysis of target site mutation in TALEN transfected Rlm11 cells is shown in Figure 4.20. In this experiment, 27 different colonies were sequenced. As before, the use of an enrichment step for detecting mutations prevents the determination of the mutation frequency. Nevertheless, we detected 20 different mutations in this analysis including deletions ranging from 1 to 22 bp, insertions of 1 bp, transitions and transversions.

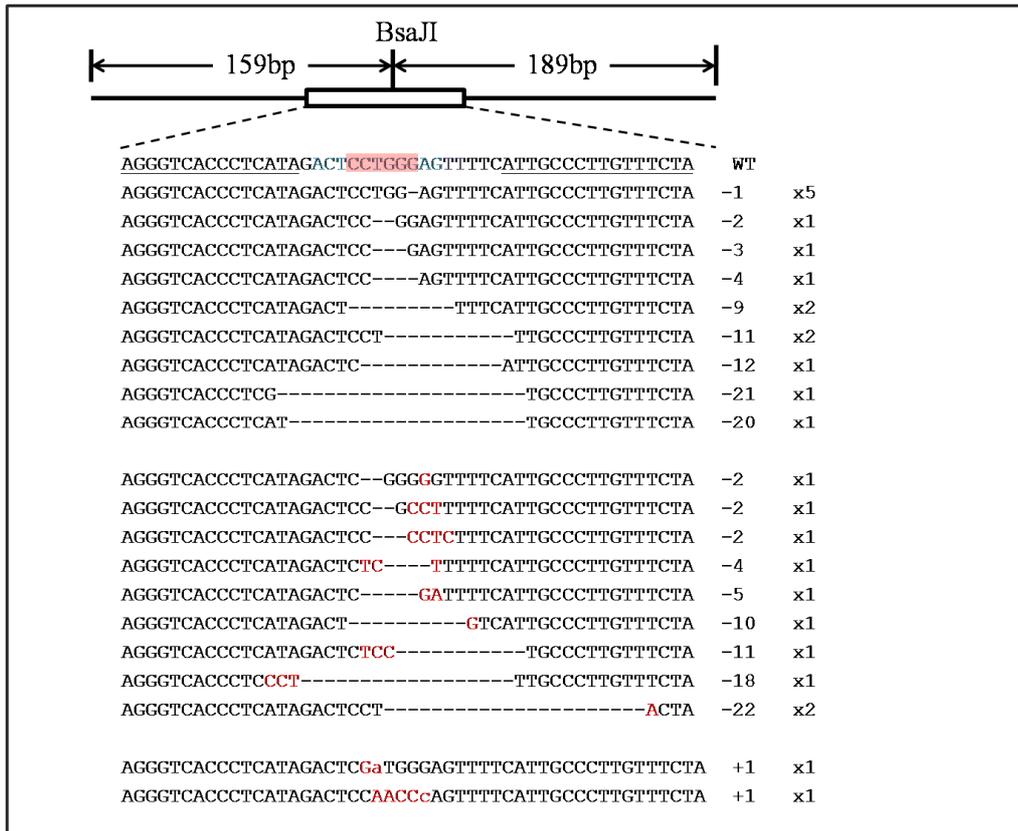


Figure 4.20 Site directed mutagenesis in RLM11 cells using the Notch binding site TALEN pair. Underlined regions of the wild type sequence show the binding sites of the TALEN monomers. Blue colored letters in the spacer region represent the Notch binding site, whereas the region highlighted in pink corresponds to the BsaJI cut site. Dashes represent deletions. Red colored letters are SNPs and small case letters in red are insertions. Numbers on the right hand side indicates number of bases modified and frequency of occurrence of this mutation.

#### 4.1.2.6 Detection of mutations at the Notch binding site TALEN target site for different backbones

To observe whether the Notch binding site TALEN pair in different backbones results in a difference in terms of DSB generation and site directed mutation efficiency, we transfected the Notch binding site TALEN pair into RLM11 cells in the pCAGT7, pCAGT7 FokI ELD/KKR and pC-Goldy backbones. These plasmids encode left and right TALEN proteins that bind exactly to the same sequence but have minor differences in the C- and N-terminal regions of the TALEN proteins as well as differences in the FokI restriction enzyme domain. Moreover the promoter for the pCAG plasmids and the pC-Goldy are slightly different [35].

We performed our modified RFLP assay for BstNI-digested genomic DNA isolated from these TALEN transfected cells. BstNI restriction enzyme digestion

generates two fragments of 188 bp and 161 bp from the wild type PCR product of 348bp. Figure 4.21 shows the agarose gel image after BstNI digest of PCR products from BstNI digested genomic DNA. The intensity of the uncut band, indicating mutation efficiency, is the highest for the TALEN pair in the pC-Goldy backbone and the lowest in pCAG-FokI mutant ELD/KKR backbone. This result is consistent with published results that show that TALENs in the pC-Goldy backbone are very efficient.

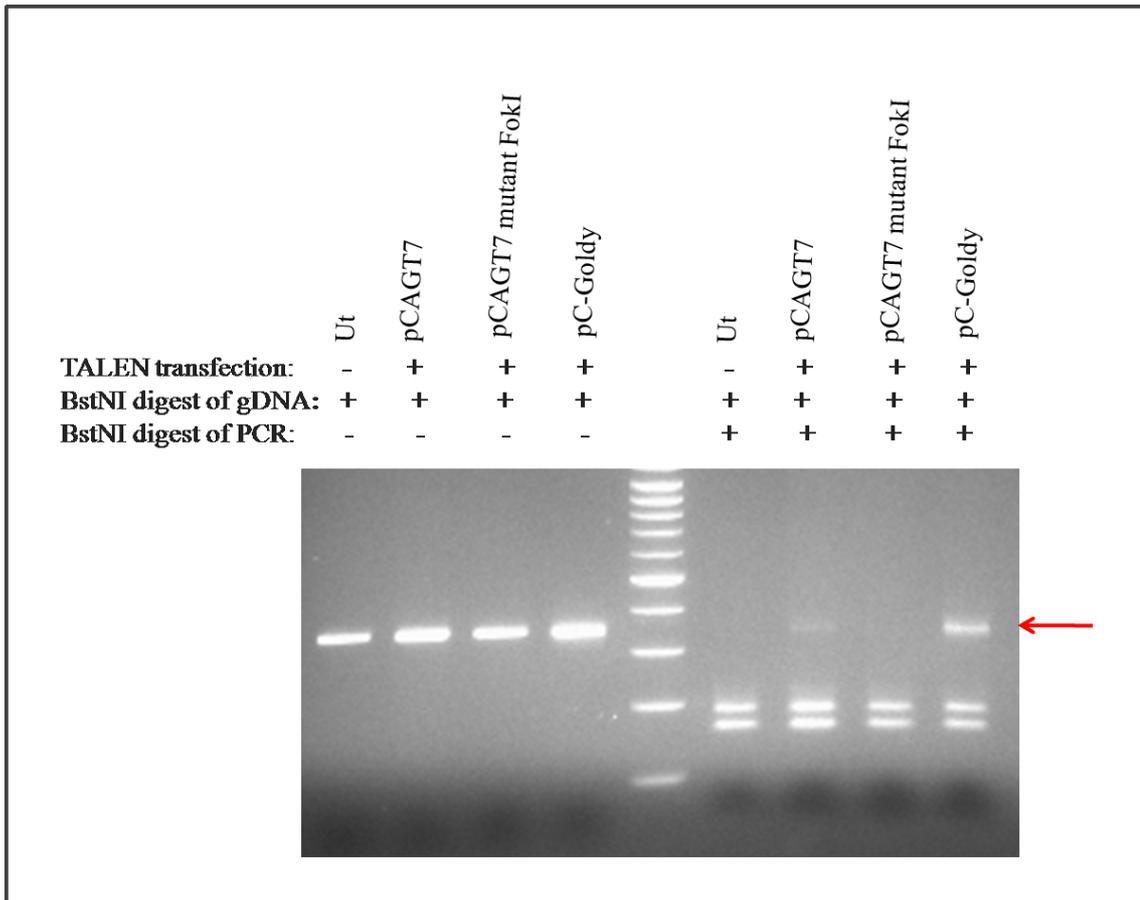


Figure 4.21 Mutation detection at the Notch binding site of RLM11 cells using a Notch binding site TALEN pair in different backbones

#### 4.1.2.7 Expression of IL7R on RLM11 cells transfected with Notch binding site TALEN pairs

Although genomic DNA was isolated from a pool of RLM11 cells and an enrichment step was included in the detection of mutations at the TALEN site, it was important to observe the phenotypic effects of these mutations on IL7R expression. The IL7R expression level of RLM11 cells transfected with Notch TALENs in pC-Goldy backbone was analyzed by flow cytometry (FACS). We observed that Notch TALEN transfection did not result in a significant change in IL7R expression (Figure 4.22).

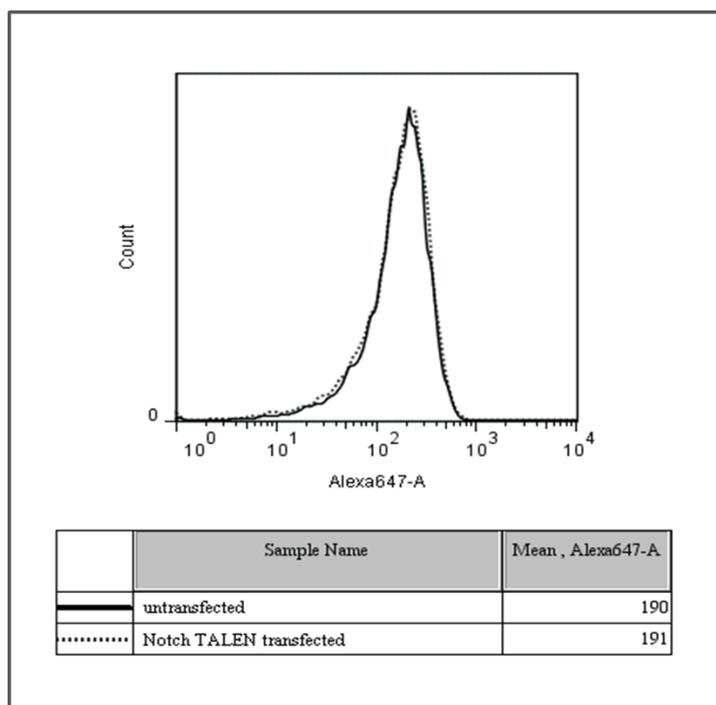


Figure 4.22 IL7R expression levels of untransfected and Notch TALEN transfected RLM11 cells. Histogram for Alexa-647 represents surface IL-7R $\alpha$  expression. Solid black histogram shows expression profile of TALEN untransfected cells whereas dotted line shows Notch TALEN transfected cells

### 4.1.3 Assembly of TALENs Targeting the NF- $\kappa$ B Binding Site of IL7R Using the Golden Gate TALEN Kit

#### 4.1.3.1 Construction of a TALEN pair targeting the IL7R ECR3 NF- $\kappa$ B binding site in the pCAGT7 backbone

A pair of TALENs targeting NF- $\kappa$ B binding site in the ECR3 region of IL7R was designed using TAL Effector Nuclease 2.0 software such that NF- $\kappa$ B left and right TALEN monomers consist of 20 and 18 RVDs, respectively and the spacer region between the TALEN binding sites is 24 bp in length.

We performed Colony PCR using primers pCR8\_F1 and pCR8\_R1 for white colonies obtained after the transformation of Golden Gate reaction #1 into *E.coli*. For the first 10 repeats in pFUS\_A plasmids, bands around 1200 bp were detected. The remaining repeats of the left NF- $\kappa$ B binding site TALEN were cloned into pFUS\_B9, producing bands around 1100 bp, whereas the repeats for the right NF- $\kappa$ B binding site TALEN were cloned into the pFUS\_B7 plasmid, generating bands around 900 bp after colony PCR. We performed AflIII-XbaI control digests for the plasmid DNA isolated

from two colonies with the expected fragment sizes after colony PCR. AflIII-XbaI digest yielded a 1048 bp-band for pFUS\_A plasmids with ten repeats, a 920 bp-band for pFUS\_B9 plasmids with nine repeats and a 720 bp band for the pFUS\_B7 plasmid with seven repeats. Figure 4.23 and Figure 4.24 show agarose gel images of the colony PCR with the pCR8\_F1 and pCR8\_R1 primers, the confirmation digests with AflIII-XbaI double digests and plasmid maps related to Golden Gate reactions A and B of left NF- $\kappa$ B binding site TALEN (NF- $\kappa$ B left TALEN) and the right NF- $\kappa$ B binding site TALEN (NF- $\kappa$ B right TALEN), respectively.

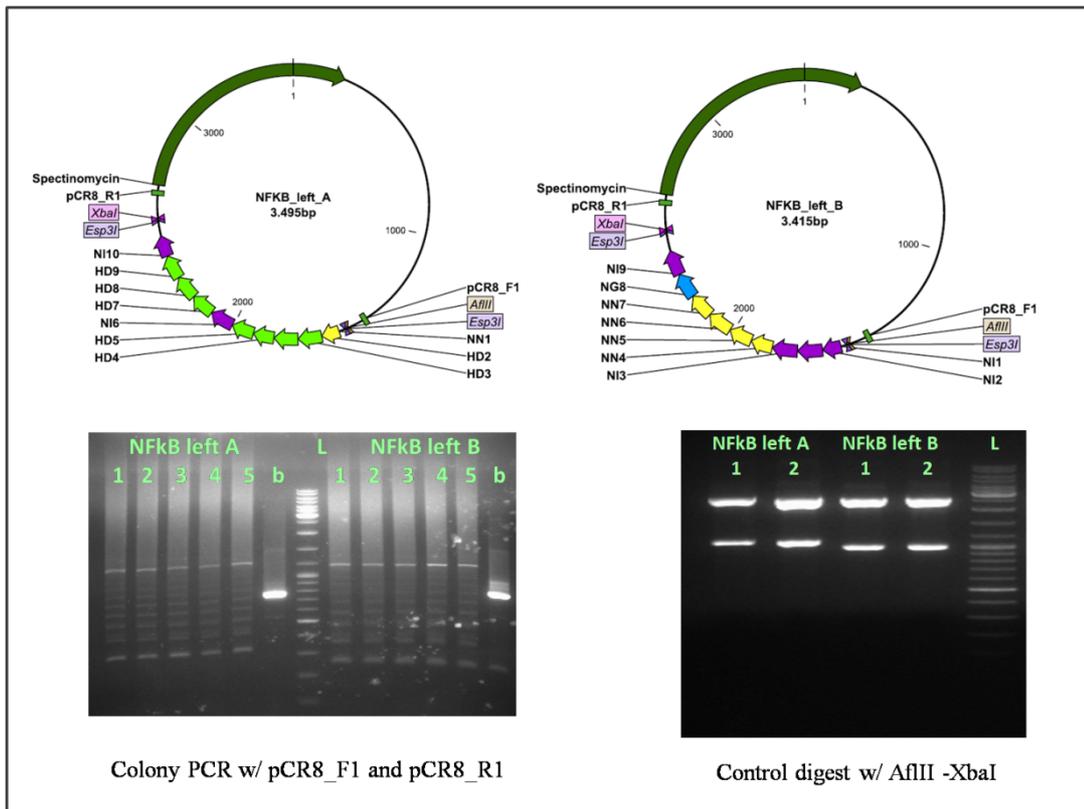


Figure 4.23 Golden Gate reaction#1 for the left NF- $\kappa$ B binding site TALEN

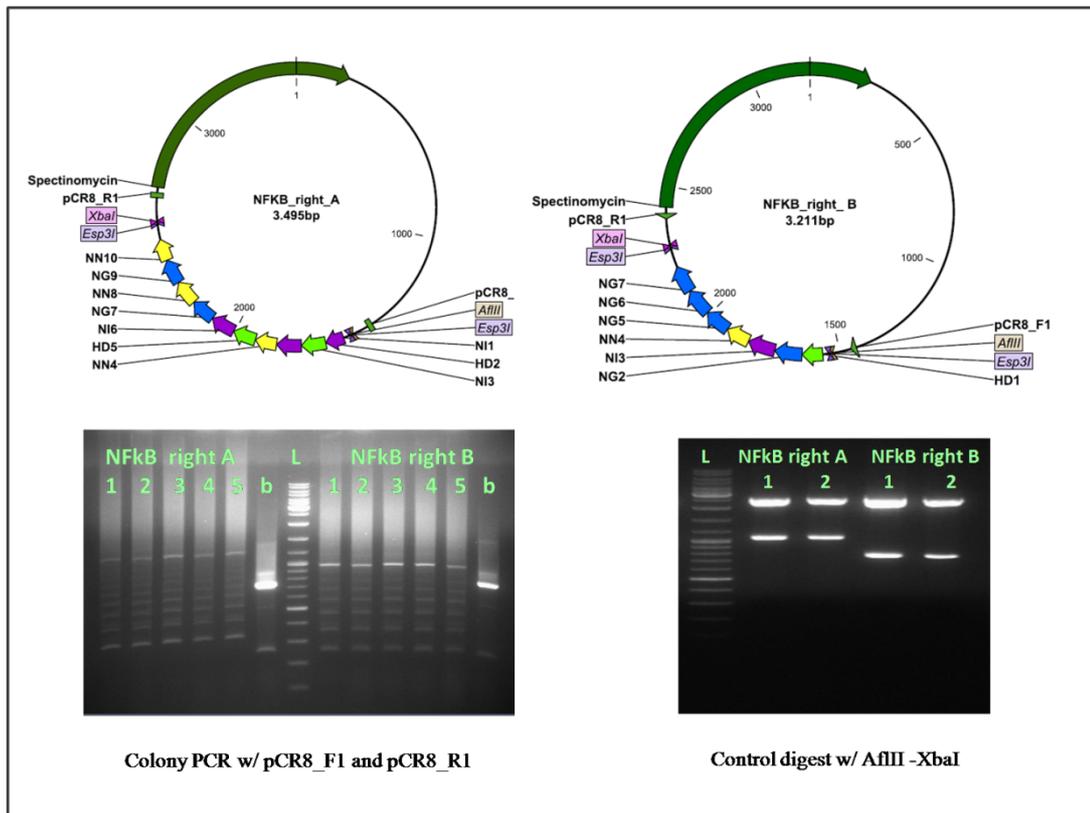


Figure 4.24 Golden Gate reaction#1 for the right NF- $\kappa$ B binding site TALEN

After joining intermediary array of repeats from reaction A and B, together with the last repeat plasmid to the pCAGT7 mammalian expression plasmid in a Golden Gate reaction #2, we performed colony PCR using the TAL\_F1 and TAL\_R2 primers. We obtained 2232 bp bands for the left NF- $\kappa$ B binding site TALEN construct with 20 repeats and a 2028 bp band for the right NF- $\kappa$ B binding site TALEN construct with 18 repeats. We also detected a smear and a ladder effect as expected, consistent with the published Golden Gate protocols. Confirmation digests with AatII and StuI resulted in a 2234 bp- and 2030 bp-band for the NF- $\kappa$ B left and right TALENs, respectively. Figure 4.25 shows agarose gel images after colony PCR with the TAL\_F1 and TAL\_R2 primers, control digest with AatII-StuI and plasmid maps of both the pCAG left NF- $\kappa$ B binding site TALEN and the pCAG right NF- $\kappa$ B binding site TALEN.

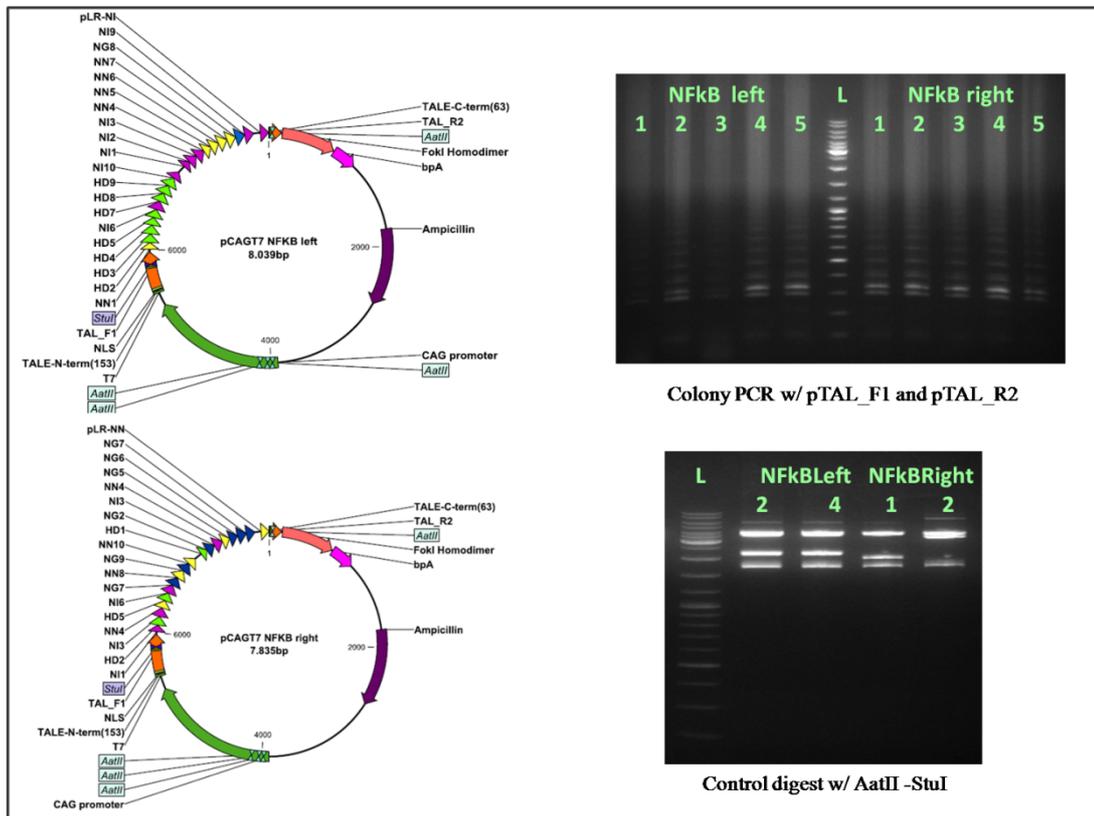


Figure 4.25 Golden Gate reaction#2 for the left and right NF- $\kappa$ B binding site TALENs in the pCAGT7 backbone (Strategy 2a).

#### 4.1.3.2 Construction of a TALEN pair targeting the IL7R ECR3 NF- $\kappa$ B binding site in the pC-Goldy backbone

As before, during the course of this study, we realized that the pC-Goldy backbone generates better TALENs as compared to the pCAG-T7 backbone. Therefore we discontinued the use of our NF- $\kappa$ B site TALENs in the pCAG-T7 backbone and generated NF- $\kappa$ B site TALENs in the pC-Goldy backbone. To do this, we repeated Golden Gate reaction #2 for the left and right NF- $\kappa$ B binding site TALEN monomers using pC-Goldy TALEN as a backbone mammalian expression plasmid. Colony PCR with TAL\_F1 and TAL\_R2 primers produced the expected 2245 bp and 2041 bp bands for the left and right NF- $\kappa$ B binding site TALENs, respectively. AatII–StuI double digests of the pC-Goldy left NF- $\kappa$ B binding site TALEN generated five bands of 3863 bp, 2234 bp, 1074 bp, 87 bp and 57 bp, of which the 2234 bp-band reflects the number of repeats assembled. The 2030 bp-band after confirmation digest of pC-Goldy right NF- $\kappa$ B binding site TALEN indicates the assembly of 18 repeats. Agarose gel images showing the colony PCR, the AatII-StuI confirmation digest and plasmid maps of pC-Goldy left and right NF- $\kappa$ B binding site TALEN plasmids is shown in Figure 4.26.

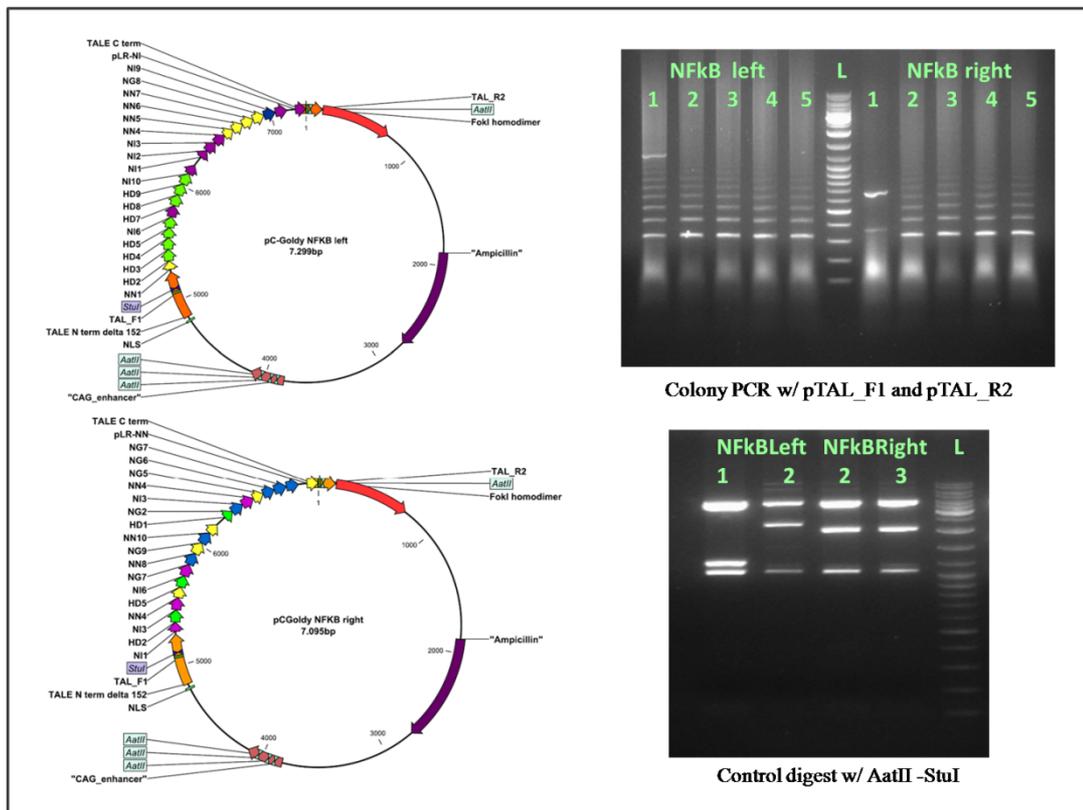


Figure 4.26 Golden Gate reaction#2 for the left and right NF- $\kappa$ B binding site TALENs in the pC-Goldy backbone (Strategy 2c)

#### 4.1.3.3 Expression of the designed NF- $\kappa$ B binding site TALEN pairs in RLM11 cells and detection of site-specific mutations

We transfected RLM11 cells with NF- $\kappa$ B binding site TALEN pairs in the pC-Goldy backbone for ectopic expression which would result in mutation at NF- $\kappa$ B binding site of the IL7R gene locus. We performed our modified restriction fragment length polymorphism (RFLP) assay to detect mutations at the target site. Figure 4.27 shows the locations of the binding sites of NF- $\kappa$ B binding site TALEN pair, the forward and reverse primers for amplification of the target site, and the restriction enzyme cut sites present in the spacer region.

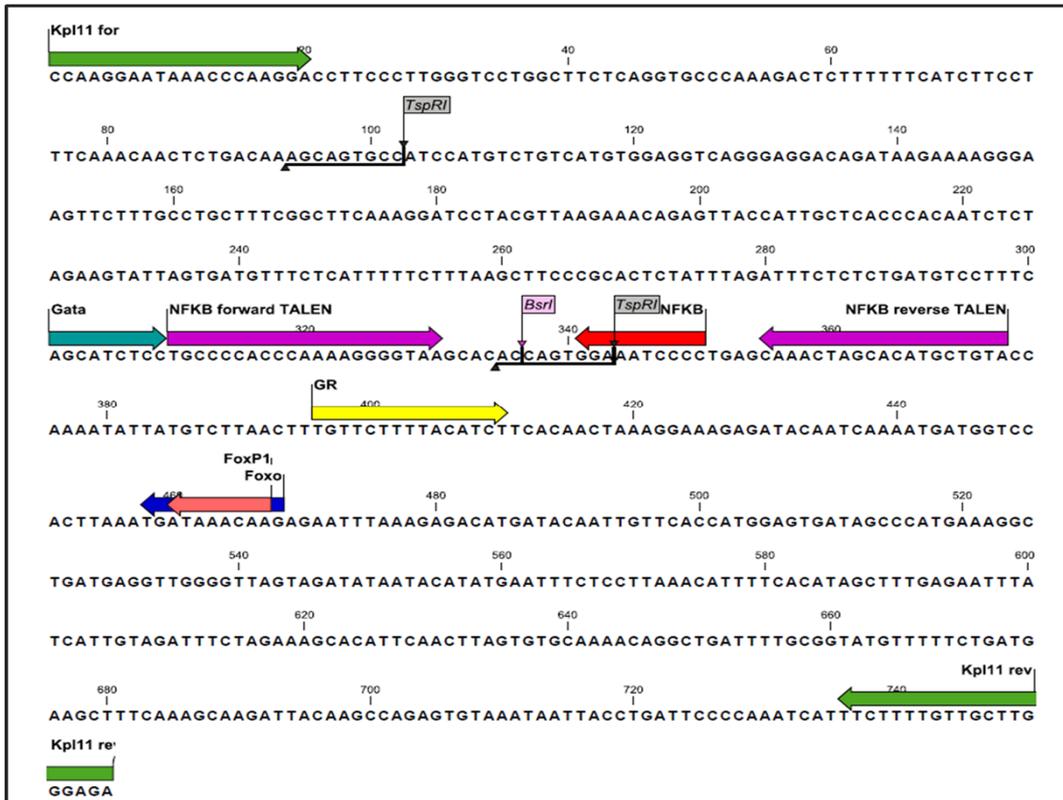


Figure 4.27 Binding site of the assembled TALEN pair targeting the NF- $\kappa$ B binding site of the IL7R enhancer region. Purple arrows indicate the left and right TALEN binding sites. Red arrow indicates the NF- $\kappa$ B binding site. Green arrows indicate the primer binding sites.

We digested a 777 bp PCR products of BsrI digested and undigested genomic DNA with BsrI to determine the presence of mutated alleles in the pool of NF- $\kappa$ B binding site TALEN transfected cells. Wild type PCR products generated of 421 bp 336 bp fragments after BsrI digestion (Figure 4.28). We extracted the uncut band and cloned it into the pJET1.2/blunt vector for sequencing.

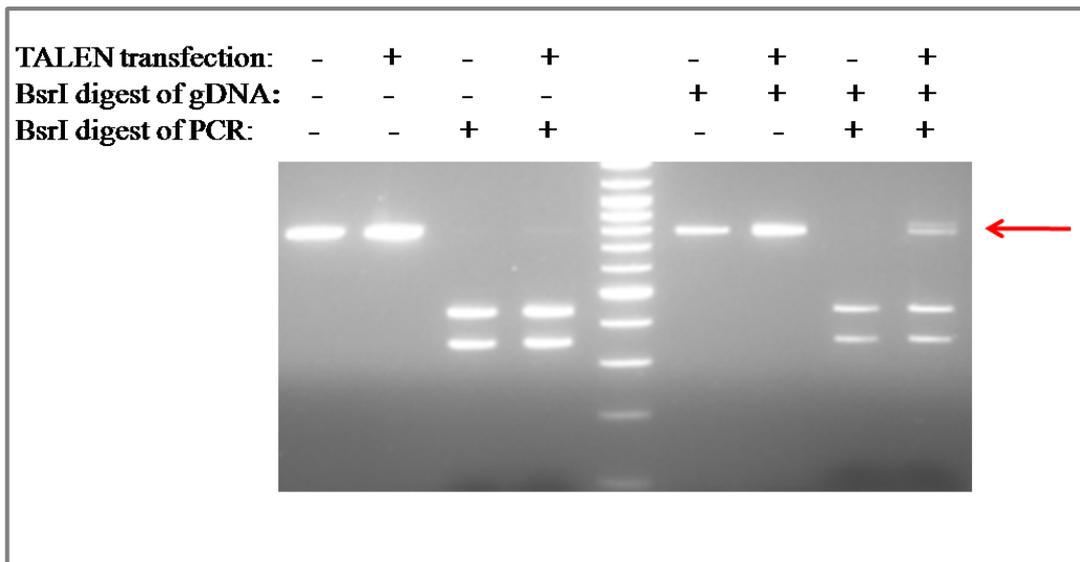


Figure 4.28 Mutation detection at NF- $\kappa$ B TALEN target site of RLM11 cells using modified RFLP assay

Sequencing results for 40 different colonies reveal the presence of 4 different mutations (Figure 4.29). Three of these contain a deletion of 15 bp. Moreover, transition and transversion mutations were detected in all mutants sequenced except one. Note that the NF- $\kappa$ B binding site was not mutated in any of the mutants sequenced.

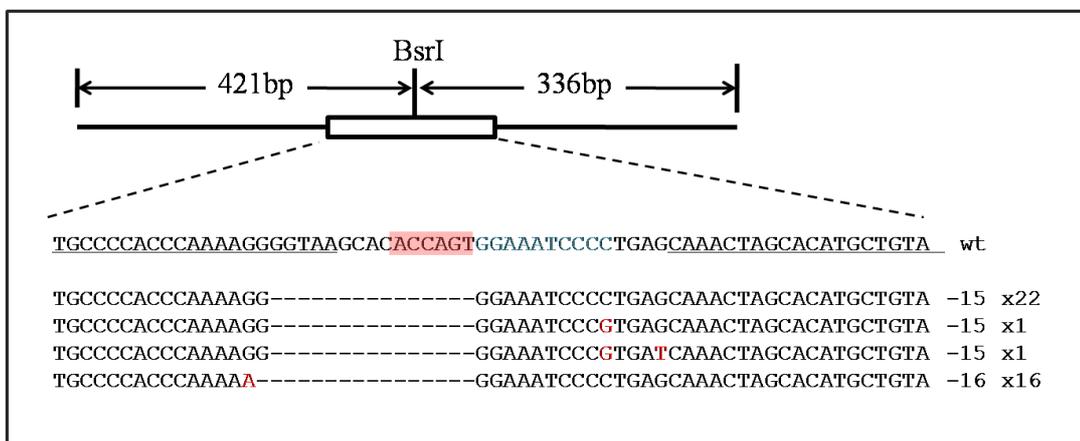


Figure 4.29 Site directed mutagenesis in RLM11 cells transfected with TALEN pair targeting NF- $\kappa$ B binding site. Blue color in wild type sequence indicates NF- $\kappa$ B binding site and underlined sequence represent binding sites of TALEN monomers. BsrI cut site is highlighted in pink. Dashes indicate deletions whereas red colored bases are for SNPs. Numbers at right-hand side gives number of bases deleted and occurrence of this mutation, respectively.

#### 4.1.3.4 Expression of IL7R on RLM11 cells transfected with NF- $\kappa$ B binding site TALEN pairs

Although the NF- $\kappa$ B site was not mutated in any of the sequenced mutants, we nevertheless decided to assess IL7R expression in these mutant RLM11 cells. We hypothesized that the large 15bp deletion would result in a block of enhancer activity and that we would observe a decrease in IL7R expression in these mutant cell containing pools of RLM11 cells. As before we assessed IL7R levels by FACS and found that expression of these surface proteins was less in mutant cell pools than that of untransfected RLM11 cells (Figure 4.30).

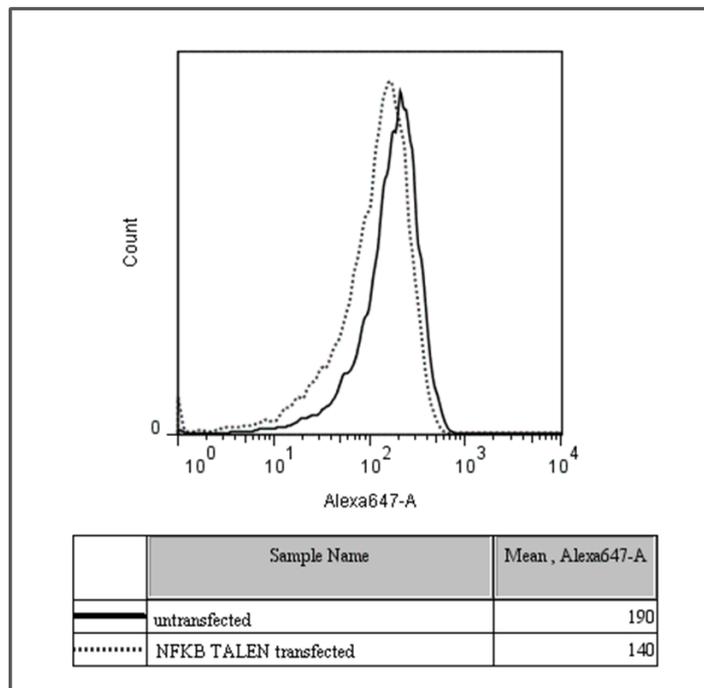


Figure 4.30 IL7R expression levels of untransfected and NF $\kappa$ B TALEN transfected Rlm11 cells. Alexa 647 mean shows the relative IL7R expression on the cell surface.

Solid black line shows the expression profile TALEN untransfected cells whereas dotted line shows of NF- $\kappa$ B TALEN transfected cells.

## 4.2 Use of TALE as Competitive Inhibitors

TNF $\alpha$  treatment results in phosphorylation and subsequent ubiquitinylation of I $\kappa$ B. Degradation of I $\kappa$ B activates the NF- $\kappa$ B complex resulting in translocation to the nucleus. A reporter plasmid containing four NF- $\kappa$ B binding sites upstream of fos promoter and GFP sequence was constructed by Pinar Onal (2007) in the Erman Lab. Induction of cells transfected with the 4 $\kappa$ B GFP constructs with TNF $\alpha$  results in the production of GFP due to the binding of nuclear translocated NF- $\kappa$ B upstream of the GFP gene. The 4XNF- $\kappa$ B binding site-fos-GFP cassette was stably integrated into the

genome of HEK293 cells by Belkıs Atasever Arslan, in the Erman Lab and this reporter cell line was named HEK293 6.1.1.

TALE DNA binding domains were constructed to competitively inhibit NF- $\kappa$ B binding to DNA. We hypothesized that expression of TALE proteins in reporter cells would inhibit the binding of nuclear translocated NF- $\kappa$ B to its binding sites in this reporter, after TNF $\alpha$  induction. Our expectation was that the level of GFP expression upon TNF treatment in cells transfected with TALE proteins would be lower than in untransfected reporter cells (Figure 4.31). We designed this system by fusing the TALE proteins to a fluorescent protein, dsRed, in order to track transfected cells.

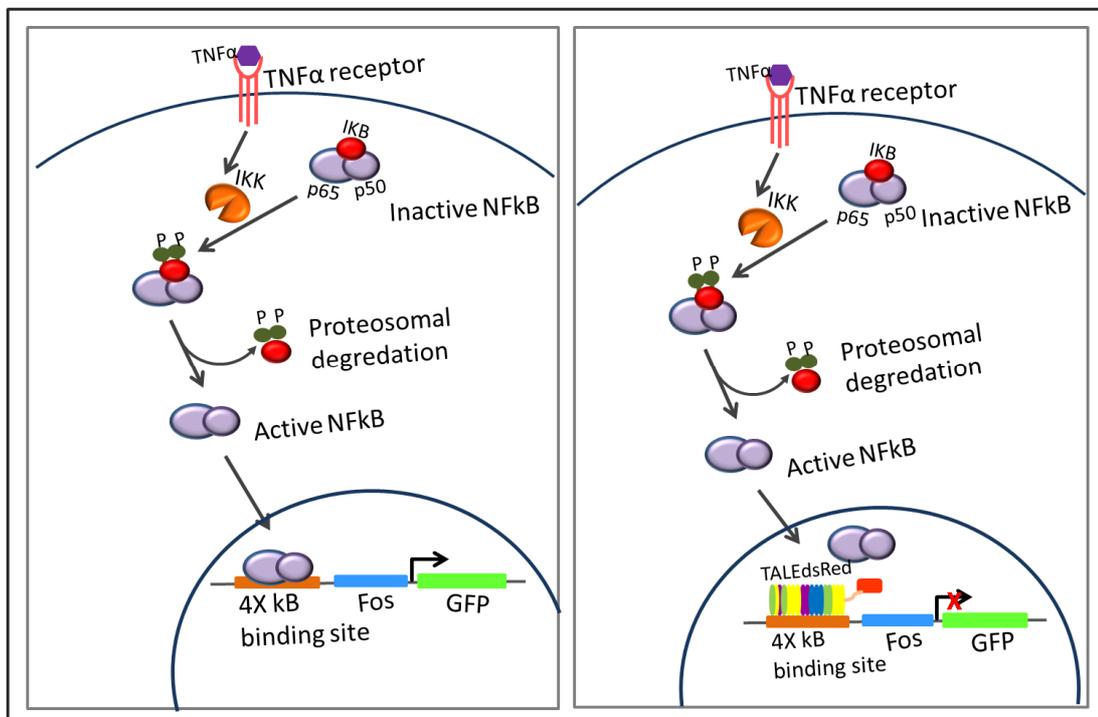


Figure 4.31 Strategy of designing TALEdsRed constructs as competitive inhibitors. TNF $\alpha$  treatment of HEK293 6.1.1 cells, containing stably integrated 4X kappaB binding site upstream of a fos promoter and GFP, results in GFP expression whereas TALEdsRed expression in these cells is expected to inhibit GFP expression.

#### 4.2.1 Construction of pCAGdsRed Plasmid

To construct TALEdsRed competitive inhibitors, the cloning strategy for backbone plasmid generation was designed such that dsRed in the pBluescript-dsRed plasmid (Plasmid was previously constructed in Erman Lab for an unrelated project.) was replaced with in the pCAGT7 plasmid (Figure 4.32).

We digested the pCAGT7 plasmid with BamHI, blunted using the Klenow enzyme and digested with SacI, which should generate fragments of 5845bp and 604bp. The 5845bp fragment corresponding to the plasmid backbone was gel extracted. The pBluescript-dsRed plasmid was digested with the SmaI-SacI restriction enzyme pair, producing bands with sizes of 2921bp and 752bp. The 752bp-band corresponding to dsRed sequence was gel extracted. SmaI generates blunt-ended fragments, making ligation to the fragment blunted after BamHI digest possible. After ligation, NotI confirmation digest was performed for the plasmid DNA extracted from colonies. NotI digestion should generate bands of 6267 bp and 330 bp for correctly ligated plasmid DNA. As a second control, PstI-Bgl2 double digestion was performed, producing three bands of 5177 bp, 1219 bp and 205 bp. Figure 4.32 shows agarose gel images of confirmation digests and a map of the constructed plasmid.

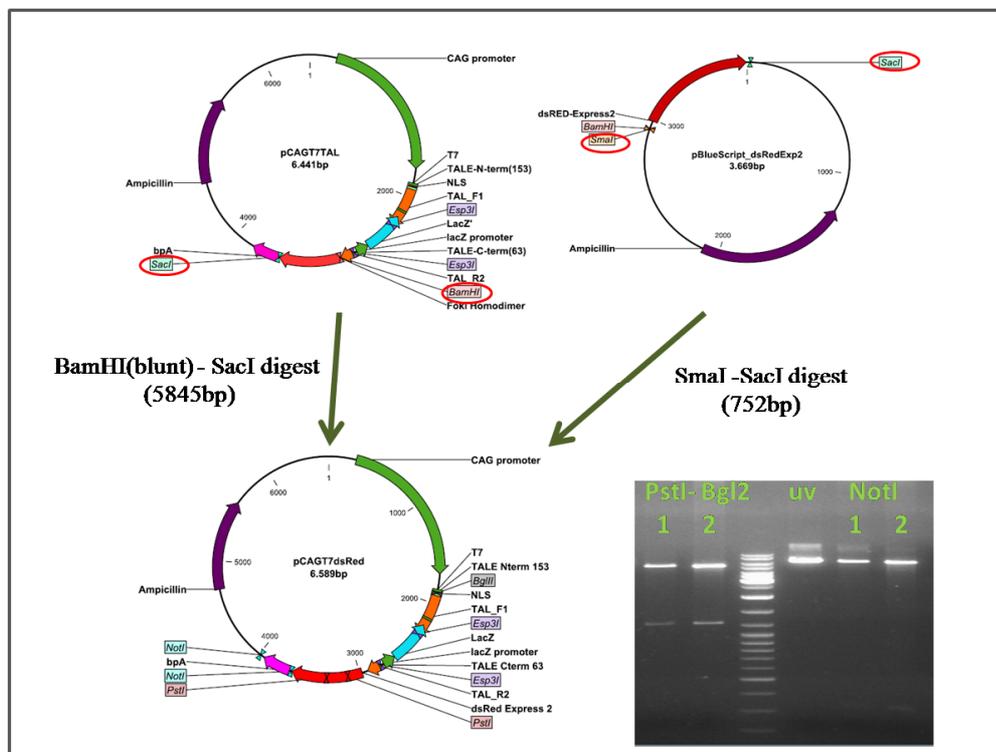


Figure 4.32 Cloning strategy for construction of the backbone plasmid with a fluorescent reporter, pCAGdsRed.

#### 4.2.2 Golden Gate TALE Assembly of the Competitive Inhibitors of NF- $\kappa$ B Binding Using the pCAGdsRed Plasmid as Backbone Vector

Six different TALEdsRed constructs that have specificities to bind to NF $\kappa$ B binding sites were designed using the Golden Gate TALEN kit. The number of repeats range from 12 to 17 such that the first 12 RVDs were same in all constructs. Therefore,

in the Golden Gate reaction#1 only one pFUS\_A plasmid containing repeat was constructed for all six TALEdsRed constructs and different pFUS\_B plasmids were generated for each. Colony PCR was performed for white colonies obtained after transformation of reaction#1 using primer pair pCR8\_F1 and pCR8\_R1. For repeats in pFUS\_A plasmids, bands around 1200 bp were detected whereas for each pFUS\_B plasmids, band sizes changed from 350 bp to 820 bp according to the number of repeats in that array plasmid. AflIII-XbaI confirmation digest was performed for colonies producing expected bands after colony PCR. A 1048 bp-band was obtained for pFUS\_A plasmid with 10 repeats. Bands from 147 bp to 622 bp were obtained for the NF- $\kappa$ B reporter B1 to NF- $\kappa$ B reporter B6, with sizes increasing approximately 100 bp for each repeat. Figure 4.33 and 4.34 show agarose gel images after colony PCR with pCR8\_F1 and pCR8\_R1, confirmation digest with AflIII-XbaI and plasmid maps of reaction A and reaction B of TALEdsRed constructs.

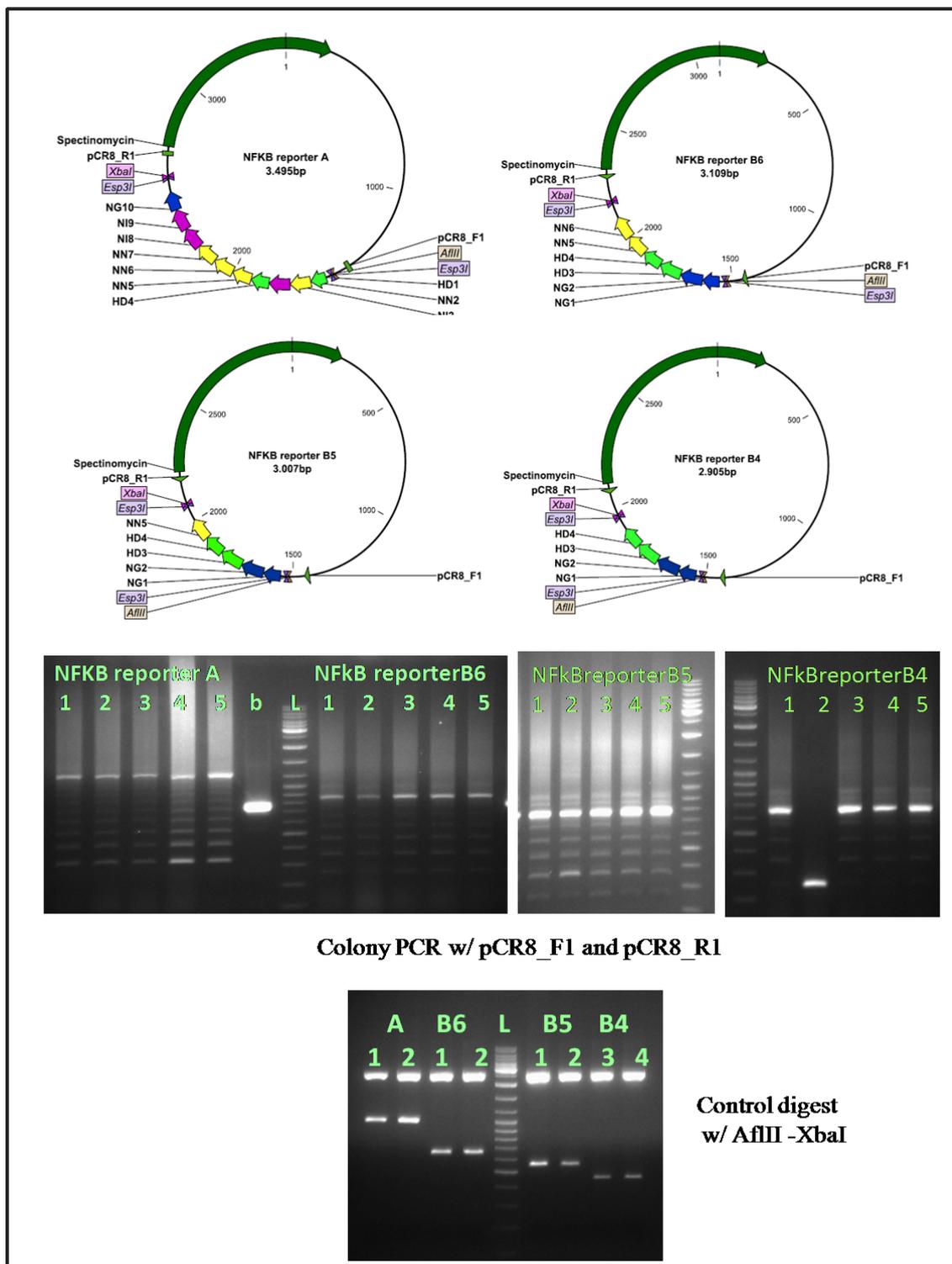


Figure 4.33 Golden Gate reaction#1 for NF-κB reporter A, B6, B5 and B4 plasmids.

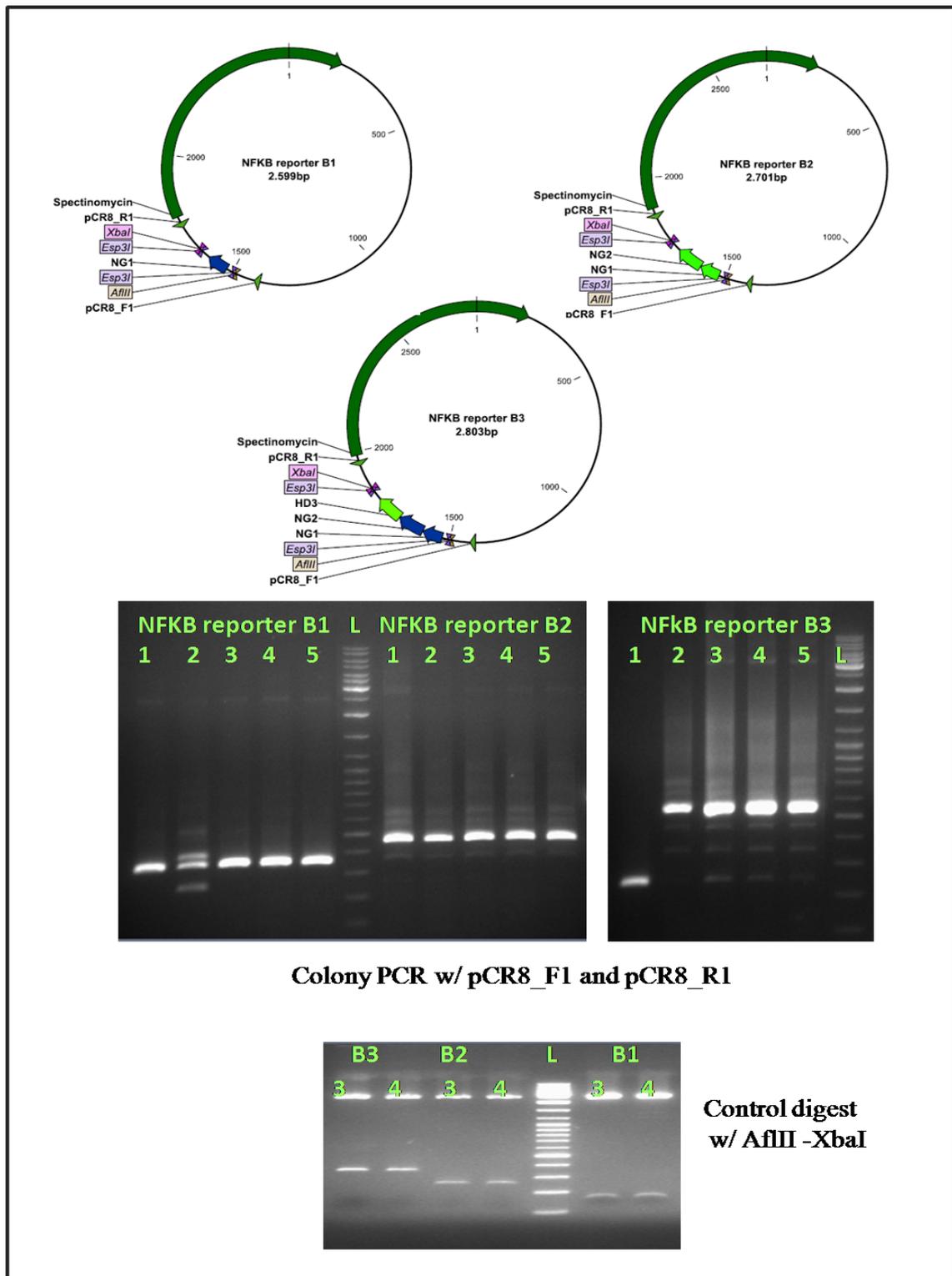


Figure 4.34 Golden Gate reaction#1 for NF-κB reporter B1, B2 and B3 plasmids.

For construction of each TALEdsRed plasmid, a pFUS\_A plasmid together with a pFUS\_B plasmid with corresponding repeats and the last repeat plasmid were joined to the pCAGT7dsRed final expression plasmid in the Golden Gate reaction #2. Colony

PCR was performed for colonies obtained after transformation using pTAL\_F1 and pTAL\_R2. The size of bands obtained after colony PCR ranged from 1400 bp to 1900 bp for plasmids TALEdsRed12 to TALEdsRed17 such that one repeat increases the size 100 bp. Confirmation digest with AatII-StuI enzyme pair was performed for colonies generating expected band sizes. The size of the bands obtained after digestion was the same with that obtained after colony PCR. Agarose gel images related to colony PCR and confirmation digests, and plasmid maps of the final TALEdsRed constructs are shown in Figure 4.35, 4.36 and 4.37.

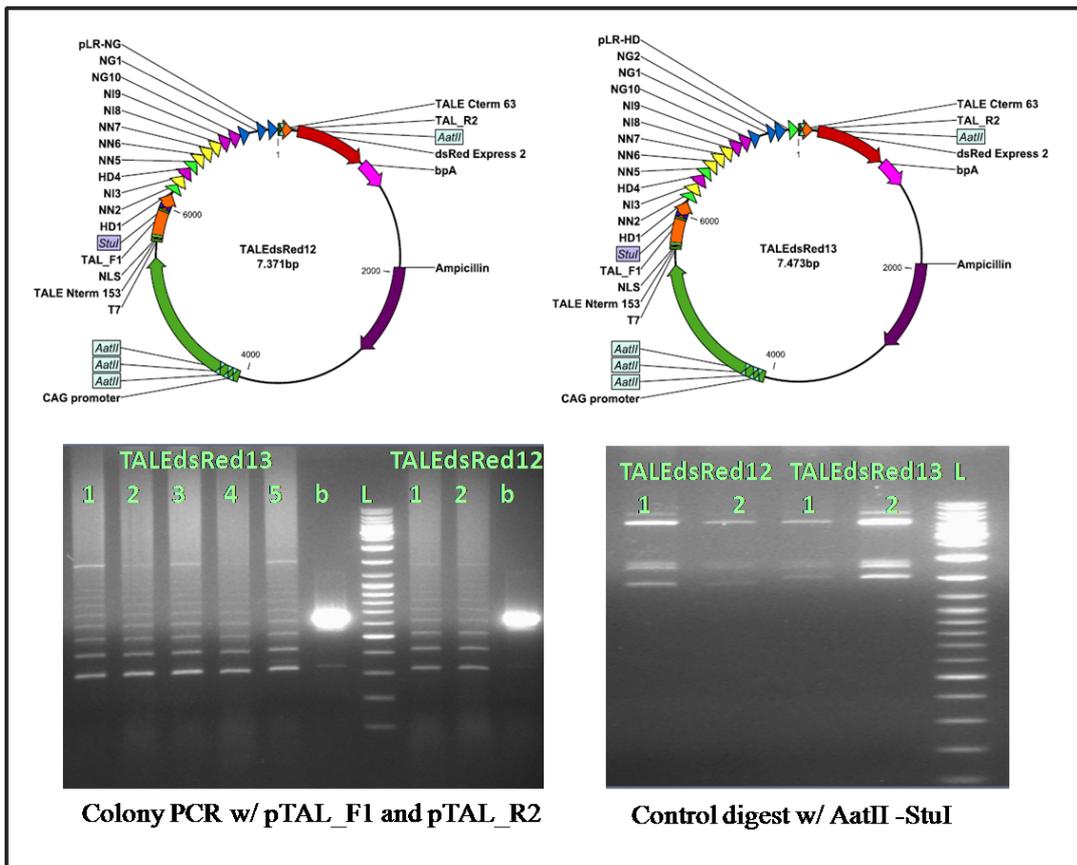


Figure 4.35 Golden Gate reaction #2 for TALEdsRed12 and TALEdsRed13 plasmid construction.

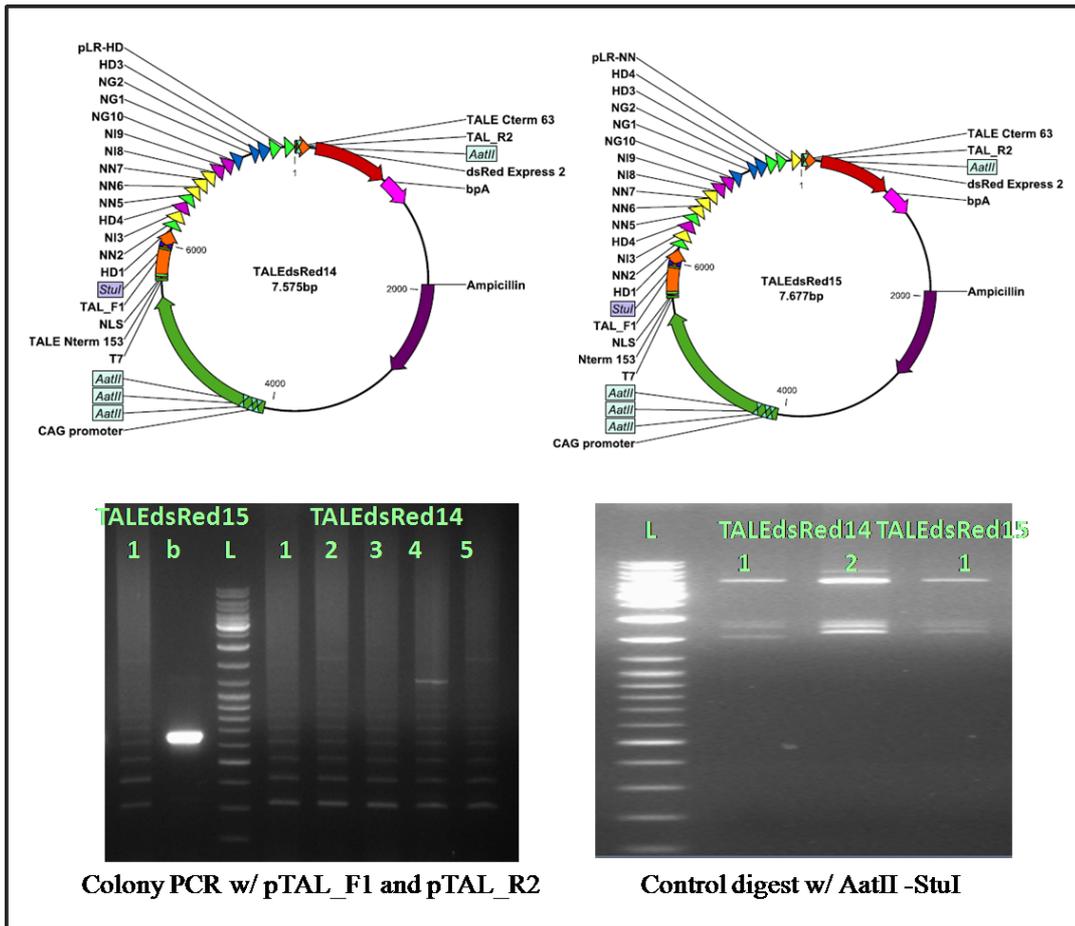


Figure 4.36 Golden Gate reaction #2 for TALEsRed14 and TALEsRed15 construction.

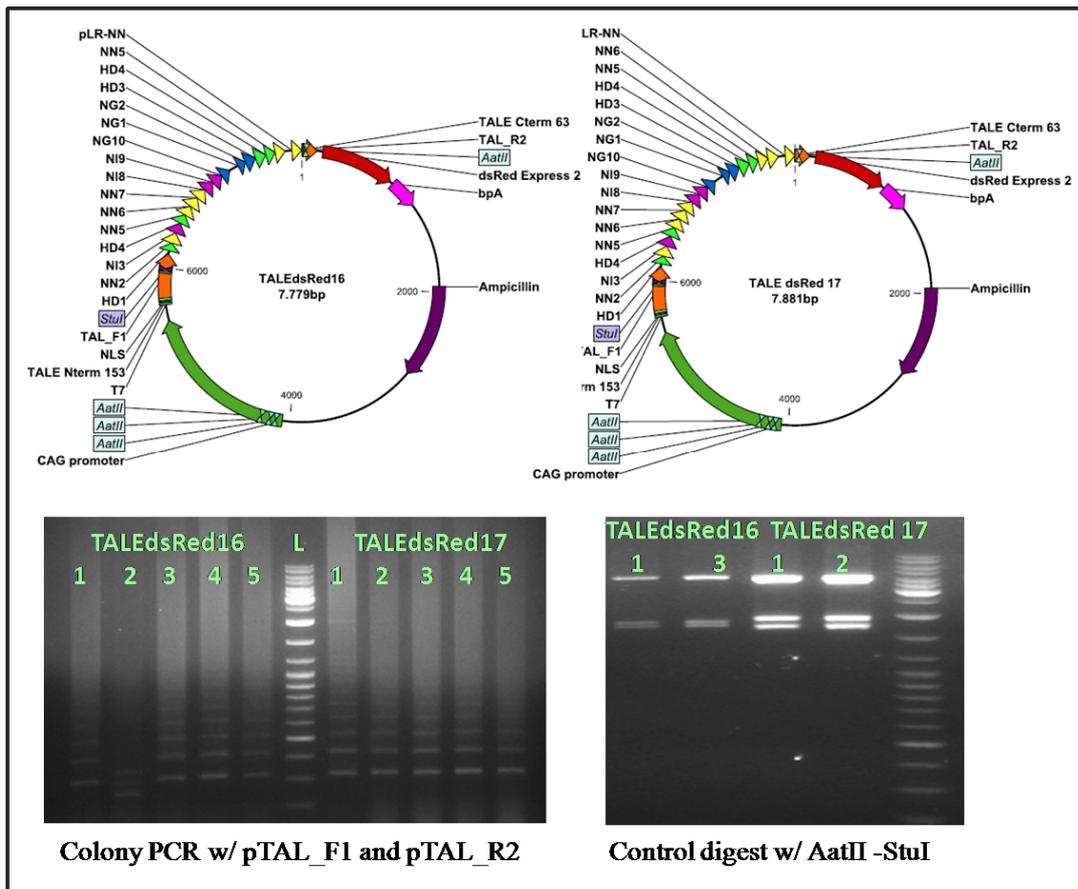


Figure 4.37 Golden Gate reaction #2 for the TALEdsRed16 and the TALEdsRed17 construction.

#### 4.2.3 Expression of TALE-dsRed Constructs in HEK293 6.1.1 Cells and the Effect on GFP Expression

To determine whether expression of TALE dsRed constructs affects GFP expression in HEK293 6.1.1 cells upon TNF $\alpha$  treatment, designed TALEdsRed constructs were transfected to these cells and TNF $\alpha$  treatment was performed approximately 9 hours before FACS analysis. Figure 4.38 shows the net GFP fluorescence means of cells transfected with TALEdsRed constructs. The DsRed positive population of all NF- $\kappa$ B transfected cells has a lower GFP fluorescence mean than that of the untransfected samples.

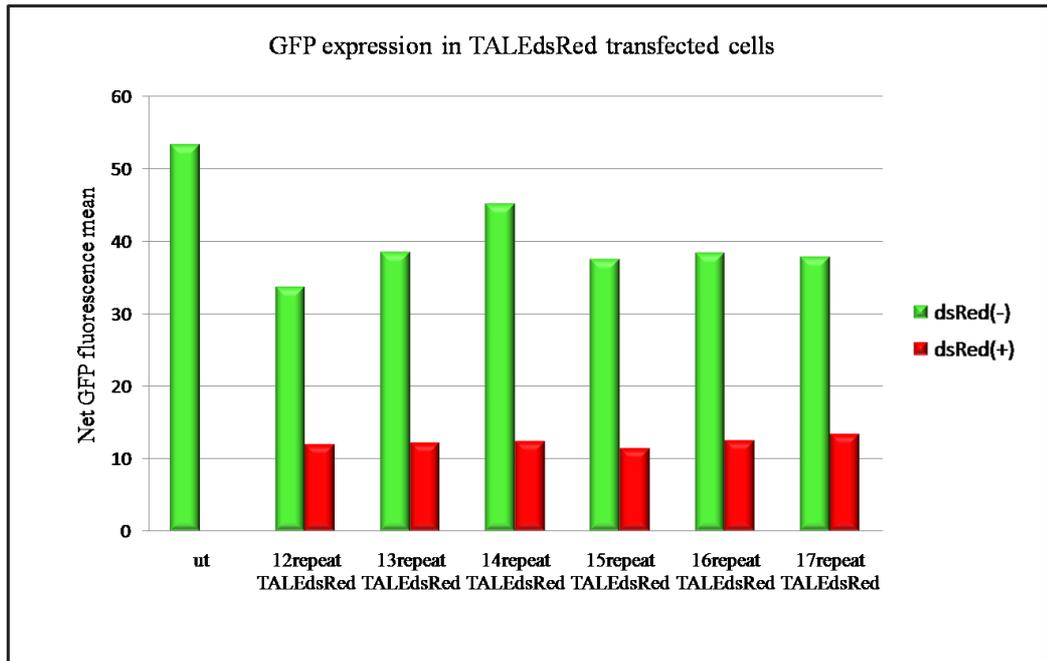


Figure 4.38 GFP expressions of TALEdsRed transfected HEK293 6.1.1 cells after  $TNF\alpha$  treatment. y-axis shows net GFP fluorescence mean. Green bars show the net GFP mean for dsRed negative cells and red bars show the net GFP mean for dsRed positive cells. ut: untransfected.

## 5. DISCUSSION

Transcription activator-like effector (TALE) proteins are recently described DNA binding proteins, isolated from the plant pathogen *Xanthomonas*[1]. The DNA binding domain of TALE proteins is composed of 34 amino acid long repeat units where the 12th and 13th residues (repeat variable di-residues, RVDs) of each repeat determine the binding specificity of the protein[3, 4]. A simple one RVD to one base code and the modularity of repeat arrays enable researchers to construct TAL effector proteins to bind desired DNA sequences. Fusion of the TALE DNA binding domain to the catalytic domain of the FokI restriction enzyme makes them an efficient tool for genome engineering as site-directed nucleases[5, 6].

Interleukin-7 signaling is crucial for development, differentiation and survival of lymphocytes. The expression of IL-7R $\alpha$  on lymphocytes changes during their development. Upstream promoter and enhancer regions of the IL-7R $\alpha$  gene locus contain binding sites for transcription factors that have roles in regulating this expression[66, 68, 69, 73]. In order to understand the roles of these transcription factors in the regulation of IL-7R $\alpha$  gene expression in lymphocytes, we mutated the binding sites of three transcription factors; namely, glucocorticoid receptor (GR), Notch and NF- $\kappa$ B, using TALEN technology. In addition, we constructed TALE-dsRed fusion proteins that competitively bind in place of NF- $\kappa$ B.

Our first TALEN pair, targeting the GR binding site of the IL-7R $\alpha$  gene locus was synthesized and purchased from a company. We cloned both the forward and reverse TALEN monomers into a retroviral vector backbone encoding fluorescent reporter proteins to create stable cell lines expressing the TALEN pair after infection. These retroviral constructs utilized internal ribosome entry site (IRES) elements between the TALEN cDNA and the fluorescent protein encoding genes. This enabled us

to express TALEN monomers and fluorescent marker proteins (eGFP and dsRed for the forward and reverse TALENs respectively) in the same cells and allowed the identification of TALEN expressing cells. However, we could not efficiently infect the NIH 3T3 cell line using these TALEN retroviral constructs. Although we achieved approximately 40% infection efficiency with virus generated from control vector backbones, demonstrating the infectibility of these NIH3T3 cells, TALEN encoding vectors (which are considerably larger) could not produce infectious virus. This indicates that the large molecular weight of TALEN constructs could be the limiting factor for virus production in the Phoenix cell line.

Another possibility is that infected and stably TALEN expressing cells could not survive due to the genotoxic effects of TALENs. But this is an unlikely possibility as we could later generate TALEN expressing (albeit not stably) cell lines. Possibly, off-target mutations may result in cells stably expressing TALENs to be at a relative disadvantage to untransfected cells or to those cells that lost TALEN expression. During the course of this study we realized that in fact stably expressing TALENs is not desired. To reduce off target effects, TALENs should only transiently be expressed in cells. After inducing double stranded breaks (DSBs) at the desired endogenous loci, the expression of TALEN constructs is no longer necessary and may be deleterious to the mutation containing cells. Therefore, we decided to express TALEN constructs ectopically.

Although transient transfection and expression could be achieved using our retroviral plasmid constructs, used for virus production, we generated new mammalian expression plasmids containing TALEN-IRES-fluorescent reporters. These constructs contained a CMV promoter, which is much stronger than the LTR of the retroviral vector construct. The transfection efficiency of Neuro-2a cells transfected with both an IRES-eGFP and IRES-dsRed TALEN expressing plasmids was around 30%. This is half of the normal transfection efficiency in this cell line. Perhaps the size of the fluorescent TALEN pair constructs limits the transfection of these cells, as in the case for our infection experiments. During the course of this study, we learned that efficient TALEN cleavage is influenced greatly by the presence of a C-terminally truncated protein. Unfortunately, the TALEN pair targeting the IL7R GR binding site was in a scaffold containing long C-terminal segments [6]. In addition, the GR TALEN pair was

synthesized using NK RVDs to target G bases. We also recently learned that NK containing RVDs have a lower affinity for G bases as compared to NN RVDs [14]. For these reasons, we discontinued our experiments with the commercially designed TALEN pair targeting the GR binding site and focused our attention to the TALEN pairs targeting other sites in the IL7R gene enhancer. We constructed these second generation TALENs to contain only NN RVDs to target G's and in the plasmid backbones (pCAGT7 and pC-Goldy) which have the appropriate C-terminal truncation necessary for efficient DNA cutting.

Thus, we assembled TALEN constructs targeting the Notch and NF- $\kappa$ B binding sites using the Golden Gate TALEN kit. The pCAGT7 plasmid that contains a CAG promoter for mammalian expression was used as a final destination plasmid in TALEN assembly. Formation of homodimers (instead of heterodimers) by TALEN monomers decreases the cutting efficiency and results in off-target cleavage. Research on FokI enzyme dimerization indicated that homodimerization could be prevented by using mutants in the dimerization interface. New TALEN backbone plasmids were developed during the course of this study which we acquired, that use mutant FokI domains for each TALEN monomer[45]. We used these mutant FokI destination plasmids when generating TALEN monomers targeting the Notch binding site. More recently, the Goldy TALEN scaffold was reported to have improved genome editing efficiency compared to other TALEN encoding backbones[35]. Goldy TALENs contain drastically truncated N and C terminal domains. This modification is likely necessary for the assembly of the TALEN protein on the target site and the appropriate orientation of the FokI domains on the spacer region. We therefore transferred both the Notch and NF- $\kappa$ B TALENs into the Goldy TALEN backbone.

Indeed, in experiments where we compared the cutting efficiency of TALENs targeting the Notch site, with identical DNA binding domains expressed in the context of pCAGT7, pCAGT7-mutant FokI and pC-Goldy TALEN backbones, we found that Goldy TALENs were the best cutters! (Figure 4.21) These experiments were conducted in RLM11 cells. The intensity of the uncut band was found to be the highest for the Notch TALEN pair in the Goldy backbone, consistent with recently published results. We found that band intensity was the lowest for TALEN pairs in the backbone with mutant FokI domains. This result is not consistent with previously published work, as

the mutant FokI domain was shown to have increased cutting efficiencies as well as higher specificity. Further evaluation of the cut site efficiency by normalizing the amount of expressed TALEN proteins may be necessary to explain this discrepancy.

In addition to scaffold optimizations during TALEN assembly, there are some critical points that have to be taken into consideration for detecting mutations generated by these proteins. First of all, in all TALEN expression experiments, we subjected cells to transient hypothermia after transfection, by incubating at 32°C instead of 37°C. Transient hypothermia was previously shown to increase mutation efficiency both in zinc finger nucleases and in TALENs[6, 82]. This effect likely increases the folding efficiency of TALEN proteins that have to adopt a super helical structure. It is also possible that decreased temperatures decreases the progression of the cell cycle, giving more time for the generated DSBs to be repaired by NHEJ. It would therefore be interesting to test the mutation profile of cells that express TALENs in the presence and absence of hypothermia. We also harvested, transfected cells 72 hours after transfection to allow for time necessary for TALEN expression, double stranded break (DSBs) generation and NHEJ mediated repair of induced DSBs. It may be necessary to optimize the time of incubation to prevent toxicity to cells that may result in an apparent increase in the mutation rate.

Detection of induced mutations was another critical rate limiting point in this project. Transfection efficiency, efficiency of TALENs to induce double stranded breaks, the capacity of the transfected cell line for DNA repair and the mutation of either one or both chromosomes, the unknown ploidy of the cancer cells being used are all parameters that influence the number of mutant alleles in genomic DNA isolated from pooled cells. All of these listed parameters may result in restriction resistant PCR products in the restriction fragment length polymorphism (RFLP) assay. In order to detect a mutated allele from a genomic DNA pool of cells, we modified the RFLP assay by first digesting the genomic DNA with the enzyme in the spacer region, before amplifying the genomic DNA by PCR and re-digestion with the RFLP enzyme. We reasoned that the digestion of genomic DNA would decrease the number of wild type alleles in the in the pool of genomic DNA to be amplified and that the relative fraction of mutant alleles would increase. Theoretically, the initial restriction enzyme digest of

genomic DNA should remove all wild type alleles from the PCR template pool, but digestion of genomic DNA may not be as efficient as digestion of purified plasmid DNA. In figures 4.19, 4.21, and 4.28, we show experiments where we PCR amplified the TALEN target region from digested genomic DNA used as a template followed by a second restriction digestion of the PCR amplicon. In this modified RFLP assay we could detect restriction resistant mutant alleles that show up as an uncut band with the same size of the undigested PCR product. Because we had to use this enrichment step, we could not quantify the mutagenesis frequency in the pool of cells that contain TALEN induced mutations [5]. All the mutations detected in this study for both the Notch and NF- $\kappa$ B TALENs involve this enrichment step.

Actually, the modified RFLP assay underestimates the number of mutations created at the TALEN target site because only mutant alleles that are restriction enzyme resistant are scored. It is likely that there are many mutant alleles generated along those that lose the restriction enzyme binding site, that do not lose this site. Because our sequencing results are derived from restriction enzyme resistant PCR amplicons cloned from the modified RFLP assay, these results are also underestimating the number of possible mutants in the pool of TALEN treated cells. Therefore we think that both the RFLP and the sequencing results we provide are an underestimation of the mutant alleles produced by these TALENs.

We tested the induction of mutations by TALENs targeting the IL7R Notch site in both the Neuro-2a and RLM11 cells using the modified RFLP assay. Neuro-2a cells were not appropriate for phenotypic analysis as they are neurons that do not express IL7R $\alpha$ . However, previous work shows that NHEJ is very efficient in this cell line, making them a good candidate for detecting TALEN induced mutations. The RLM11 cell line is IL7R $\alpha$  positive and the phenotypic effect of induced mutations could be analyzed phenotypically. Our sequencing results provide evidence for mutant alleles that have deletions, insertions and single nucleotide polymorphisms at the target site. The most frequent mutation we detected was a deletion of one nucleotide. Indeed the same nucleotide was deleted in both the RLM11 and Neuro-2a cell lines. This may indicate that during NHEJ, mutations occur in an additive manner, starting with one base pair deletion and evolving to have various additional mutations. In addition, the frequency of mutations with 4 bp and 9 bp deletions were the same for both cell lines.

Although the number of colonies screened for the Neuro-2a cell line is more than that for RLM11 cells, the number of different mutations detected was higher in the RLM11 cell line. In fact the number of observed SNPs was also higher in RLM11 cells. This result may indicate that IL7R $\alpha$  expression may be priming the genomic locus to become more susceptible to mutation. It is well known that transcriptionally active gene loci are devoid of histone proteins. Thus it is possible that the target site is more available for TALEN binding and DSB generation in the IL7R expressing RLM11 cell line compared to non-expressing Neuro-2a cells. In a recent study, it was demonstrated that the use of constructed TALE proteins together with epigenetic inhibitors enhanced transcriptional activation rate of epigenetically silenced genes [83]. Moreover, it was shown that the TALEN induced mutation rate was negatively correlated with the number of CpG targets in the target site[34]. Our results that indicate a higher frequency of mutations in the target site in cells expressing the IL7R are consistent with these results.

To analyze the mutations generated by the NF- $\kappa$ B TALENs transfected to RLM11 cells in the Goldy backbone, we cloned and sequenced the uncut band obtained after the modified RFLP assay. We found four different mutations with deletions of either 15 or 16 nucleotides and one or two SNPs were detected in the 40 colonies screened. As expected, the restriction site in all sequenced clones was destroyed. Surprisingly, the NF- $\kappa$ B binding site was not destroyed in any of the mutants. This result seems contradictory to our finding that IL7R expression was decreased in these clones. It is possible that yet another transcription factor right next to the NF- $\kappa$ B binding site is present in this enhancer, and that this transcription factor is necessary for IL7R expression. Although the NF- $\kappa$ B binding site was intact, a 15 bp-deletion corresponds to one and half turn of the DNA helix and this modification may in fact result in a loss of a necessary interaction between NF- $\kappa$ B and close-by nuclear factors binding to the enhancer. It is well known that enhancers function by multiple transcription factors binding in close proximity on the DNA. The concept of an enhanceosome requires that these DNA binding factors fit like pieces of a puzzle and that any alteration in the spacing between the binding sites dramatically affects transcription factor interaction and enhancer activity.

Although we observed a small decrease in the IL-7R $\alpha$  expression in NF- $\kappa$ B binding site TALEN transfected RLM11 cells, expression was unchanged in Notch binding site TALEN transfected cells. Although we do not know the frequency, both TALENs were capable of generating mutated cells in the pool of transfected cells. The use of a heterogeneous cell pool to assess IL-7R $\alpha$  expression on Notch site TALEN transfected cells may not be sensitive enough to detect changes in expression. Assuming that the NF- $\kappa$ B site TALEN transfected cell pool has a similar complexity and results in a detectable decrease in IL-7R $\alpha$  expression we could conclude that mutations in the Notch binding site (RBP-Jk) alone does not affect IL-7R expression. Current dogma indicates that Notch needs to be activated, cleaved from the plasma membrane, translocated to the nucleus, associate with the RBP-Jk protein and displaces previously bound suppressors to activate transcription. These cellular events may not be happening in the chosen cell line mutant. Thus the same mutants that do not affect IL7R expression may in fact be important in other cell types.

Although mutations were detected in TALEN target regions and the phenotypic effect of mutations was demonstrated on IL-7R $\alpha$  expression, selecting TALEN transfected cells is important to quantify TALEN mutagenesis frequency. It is possible to select high TALEN expressing cells by co-transfecting TALEN pairs with a GFP plasmid and sorting GFP positive cells by flow cytometry[84]. However, as indicated in a recent study, sorted cells sometimes cannot form colonies due to damage from the strong laser and hydrostatic pressure they are exposed to in the FACS machine. Sorting of TALEN transfected cells can also be achieved by magnetic separation or antibiotic selection using reporter systems in which a selectable marker is expressed only when mutations are generated at the TALEN target site [85]. In fact in data not included in this thesis, we attempted to sort TALEN transfected and IL7R low expressing cells by magnetic bead separation. Unfortunately we could not detect any mutations in these sorted populations. Ultimately, in order to assign a single phenotype to a single cell with a single mutation, it is necessary to generate clonal populations from sorted cell pools. We are currently attempting to single cell clone Notch and NF- $\kappa$ B site mutated cells from our pools.

In the final section of this thesis, we generated six TALE repeat arrays that we thought could function as competitive inhibitors for an NF- $\kappa$ B binding site. Rather than

being fused to a DNA cleaving FokI restriction enzyme domain, these TALE DNA binding domains were fused to a DsRed fluorescent protein encoding domain. These constructs were expressed in the context of the FokI sequence deficient pCAGT7 derived mammalian expression plasmid named pCAGdsRed. We generated and transfected six TALEdsRed constructs with different repeat numbers into the HEK293 6.1.1 cell line, constructed previously by Pinar Onal and Belkis Atasever in our lab. This reporter cell line has stably integrated four NF- $\kappa$ B binding sites controlling the human c-Fos promoter driving the expression of an eGFP gene. This cell line responds to TNF- $\alpha$  treatment by upregulating GFP expression as a result of NF- $\kappa$ B nuclear translocation. We found that our TALE-dsRed constructs suppressed GFP expression upon TNF- $\alpha$  induction. This effect was specific to those cells that were dsRed positive indicating that the TALE-dsRed bound to the reporter construct NF- $\kappa$ B sites in place of NF- $\kappa$ B. In fact GFP expression of dsRed positive cells was more than 4-fold decreased compared to untransfected cells.

It is known that dsRed proteins can result in tetramerization of fusion proteins. Our reporter cell line contains four NF- $\kappa$ B sites integrated into its genome. We have not yet assessed whether our TALEdsRed proteins form tetramers and whether this putative interaction plays a role in DNA binding. On the other hand, dsRed tetramerization before the TALE proteins bind DNA may be decreasing the affinity of these fluorescent transcription factors to the DNA and perhaps limiting the NF- $\kappa$ B competitive inhibition activity. In such a scenario, it may be more advantageous to express the TALE DNA binding domain independently of the fluorescent protein domain. We note that our TALEdsRed constructs were assembled in a plasmid optimized for TALEN induced genome modification. This plasmid has been optimized for DNA cleavage by a C-terminal truncation. Earlier reports show that the optimal TALE DNA binding domain structure for cleavage and gene regulation activities differs [6]. Thus our TALEdsRed constructs may require more optimization for a more robust DNA binding competitive effect.

An important control for our competitive binding experiment, involves the demonstration that a TALE dsRed construct that does not have specificity to NF- $\kappa$ B does not inhibit NF- $\kappa$ B activity. We are currently constructing such a construct. We also note that, TNF- $\alpha$  treatment was initiated 30 hours after the transfection of the

TALeDsRed constructs and that we analyzed these cells after a 9 hour TNF- $\alpha$  treatment. GFP expression in untransfected cells was around 60%, although this was expected to be 100%, as this is a stable cell line. This indicates that TNF- $\alpha$  treatment should be optimized, perhaps by using confocal microscopy to show that a majority of treated cells express GFP and TALeDsRed proteins. Confocal microscopy can also be used to demonstrate that TALeDsRed and NF- $\kappa$ B proteins are co-localized in the nucleus at the same time. Under these controlled settings, an optimal transcription inhibition activity of our constructed TALE inhibitor proteins can be assessed.

## 6. CONCLUSION

In this study, our main aim was to create mutations at the nuclear factor binding sites of the endogenous IL-7R $\alpha$  gene locus using TALEN technology. This line of work is important for understanding the roles of these transcription factor proteins for the transcription regulation of this immunologically important gene. We used three different strategies to design and construct different TALEN proteins targeting three sites in the IL7R gene. The literature indicates that these sites are bound by the transcription factors, GR, Notch and NF- $\kappa$ B. While we were not successful in generating active TALENs using the first two strategies, the final strategy of expressing TALENs in the context of the pC-Goldy expression system yielded active TALENs and we were able to detect INDELS at the targeted sites. We detected the induction of mutations at the targeted sites using a modified RFLP assay and subsequently sequenced PCR products from these regions to demonstrate that INDELS were actually being generated specifically at the target site. We do not currently know the off target specificity of these TALENs and whether mutations were incorporated at different sites of the genomes of these targeted cells. We analyzed the effect of these mutations on IL-7R $\alpha$  expression in a heterogeneous mutated cell pool by flow cytometry and found that mutations generated by the TALEN targeting the NF- $\kappa$ B binding site indeed reduced IL-7R $\alpha$  expression. On the other hand, mutations induced by the TALEN targeting the Notch binding site did not detectably affect IL-7R $\alpha$  expression. We did not assess the mutagenic potential of the TALEN targeting the GR binding site. These experiments show that TAL effector nucleases can be used for mutation of endogenous genomic transcription enhancers to conduct structure-function studies on these sequences in the endogenous genomic context. Currently most transcription enhancer structure uses the luciferase reporter plasmid expression system which is admittedly an overexpression system and may not reflect the true activity of endogenous enhancers. In this study we

demonstrate that individual transcription factor binding sites in endogenous enhancers can be mutated to affect transcriptional enhancer activity to study function.

In this study we also assessed transcription factor activity using a second approach. In addition to using TALENs for site directed mutagenesis of enhancers, we constructed plasmids encoding TALEdsRed fusion proteins to create competitive binders of an NF- $\kappa$ B transcription factor binding site. To achieve this aim we used a previously generated stable cell line (named HEK293 6.1.1), in which TNF- $\alpha$  treatment results in GFP expression due to NF- $\kappa$ B nuclear translocation. This cell line contains a stably integrated construct containing four NF- $\kappa$ B binding sites controlling a minimal human c-Fos gene promoter driving the expression of an eGFP gene. We demonstrated by flow cytometry that TALEdsRed expression resulted in a reduction of GFP expression in TNF- $\alpha$  treated cells. This part of the study demonstrates that TAL effector proteins can be used as competitive inhibitors of nuclear factors to inhibit gene transcription.

## REFERENCES

1. Boch J and Bonas U (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol* 48:419-36. doi: 10.1146/annurev-phyto-080508-081936
2. Chen K and Gao C (2013) TALENs: Customizable Molecular DNA Scissors for Genome Engineering of Plants. *J Genet Genomics* 40:271-9. doi: 10.1016/j.jgg.2013.03.009
3. 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* 326:1509-12. doi: 10.1126/science.1178811
4. Moscou MJ and Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. *Science* 326:1501.
5. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ and Voytas DF (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39:e82. doi: 10.1093/nar/gkr218
6. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Zhang HS, Holmes MC, Zhang L, Gregory PD and Rebar EJ (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29:143-8. doi: 10.1038/nbt.1755
7. Zhang Y, Zhang F, Li X, Baller JA, Qi Y, Starker CG, Bogdanove AJ and Voytas DF (2013) Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiol* 161:20-7. doi: 10.1104/pp.112.205179
8. Mak AN, Bradley P, Cernadas RA, Bogdanove AJ and Stoddard BL (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335:716-9. doi: 10.1126/science.1216211
9. Deng D, Yan C, Pan X, Mahfouz M, Wang J, Zhu JK, Shi Y and Yan N (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335:720-3. doi: 10.1126/science.1215670

10. Weber E, Gruetzner R, Werner S, Engler C and Marillonnet S (2011) Assembly of designer TAL effectors by Golden Gate cloning. *PLoS One* 6:e19722. doi: 10.1371/journal.pone.0019722
11. Cong L, Zhou R, Kuo YC, Cunniff M and Zhang F (2012) Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun* 3:968. doi: 10.1038/ncomms1962
12. Schmid-Burgk JL, Schmidt T, Kaiser V, Honing K and Hornung V (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat Biotechnol* 31:76-81. doi: 10.1038/nbt.2460
13. Mercer AC, Gaj T, Fuller RP and Barbas CF (2012) Chimeric TALE recombinases with programmable DNA sequence specificity. *Nucleic Acids Research* 40:11163-11172. doi: 10.1093/nar/gks875
14. Christian ML, Demorest ZL, Starker CG, Osborn MJ, Nyquist MD, Zhang Y, Carlson DF, Bradley P, Bogdanove AJ and Voytas DF (2012) Targeting G with TAL effectors: a comparison of activities of TALENs constructed with NN and NK repeat variable di-residues. *PLoS One* 7:e45383. doi: 10.1371/journal.pone.0045383
15. Streubel J, Blücher C, Landgraf A and Boch J (2012) TAL effector RVD specificities and efficiencies. *Nat Biotechnol* 30:594-7. doi: 10.1038/nbt.2304
16. Ma S, Zhang S, Wang F, Liu Y, Xu H, Liu C, Lin Y, Zhao P and Xia Q (2012) Highly Efficient and Specific Genome Editing in Silkworm Using Custom TALENs. *PLoS One* 7:e45035. doi: 10.1371/journal.pone.0045035
17. Joung JK and Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* 14:49-55. doi: 10.1038/nrm3486
18. Huang P, Xiao A, Zhou M, Zhu Z, Lin S and Zhang B (2011) Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol* 29:699-700. doi: 10.1038/nbt.1939
19. Sander JD, Cade L, Khayter C, Reyon D, Peterson RT, Joung JK and Yeh JR (2011) Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol* 29:697-8. doi: 10.1038/nbt.1934

20. Reyon D, Khayter C, Regan MR, Joung JK and Sander JD (2012) Engineering designer transcription activator-like effector nucleases (TALENs) by REAL or REAL-Fast assembly. *Curr Protoc Mol Biol* Chapter 12:Unit-Un15. doi: 10.1002/0471142727.mb1215s100
21. Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD and Joung JK (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* 30:460-5. doi: 10.1038/nbt.2170
22. Briggs AW, Rios X, Chari R, Yang L, Zhang F, Mali P and Church GM (2012) Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res* 40:e117. doi: 10.1093/nar/gks624
23. Wang Z, Li J, Huang H, Wang G, Jiang M, Yin S, Sun C, Zhang H, Zhuang F and Xi JJ (2012) An Integrated Chip for the High-Throughput Synthesis of Transcription Activator-like Effectors. *Angewandte Chemie International Edition* 51:8505-8508. doi: 10.1002/anie.201203597
24. Doyle EL, Boohar NJ, Standage DS, Voytas DF, Brendel VP, Vandyk JK and Bogdanove AJ (2012) TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res* 40:W117-22. doi: 10.1093/nar/gks608
25. Perez EE, Wang J, Miller JC, Jouvenot Y, Kim KA, Liu O, Wang N, Lee G, Bartsevich VV, Lee Y-L, Guschin DY, Rupniewski I, Waite AJ, Carpenito C, Carroll RG, S Orange J, Urnov FD, Rebar EJ, Ando D, Gregory PD, Riley JL, Holmes MC and June CH (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotech* 26:808-816. doi:10.1038/nbt1410
26. Pavletich NP and Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252:809-17.
27. Durai S, Mani M, Kandavelou K, Wu J, Porteus MH and Chandrasegaran S (2005) Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res* 33:5978-90. doi: 10.1093/nar/gki912
28. Cathomen T and Joung JK (2008) Zinc-finger nucleases: the next generation emerges. *Mol Ther* 16:1200-7. doi: 10.1038/mt.2008.114

29. Sun N and Zhao H (2013) Transcription activator-like effector nucleases (TALENs): A highly efficient and versatile tool for genome editing. *Biotechnology and Bioengineering* 110:1811-1821. doi: 10.1002/bit.24890
30. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ and Voytas DF (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186:757-61. doi: 10.1534/genetics.110.120717
31. Li T, Huang S, Zhao X, Wright DA, Carpenter S, Spalding MH, Weeks DP and Yang B (2011) Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res* 39:6315-25. doi: 10.1093/nar/gkr188
32. Sakuma T, Hosoi S, Woltjen K, Suzuki K, Kashiwagi K, Wada H, Ochiai H, Miyamoto T, Kawai N, Sasakura Y, Matsuura S, Okada Y, Kawahara A, Hayashi S and Yamamoto T (2013) Efficient TALEN construction and evaluation methods for human cell and animal applications. *Genes Cells* 18:315-26. doi: 10.1111/gtc.12037
33. Hisano Y, Ota S, Arakawa K, Muraki M, Kono N, Oshita K, Sakuma T, Tomita M, Yamamoto T, Okada Y and Kawahara A (2013) Quantitative assay for TALEN activity at endogenous genomic loci. *Biol Open* 2:363-7. doi: 10.1242/bio.20133871
34. Chen S, Oikonomou G, Chiu CN, Niles BJ, Liu J, Lee DA, Antoshechkin I and Prober DA (2013) A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. *Nucleic Acids Res* 41:2769-78. doi: 10.1093/nar/gks1356
35. Bedell VM, Wang Y, Campbell JM, Poshusta TL, Starker CG, Krug RG, 2nd, Tan W, Penheiter SG, Ma AC, Leung AY, Fahrenkrug SC, Carlson DF, Voytas DF, Clark KJ, Essner JJ and Ekker SC (2012) In vivo genome editing using a high-efficiency TALEN system. *Nature* 491:114-8. doi: 10.1038/nature11537
36. Dahlem T, Hoshijima K, Jurynek M, Gunther D, Starker C, Locke A, Weis A, Voytas D and Grunwald D (2012) Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS Genet* 8:e1002861. doi:10.1371/journal.pgen.1002861

37. Lei Y, Guo X, Liu Y, Cao Y, Deng Y, Chen X, Cheng C, Dawid I, Chen Y and Zhao H (2012) Efficient targeted gene disruption in *Xenopus* embryos using engineered transcription activator-like effector nucleases (TALENs). *Proc Natl Acad Sci U S A* 109:17484 - 17489. doi: 10.1073/pnas.1215421109
38. Carlson DF, Tan W, Lillico SG, Stverakova D, Proudfoot C, Christian M, Voytas DF, Long CR, Whitelaw CBA and Fahrenkrug SC (2012) Efficient TALEN-mediated gene knockout in livestock. *Proceedings of the National Academy of Sciences* 109:17382-17387. doi: 10.1073/pnas.1211446109
39. Sung YH, Baek IJ, Kim DH, Jeon J, Lee J, Lee K, Jeong D, Kim JS and Lee HW (2013) Knockout mice created by TALEN-mediated gene targeting. *Nat Biotechnol* 31:23-4. doi: 10.1038/nbt.2477
40. Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC, Zeitler B, Cherone JM, Meng X, Hinkley SJ, Rebar EJ, Gregory PD, Urnov FD and Jaenisch R (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* 29:731-4. doi: 10.1038/nbt.1927
41. Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T and Cathomen T (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* 39:9283-93. doi: 10.1093/nar/gkr597
42. Ousterout DG, Perez-Pinera P, Thakore PI, Kabadi AM, Brown MT, Qin X, Fedrigo O, Mouly V, Tremblay JP and Gersbach CA (2013) Reading Frame Correction by Targeted Genome Editing Restores Dystrophin Expression in Cells From Duchenne Muscular Dystrophy Patients. *Mol Ther*. doi: 10.1038/mt.2013.111
43. Ma AC, Lee HB, Clark KJ and Ekker SC (2013) High Efficiency *In Vivo* Genome Engineering with a Simplified 15-RVD GoldyTALEN Design. *PLoS ONE* 8:e65259. doi: 10.1371/journal.pone.0065259
44. Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I, Beausejour CM, Waite AJ, Wang NS, Kim KA, Gregory PD, Pabo CO and Rebar EJ (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* 25:778-85. doi: 10.1038/nbt1319

45. Cade L, Reyon D, Hwang WY, Tsai SQ, Patel S, Khayter C, Joung JK, Sander JD, Peterson RT and Yeh J-RJ (2012) Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Research* 40:8001-8010. doi: 10.1093/nar/gks518
46. Ota S, Hisano Y, Muraki M, Hoshijima K, Dahlem TJ, Grunwald DJ, Okada Y and Kawahara A (2013) Efficient identification of TALEN-mediated genome modifications using heteroduplex mobility assays. *Genes Cells* 18:450-8. doi: 10.1111/gtc.12050
47. Wefers B, Meyer M, Ortiz O, Hrabé de Angelis M, Hansen J, Wurst W and Kühn R (2013) Direct production of mouse disease models by embryo microinjection of TALENs and oligodeoxynucleotides. *Proceedings of the National Academy of Sciences* 110:3782-3787. doi: 10.1073/pnas.1218721110
48. Al-Rawi MAA, Mansel RE and Jiang WG (2003) Interleukin-7 (IL-7) and IL-7 receptor (IL-7R) signalling complex in human solid tumours. *Histol Histopathol* 18:911-923.
49. Jiang Q, Li WQ, Aiello FB, Mazzucchelli R, Asefa B, Khaled AR and Durum SK (2005) Cell biology of IL-7, a key lymphotrophin. *Cytokine & Growth Factor Reviews* 16:513-533. doi: 10.1016/j.cytogfr.2005.05.004
50. Kittipatarin C and Khaled AR (2007) Interlinking interleukin-7. *Cytokine* 39:75-83. doi: 10.1016/j.cyto.2007.07.183
51. Cosenza L, Gorgun G, Urbano A and Foss F (2002) Interleukin-7 receptor expression and activation in nonhaematopoietic neoplastic cell lines. *Cellular Signalling* 14:317-325. doi: 10.1016/S0898-6568(01)00245-5
52. González-García S, García-Peydró M, Alcain J and Toribio M (2012) Notch1 and IL-7 Receptor Signalling in Early T-cell Development and Leukaemia. In: Radtke F (ed) *Notch Regulation of the Immune System*, Springer Berlin Heidelberg, pp. 47-73
53. Ceredig R and Rolink AG (2012) The key role of IL-7 in lymphopoiesis. *Seminars in Immunology* 24:159-164. doi: 10.1016/j.smim.2012.02.004
54. Peschon JJ, Morrissey PJ, Grabstein KH, Ramsdell FJ, Maraskovsky E, Gliniak BC, Park LS, Ziegler SF, Williams DE, Ware CB, Meyer JD and Davison BL (1994) Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice. *The Journal of Experimental Medicine* 180:1955-1960. doi: 10.1084/jem.180.5.1955

55. Corcoran AE, Riddell A, Krooshoop D and Venkitaraman AR (1998) Impaired immunoglobulin gene rearrangement in mice lacking the IL-7 receptor. *Nature* 391:904-907. doi: 10.1038/36122
56. Johnson K, Hashimshony T, Sawai CM, Pongubala JMR, Skok JA, Aifantis I and Singh H (2008) Regulation of Immunoglobulin Light-Chain Recombination by the Transcription Factor IRF-4 and the Attenuation of Interleukin-7 Signaling. *Immunity* 28:335-345. doi: 10.1016/j.immuni.2007.12.019
57. Kikuchi K, Lai AY, Hsu C-L and Kondo M (2005) IL-7 receptor signaling is necessary for stage transition in adult B cell development through up-regulation of EBF. *The Journal of Experimental Medicine* 201:1197-1203. doi: 10.1084/jem.20050158
58. Mackall CL, Fry TJ and Gress RE (2011) Harnessing the biology of IL-7 for therapeutic application. *Nat Rev Immunol* 11:330-342. doi: 10.1038/nri2970
59. Hong C, Luckey MA and Park J-H (2012) Intrathymic IL-7: The where, when, and why of IL-7 signaling during T cell development. *Seminars in Immunology* 24:151-158. doi: 10.1016/j.smim.2012.02.002
60. Yu Q, Erman B, Park JH, Feigenbaum L and Singer A (2004) IL-7 receptor signals inhibit expression of transcription factors TCF-1, LEF-1, and ROR $\gamma$ mat: impact on thymocyte development. *J Exp Med* 200:797-803. doi: 10.1084/jem.20032183
61. Singer A, Adoro S and Park J-H (2008) Lineage fate and intense debate: myths, models and mechanisms of CD4- versus CD8-lineage choice. *Nat Rev Immunol* 8:788-801. doi: 10.1038/nri2416
62. Puel A, Ziegler SF, Buckley RH and Leonard WJ (1998) Defective IL7R expression in T-B+NK+ severe combined immunodeficiency. *Nat Genet* 20:394-397. doi: 10.1038/3877
63. Shochat C, Tal N, Bandapalli OR, Palmi C, Ganmore I, te Kronnie G, Cario G, Cazzaniga G, Kulozik AE, Stanulla M, Schrappe M, Biondi A, Basso G, Bercovich D, Muckenthaler MU and Izraeli S (2011) Gain-of-function mutations in interleukin-7 receptor- $\alpha$  (IL7R) in childhood acute lymphoblastic leukemias. *The Journal of Experimental Medicine* 208:901-908. doi: 10.1084/jem.20110580
64. Bradley LM, Haynes L and Swain SL (2005) IL-7: maintaining T-cell memory and achieving homeostasis. *Trends in Immunology* 26:172-176. doi: 10.1016/j.it.2005.01.004

65. DeKoter RP, Lee H-J and Singh H (2002) PU.1 Regulates Expression of the Interleukin-7 Receptor in Lymphoid Progenitors. *Immunity* 16:297-309. doi: 10.1016/S1074-7613(02)00269-8
66. DeKoter RP, Schweitzer BL, Kamath MB, Jones D, Tagoh H, Bonifer C, Hildeman DA and Huang KJ (2007) Regulation of the Interleukin-7 Receptor  $\alpha$  Promoter by the Ets Transcription Factors PU.1 and GA-binding Protein in Developing B Cells. *Journal of Biological Chemistry* 282:14194-14204. doi: 10.1074/jbc.M700377200
67. Egawa T, Tillman RE, Naoe Y, Taniuchi I and Littman DR (2007) The role of the Runx transcription factors in thymocyte differentiation and in homeostasis of naive T cells. *The Journal of Experimental Medicine* 204:1945-1957. doi: 10.1084/jem.20070133
68. Vosshenrich CAJ, Garcia-Ojeda ME, Samson-Villeger SI, Pasqualetto V, Enault L, Goff OR-L, Corcuff E, Guy-Grand D, Rocha B, Cumano A, Rogge L, Ezine S and Di Santo JP (2006) A thymic pathway of mouse natural killer cell development characterized by expression of GATA-3 and CD127. *Nat Immunol* 7:1217-1224. doi:10.1038/ni1395
69. Kerdiles YM, Beisner DR, Tinoco R, Dejean AS, Castrillon DH, DePinho RA and Hedrick SM (2009) Foxo1 links homing and survival of naive T cells by regulating L-selectin, CCR7 and interleukin 7 receptor. *Nat Immunol* 10:176-184. doi: 10.1038/ni.1689
70. Kopan R (2012) Notch signaling. *Cold Spring Harb Perspect Biol* 4. doi: 10.1101/cshperspect.a011213
71. Bray SJ (2006) Notch signalling: a simple pathway becomes complex. *Nat Rev Mol Cell Biol* 7:678-689. doi:10.1038/nrm2009
72. Radtke F, Wilson A, Stark G, Bauer M, van Meerwijk J, MacDonald HR and Aguet M (1999) Deficient T Cell Fate Specification in Mice with an Induced Inactivation of Notch1. *Immunity* 10:547-558. doi:10.1016/S1074-7613(00)80054-0
73. Gonzalez-Garcia S, Garcia-Peydro M, Martin-Gayo E, Ballestar E, Esteller M, Bornstein R, de la Pompa JL, Ferrando AA and Toribio ML (2009) CSL-MAML-dependent Notch1 signaling controls T lineage-specific IL-7R $\alpha$  gene expression in early human thymopoiesis and leukemia. *J Exp Med* 206:779-91. doi: 10.1084/jem.20081922

74. Napetschnig J and Wu H (2013) Molecular Basis of NF- $\kappa$ B Signaling. *Annual Review of Biophysics* 42:443-468. doi:10.1146/annurev-biophys-083012-130338
75. Hayden MS and Ghosh S (2012) NF- $\kappa$ B, the first quarter-century: remarkable progress and outstanding questions. *Genes & Development* 26:203-234. doi: 10.1101/gad.183434.111
76. Tian B, Nowak DE, Jamaluddin M, Wang S and Brasier AR (2005) Identification of Direct Genomic Targets Downstream of the Nuclear Factor- $\kappa$ B Transcription Factor Mediating Tumor Necrosis Factor Signaling. *Journal of Biological Chemistry* 280:17435-17448. doi: 10.1074/jbc.M500437200
77. Smoak KA and Cidlowski JA (2004) Mechanisms of glucocorticoid receptor signaling during inflammation. *Mechanisms of Ageing and Development* 125:697-706. doi: 10.1016/j.mad.2004.06.010
78. De Bosscher K, Vanden Berghe W, Vermeulen L, Plaisance S, Boone E and Haegeman G (2000) Glucocorticoids repress NF- $\kappa$ B-driven genes by disturbing the interaction of p65 with the basal transcription machinery, irrespective of coactivator levels in the cell. *Proceedings of the National Academy of Sciences* 97:3919-3924. doi: 10.1073/pnas.97.8.3919
79. Lee H-C, Shibata H, Ogawa S, Maki K and Ikuta K (2005) Transcriptional Regulation of the Mouse IL-7 Receptor  $\alpha$  Promoter by Glucocorticoid Receptor. *The Journal of Immunology* 174:7800-7806.
80. Ligons DL, Tuncer C, Linowes BA, Akcay IM, Kurtulus S, Deniz E, Atasever Arslan B, Cevik SI, Keller HR, Luckey MA, Feigenbaum L, Möröy T, Ersahin T, Atalay R, Erman B and Park J-H (2012) CD8 Lineage-specific Regulation of Interleukin-7 Receptor Expression by the Transcriptional Repressor Gfi1. *Journal of Biological Chemistry* 287:34386-34399. doi: 10.1074/jbc.M112.378687
81. Silverstone A, Sun L, Witte ON and Baltimore D (1980) Biosynthesis of murine terminal deoxynucleotidyltransferase. *Journal of Biological Chemistry* 255:791-796.
82. Doyon Y, Vo T, Mendel M, Greenberg S, Wang J, Xia D, Miller J, Urnov F, Gregory P and Holmes M (2011) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods* 8:74 - 79. doi:10.1038/nmeth.1539

83. Bultmann S, Morbitzer R, Schmidt CS, Thanisch K, Spada F, Elsaesser J, Lahaye T and Leonhardt H (2012) Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res* 40:5368-77. doi: 10.1093/nar/gks199
84. Hu R, Wallace J, Dahlem TJ, Grunwald DJ and O'Connell RM (2013) Targeting human microRNA genes using engineered Tal-effector nucleases (TALENs). *PLoS One* 8:e63074. doi: 10.1371/journal.pone.0063074
85. Kim H, Kim M-S, Wee G, Lee C-i, Kim H and Kim J-S (2013) Magnetic Separation and Antibiotics Selection Enable Enrichment of Cells with ZFN/TALEN-Induced Mutations. *PLoS ONE* 8:e56476. doi: 10.1371/journal.pone.0056476

## APPENDIX

### APPENDIX A: Chemicals Used in the Study

Chemicals	Supplier Company
Acetic Acid	Sigma, Germany
Agarose	Sigma, Germany
Ampicillin sodium salt	Cellgro, USA
Boric acid	Molekula, UK
Calcium chloride	Sigma, Germany
Distilled water	Milipore, France
DMEM	Gibco, USA
DMSO	Sigma, Germany
EDTA	Sigma, Germany
Ethanol	Sigma, Germany
Ethidium bromide	Sigma, Germany
Fetal Bovine Serum (FBS)	Lonza, Switzerland
Glycerol	Sigma, Germany
HBSS	Gibco, USA
Hydrochloric Acid	Sigma, Germany
Isopropanol	Sigma, Germany
Kanamycin Sulfate	Gibco, USA
LB Agar	BD, USA
LB Broth	Sigma, Germany
L-glutamine	Hyclone, USA

Liquid nitrogen	Karbogaz, Turkey
2-mercaptoethanol	Sigma, Germany
Penicillin-Streptomycin	Sigma, Germany
PIPES	Sigma, Germany
PBS	Sigma, Germany
RNase A	Roche, Germany
RPMI 1640	Gibco, USA
SDS	Sigma, Germany
Sodium Azide	Amresco, USA
Spectinomycin	Sigma, Germany
Tetracyclin	Sigma, Germany
Tris base	Sigma, Germany
Trypan Blue	Fluca, Germany
Trypsin-EDTA	Gibco, USA

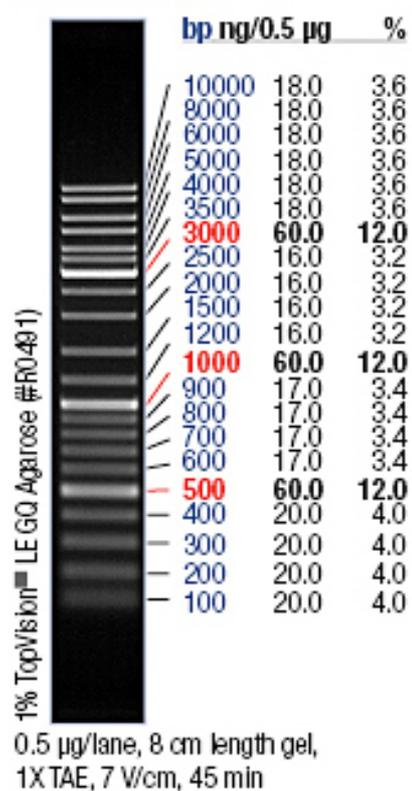
## APPENDIX B: Equipment Used In the Study

Equipment	Company
Autoclave	Priorclave, UK
Balance	Sartorius, BP221S, Germany Schimadzu, Libror EB-3200 HU, Japan
Centrifuge	Hitachi, Sorvall RC5C Plus, USA Eppendorf, 5415D, Germany Eppendorf, 5418R, Germany Beckman Coulter, Allegra®X-15R, USA
CO <sub>2</sub> Incubator	Binder, Germany
Deepfreeze	-80°C, Forma, Thermo Electron Corp., USA -20°C, Bosch, Turkey
Distilled Water	Millipore, Elix-S, France
Electrophoresis Apparatus	Biorad Inc., USA
Electroporator	Invitrogen, Neon Transfection Systems, USA
Flow Cytometer	BD FACS Canto, USA
Gel Documentation	Biorad, UV-Transilluminator 2000, USA
Heater	Thermomixer Comfort, Eppendorf, Germany
Hematocytometer	Hausser Scientific, Blue Bell Pa., USA
Ice Machine	Scotsman Inc., AF80, USA
Incubator	Memmert, Modell 300, Germany
Laminar Flow	Kendro Lab. Prod., Heraeus, HeraSafe HS12, Germany
Liquid Nitrogen Tank	Taylor-Wharton, 3000RS, USA

Magnetic Stirrer	Stuart™,SB162, UK
Microliter Pipettes	Gilson, Pipetman, France
Microscope	Olympus IX70, Japan Olympus CK40, Japan
Microwave Oven	Bosch,Turkey
pH meter	Mettler Toledo, S220 SevenCompact™ pH/Ion, USA
Refrigerator	Bosch,Turkey
Shaker Incubator	New Brunswick Sci., Innova 4330, USA
Spectrophotometer	Amersham Biosciences, UK
Thermocycler	Eppendorf, Mastercycler Gradient, Germany
Vortex	Velp Scientifica,Italy

## APPENDIX C: DNA Molecular Weight Marker

Gene Ruler™ DNA Ladder Mix  
Fermentas, Germany



**APPENDIX D: FACS Analysis of GFP and dsRed Expression Levels in NIH3T3 Cell Line Infected with ECR3 GR Binding Site TALEN plasmids**

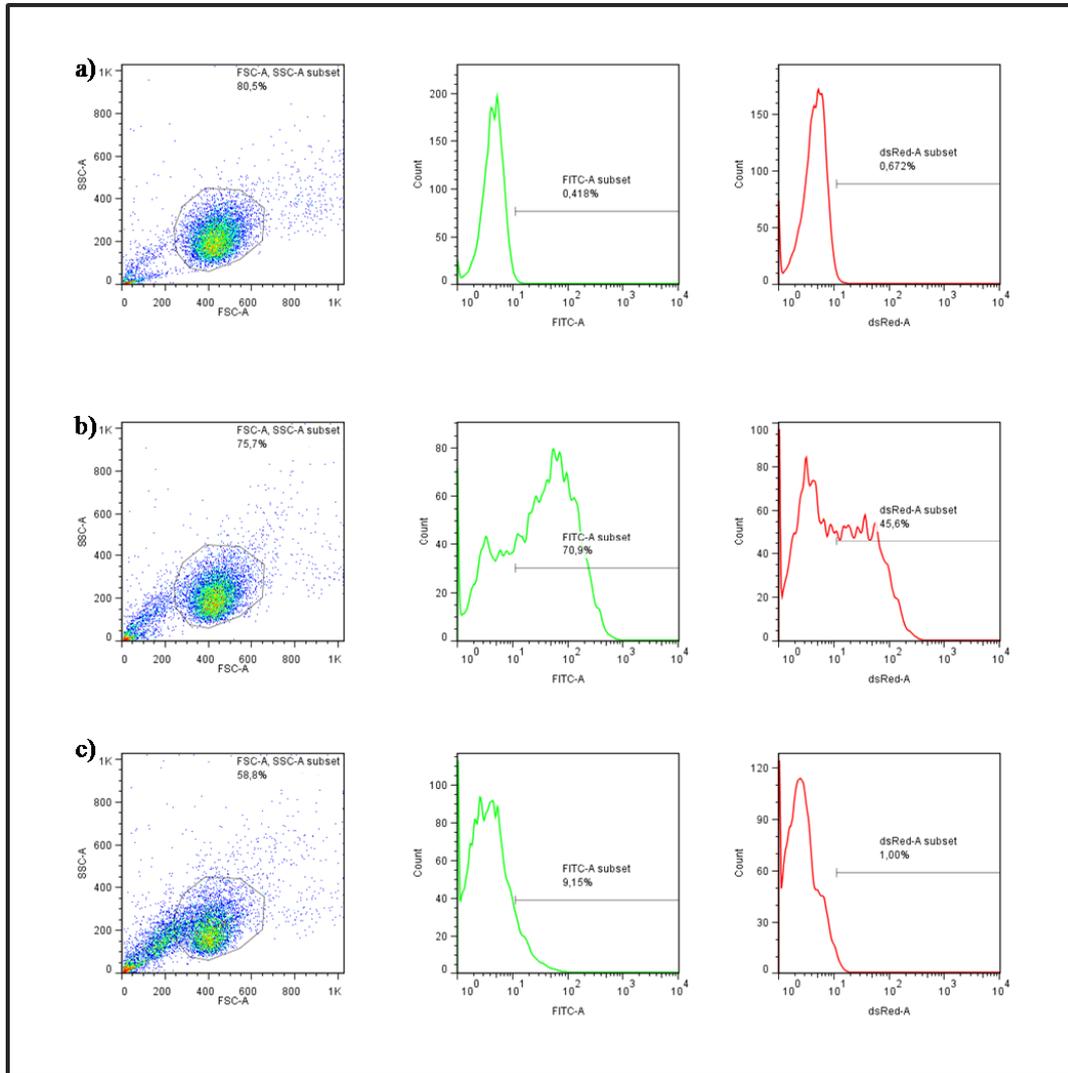


Figure D.1 FACS analysis of GFP and dsRed expression in ECR3 GR binding site TALEN infected NIH3T3 cell line. The first column shows side scatter vs. forward scatter dot plots, the second column shows histograms of GFP expression and the third column shows histograms of dsRed expression. . a) Uninfected NIH3T3 b) pMIGII TcRαβ-IRESdsRed and pMIGII-IRESdsRed infected NIH3T3 cells c) Left GR binding site TALEN-IRESGFP and right GR binding site TALEN-IRESdsRed infected NIH3T3 cells. Forward scatter, shows the size of the cells and side scatter shows the granularity of cells. The analysis was restricted to the live cells falling into the gate shown in side scatter vs. forward scatter dot plots. The percentage of cells falling into each gate is indicated under the label for each gate.

**APPENDIX E: FACS Analysis of GFP and dsRed Expression Levels in Neuro-2a Cell Line Transfected with CMV ECR3 GR Binding Site TALEN Plasmids**

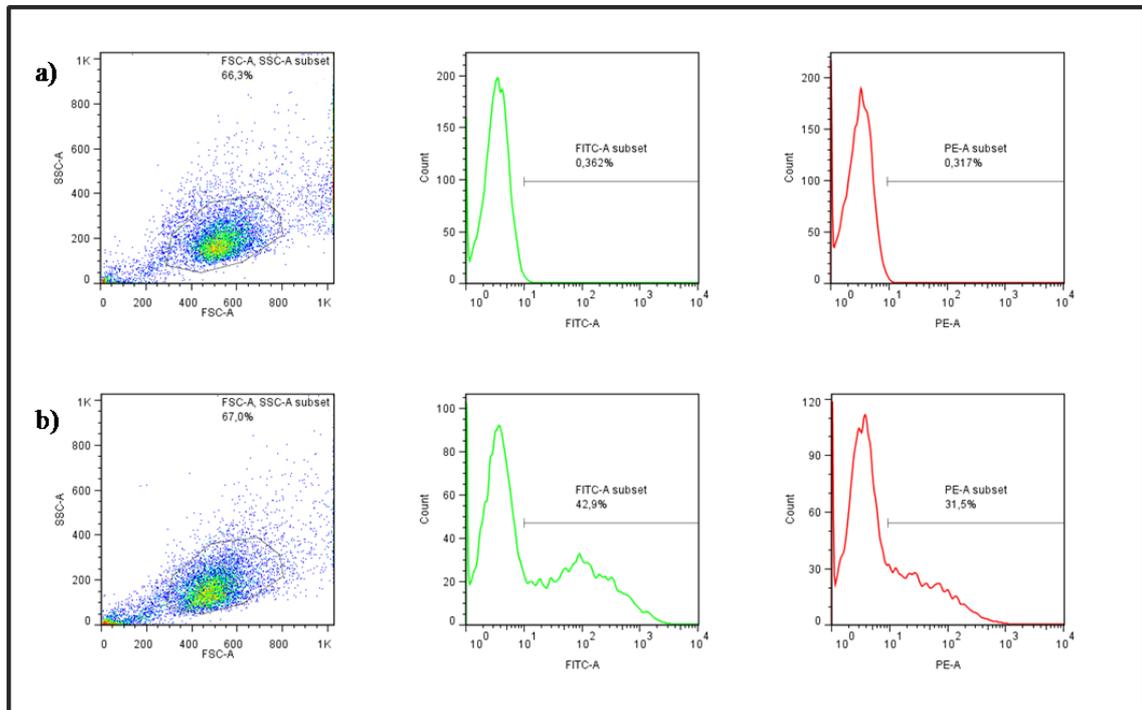


Figure E.1 FACS analysis of GFP and dsRed expression in CMV ECR3 GR binding site TALEN transfected Neuro-2a cell line. The first column shows side scatter vs. forward scatter dot plots, the second column shows histograms of GFP expression and the third column shows histograms of dsRed expression. . a) Untransfected Neuro-2a b) CMV left GR binding site TALEN-IRESGFP and CMV right GR binding site TALEN-IRESdsRed co-transfected Neuro-2a cells. Forward scatter, shows the size of the cells and side scatter shows the granularity of cells. The analysis was restricted to the live cells falling into the gate shown in side scatter vs. forward scatter dot plots. The percentage of cells falling into each gate is indicated under the label for each gate.

**APPENDIX F: FACS Analysis of IL7R $\alpha$  Expression Levels on RLM11 Cell Line Transfected with Constructed TALEN Plasmids**

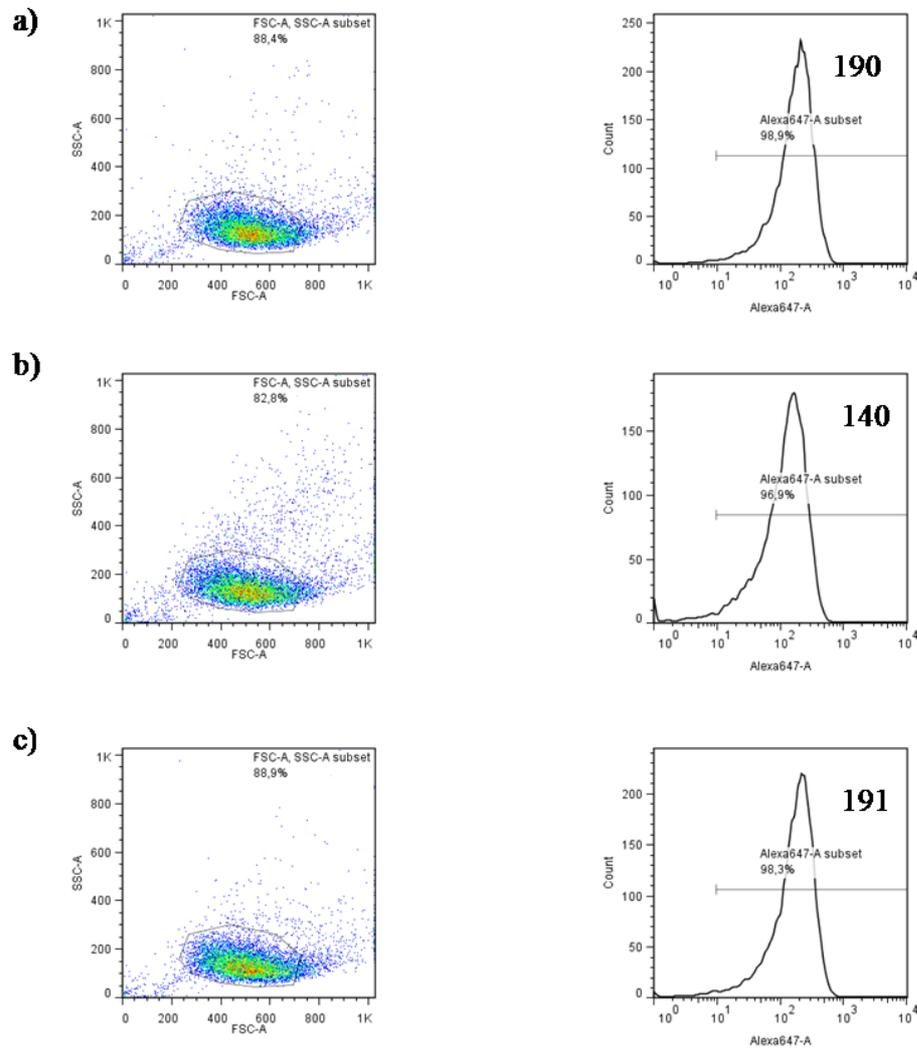


Figure F.1 FACS analysis of IL7R $\alpha$  expression on the RLM11 cell line transfected with TALEN expression plasmids targeting NF- $\kappa$ B binding site and Notch binding site of IL-7R enhancer region. The first column shows side scatter vs. forward scatter dot plots, the second column shows histograms of IL-7R $\alpha$  expression. a) untransfected RLM11 b) pC-Goldy NF- $\kappa$ B binding site TALEN pair transfected RLM11 cells c) pC-Goldy Notch binding site TALEN pair transfected RLM11 cells. Forward scatter, shows the size of the cells and side scatter shows the granularity of cells. The analysis was restricted to the live cells falling into the gate shown in side scatter vs. forward scatter dot plots. The bold numbers inside the graph indicates the mean Alexa 647 fluorescence intensity of the cells, indicating IL7R expression. IL7R $\alpha$  was detected with biotinylated CD127 antibodies followed by Alexa647 labeled streptavidin. The percentage of cells falling into each gate is indicated under the label for each gate.

**APPENDIX G: Representative Sequence Analysis of Mutations Induced at Notch Binding Site TALEN Target Sites of Neuro-2a Cells**

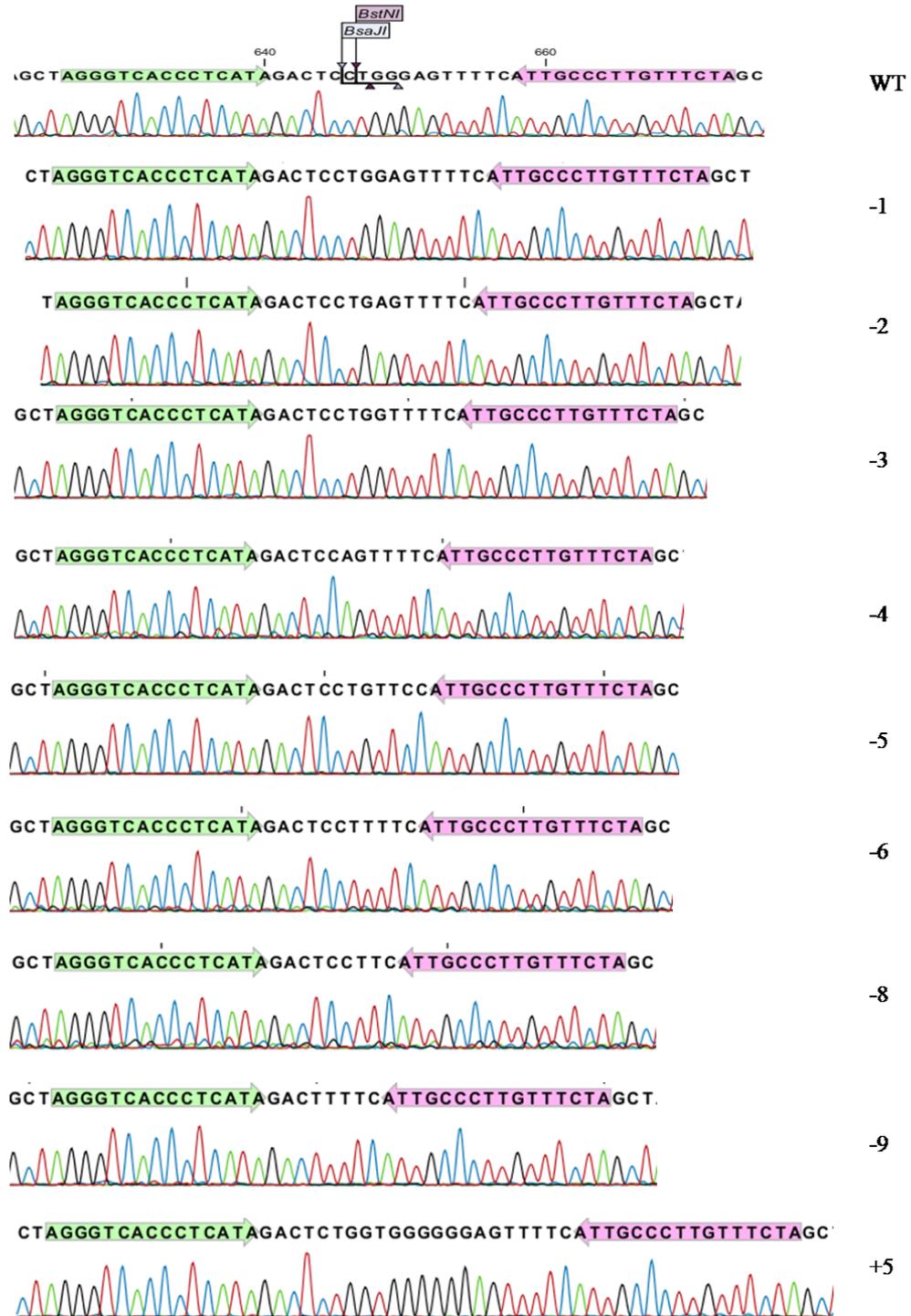
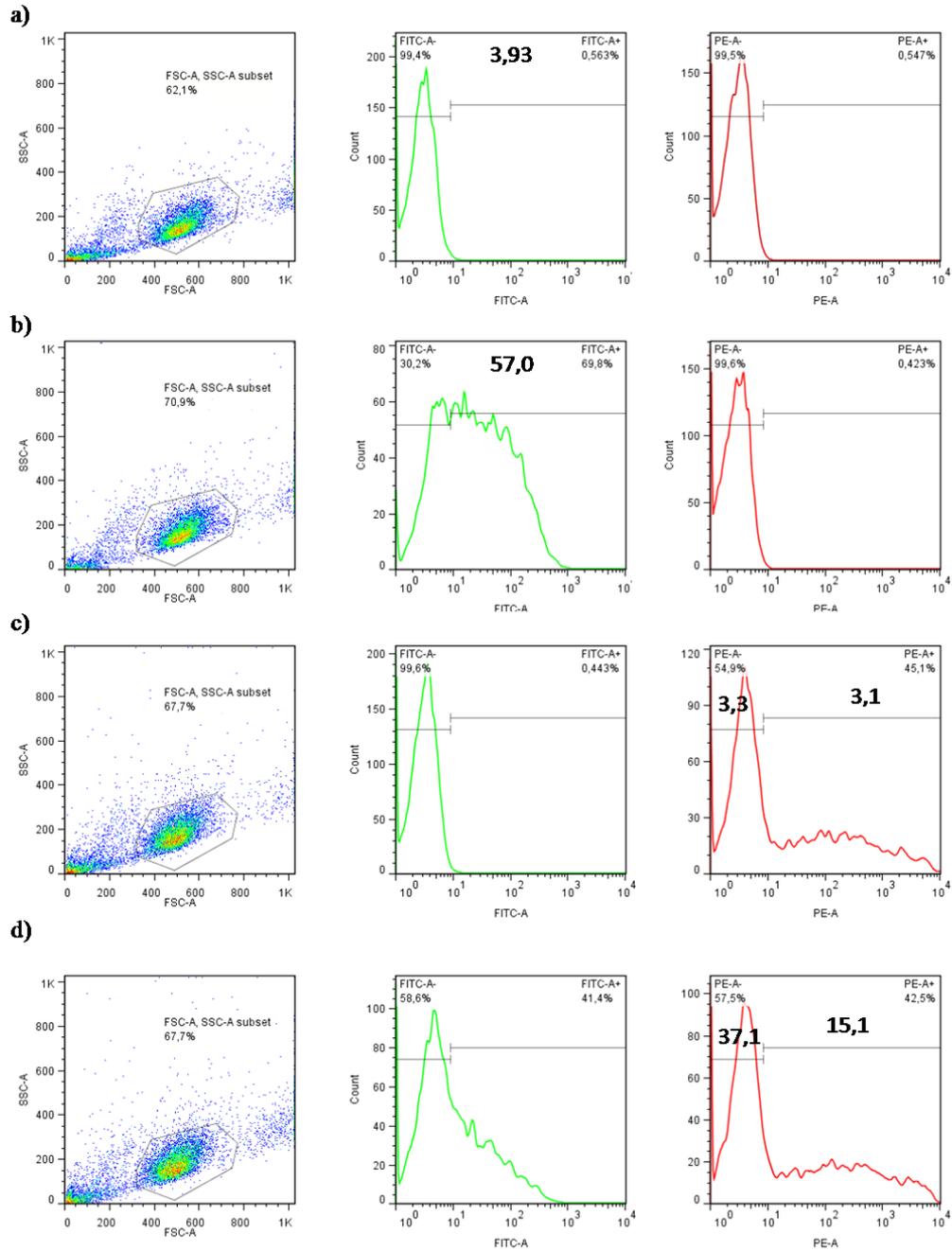
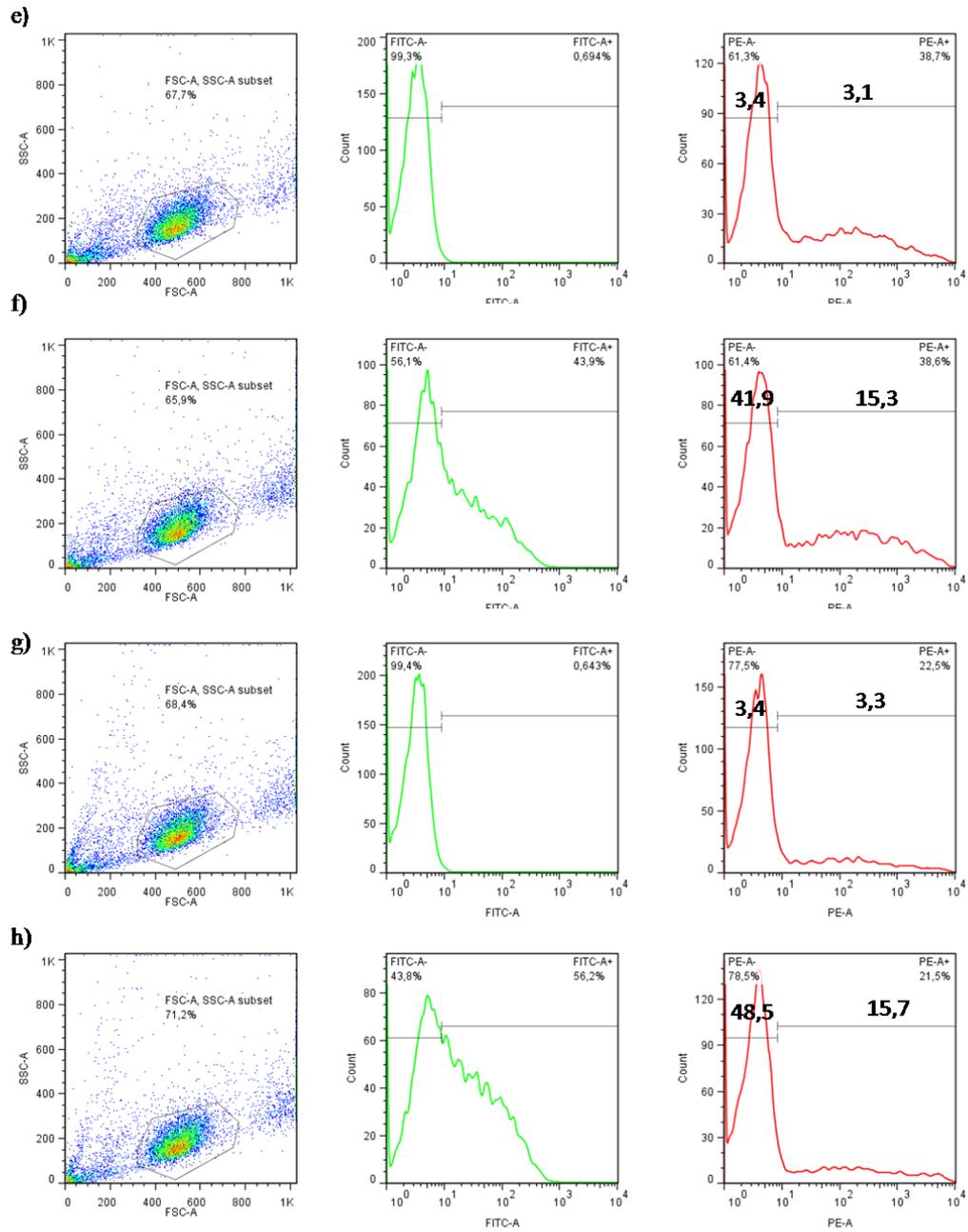
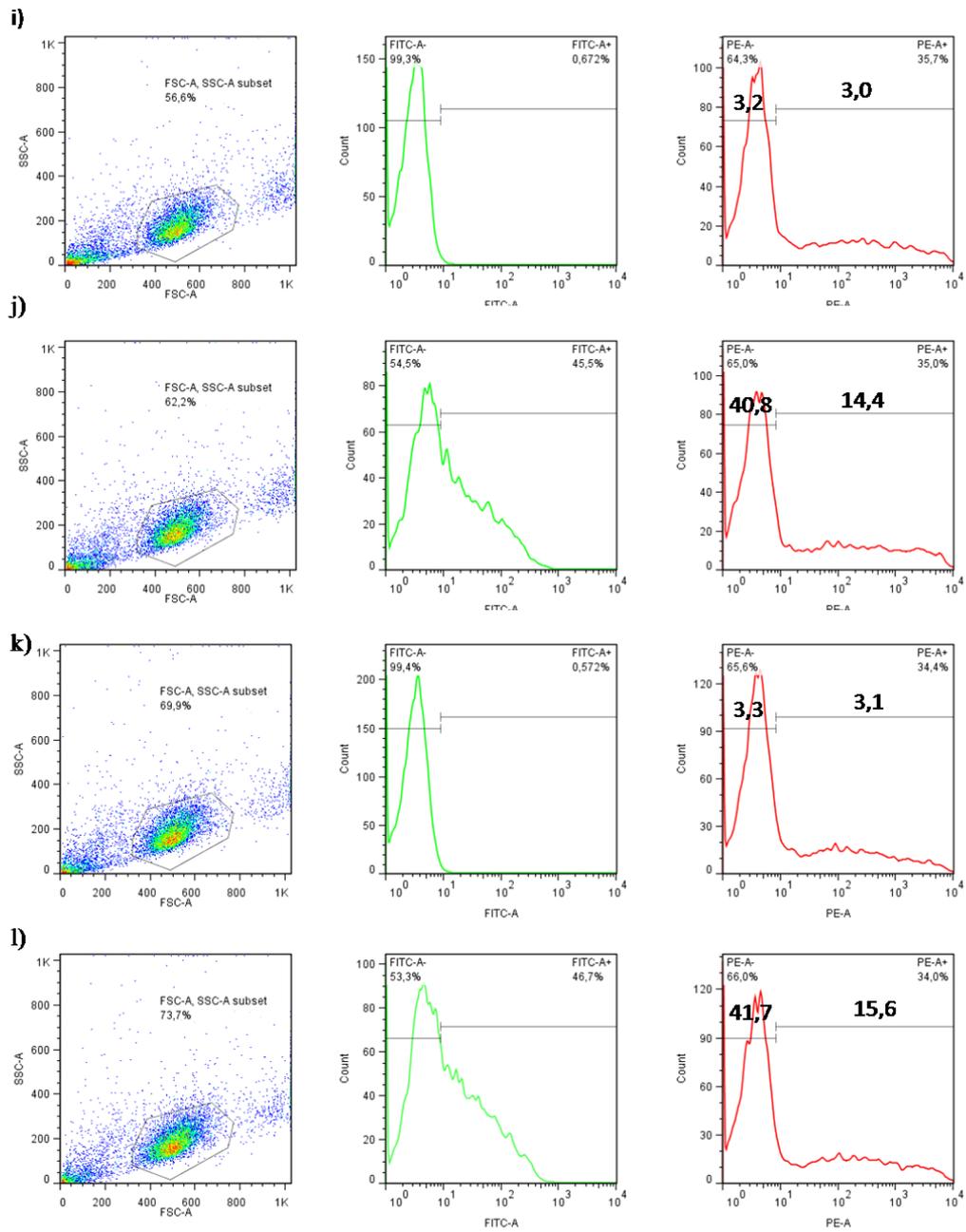


Figure G.1 Sequencing analysis of the Notch binding site TALEN transfected Neuro-2a cells. Green arrow represents the binding site of left Notch TALEN. Red arrow represents the binding site of right Notch TALEN. The numbers on the left indicates the number of nucleotides inserted or deleted. WT corresponds to the sequence of Notch binding site TALEN target site of the untransfected Neuro-2a cells.

## APPENDIX H: FACS Analysis of GFP Expression Levels in HEK293 6.1.1 Reporter Cell Line Transfected with TALEdsRed Expression Plasmids







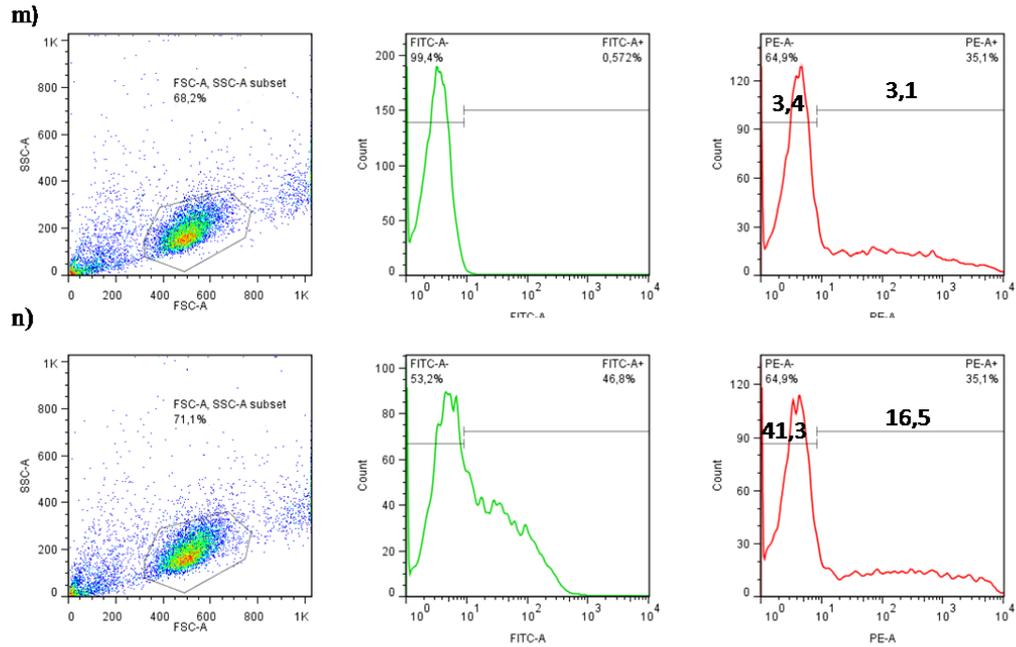


Figure H.1 FACS analysis of GFP expression in transfected HEK293 6.1.1 reporter cell line. The first column shows side scatter vs. forward scatter dot plots, the second column shows histograms of GFP expression and the third column shows histograms of dsRed expression. a) untransfected TNF untreated b) untransfected TNF treated c) TALEdsRed12 transfected TNF untreated d) TALEdsRed12 transfected TNF treated e) TALEdsRed13 transfected TNF untreated f) TALEdsRed13 transfected TNF treated g) TALEdsRed14 transfected TNF untreated h) TALEdsRed14 transfected TNF treated i) TALEdsRed15 transfected TNF untreated j) TALEdsRed15 transfected TNF treated k) TALEdsRed16 transfected TNF untreated l) TALEdsRed16 transfected TNF treated m) TALEdsRed17 transfected TNF untreated n) TALEdsRed17 transfected TNF treated HEK293 6.1.1 cells. Forward scatter, shows the size of the cells and side scatter shows the granularity of cells. The analysis was restricted to the live cells falling into the gate shown in side scatter vs. forward scatter dot plots. The bold numbers inside the graph indicates the mean GFP fluorescence intensity of the cells in the gates. The percentage of cells falling into each gate is indicated under the label for each gate.