



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Editing the mouse genome to understand the regulation of milk proteins



Akshay Joshi

Thesis presented for the degree of Doctor of Philosophy
University of Edinburgh, June 2019

Declaration

This thesis and the work presented in it is, unless otherwise stated, my own work. The work presented here has not been submitted for any other degree or professional qualification.

A handwritten signature in red ink, reading "Akshay Joshi". The signature is written in a cursive style with a long horizontal stroke extending to the right.

Akshay Joshi, June 2019

Acknowledgements

First of all, I thank Professor Bruce Whitelaw for the wonderful opportunity to work at The Roslin Institute and his supervision during my PhD.

I am grateful to Dr. Simon Lilllico and Dr. Chris Proudfoot for always guiding me in the right direction with a lot of patience and their help and support throughout my PhD.

I thank Dr. Dharmesh Harwani for being a huge inspiration in my life.

I thank Dr. Satish Kumar for his help prior and during my PhD and for being a great collaborator to work with.

I thank my PhD committee chair Dr. Peter Hohenstein for his support.

In the Whitelaw group, I thank my friends, present and past lab members, Sarah, Claire, Gus, Hamish, Slyvia and Spring for their help and for being amazing lab mates. A special mention to Dina for always encouraging me especially during my 'those days of PhD' when things do not work as expected.

I thank Dadu and Rakhi, who has been massive support during my PhD and Prasun for some good chats during lunches.

My friends in India, Vinu and Aarti, have always encouraged me from far off.

In the end, I thank my family, who is slightly more excited about my PhD than me, for their love and for believing in me.

Abstract

Dairy industries not only contribute to global food security but also have a major economic role worldwide. Milk is an important source of proteins, lactose, fat, and minerals, as such, understanding the genetic regulation of milk components is crucial to improve dairy production and livestock health. The Caseins, the predominant milk proteins, consist of four components (α S1, α S2, β and κ casein) that exhibit different functions during lactation. Recently a comparative analysis of the casein locus identified an evolutionarily conserved region (ECR) in the mouse genome, predicted to be a potential regulatory element. CRISPRs are programmable molecular scissors that generate targeted double stranded breaks in DNA which are subsequently repaired by non-homologous end joining or homology directed repair. To understand the role of ECR in regulation of casein locus, an ECR knockout mouse has been generated using CRISPRs. Similarly, the entire casein locus of mouse was attempted to be removed in order to redesign this locus with proteins of interest. Lactose is a milk sugar, the concentration of which is regulated by α lactalbumin which in turn is controlled by a transcription activator protein called STAT5, which binds to α lactalbumin promoter at a consensus binding site resulting in transcription activation. The STAT5 binding site the mouse α lactalbumin promoter have been altered to check its expression in mouse mammary epithelial (HC11) cells using the reporter assay. The mouse *lalba* promoter without a STAT5 binding site showed the maximum luciferase activity suggesting its role as a limiting factor in the *lalba* gene regulation.

Lay summary

Milk is essential for humans in dairy and for livestock in nourishment of their offspring. The quality and quantity of livestock milk can be improved by understanding the regulation of its major component such as Caseins and Lactose.

In this thesis, we attempted to unravel milk components in the mouse such as an evolutionary conserved region (ECR) in the casein locus, the entire casein locus and the STAT5 transcription factor binding site in the mouse *lalba* promoter. The 147 bp mouse ECR was initially targeted in mouse embryonic fibroblast cells using CRISPRs and subsequently ECR knock-out mice were generated by injecting CRISPR plasmids in mouse zygotes via pronuclear microinjections. These mice are currently being phenotyped in our collaborator's laboratory in India. Similarly, the 266.4kb mouse casein locus was deleted in mouse embryonic fibroblast cells using CRISPRs followed by attempts to generate casein knock-out mice by pronuclear microinjections of CRISPR plasmids.

The STAT5 transcription factor binding site in the *lalba* gene promoter of mouse was altered by creating *lalba* promoter with a STAT5 tetramer transcription factor binding site and another with no STAT5 transcription factor binding site using fusion PCR. The activity of the wild type and both the altered promoters was analysed in the HC11 cells using reporter assay. The results showed the maximum activity in mouse *lalba* promoter lacking the STAT5 transcription factor binding site, subsequently, the same alteration were attempted to create in the *lalba* promoter of HC11 cells using CRISPRs and homology directed repair.

Table of contents

Declaration	1
Lay summary	4
Chapter 1: General Introduction	1
1.1 Evolution of mammals	1
1.1.1 Monotremes	2
1.1.2 Marsupials	4
1.1.3 Eutherians	5
1.2 Lactation in eutherians	7
1.3 Domestication	8
1.4 Domestication of livestock	11
1.5 Dairy Livestock	14
1.5.1 Cattle	14
1.5.2 Buffalo	15
1.5.3 Goat	16
1.5.4 Sheep	16
1.6 Milk constituents	17
1.7 Dairy industry	22
1.8 Transgenesis	25
1.9 Gene editors	28
1.9.1 Zinc finger nucleases (ZFN)	29
1.9.2 Transcription activator-like effector nucleases (TALENs)	32
1.9.3 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)	34
1.10 Methods of transgene delivery	37
2 Chapter 2 Materials and methods	40
2.1 General microbiological technique	40
2.1.1 Culture medium and antibiotics solutions	40
2.1.2 Transformation	40
2.1.3 Plasmid miniprep	41
2.1.4 Plasmid Maxiprep	41
2.1.5 Nanodrop spectrophotometer	42
2.1.6 Plasmid digestion	43

2.1.7	Plasmid ligation	43
2.2	Molecular biology techniques	43
2.2.1	Polymerase chain reaction (PCR).....	43
2.2.2	Isolation/ collection of pure DNA after PCR	44
2.2.3	Gel electrophoresis	45
2.2.4	Gel documentation.....	45
2.2.5	DNA Sequencing	45
2.2.6	Luciferase assay	46
2.3	Cell culture methods.....	46
2.3.1	Mouse embryonic fibroblast (MEF) cell culture.....	46
2.3.2	HC11 (mouse mammary epithelial) Cell culture	46
2.3.3	Cryopreservation of cultured cells.....	46
2.3.4	Recovery of cryopreserved cells.....	47
2.3.5	Transfection by electroporation.....	47
2.3.6	Transfection by lipofection	48
2.3.7	Fluorescent activated cell sorting (FACS).....	48
2.3.8	Isolation of DNA from animal cells	48
2.3.9	Isolation of DNA from animal tissues	49
2.4	Protein biology methods	49
2.4.1	Western blotting	49
2.5	CRISPR constructs.....	51
2.5.1	Designing CRISPR guides.....	51
2.5.2	Cloning guide oligonucleotides into <i>pSL66</i> plasmid.....	51
2.5.3	Plasmid safe nuclease treatment.....	52
2.5.4	Colony PCR/ Bacterial PCR.....	53
2.5.5	T7 Endonuclease assay.....	53
2.5.6	HDR template designing	54
2.6	Tissue Histology	55
	Collection of tissue samples	55
	Tissue Processing	55
3	Chapter 3 Targeting the ECR, <i>Odam</i> and ECR-<i>Odam</i> in mouse	59
3.1	Introduction.....	59
3.2	Aims.....	64
3.3	Results.....	64

3.3.1	Designing and validating CRISPRs activity to target ECR, <i>Odam</i> and ECR- <i>Odam</i> in mouse embryonic fibroblast (MEF) cells	64
3.3.2	CRISPR cutting efficiency in MEF cells	70
3.3.3	Deleting ECR, <i>Odam</i> and ECR- <i>Odam</i> in mouse embryos.....	78
3.4	Discussion	92
4	Chapter 4: Deleting the entire casein locus in mouse	99
4.1	Introduction	99
4.2	Aims:.....	104
4.3	Results:.....	104
4.3.1	Design and validation of CRISPRs targeting the mouse casein locus	104
4.3.2	Assessing CRISPR cutting activity in MEF cells	108
4.3.3	Targeting the entire casein locus in mouse zygotes	114
4.3.4	Culturing mouse zygotes injected with CRISPRs targeting the entire casein locus	124
4.4	Discussion	126
5	Chapter 5: Modifying the alpha-lactalbumin promoter	135
5.1	Introduction	135
5.2	Aims.....	139
5.3	Results:.....	139
5.3.1	Analysing the <i>lalba</i> promoter in mammals	139
5.3.2	Amplifying the mouse <i>lalba</i> promoter.....	144
5.3.3	Creating mouse <i>lalba</i> promoters with altered STAT5 binding site using fusion PCR for reporter assay.....	152
5.3.4	Assessing expression from the altered <i>lalba</i> promoter in HC11 cells using luciferase assay	166
5.3.5	Modifying the WT STAT5 binding site of mouse <i>lalba</i> promoter in HC11 cells using CRISPR/Cas9 and homology-directed repair (HDR).....	171
5.4	Discussion	181
6	Chapter: 6 General discussion	189
7	Chapter 7 Appendices	199
8	Chapter 8- Bibliography	213

List of figures

Figure 1. 1: Evolution of mammals, birds and reptiles.....	3
Figure 1. 2: Mammary gland of a Tammar Wallaby.....	4
Figure 1. 3: Initiation of the mammary gland development during embryogenesis in the mouse.....	6
Figure 1. 4: Development of the mammary bud	7
Figure 1. 5: History of domestication of livestock animals	13
Figure 1. 6: Gene editing of targeted DNA	28
Figure 1. 7: Structure of zinc fingers	30
Figure 1. 8: Assembly of the zinc finger nuclease	31
Figure 1. 9: Assembly of TALENs.....	33
Figure 1. 10: The front and lateral structure of zinc finger and TAL protein and its binding to the target DNA.....	34
Figure 1. 11: Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs).....	36
 Figure 2. 1: Image of wet transfer set up for western blotting.....	 50
 Figure 3. 1: Comparison of the mammalian casein locus.....	 59
Figure 3. 2: Evolution of the casein locus from the early tetrapod to humans	61
Figure 3. 3: Location of ECR and the <i>Odam</i> gene in the casein locus of mammals:	62
Figure 3. 4: CRISPRs targeting mouse ECR and <i>Odam</i>	65
Figure 3. 5: Colony PCR of four bacterial clones of CRISPR number 1 targeting the mouse ECR	66
Figure 3. 6: Sequencing of CRISPRs targeting upstream of the mouse ECR	67
Figure 3. 7: Transfection of CRISPRs targeting mouse ECR and <i>Odam</i> and accessing NHEJ induction and cutting of the target site by T7 endonuclease assay	71
Figure 3. 8: Targeting ECR in MEF cells	72

Figure 3. 9: Sequencing of DNA isolated from ECR targeted MEF cells	73
Figure 3. 10: Targeting <i>Odam</i> in MEF cells	74
Figure 3. 11: Sequencing of DNA isolated from <i>Odam</i> targeted MEF cells..	75
Figure 3. 12: Deletion of ECR- <i>Odam</i> in MEF cells	76
Figure 3. 13: Sequencing of DNA isolated from ECR and <i>Odam</i> targeted MEF cells	77
Figure 3. 14: Pronuclear microinjections and transfer of injected zygotes into a recipient mouse.....	79
Figure 3. 15: Schematic of primer pairs to genotyping mouse embryos	80
Figure 3. 16: Checking primer pairs for genotyping of ECR and <i>Odam</i> targeted mouse pups	81
Figure 3. 17: Genotyping an ECR and <i>Odam</i> targeted mouse embryo	82
Figure 3. 18: Sequencing of an ECR targeted mouse embryo	83
Figure 3. 19: Mouse tooth development analysis	85
Figure 3. 20: Genotyping of ECR and <i>Odam</i> targeted mouse pups	87
Figure 3. 21: Amplifying the mouse <i>Odam</i> downstream target site	90
Figure 3. 22: Sequencing of <i>Odam</i> downstream target site.....	91
Figure 4. 1: Organisation of the casein locus in mammals	100
Figure 4. 2: Casein micelles structure.....	101
Figure 4. 3: Mice nursed by wild-type and α casein deficient dams.....	102
Figure 4. 4: Toluidine blue staining of mammary gland Sections of wild-type and <i>Csnk</i> / mice at mid-pregnancy (16.5 days post coitum)	103
Figure 4. 5: CRISPRs targeting the mouse casein locus.....	105
Figure 4. 6: Colony PCR of four bacterial clones of CRISPR number 10 targeting the mouse casein locus upstream	106
Figure 4. 7: Sequencing of CRISPRs targeting upstream of the mouse casein locus.....	107
Figure 4. 8: Transfection of CRISPRs targeting the mouse casein locus and assessment of NHEJ induction at the target site by T7 endonuclease assay	109

Figure 4. 9: Schematics of primer positions flanking the casein locus and cut sites	110
Figure 4. 10: Deletion of the casein locus in MEF cells	111
Figure 4. 11: Sequencing of DNA isolated from casein locus targeted MEF cells.....	113
Figure 4. 12: Schematics of primer organisation to assess casein locus deletion	115
Figure 4. 13: PCR to evaluate casein deletion using the three primer pairs.....	117
Figure 4. 14: PCR screening of casein locus targeted mouse pups.....	118
Figure 4. 15: Sequencing of a casein locus targeted mouse pup and the analysis of deletion	119
Figure 4. 16: Genotyping with casein locus targeted mouse pups	121
Figure 4. 17: Genotyping of the casein locus targeted mouse pups.....	123

Figure 5. 1: Maximum frequency matrix of the mouse STAT5 dimer binding site throughout the genome	141
Figure 5. 2: The mouse <i>lalba</i> gene and its promoter containing the STAT5 transcription factor binding site	145
Figure 5. 3: PCR amplification of mouse <i>lalba</i> promoter	146
Figure 5. 4: Sequencing of STAT5 binding site in the <i>lalba</i> promoter of mouse	147
Figure 5. 5: Media composition used during the differentiation of HC11 cells	149
Figure 5. 6: Western blot analysis of mouse alpha-lactalbumin and β casein	150
Figure 5. 7: Gel image of digested zbLac vector	152
Figure 5. 8 : Creating the WT and altered mouse <i>lalba</i> promoters with <i>HindIII</i> and <i>XmaI</i> restriction overhangs	154
Figure 5. 9: PCR to generate the LacP promoter fragment.....	155
Figure 5. 10: Creating the LacT promoter fragments.....	157
Figure 5. 11: Gel image of Fusion PCR creating the LacT promoter fragment	158

Figure 5. 12: Creating the LacD promoter fragments	159
Figure 5. 13: Gel image of Fusion PCR creating the LacD promoter fragment	160
Figure 5. 14: Gel image of digested zbP, zbT and zbD vectors.....	161
Figure 5. 15: Gel image of digested <i>pGL3B</i> plasmid	162
Figure 5. 16: Gel image of digested LacP, LacT and LacD vectors.....	163
Figure 5. 17: Sequencing of STAT5 binding site in LacP, LacT and LacD vectors	163
Figure 5. 18: CRISPRs targeting location on WT STAT5 binding site in mouse <i>alba</i> promoter	171
Figure 5. 19: Colony PCR of four bacterial clones containing Lac1 CRISPR oligo cloned in the <i>pSL66</i> plasmid	172
Figure 5. 20: CRISPRs number Lac1, Lac2 and Lac3 cloned in the <i>pSL66</i> plasmid.....	173
Figure 5. 21: Transfection of CRISPRs targeting the WT STAT5 binding site in mouse <i>alba</i> promoter and assessment of NHEJ induction	174
Figure 5. 22: Co-transfection of CRISPRs and HDR templates into HC11 cells	176
Figure 5. 23: Restriction digestion of DNA isolated from CRISPRs and HDR template transfected HC11 cells	177
Figure 5. 24: Co-transfection of CRISPRs and HDR template into HC11 cells	178
Figure 5. 25: Restriction digestion of DNA isolated from CRISPRs and HDR template transfected HC11 cells	179
Figure 5. 26: Co-transfection combinations of CRISPR plasmids and HDR templates into HC11 cells	180
Figure 5. 27: Restriction digestion of DNA isolated from CRISPRs and HDR template transfected HC11 cells	181

List of Tables

Table 1. 1: Major dairy cattle breeds in the world	15
Table 1. 2: Composition of milk in different species.....	17

Table 3. 1: Comparison of the ECR in different mammals	63
Table 3. 2 : Transfection optimisation of mouse embryonic fibroblast cells..	69
Table 3. 3: Litter size of five recipient females transferred with EO: 1 buffer injected zygotes	80
Table 3. 4: Concluding genotype of ECR and <i>Odsm</i> targeted mouse embryos	81
Table 3. 5: Genotyping of microinjection buffer EO: 1 injected zygotes	84
Table 3. 6: Litter size for buffer EO:1 injected zygotes	87
Table 3. 7: Genotyping of microinjection buffer EO: 1 injected zygotes	88
Table 3. 8: Litter size for buffer EO: 2 injected zygotes	89
Table 3. 9: Genotyping of microinjection buffer EO:2	89
Table 4. 1: The combinations of CRISPR plasmids and their expected sized of PCR product after casein locus deletion in MEF cells.	111
Table 4. 2: Predicting genotype of the casein locus targeted mouse pups.	116
Table 4. 3: Genotyping of the casein locus targeted mouse pups.....	120
Table 4. 4: Litter size and genotyping of the casein locus targeted mouse pups:	122
Table 4. 5: Litter size and genotyping of the casein locus targeted mouse pups	123
Table 4. 6: Analysing cultured mouse zygotes at 72h post microinjections.	125
Table 5. 1: The <i>lalba</i> promoter sequence retrieval and its alignment among selected mammals.	140
Table 5. 2: The STAT5 dimer binding site and lactose concentration of milk in mammals:	142
Table 5. 3: Alignment of putative STAT5 tetramer binding sites in the <i>lalba</i> promoter of mammals.....	143
Table 5. 4: Transfection optimisation of HC11 cells.....	163

List of graphs

Graph 1. 1: The casein content in milk of selected livestock (in %) (FAO, 2013).....	18
Graph 1. 2: The lactose content in milk of selected livestock (in g/100g milk) (FAO, 2013).	19
Graph 1. 3: The fat content in milk of selected livestock (in %) (FAO, 2013).	22
Graph 1. 4: Current and predicted per capita consumption of processed and fresh dairy products	25
Graph 5. 1: Quantitation of alpha-lactalbumin bands detected by western blot	151
Graph 5. 2: Luciferase assays of undifferentiated and differentiated HC11 cells transfected with LacP, LacT and LacD vectors	167
Graph 5. 3: A comparison in the mean luciferase activity and level of significance between Lac plasmids and their controls in undifferentiated and differentiated HC11 cells	169
Graph 5. 4: A comparison in the expression of LacP, LacT and LacD and their positive controls in undifferentiated and differentiated HC11 cells	170

List of Charts

Chart 1. 1: Global net production value of livestock products	14
Chart 1. 2: World milk production data	23

Abbreviations

AP1	Activator protein 1
BLG	Beta-lactalbumin
bp	Base pair(s)
BRF	Biological research facility
Cas9	CRISPR associated protein 9
CEBP/b	CCAAT/enhancer-binding protein beta
CEL1	Surveyor nuclease assay
CM	Casein micelle
CRISPR	Clustered regularly interspaced short palindromic repeats
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
E	Embryonic day
ECR	Evolutionary conserved region
ES cell	Embryonic stem cell
FACS	fluorescence-activated cell sorting
FAO	Food and agriculture organisation
FBS	Fetal bovine serum
FDC	Four-disulfide core
<i>FDCSP</i>	Follicular dendritic cell secreted protein
GFP	Green fluorescent protein
gRNA	Guide RNA
HDR	Homology directed repair
Het	Heterozygous
HIV	Human immunodeficiency virus
Hom	Homozygous
HRXCT	High-resolution X-ray computed tomography
<i>HTN1</i>	Histatin1
<i>HTN2</i>	Histatin2
IGs	Immunoglobulins
IRL-1	Insulin like factor 1

KO	Knock-out
LB	Lysogeny broth
LF	lactoferrin
MEF	Mouse embryonic fibroblast
MFG	Milk fat globule
ml	Millilitre
mM	millimolar
mRNA	Messenger RNA
NHEJ	Non-homologous end joining
<i>Odam</i>	Odontogenic ameloblast associated
OT	Oxytocin
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
RPMI	Roswell Park Memorial Institute
SCPPs	Secretory calcium-binding phosphoproteins
SLPI	Secretory leukocyte protease inhibitor
SOB	Super optimal broth
<i>SPARC</i>	Secreted protein, acidic, cysteine-rich
STAT3	Signal transducer and activator of transcription 3
STAT5	Signal transducer and activator of transcription 5
<i>STATH</i>	Statherin
T7EI	T7 endonuclease I
TAE	Tris-acetate-EDTA
TALEN	Transcription factor like effector nuclease
TBS-T	Tris buffer saline-tween
TE	Tris-EDTA
TF	Transcription factor
TF IIIA	Transcription factor IIIA
TFBS	Transcription factor binding site
UK	United kingdom
USA	United states of America

WAP	Whey acidic protein
WT	Wild type
ZFN	Zinc finger nuclease
μg	Microgram
μL	Microliter
μm	Micrometre
GAS	Gamma interferon activation site
TSS	Transcription start site

Chapter 1: General Introduction

1.1 Evolution of mammals

Carl Linnaeus proposed the scientific classification of class Mammalia in 1758 with an emphasis on lactation as a hallmark for identification of mammals along with their most noticeable anatomical feature, the mammary gland. The other distinctive characters that separates mammals from other vertebrates such as reptiles, amphibians and birds are the presence of hairs and fur on the skin, composed of alpha-keratins arranged in a helical structure (Dhouailly, 2009) and the presence of three small bones, the incus, the malleus and the stapes, in the middle ear (Anthwal et al., 2013). The myocardial muscles of the mammalian heart are arranged in a spiral-shaped rope and control heart pumping by a torsion movement (Stephenson et al., 2017). Another peculiar mammalian characteristic is their dual chambered body structure divided by a respiratory muscle, the diaphragm, into the chest and the abdominal region (Pickering and Jones, 2002).

The evolutionary origin of mammals dates back to the early Palaeozoic period when the amniotes were evolved from their ancestors, the early tetrapod, with a specialised ability to protect their foetus by a membrane called the amnion. The amniotes were split into the Synapsids and the Sauropsids, about 315 million years (Myr) ago (Ofstedal, 2012). The Synapsids gave rise to the Therapsids (mammal-like reptile) at the beginning of Permian era, and these evolved during the Permian, Triassic and the Jurassic period, diverging into the Prototherian and the Therian mammals about 166 Myr ago (Clemens, 1981). The Prototherian mammals had distinctive characteristics similar to the reptiles and are represented by the present-day monotremes (Osborn and Hillman, 1979) (Figure 1. 1).

The Therian mammals split about 148 Myr ago into the marsupials and the eutherians. Despite some differences, they both share similar physiology and anatomy. The most noticeable difference between these two groups is their mode of nutrient transfer to the embryos. In the eutherians, the uterus

possesses a specialised organ, the placenta, which provides nutrients to the embryos whereas in the marsupials nutrient transfer is achieved via the yolk sac (Amoroso, 1959) (Renfree, 2010).

Other major differences exist in their reproduction anatomy and nurturing strategy. The reproductive system of a female marsupial is different from that of a eutherian, having duplicate reproductive organs (two each of uterus, cervix and vagina) as compared to only one reproductive organ set in the eutherians. Only one of the duplicate sets can be fertilised at one time, and another set is used for the birth of the joey (Tyndale-Biscoe and Hearn, 1981), (Renfree et al., 1996). The marsupials have a very brief gestation period as compared to the eutherians, and their blastocyst lacks the inner cell mass (ICM) (Tyndale-Biscoe et al., 1988). The marsupials also possess more number of incisor teeth and presence of an external pouch which is used to provide safety to their suckling infants (Jacobs et al., 1989).

1.1.1 Monotremes

The earliest split in mammalian phylogeny established the monotremes, which separated the Prototherian from the Therian mammals about 166myr ago. There are only two extant species of the monotremes, the platypus are found exclusively in Australia whereas the echidna is found in Australia and Papua new Guinea (Grutzner et al., 2008). The monotremes are oviparous mammals and considered as the most primitive because of the similarity in their mode of reproduction with reptiles (Hughes, 1993).

Due to the low representation in the fossil record, the evolutionary history of the monotremes is poorly documented. High-resolution X-ray computed tomography (HRXCT) has revealed platypus as the most primitive monotreme and a direct descendant of the Prototherian and Therian split. The divergence of platypus from the echidna is estimated 17-80mya (Rowe et al., 2008).

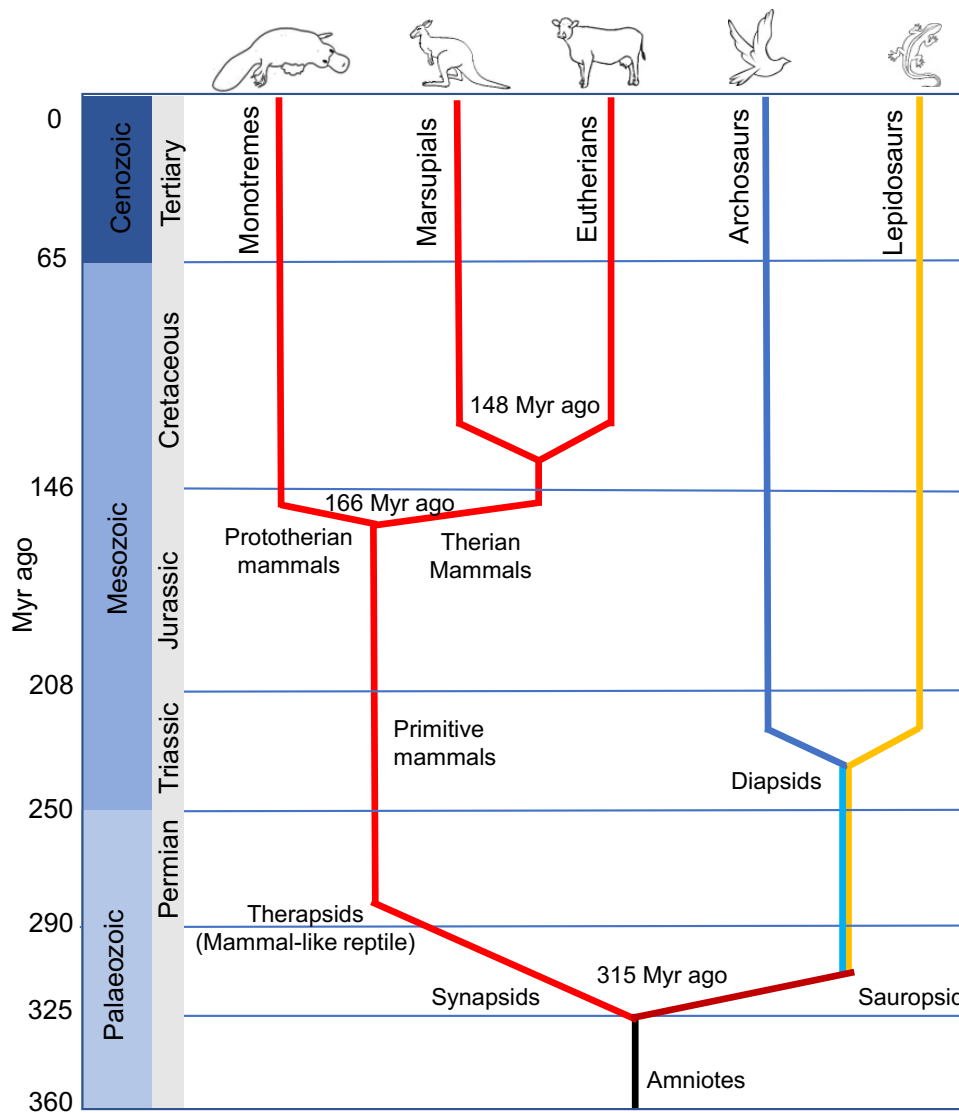


Figure 1. 1: Evolution of mammals, birds and reptiles: The Amniotes diverged into the Synapsids and the Sauropsid about 315 Myr ago. The Synapsids split into the Prototherian and the Therian mammals in the late Jurassic era. The Prototherians evolved into the monotremes, while the Therian mammals led to the origin of the marsupials and the Eutherians (adapted from Osborn and Hillman, 1979).

The mammary gland of the monotremes was derived from a modified sweat gland whose primary function was likely to provide a source of moisture to protect its eggs from desiccation during incubation. The role of this primordial mammary gland subsequently evolved from the homeostasis of the laid egg into a nutrient delivery system for the newly hatched offspring (Blackburn, 1992), (Ofstedal, 2002)

1.1.2 Marsupials

The marsupials are differentiated from the monotremes and the eutherians based on their nurturing strategy. These are defined by the presence of a permanent pouch on the female which provides safety to its infant during lactation (Deakin et al., 2013). Due to the absence of placenta, the infant marsupials are poorly developed at birth and are nurtured inside the pouch of their mother. Within a few minutes of birth, the marsupial offspring crawl into the mother's pouch and attached to one of the teats (Sharman, 1970) (Figure 1. 2).



Figure 1. 2: Mammary gland of a Tammar Wallaby: The infant joey is attached to the smaller nipple. The long nipple is still being suckled by an older joey that is now staying outside of the pouch (Nicholas et al., 2012).

The female marsupial alters the composition of milk throughout the lactation period. The level of all major components in milk changes depending on the nutritional requirements of the young, allowing the female to have overlapping lactation periods (Robert and Braun, 2012).

For example, the tammar wallaby has three phases of lactation. In the first phase, the infant joey is attached to one of the teats and the milk is very low in fats and proteins whereas carbohydrate levels are high. In the second phase, when the young joey is intermittently present in the pouch, the carbohydrate content decreases with an increase in the proteins and lipids. In the third phase, during weaning of the joey, the milk is very high in fats and moderate in proteins whereas the carbohydrate content is low. It is common for a mother to produce phase one and phase three milk at the same time to support both a newborn and a more mature joey at concurrently (Tyndale-Biscoe et al., 1988), (Sharp et al., 2017). Similarly, in the kangaroo, teats are allocated to specific infants. Milk with different compositions is secreted simultaneously by different teats according to the age of the suckling infant (Sharman and Calaby, 1964).

1.1.3 Eutherians

Today the eutherians are the most abundant group of mammals regarding both numbers and species. They are found on all continents and exhibit significant differences in size, ranging from the tiny bumblebee bat (1 g) to the blue whale (100 tonnes) (Luo, 2007), and can be either terrestrial, aquatic or aerial showing a wide diversity of habitats.

The eutherians are differentiated from the marsupials by the presence of the placenta and the higher body mass at birth (Archibald et al., 2001). Their gestation period varies from 21- 660 days in the mouse and elephant respectively which leads to the birth of an anatomically advanced infant as compared to the marsupials (Lillegraven et al., 1987). However, the average length of the lactation period is shorter in the eutherians as compare to the monotremes and marsupials (~50 d, 675 species). E.g. the body mass of a tammar wallaby is similar to a sheep. However, its lactation spans over 350 days as compared to 180- 200 days lactation period in sheep.

Figure 1.3 shows mammary gland development during embryogenesis in the mouse, which begins with the formation of five pairs of mammary

placodes at embryonic day 11.5 (E: 11.5). These placodes grow toward the chest and determine the positioning of the nipples. During E: 13.5, the placodes form a blub shaped structure and start penetrating into the mesenchyme. The mammary mesenchyme then goes through differentiation, and mammary buds appears on E: 15.5, invading into the underlying mammary fat pad. This is followed by the appearance of the first rudimentary gland on E: 18.5 (Cowin and Wysolmerski, 2010), (Gjorevski and Nelson, 2011), (Paine and Lewis, 2017).

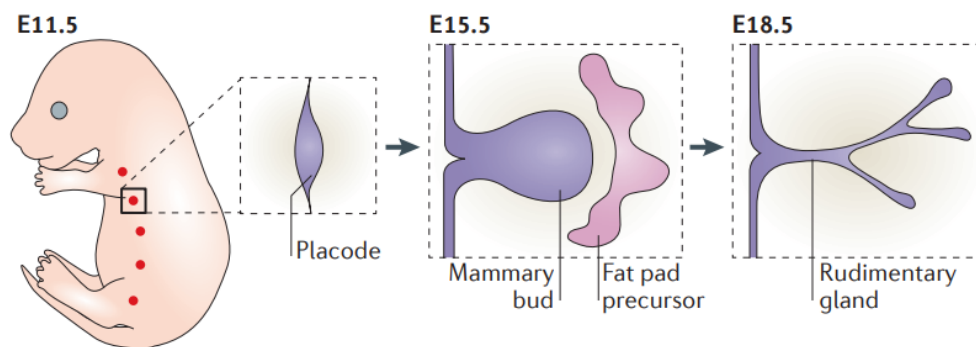


Figure 1. 3: Initiation of the mammary gland development during embryogenesis in the mouse: The positioning on five pairs of placodes and the formation of the rudimentary mammary gland from the mammary bud (Gjorevski and Nelson, 2011).

In the eutherian males, the mammary gland remains rudimentary and inactive throughout life, whereas in the females it continuously grows until the completion of puberty, after that goes through metabolic cycles during and after each pregnancy. In the adult virgin mouse, the ductal morphogenesis of mammary gland is instigated by the growth hormone along with ovarian and liver hormones such as estrogen (Silberstein and Daniel, 1987) (Mueller et al., 2002), progesterone (Stingl, 2011) and Insulin like factor 1 (IGF-1) (Rowzee et al., 2009) (Hyuk song, 2016). The rudimentary gland elongates leading to an extensive branching of the terminal end bud. However, it is inactive until the first pregnancy when differentiation occurs in response to a combined effect of progesterone and prolactin results in the development of alveolar cells (Macias

and Hinck, 2012). The alveologenesis is further fuelled by IGF-1 leading to the lactogenesis until weaning.

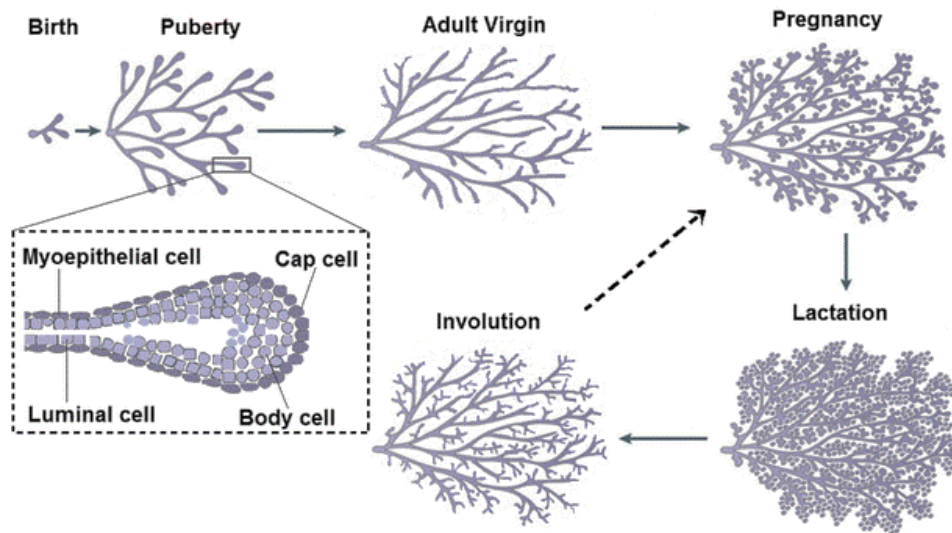


Figure 1. 4: Development of the mammary bud: Elongation and branching of the mammary bud until first pregnancy and the beginning of a metabolic cycle that involves lactation and involution (Paine and Lewis, 2017).

One of the major factor involved in the involution of the mammary gland is the signal transducer and activator of transcription 3 (STAT3) signalling (Chapman et al., 1999). The involution is followed by remodelling of mammary gland which goes through the same cycle in the next pregnancy (Rudolph et al., 2003).

1.2 Lactation in eutherians

Lactation is a process of transporting nutrition from the mother to its infant in the form of milk. Over the course of lactation, milk is produced and secreted by the mammary gland of the mother, and the survival of infant mammals is dependent on the lactation ability of their mother (Veltmaat et al., 2003) (Ofstedal, 2012). At the beginning of the lactation cycle the mammary gland secretes colostrum, which is high in antibodies and sugars, providing energy and immunity to the infant. In other phases of lactation, the levels of protein

and fat increase while the antibody and sugar content reduces.

The lactation period and its volume vary among the species. The duration of lactation differs from a few days to a few years in different mammals. The shortest lactation span is observed in the hooded seals as an example of adaptive breeding. These seals usually breed on unstable ice packs which gives a short period for nursing its infant. The lactation lasts only for 3-5 days, and rapid growth of seal pups is achieved by feeding the nutritional rich milk in a high amount (Lydersen and Kovacs, 1996).

1.3 Domestication

The origin of domestication was primarily described by Darwin (1859; 1868) who considered it a process of breeding wild animals in a confined environment that was entirely controlled by the domesticator and might have caused variations in specific body organs, allowing the animals to attain better plasticity. All domesticated animals today are herbivorous or omnivorous and were mainly domesticated for food, transport, companionship and wool. The domestication of wild terrestrial animal started with wolves more than 15,000 years ago resulted in first appearance of dogs in Eurasia (Larson et al., 2012).

The process of domestication starts with taming of an animal in a controlled environment, however not all tamed animals eventually becomes domesticated. For e.g. ancient Egyptians kept wild antelope, monkey, crocodile, cat, panthers and hyena as pets around 2000- 3000 years BC (Galton, 1865) although none of them turned into a domesticated animal except cat, which according to archaeological data, were already tamed about 10,000 years ago in Mediterranean countries for pest control (Driscoll et al., 2009).

Francis Galton (1865) considered domestication a process of strict conditions with several limiting factors. The primary condition for a wild animal to be attracted by human was its usefulness whether its dog for companionship or sheep for wool, the primarily domesticated animals were beneficial to humans. The other condition includes ease of breeding in captivity and an

ability of flourishing in limited resources. One of the major factor involved in establishment of this mutual relationship is animal's fondness toward human for e.g. Horse, donkey and zebra are closely related and members of genus *Equus*. While the formers were domesticated 5,500 years ago, the latter, zebra due to its combative behaviour toward humans, could never be tamed or domesticated. Similarly, the herbivores animals has always needed protection from their wild predators, here, one exception emerges as elephant which due to its large size never needed human for protection and thus probably never been domesticated but only tamed. These conditions were further discussed by Jared Diamond in his book '*guns, germs and steel*' who suggested that the growth rate of animal must be faster than its domesticator and there must be social structure where human can be identified and accepted as a leader and protector. These factors makes it Clear to understand why some animal could never be tamed or domesticated. Based on geographical presence, domesticated mammals are distinguished into two categories, the widespread major five which includes sheep, goat, cow, pig and horse and the finite minor nine consist of Arabian and Bactrian camel, llama and alpaca, donkey, deer, reindeer, water buffalo, Bali cattle and mithun. Some of domesticated birds are chicken, duck, goose and turkeys and insects includes honey bee and silk moth (Diamond, 1997).

The beginning of domestication suggests that the interactions between early human and wild animals perhaps remoulded their psychology which resulted in development of affection towards each other. One of the most interesting study that explains the reason and the need of social bonding between wild animals and the humans is regarding a pituitary hormone, neuropeptide oxytocin (OT), which promotes social bonding and empathy (Heinrichs and Domes, 2008), (Grinevich et al., 2014). According to Y E Herbeck, interactions between the wild animals and humans might have increased OT level in their body fluids. The elevated level of OT resulted in reduced stress and aggression in both humans and animals, encouraging persistent contact between them (Herbeck YE, 2017).

The beginning of human-animal interactions improved their bio-behavioural profile which includes an enhanced positive behaviour and their ability to learn and understand (Hennessy et al., 2015). The relaxed animals were more likely to breed well, had fewer apprehensions of captivity and were less harmful to the owners and each other. Similarly, humans were benefitted with various advantages these docile animals offered, i.e. companionship and a source of food and fibre.

Dogs were the first animals domesticated by humans about 15,000 years ago as a result of interactions between humans and wolves in the nomadic period when humans were hunters and gatherers. The beginning of this mutual relationship perhaps established when the wolves scavenged on the waste of human hunt and in return, humans were benefitted by warning of dangerous predators from wolf's howling and barking. Over a period, a group of wolves might have developed a trait for social behaviour such as an ability to gaze, giving rise to dogs (Scott et al., 1967), (Heberlein et al., 2016).

Glutamine is one of the amino acids associated with brain neurotransmission, mutations in the genes involved in glutamine metabolism might have resulted in synaptic plasticity in early domesticated dogs which includes reduced fear response and their ability to learn and remember (Li et al., 2014). The gazing behaviour of dogs was found to be improved in dogs injected with OT intranasally whereas it did not affect hand raised wolves (Nagasawa et al., 2015). The experiment shows that gazing ability of dogs is gained over time by co-evolution with humans.

The recently coined term domestication syndrome links docility with a deficiency in neural crest cells during embryonic development of vertebrates. These stem cells give rise to many tissues including adrenal cortex, brain, ganglia and are a precursor of bone and cartilage formation, as well as jaw and tooth development. A reduction in the number of these cells in dogs could underlay the observation of the lower level of adrenal hormones and anatomical variations like shortened snout, floppy ears, reduced teeth size and shortened tail when compared to the wolf (Wilkins et al., 2014).

The successful domestication of wild animals by human intervention was demonstrated by Dmitri K Belayev in Russia. The aim was to understand whether selective breeding of wild animals based purely on a behavioural trait could ultimately lead to the taming of these wild animals. The experiments were performed on the wild silver fox by breeding individuals based on a single character of tameness without involving any training or medical interference. The foxes were tested for their behaviour toward human from the age of one month until puberty; the final tameness was evaluated after achieving sexual maturity. The least aggressive foxes were bred with each other and as a control foxes with opposite behavioural traits were also bred to ensure that tameness resulted from genetic selection and not just by coincidence. After over 40 breeding cycles the wild foxes resulted in a group of docile foxes. These newly domesticated foxes were welcoming toward humans and showed behavioural traits of a pet dog.

The primary observation was the duration of the social bond formation between humans and pups of newly domesticated foxes which were found to be similar to the human and dog's pup. The significant anatomical changes like lightened coat colour, floppy ears, shortened and curved tails were observed including some hormonal and physiological changes, for example, a decreased level of corticosteroids. The outcome of experiments conducted by Dmitri K Belayev and the continuous research in the past few decades showed interesting observations such as early sexual maturity, slightly improved litter size and an increase in breeding cycles in domesticated foxes (Trut, 1999). These findings are helpful to understand how careful artificial selection of livestock could be applied to improve their productivity.

1.4 Domestication of livestock

The domestication of the livestock started about 10,000- 11,000 years ago with tamed sheep and goat (Naderi et al., 2008) (Figure 1.5). Wild aurochs (*Bos primigenius*), the ancestor of modern-day cattle, were domesticated about 10,500 years ago. The breeding of aurochs gave rise to much smaller zebu cattle (*Bos indicus*) in the Indian continent and independently to the

taurine cattle (*Bos Taurus*) in Eurasia. The physical characteristics that differentiate the *Bos indicus* from the *Bos taurus* are the presence of a large dewlap, hump and the physiological ability to survive in arid regions. The pronounced hump in the *Bos indicus* works as a potential energy storage organ like other humped animals and aids their survival in the arid regions.

The water buffalo are of two types, river and swamp; both are believed to be domesticated as independent events from their ancestor *Babulas arnee* (Kumar et al., 2007). The evidence of their domestication indicates it might have happened about 5,000 years ago in the Indus valley and Mesopotamia civilization (Cockrill, 1981). Their domestication in the Indian subcontinent prior to mature Indus Valley civilization was confirmed by the examination of mitochondrial DNA collected from India, Egypt, Pakistan and Iran (Nagarajan et al., 2015).

The early domestication of the sheep and goat suggests that their relatively smaller size might have helped the domesticator to control and tame these animals without inconvenience. The cattle and buffalo were mainly domesticated for dairy whereas the other animals were domesticated based on the specific needs, for example; domestication of chicken for eggs and meat and horse and camel for transport, trade and migration. Chart 1.1 shows the global net production value of cattle milk, cattle meat and chicken meat in 2015 which is proposed to gradually increase in the next decades.

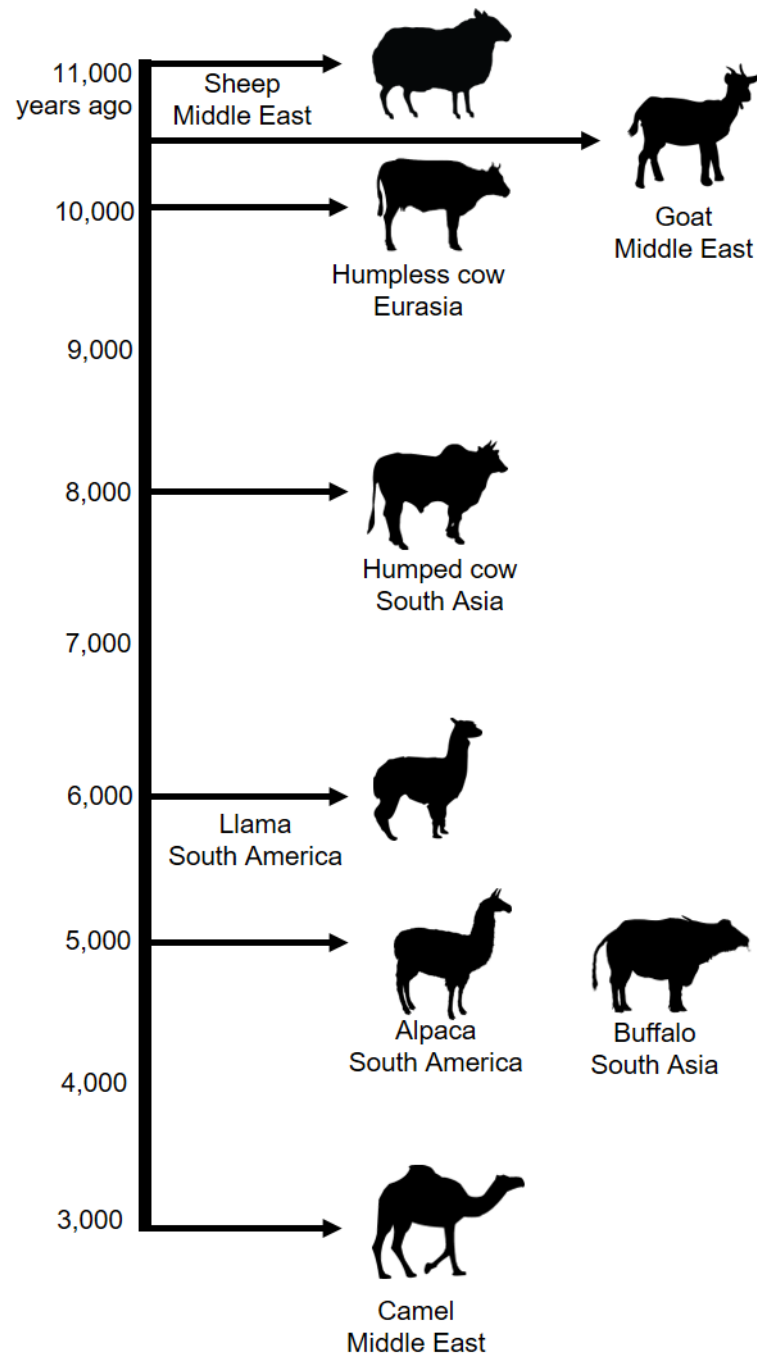


Figure 1. 5: History of domestication of livestock animals: A timeline of livestock domestication which started 11,000 years ago with sheep and one of the most recently domesticated one humped camel 3,000 years ago in the Middle East.

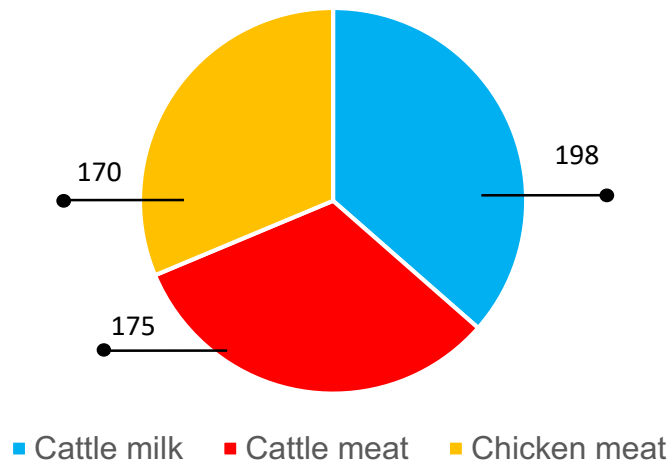


Chart 1. 1: Global net production value of livestock products: The global value of cattle milk is highest among all with 198 billion dollars in 2015 whereas cattle meat and chicken meat had 175 and 170 billion dollars respectively in the same year (in billion \$).

1.5 Dairy Livestock

1.5.1 Cattle

Dairy cattle occupy a unique niche among the milk-producing livestock as they have undergone the most stringent selection specifically for milk production. A cattle breed can be defined as a group of animals that exhibit similar physical and genetic features such as the size, colour or milking potential. The two cattle sub-species (the *Bos indicus* and the *Bos taurus*) have given rise to almost all major cattle breeds globally. The most popular dairy cattle breeds are listed in table 1.1.

The major breeds that belong to the *Bos indicus* have originated in South Asian countries are Sahiwal, Sindhi, Gyr, and Rathi. These are moderate yielding dairy cattle. The major characteristics are adaption in an arid region, and they can also be used as draught animals. The taurine cattle are Holstein Friesian, Jersey, Guernsey, Ayrshire and Brown Swiss; these are moderate to high yielding cattle. The Holstein Friesian is most significant dairy cattle breed in the world as it comprises over 35% dairy cattle worldwide (O'Mahony, 1998), (Hall et al., 2011).

Species	Breed	Milk yield per day (in kg)	Origin
<i>Bos indicus</i>	Sahiwal	8- 10	South Asia
	Sindhi	10-12	South Asia
	Rathi	16- 18	South Asia
	Gyr	17-19	South Asia
<i>Bos taurus</i>	Jersey	16-25	Norman mainland
	Guernsey	18-26	Guernsey
	Ayrshire	19-27	Scotland
	Brown swiss	21-29	Switzerland
	Holstein-Friesian	25- 35	Netherlands

Table 1. 1: Major dairy cattle breeds in the world: The economically important breeds of dairy cattle, their daily yield and origin.

1.5.2 Buffalo

Buffalo is the second largest milk-producing ruminant in the world. It contributes to about 11.1% of the world milk production in 2006- 2009, mainly in the south Asian countries (FAO, 2013) .The water buffalo has historically been categorised as either swamp or river buffalo based on morphological, behavioural and geographical criteria.

River buffalo are found in the Indian subcontinent, southwest Asia and Mediterranean countries and mainly used for milk production (Wang et al., 2007). Their milk yield is about 6-7 kg per day. The major river buffalo breed is Murrah, Mehsana, Surti, Nagpuri and Jarrarabadi found mainly in India and Pakistan (Chantalakhana, 2000).

Historically, swamp buffalo produced very little milk and were used predominantly as draught animals. By crossing this breed with river buffalo productivity has been improved to about 5- 6.5 kg per day. The swamp buffalo

are mainly found throughout South Asia and China where they are primarily used as the draught animals, providing over one-quarter of farm labour in various Asian countries (Nanda et al., 2003).

1.5.3 Goat

Goats were domesticated about 10,000 years ago in southwest Asia, during the same period and region as sheep. Goats accounted for 2.4% of global milk production in 2010 (FAO, 2013). India is the leading producer of the goat milk (30%) followed by Bangladesh (17%) and Sudan (11%). Goats are easy to rear due to their small size and adaptability. Goat farming is prevalent in the arid and semi-arid regions, and it is widely carried out in South Asia, west and central Africa. Genetic selection and breeding programs have resulted in improved yields and an extended lactation period. The average milk production is about 2.5- 3.5kg per day and about 140kg per year (Turck, 2013).

Goat farming is preferred in various developing countries due to their affordable maintenance as compared to the cattle. Their small size makes them easy to rear with minimum need for management. They are capable of surviving on forage and do not require specialised housing. Goat farming is economically significant in some Mediterranean countries where the milk is used to make cheese and yoghurt.

1.5.4 Sheep

Dairy sheep account for a total of 1.3% of world milk production. Their milk is mainly produced in the Mediterranean region for production of cheese and other dairy products which account for about 61% of the world's total sheep dairy (FAO, 2013). Sheep are believed to be the most ancient of the livestock species, mainly domesticated for wool, and have not undergone any significant improvement in either milk yield or lactation length during this period. The global use of the dairy sheep is nominal since they do not hold substantial marketable value. There are some other milks producing livestock animals, including camel, moose, reindeer, mithun, donkey, llama, alpaca, and

yak. These are mostly milked in specific regions throughout the world where these animals are found in abundance. These milk are usually produced and consumed on-farm and not favoured in large-scale dairy industries due to limited availability.

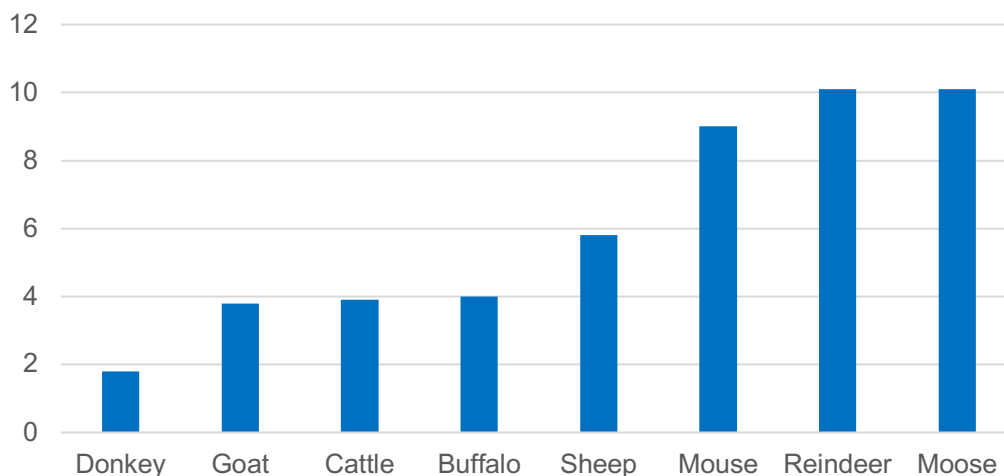
1.6 Milk constituents

Milk contains all the essential elements that an infant requires for its growth and development. Although the general composition of milk is similar in all the livestock, the concentration of milk constituents varies among species (Table 1. 2). Milk exhibits major common components such as the proteins (caseins, whey), fat and lactose (Lee et al., 1962), (Ballard and Morrow, 2013).

The major milk proteins can be categorised as caseins and whey proteins with casein accounting for ~80% of total milk proteins (Rijnkels et al., 1995), these are a cluster of proteins expressed by the casein locus in mammals. These are synthesised in the mammary epithelial cells and secreted through the apical surface into the mammary duct by exocytosis (Franke et al., 1976). Graph 1. 1 shows casein content in milk of selected livestock.

	Water	Protein	Fat	Lactose
Human	87.5	1.0	4.4	6.9
Cattle	87.8	3.3	3.3	4.7
Goat	87.7	3.4	3.9	4.4
Buffalo	83.2	4.0	7.5	4.4
Sheep	82.1	5.6	6.4	5.1

Table 1. 2: Composition of milk in different species: A comparison of major milk components such as water, proteins, fat and lactose in human and livestock (in %).



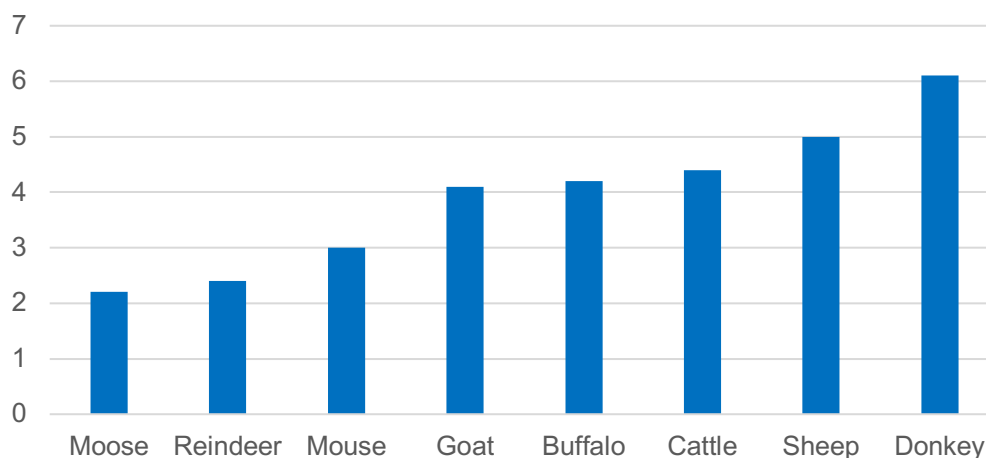
Graph 1. 1: The casein content in milk of selected livestock (in %) (FAO, 2013).

The whey proteins in milk comprise several proteins such as alpha-lactalbumin, beta-lactalbumin (BLG), whey acidic proteins (WAP), lactoferrin (LF), transferrin (TF) and immunoglobulins. Some proteins such as alpha-lactalbumin, BLG and WAP are synthesised in the mammary epithelial cells whereas LF, TF and immunoglobulins are transferred directly from the blood.

Lactose is a milk sugar found in a large number of mammalian species. Its expression is controlled by its precursor, alpha-lactalbumin which enables lactose synthase to form a catalytic motif with beta 1,4-galactosyltransferase which leads to the transport of galactose moiety to glucose, resulting in the formation of lactose (Brodbeck and Ebner, 1966). The colostrum usually contains high concentration of the lactose as it is an essential energy source for infants in all mammalian species, however, the alpha-lactalbumin expressing gene (*alba*) is poorly expressed in some species such as the fur seal which has high milk fat content to increase body mass of seal pups, required for their survival (Sharp et al., 2008).

Lactogenic hormones such as prolactin and insulin control the regulation of alpha lactalbumin. Its expression starts during the last phase of gestation and early stage of parturition when mammary gland differentiates and start producing milk (Robinson et al., 1994), (Nicholas *et al.*, 1997). The

lactose concentration in milk of selected livestock is shown in Graph 1. 2.



Graph 1. 2: The lactose content in milk of selected livestock (in g/100g milk) (FAO, 2013).

Beta- lactalbumin (BLG) is the most abundant whey protein in the milk of most mammals. Its contribution to total milk proteins is about 10% and about 50% of the total whey proteins (Loveday et al., 2009). It is absent from the milk of some mammals such as human and rodents and therefore BLG in milk from other species can act as an allergen in humans. However, it is present in the milk of ruminant and most mammalian species (Kontopidis et al., 2004).

The bovine BLG usually occurs as a dimer and shows resistance to acid hydrolysis and proteases, remaining largely intact after digestion. Due to these properties, it can behave as an antigen in the body of its recipient (Tsabouri et al., 2014). The antigenic property of BLG can thus result in an intestinal allergic reaction causing diarrhoea in infants (Gall et al., 1996).

Lysozyme is an immune protein that is present in many body fluids such as saliva, tears and milk of all the mammals including human. It causes bacterial lysis by breaking peptidoglycan links present in the bacterial cell wall (Masschalck and Michiels, 2003). Lysozyme shows over 40% similarity in amino acids with alpha-lactalbumin and its concentration in milk differs greatly between species; human milk contains 400 ug/ml lysozyme whereas the cow and goat milk contains 0.13 ug/ml and 0.25 ug/ml lysozyme respectively (Maga

et al., 2012).

The whey acidic protein (WAP) is an important whey protein found in the milk of some rodents and ruminants except sheep and cattle whereas it is also absent in the milk of human (Demmer et al., 2001). The WAP is a protein family composed of proteins with a four-disulfide core (FDC) motif, such as some inhibitor proteins SLPI (secretory leukocyte protease inhibitor) and elafin (elastase-specific inhibitor) (Scott et al., 2011), which shows homology with the protease and ATPase inhibitors therefor WAP has been termed as a protease inhibitor.

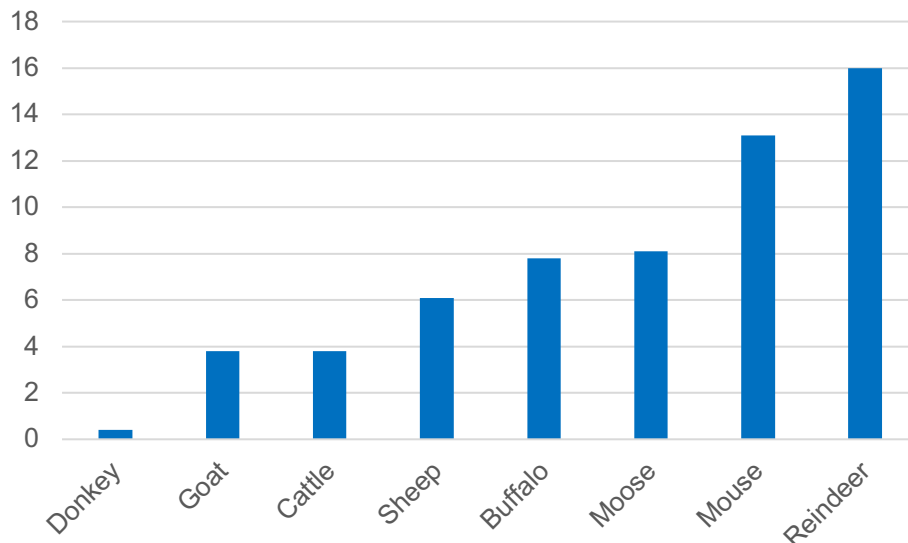
The gene encoding WAP in the pig, mouse and rabbit starts expressing during late pregnancy and continues until the last phase of lactation (Hennighausen and Sippel, 1982). The WAP deficient milk was produced in a mouse lacking both alleles of this gene. It resulted in a healthy pregnancy with normal mammary development and had no effects on the concentration of other milk components such as caseins and alpha-lactalbumin. The WAP deficient milk was found to be inessential as a protease inhibitor. The pups feeding on the WAP knockout mouse were slightly malnourished during the late phase of lactation probably due to reduced total milk proteins concentration. These observations suggest that the WAP is not essential for mammary development however it is required to sustain nourishment of pups (Triplett et al., 2005).

Lactoferrin (LF) and transferrin (TF) are iron binding antimicrobial proteins directly transported by blood into the mammary gland. LF was first reported in the 1940s when it was extracted from human whey proteins by column chromatography. Subsequent studies showed many similarities between LF and a serum iron-binding protein TF (Brock, 1980). The presence and concentration of LF and TF vary between mammals. The milk of mouse, horse, cow, and goat and pig contains both TF and LF as compared to the rat, dog and rabbit which has abundant TF but very scarce amount of LF in their milk (Masson and Heremans, 1971). LF and TF share many structural and functional similarities including the 3D structure and iron binding sites, but these can be differentiated based on their physiological origin and iron binding

capacity. TF is synthesised in the liver and transported to the mammary gland via plasma (Baker et al., 2002), whereas LF is synthesised in the extracellular fluids by neutrophils and glandular epithelial cells. It is present in the milk, tears, saliva and airway mucus where it contributes to the innate immunity of the body (Peterson et al., 2000).

Immunoglobulins (IGs) are also transferred to milk from the blood. In many mammalian species, the milk Igs are a primary source of antibodies that form an immunological link between the mother and neonatal infant (Barrington and Parish, 2001). IgG is predominantly found in the colostrum, reducing to a lower steady level throughout the remaining lactation. In humans, IgG is directly transported from the mother to embryo since maternal blood is in direct contact with the foetus by the hemochorial placenta. Species with cotyledonary, zonary or diffuse placenta such as cattle, dogs and pigs respectively cannot directly transport IgG to the developing infant during pregnancy, so the only way for transfer is via lactation.

Milk fat is a source of energy and essential fatty acids for the mammals. It carries crucial fat-soluble vitamins such as vitamin A and vitamin D (German and Dillard, 2006). The milk fat globule (MFG) is a complex structure containing various fatty acids and other lipid molecules. Its concentration in selected livestock is shown in Graph 1.3 which varies from 0.2% of milk in the donkey to 16.0 % of milk in reindeer.



Graph 1. 3: The fat content in milk of selected livestock (in %) (FAO, 2013).

1.7 Dairy industry

The first revolution in the human history started 10000BC ago in eastern Europe and was referred as the Neolithic revolution (Gordon Childe, 1942). This marked the transition of early humans from a hunter/gatherer to an agrarian society, followed by the domestication of sheep, goat and cattle around 8000BC. These were major landmarks in the establishment of human civilisation, in the next few millennia, more animal species were domesticated.

Pasteurization was invented by Louis Pasteur in 1864 as a method to preserve wine. This process of heating at low temperature for a long time, 63°C for 30 minutes, was applied to milk to destroy pathogens responsible for food poisoning. The modified procedure of heating at high temperature for a short time, 72°C for 15 seconds, was used to produce first commercially pasteurized milk (Currier and Widness, 2018). However, it was not mandatory until 1908 in Chicago. During the same period, the first large-scale milk production was started in the USA. The development of sophisticated milk processing machines, for example, the first industrial production of cheese, butter and yoghurt in Europe in 1815, 1876 and 1919 respectively increased the production of dairy products worldwide by the late 20th century (Tunick, 2009),

(Barbano, 2017).

Raw milk can either be consumed directly after pasteurisation or can be used for the production of value-added dairy products. Out of all milk-producing livestock, cattle milk is majorly produced and used in the dairy industries. In 2013, over 85% of total milk produced in the world was from cattle, followed by buffalo (11%), goat (2%), sheep (1%) and camel (0.4 %). India was the biggest producer of milk with a total of 18% of total dairy worldwide in 2013. In the past few decades, the production of milk has increased every year. In 2015, more than 800 million tonnes of the cattle milk was produced worldwide as compared to over 700 million tonnes in 2005. Similarly, the production of buffalo and the other milk have also been increasing over the years globally (Chart 1. 2).

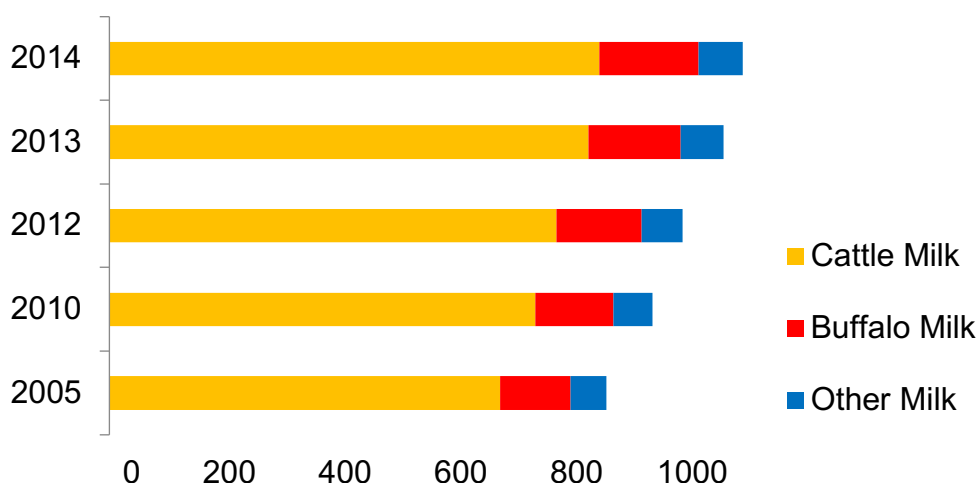


Chart 1. 2: World milk production data: A growth in the global production of cattle, buffalo and other milk from the year 2005 to 2014 (in Million Tonnes) (Summit daily, Sept 2015).

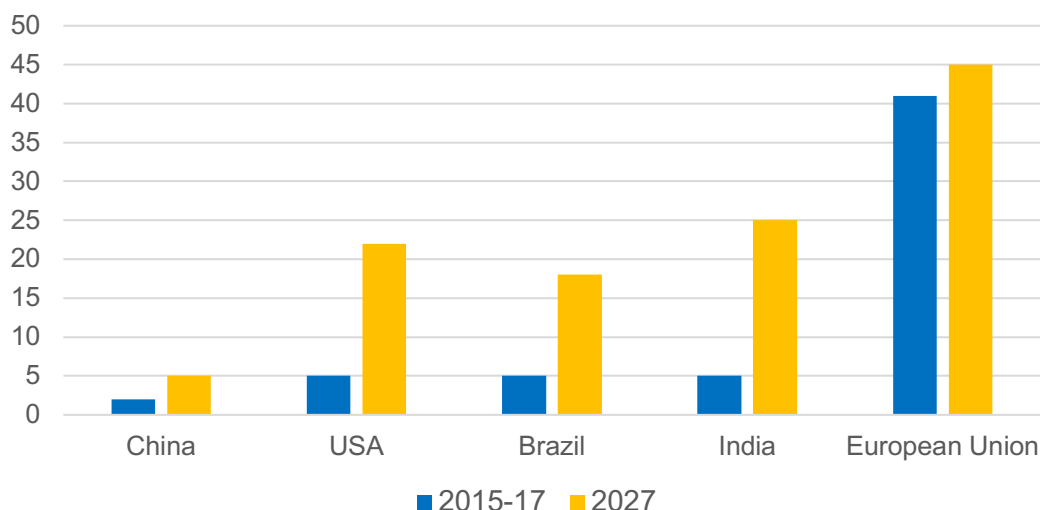
Yoghurt is a fermented milk product that is produced by acidification of milk where *Lactobacillus* forms lactic acid from milk sugars (lactose) and coagulate caseins; the process does not result in loss of proteins and fats. The worldwide production of yoghurt has increased from 30.3 million tonnes in 2010 to 40.7 million tonnes in 2015. China is the leading producer of yoghurt with 3.6 million tonnes in 2010 and 6.0 million tonnes in 2015 (world yoghurt

market report, 2013).

Butter and ghee are made by separating fats from the proteins and water content of milk. Buffalo milk is often preferred for butter production due to its high fat concentration. The Indian ghee is made mainly from cattle and buffalo milk whereas in the Mediterranean it is produced from the milk of goat, sheep and camel (Narvhus *et al.*, 1998). The major difference between butter and ghee is their fat to water ratio; butter contains 80% fat with the remainder being water whereas ghee is 99% fat with no water content and has a shelf life up to 2 years. India is the primary producer of butter and ghee. In 2014, worldwide butter and ghee production were 21.1 million metric tons. The consumption of butter is increasing gradually with 8.6 million MT in 2010, expected to reach 11.23 million MT by 2020 (world butter market report, 2000-2020).

Cheese is made from the milk caseins, and thus it is a significant source of calcium. The production of cheese is a process of milk coagulation either enzymatically or by acidification, results in the separation of proteins (mainly casein) and fat from the water. (Gouin, 1997). The global sale of cheese is about \$80.5 billion worldwide; it is expected to rise by \$110.5 billion in 2021 (Globe news wire). The global sales of dairy products have reached over \$600 billion which plays a crucial role in the world economy. The world milk production has an average annual growth rate of 1.8% and is predicted to maintain this rate in the next ten years. With a continuously growing world population, the demand for food sources is also increasing. It has been predicted that 69% more food will be required by 2050 to feed an estimated world population of 9.6 Billion (World Resources Institute, 2013). Graph 1. 4 shows the current per capita consumption of the fresh and processed milk products in major parts of the world and its forecasted increment by the year 2027. The elevation in the dairy requirement is a clear indicator of the growing

need for improved global dairy shortly.



Graph 1. 4: Current and predicted per capita consumption of processed and fresh dairy products: A comparison of current per capita fresh and milk products consumption in different countries and predicted an increase in 2027 (in kg) (FAO, 2013).

Milk can be used in controlling malnutrition in the pregnant woman and children, mainly in the underdeveloped countries. Dairy is the third largest provider of proteins and fat and the fifth largest provider of energy globally in human diets. Dairy is also linked to the livelihood and employment, with agriculture directly or indirectly employing about 28% of the world population. Dairy animals provide subsistence farmers with a guaranteed source of food and income. In more developed economies the dairy industry is also a major employer; for example, In the Netherlands and Australia, dairy farms employ 12,900 and 39,000 workers respectively. According to the global dairy platform, over 133 million dairy farms worldwide has 600 million resident workers and creates 400 million additional full-time jobs (FAO, 2013).

1.8 Transgenesis

Pioneering work on the culture of mammalian embryos such as rat and mouse in 1938 and Guinea pig and rabbit in 1961 (Biggers, 1987) set the scene for Rudolf Jaenisch to use a DNA virus to transduce the early stage mouse embryo (Jaenisch and Mintz, 1974). However, it was not until much

later that R D Palmiter succeeded in using naked DNA to introduce a transgene and demonstrate germline transmission (Palmiter et al., 1985).

The mouse was adopted as the main model organism for mammalian transgenesis research due to its small size, short breeding cycle and similar genetics, anatomy and physiology to many mammalian species including humans (Perlman, 2016). Along with the use of mouse model in studying genetics, and infection biology, the biology of its mammary gland caught the attention of many scientists. Sheep beta-lactoglobulin was successfully expressed in the milk of a transgenic mouse in 1987 (Simons et al., 1987), followed by several human transgenes such as CD4 protein (Yu et al., 1989), tissue plasminogen activator (Gordon et al., 1992), lactoferrin (Platenburg et al., 1994) and the human growth hormone (Devinoy et al., 1994).

During the same period, the use of transgenic livestock such as pig, sheep, goat and cattle as bioreactors to express human proteins such as insulin, plasminogen activator, alpha-1 antitrypsin, and coagulation factors VIII and IX in the milk was broadly discussed (Simons et al., 1987), (Janne et al., 1991) primarily because of the potential for producing large quantity of proteins and the availability of established methods for the extraction and purification of these proteins from milk. Over the last two decades, multiple exogenous proteins have been produced in the livestock mammary gland. Some of the early examples include the expression of human coagulation factor IX in the milk of transgenic sheep (Simons et al, 1988) and the expression of mouse whey acidic protein in pig milk (Shamay et al., 1991).

The potential of utilising the mammary gland as a bioreactor has fuelled sustained research activity in this area, resulting in expression of some milk and therapeutic proteins in the milk to improve either lactation in livestock or for treatment of diseases.

The milk composition was altered by adding extra copies of bovine beta and kappa casein in cattle which resulted in a nearly double concentration of the beta and kappa casein in transgenic cattle (Brophy et al., 2003). The

altered milk had reduced casein micelles with low short and medium chain fatty acids and higher palmitic and sialic acid and macronutrients whereas lactose concentration of milk remained unchanged (Laible et al., 2016). This altered milk shed light on the regulation of milk components and how milk composition could be systematically adjusted by adding or removing its components.

Pompe disease is caused by deposition of glycogen in the target organ due to the absence of an enzyme alpha-glucosidase, resulting in the organ failure, although it is rare yet a fatal disease (Kishnani et al., 2006). Transgenic rabbit expressing human alpha-glucosidase in milk were generated by transgenesis. The alpha-glucosidase was extracted from milk using column chromatography and was administered to patients of different ages (1 month to 6 months). The initial alpha glucosidase dosage was 15- 20 mg/kg of the patient, later increased to 40 mg/kg in the treatment period spanning 2-3 months. This resulted in a reduction in glycogen storage in muscles of all patients with improved tissue morphology (Van den Hout et al., 2004).

Antithrombin III is a therapeutic protein produced in the liver, its deficiency in humans is either hereditary or acquired which results in loss of regulation in blood coagulation. Transgenic goats expressing >1 g/litre human Antithrombin III in milk were generated (Edmunds et al., 1998b). In 2009, this became the first registered recombinant protein product in the mammary gland of a transgenic animal (Kues and Niemann, 2011).

The transgenic goats expressing human lysozyme, an antimicrobial enzyme, were produced to enhance the antimicrobial properties of milk (Maga et al., 2006). Later it was demonstrated that the milk containing antibacterial proteins such as lactoferrin and lysozyme could reduce mastitis in cows and improved digestion and health in the piglets with increased serum metabolites level (Cooper et al., 2013), (Cooper et al., 2014). The analysis of antimicrobial activity and physicochemical property of transgenic milk expressing human lysozyme showed its activity against gastrointestinal pathogen and similar physicochemical properties to non-transgenic milk thus it was found to be safe for human consumption (Carneiro et al., 2018). However, the development of

gene editing tools in early 2000 has opened new doors in transgenesis offering efficacy with precise gene targeting in the livestock.

1.9 Gene editors

The discovery of programmable nucleases has bought a revolution in creating genetically engineered animals by removing or inserting a gene of interest. These nucleases offer manipulation of the DNA by insertion, deletion or replacement of a single base to megabase pairs to alter the genome; the technique is called gene editing (Figure 1. 6).

Gene editors such as Zinc finger nucleases (ZFN), Transcription activator-like effector nucleases (D'Arco et al.) and newly discovered Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) are highly competent in achieving gene editing. All gene editors can be used to create a single or a double-stranded break (DSB) in the targeted DNA strand to activate or repress gene expression.

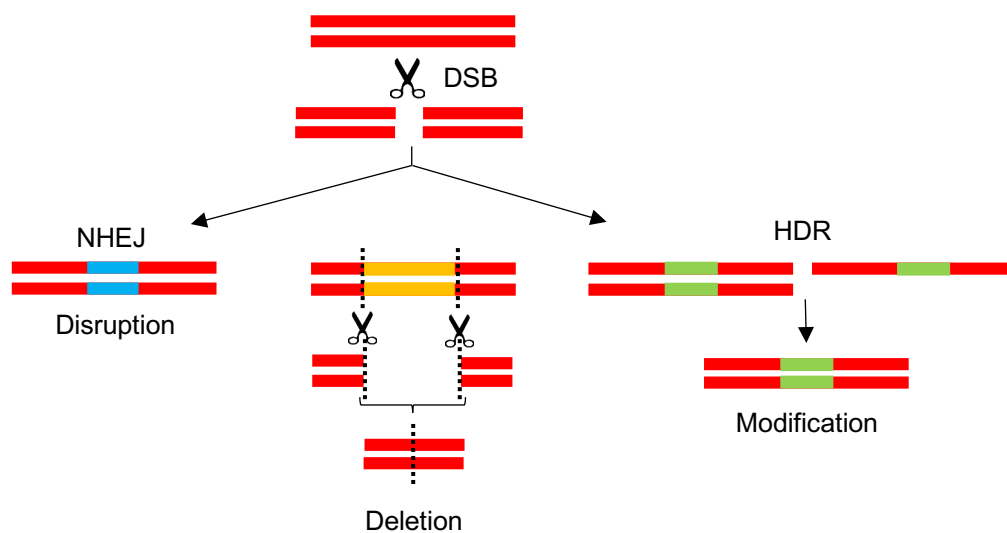


Figure 1. 6: Gene editing of targeted DNA: Creating a double-stranded break (DSB) to achieve disruption, deletion or modification in the target DNA.

To create a gene knockout, a disruption in the DNA sequence can be made by using a single gene editor which creates a DSB; the targeted DNA is repaired by non-homologous end joining (NHEJ). Similarly, the desired gene

can be deleted using two gene editors with the ends then repaired by NHEJ, here a predictable NHEJ joint can be incorporated by supplying a homology-directed repair (HDR template). To create a knock-in, a DSB is created, and a repair template is provided for HDR. This incorporates desired DNA into the target site.

1.9.1 Zinc finger nucleases (ZFN)

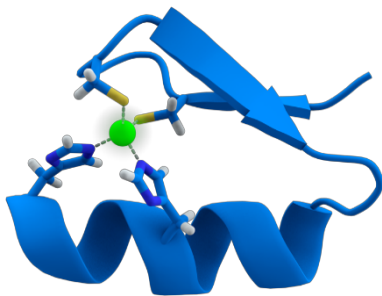
In 1983, the DNA binding ability of zinc ions was discovered in transcription factor IIIA (TF IIIA) of the *Xenopus* oocytes (Hanas et al., 1983). A single TF IIIA molecule was found to contain 7-11 zinc atoms in an independent manner, as each zinc ions was individually stabilised by two histidine residues and two cysteine residues (C2H2) attached to an α helix and β sheet making a zinc finger motif (Figure 1. 7A). These zinc finger motifs bind to 5S RNA gene in the *Xenopus* oocytes and are responsible for its transcription (Wang et al., 1985). A C2H2 zinc finger protein Zif28 was isolated from a mouse immediate early protein, whose crystalline structure showed that each zinc finger binds to 3 bp of target DNA (Pavletich and Pabo, 1991).

The first demonstration of DNA targeting activity of zinc fingers was shown in 1996 by fusing two zinc finger proteins to an endonuclease FokI domain (Kim et al., 1996) followed by creating a double-stranded break using the similar binding and cleaving domains (Smith et al., 2000). Over the years, several structural variations in zinc finger motifs have been discovered, some of the examples are zinc ribbon structure of *Pyrococcus furiosus* (Zhu et al., 1996) and Gag- knuckles structure of human HIV type I nucleocapsid protein (Guo et al., 2000). These structurally different zinc fingers are small proteins involved in various cellular processes such as gene regulation, DNA repair, replication and translation by interacting with other proteins, DNA, RNA and small ligands (Krishna et al., 2003).

The DNA-protein binding of zinc fingers depends on its structure e.g. in C2H2 zinc fingers, the binding domains form a simple $\beta\beta\alpha$ structure that folds around a zinc ion, making contact with the major groove of the target DNA

(Figure 1.7B) (Pavletich and Pabo, 1991), (Lee et al., 1989), (Pabo et al., 2001).

A.



B.

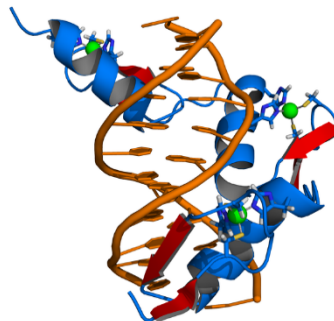


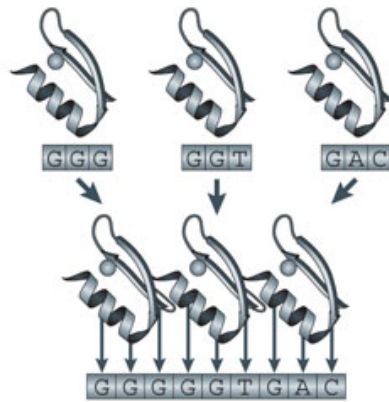
Figure 1. 7: Structure of zinc fingers: A. C2H2 zinc finger motif composed of α helix and β sheet, the zinc ion (green) is stabilised by two histidine residues and two cysteine residues. B. The zinc finger proteins-- DNA binding through the major groove of DNA (Illustration courtesy Splettstoesser T, SciStyle).

The FokI is a bacterial restriction endonuclease composed of two different domains. A DNA recognition and binding and an endonuclease domain. A chimeric FokI was prepared by replacing the native DNA binding domain to homeobox genes linked to FokI cleaving domain, resulting in a successful nuclease activity (Kim et al., 1996). This creation of artificial nuclease paved a way to construct programmable gene editing tool by combining specific DNA binding domain of zinc fingers to cleavage domain of FokI.

Figure 1. 8A shows the zinc finger nuclease (ZFN) assembly containing three C2H2 zinc finger motifs each recognising 3 bp target DNA. To design the DNA binding domain of ZFN, first of all, DNA bases of the target sequence are analysed in triplet manner, and their compatible zinc fingers are fused to create an array. A total of two binding domains are made to bind each strand of target DNA (Figure 1. 8A). These zinc fingers arms are fused with FokI endonuclease

to create a zinc finger nuclease assembly (Figure 1. 8B).

A.



B.

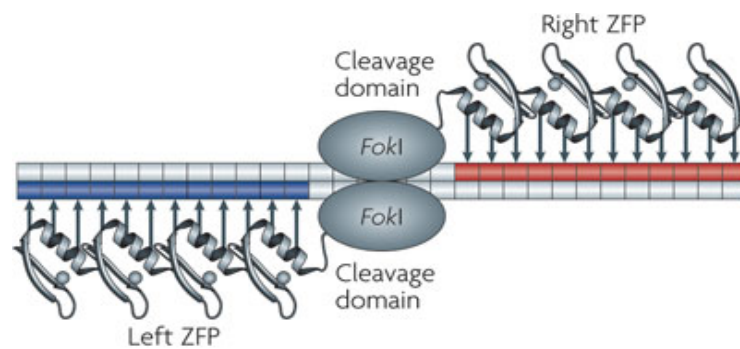


Figure 1. 8: Assembly of the zinc finger nuclease: A. Array of one arm of zinc finger binding domain B. Fusion of FokI nuclease between two zinc finger proteins array (ZFP) (Urnov et al., 2010).

The zinc finger DNA binding motif occurs in many proteins that regulate eukaryotic gene expression. The crystallised structure of Zif268 (zinc finger protein 225) helped in understanding DNA recognition by these proteins and provided a way to engineer novel DNA binding motifs (Pavletich and Pabo, 1991), (Gersbach et al., 2014). The technique was further developed by generating the zinc finger library consisting of targeting domains. This library was prepared by using approaches such as the phage display system, sequence-specific DNA recognition and random mutagenesis to alter the binding specificity of Zif268 (Rebar and Pabo, 1994), (Jamieson et al., 1994). The benefits of ZFN includes their compatibility with a variety of mammalian

somatic cell lines and their mutations are permanent and heritable.

Some of the successful use of ZFN includes its use in gene therapy studies, e.g. Glioblastoma, a type of brain cancer, was successfully suppressed using ZFN by modifying glucocorticoid receptor gene in glioblastoma-specific cytolytic T-lymphocytes. This averted the suppression of corticosteroids for treatment of a brain tumour (Till et al., 2008). CD4 cells are the main target of human immune deficiency virus (HIV), ZFNs were used to create edited CD4 cells resistant to HIV by targeting its binding co-receptor CCR5 (Han et al., 2008), the use of edited CD4 in humans showed some positive results in gene therapy against HIV (Tebas et al., 2014).

1.9.2 Transcription activator-like effector nucleases (TALENs)

Structurally, TALENs are similar to ZFNs. The binding domain of TALENs was discovered in the form of TAL (transcription activator-like) effector protein from a gram-negative bacterial pathogen *Xanthomonas*, which utilises these proteins to modify host gene expression in more than 200 different plant families (Boch and Bonas, 2010). The existence and transcription regulation activity of TALE proteins was first recognised during infection of *Xanthomonas* in rice plant (Gu et al., 2005), (Makino et al., 2006), these proteins bind to a specific gene promoter in the host plant and control the gene that facilitates infection.

The TAL effector (TALE) proteins are composed of two domains each for nuclear localisation signalling and target DNA binding respectively (Schornack et al., 2008). The DNA binding domain of TALE proteins contains 34 amino acid tandem repeats making a monomer, where the 12th and 13th amino acids of this monomer recognises a single base pair in target gene (Nemudryi et al., 2014). However, the only requirement for binding of TALE proteins to DNA is that the 5' end of target DNA must be thymine, whose absence can lead to >1000 fold decrease in TALE activity (Lamb et al., 2013). The recognition and binding ability of TALE proteins were used to engineer programmable nuclease TALENs by fusing it with FokI cleavage domain. The

FokI nuclease, as previously described, is composed of separable DNA binding and cleavage domains. The target DNA binding domain of TALENs is composed by linking 18- 24 specific DNA binding TALE proteins. The binding domain C and N terminal that are responsible for nuclear localisation signalling and target DNA binding as shown in Figure 1. 9A.

Two DNA binding domains are fused with a DNA cleaving domain, i.e. FokI with a distance of 14 to 20 bp. This assembly of two DNA binding modules and FokI generates double-stranded break in the target DNA Figure 1. 9B.

The cutting activity of the TALEN was initially used to edit target genes in the cultured human cells, human pluripotent stem cells (hPSCs), somatic zebrafish cells and nematode species (Hockemeyer et al., 2011), (Sander et al., 2011), (Wood et al., 2011), (Miller et al., 2011). However, the technique has improved over the years, and successful gene editing in livestock such as pig, cattle and sheep has already been achieved (Carlson et al., 2012), (Lillico et al., 2013). The Myostatin gene (growth differentiation factor 8), a negative regulator for muscle growth in the bovine and ovine was successfully targeted using TALEN mRNA in sheep and cattle (Proudfoot et al., 2015).

A.



B.

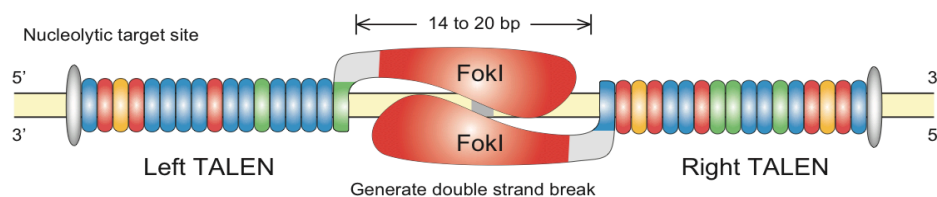


Figure 1. 9: Assembly of TALENs: A. repeat module composed of individual TALE proteins to bind at a specific sequence. B. Fusion of FokI nuclease between two repeat modules of TALENs with a distance of 14- 20 bp. Adapted from (Eid and Mahfouz, 2016)

The ZFN and TALENs both are engineered nucleases where binding proteins facilitate DNA targeting, zinc finger proteins in ZFN and TALE proteins in TALENs. The cleaving domain of both is composed by fusing FOKI nuclease between two binding arms. However, both the proteins comprise differences in target DNA recognition and binding. Each C2H2 zinc finger recognises 3 bp of target DNA.

Figure 1.10 shows the front and lateral view of zinc finger proteins and binding of one zinc finger to 3 bp of DNA as compared to the complex structure of TALE proteins and its binding to 1 bp of target DNA.

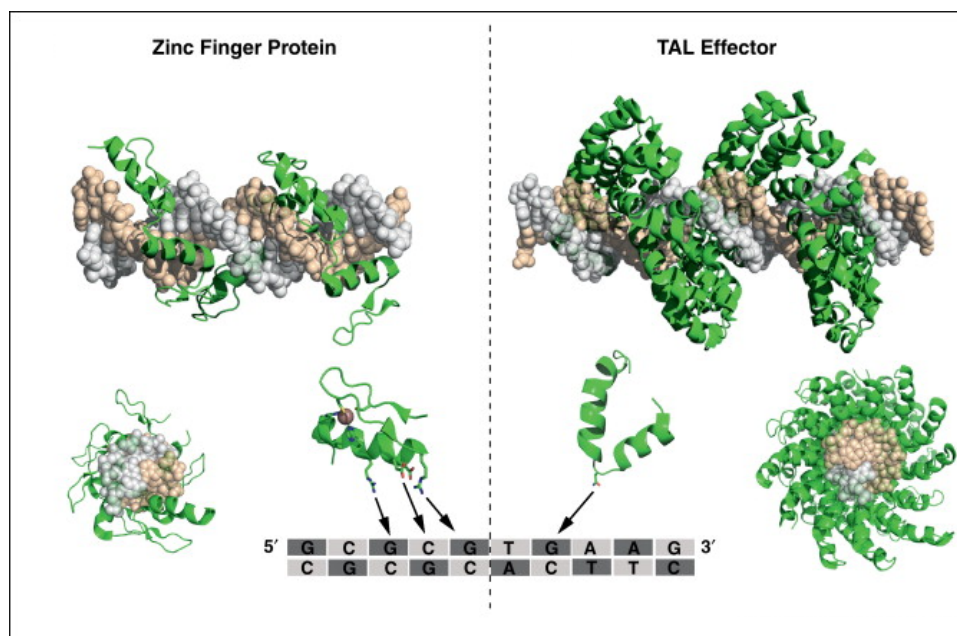


Figure 1. 10: The front and lateral structure of zinc finger and TAL protein and its binding to the target DNA: one zinc finger motif binds to 3bp target DNA whereas TALE proteins binds to the 1bp DNA (Beumer et al., 2013).

1.9.3 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)

CRISPRs are the most recent addition in genome editing tools. Its discovery has led to the expansion of the gene editor family as it provides a reliable alternative to previously discovered nucleases ZFN and TALENs.

In 1987, repeated DNA sequences were discovered in the *iap* gene in

Escherichia coli (Ishino et al., 1987), however, these interspaced repeats of 21- 37 bp DNA in bacterial and archaeal genomes were referred as clustered regularly interspaced short palindromic repeats (CRISPR) in 2002 (Jansen et al., 2002). CRISPR sequences were detected in a number of bacterial genes, and an extrachromosomal DNA of bacteriophage origin was found the interspaced DNA (spacers) (Bolotin et al., 2005), soon it was recognised as an adaptive immune response of prokaryotes (Makarova et al., 2006), (Barrangou et al., 2007). This was followed by targeting and cleaving phage and plasmid DNA by CRISPRs (Garneau et al., 2010). The presence of an endonucleases enzyme, Cas9 was discovered in *Streptococcus pyogenes*, and its activation by two different RNAs, a crRNA from the spacer DNA and a small complementary tracrRNA were recognised (Deltcheva et al., 2011).

In 2012, the CRISPRs/Cas9 guided gene editing was achieved by creating a double-stranded break in the target genome (Jinek et al., 2012). Similarly, CRISPRs/Cas9 was used to edit human and mouse cells (Cong et al., 2013). The role of the protospacer adjacent motif (PAM) in DNA recognition was demonstrated, these PAM sites differ in bacterial strains, e.g. 5'-NGG-3' for spCas9 from *Streptococcus pyogenes* (Jinek et al., 2014) and 5'-NNGRRT-3' for saCas9 from *Staphylococcus aureus* (Nishimasu et al., 2014).

To design the CRISPRs guide, first of all, the PAM sites are located in the target genome depending on the source of the Cas9 endonuclease. Usually, a 20 bp gRNA is designed at the target site upstream to a specific PAM site. Figure 1. 11 shows the binding of gRNA complex to the targeting DNA. The gRNA complex is composed of tracrRNA and crRNA and is specific for spCas9 with NGG PAM site. The binding of crRNA to the target site activates spCas9, leading to the generation of a double-stranded break.

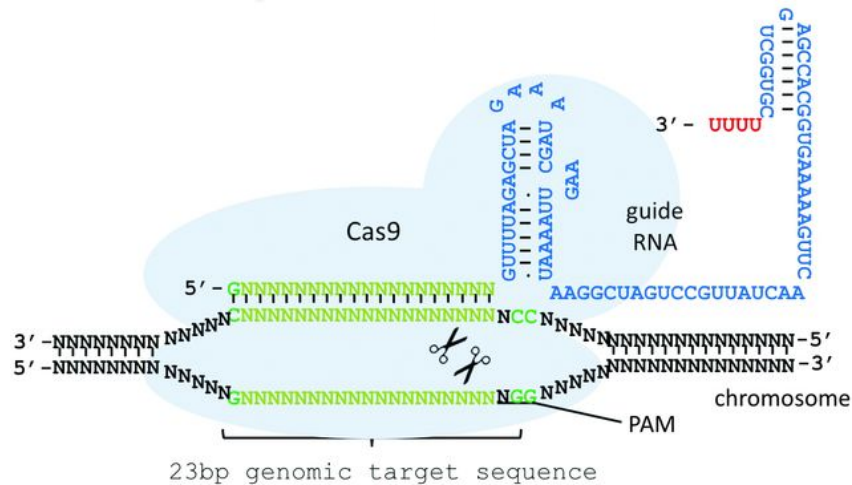


Figure 1. 11: Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs): A. The binding of gRNA complex to the 20 bp target site upstream to the PAM from spCas9.

There are now several examples whereby CRISPRs have been used to delete regions in animal cells and livestock genomes. The hypoxanthine phosphoribosyl transferase (HPRT) gene in human embryonic kidney (HEK293FT) cells were targeted by designing four guide RNAs targeting intron number one to four of this gene. The co-transfection of different combinations of this gRNA in HEK293FT cells resulted in the deletion of DNA from 1.79kb to 1000kb with deletion frequency ranging from 1- 10% (He et al., 2016).

Myostatin is a protein produced by myocytes, it is encoded by the MSTN gene and plays a role in the inhibition of muscle growth and development. The MSTN knockout sheep were successfully produced using CRISPRs with 50% deletion efficiency. The enhancement in the muscle growth was determined by the body weight of mutant lambs, which was found to be significantly higher as compared to wild-type lambs from 15-60 days after birth (Crispo et al., 2015).

The BLG, an allergen to humans, was deleted in goats by targeting the first exon of this gene using CRISPRs. The highest targeting efficiency 28.6% was acquired by zygote microinjection of two gRNA at the concentration of 50 ng/ul each. Milk was collected from 8 months old hormonally induced goats, resulted in a very slight drop in total protein concentration and secreted of the more humanised milk (Zhou et al., 2017). These examples implicate the

advancements in CRISPRs technology, and its use in farm animal transgenesis is gaining momentum and are being successfully applied to various livestock.

In the past few years, a variety of CRISPRs function has been demonstrated viz. In 2015, the idea of generating human organs in transgenic pigs using CRISPRs and stem cell biology (Feng et al., 2015) and the possibilities of making edited mosquitoes that are a poor host for malaria pathogens using CRISPRs gene drive (Gantz and Bier, 2015), were broadly discussed. In the same year, the first CRISPRs edited human embryo was reported in China (Liang et al., 2015). The year 2016 began with major approvals that expended the utility of CRISPRs, for example, UK scientist was approved to edit human embryo, and a gene-editing program was launch by the National Institute of Health (Zhao et al.), USA to cure patients. The use of CRISPRs to eliminate HIV-1 provirus in mice and edit human zygote to prevent heritable cardiac disease (Yin et al., 2017), (Ma et al., 2017) was demonstrated in 2017, followed by elimination of HIV from peripheral blood of mice using CRISPRs in 2018 (Bella et al., 2018).

1.10 Methods of transgene delivery

The mode of a transgene delivery plays a vital role in efficiency in the generation of an edited animal. The delivery of naked DNA into an animal zygote is most commonly carried out by pronuclear microinjection. Gordon and Ruddle invented the protocol of pronuclear injection of recombinant plasmid vector into mouse zygotes by using microneedles to inject plasmid suspension into mouse zygotes grasped with a holding pipette (Gordon et al., 1980). In the subsequent year, they demonstrated successful germline transmission using the same method. Two recombinant plasmids containing two different transgenes into pBR322 vector backbone were injected in mouse zygotes. The higher copy number of plasmids injected was associated with a higher rate of transformation however compromising the zygote mortality (Gordon and Ruddle, 1981). This protocol has been improved in the past few decades with very few changes and provide the most efficient way of transgene delivery with

a very high success rate (Wall, 2001). In recent years, various detailed protocols of injecting CRISPR components into mouse zygotes via pronuclear microinjections has been published to generate edited mice (Harms et al., 2014) e.g. mutation in the IRX3 and IRX5 genes in the mouse were created by pronuclear microinjection of CRISPR components (Terao et al., 2016).

Electroporation is a process where electric pulses are used to weaken zona pellucida of a zygote making it permeable for DNA intake. This technique has also been improved over the year and provides a reliable alternative for pronuclear microinjections. The rat zygotes were electrophoresed with gRNA and Cas9 protein resulted in a higher mutation rate. The same gRNA/Cas9 with ~100 bp single-stranded DNA repair template showed high knock in frequency when compared to the long single stranded and double stranded repair template which did not result in successful editing (Remy et al., 2017). Similarly, electroporation of Cas9 protein and gRNA with and without 150 bp repair template into mouse zygotes generated 100% edited pups. The higher knock in and knock out frequency was achieved by modifying the traditional electroporation protocol. E.g. treatment of mouse zygotes with acidic Tyrode's solution to improve permeability of zona pellucida was found to be inessential in this method, and the use of paired gRNA was found to be less toxic to the mouse zygotes (Troder et al., 2018).

The use of viral vectors to create transgenic animals is another efficient technique however these systems come with some limitations. The retroviral-mediated gene transfer was performed by transfer of a bacterial gene cloned in a recombinant retroviral vector into mice zygotes in 1985 (Jahner et al., 1985), (van der Putten et al., 1985). These vectors do not integrate into the host genome and but they are difficult to concentrate and usually show low transfection efficiency (Gardlik et al., 2005). The presence of zona pellucida outside the zygote is one of the major obstacles in the entry of a virus. This limitation can be overcome by using adeno-associated virus (AAV), a study has shown that these virus are permeable to zona pellucida of the mouse embryos (Botquin et al., 1994). The use of AAV offers high efficacy and

persistent gene expression. They have a packaging capacity of 4-5kb insert and remain as episomal DNA in the host (Gardlik et al., 2005), (Samulski and Muzyczka, 2014). Recombinant AAV vector expressing spCas9 and sgRNA was used to target the second exon of tryosinase gene, essential for melanin synthesis in the mouse. The pups were visually screened for eye pigmentation and coat colour resulted in 100% edited pups with high genome copy of the vector (Yoon et al., 2018).

As described in this chapter, the dairy industry is projected to grow significantly in the next few decades with the establishment of mega-dairy farms throughout the world. However, a thorough knowledge of mammary gland regulation is required to improve its mammary capacitisatation and milk quality in livestock.

The recently developed CRISPRs is a very promising gene editing tool which could help in better understanding of milk regulation in mammals, fulfilling current requirements of dairy industries such as-

- Improving milk yield from cattle (and potentially other livestock)
- Production of milk with an increased shelf life
- Creation of milk lacking major allergens
- Creation of milk with a tailored composition for specific dairy product production- e.g. high casein content for cheese production and high-fat content for butter/ ghee production.

In the past two decades, different components of the milk such as caseins, alpha-lactalbumin and BLG have been investigated in both model and livestock animals using transgenic approaches. In this thesis, we emphasize on evaluating two major components of milk i.e. casein and lactose which not only hold a commercial but also nutritive value of milk directly affecting suckling's growth and development.

2 Chapter 2 Materials and methods

2.1 General microbiological technique

2.1.1 Culture medium and antibiotics solutions

Bacteria were grown either in Luria Bertani (LB) media or on LB agar plates made using the following recipe.

Component	Weight (g)
Yeast extract	0.5
NaCl	1
Tryptophan	1
Agar (optional)	1.5
Water	100 ml

Bacterial plates were prepared by pouring melted LB agar in 90mm single vent petri dish and stored at 4°C. The stock solutions of Carbenicillin and Kanamycin (Thermo Fisher Scientific) were prepared at a concentration of 100mg/ml and 50mg/ml respectively and stored at -20°C. Carbenicillin containing media had working concentration of 100 µg/ml whereas Kanamycin media had final concentration of 50 µg/ml.

2.1.2 Transformation

DH5α competent cells (Thermo Fisher Scientific) were used for transformation. For each transformation, 25 µl of competent cells were thawed on ice, 5-10 ng DNA was added and the tubes gently flicked to mix DNA evenly. The tubes were then incubated on ice for 10-15 minutes followed by heat shock at 42°C for 45 seconds. After transformation, the tubes were again incubated on ice for another 10 minutes and 100 µl of SOB medium was then added to each tube and incubated at 37°C in shaking incubator for 60 minutes. After incubation, 50 µl culture was spread on a LB plate containing appropriate antibiotics and incubated overnight at 37°C.

In the case of TOPO TA (Invitrogen) and Zero Blunt (Invitrogen) cloning,

100µl super optimal broth (SOB) medium was added to the transformed cells after recovery on ice and incubated at 37°C for 1 hour. 10-20 µl cultured SOB medium was spread on LB plates with Carbenicillin for TOPO TA and Kanamycin for Zero blunt plasmid.

2.1.3 Plasmid miniprep

For small scale plasmid isolation, the QIAprep spin miniprep kit (Qiagen) was used. 5ml of LB media with antibiotics was inoculated with a single bacterial colony and incubated in a shaking incubator at 37°C, 200 RPM overnight. On the following day, culture medium was spun at 8,000 RPM for 5 minutes and supernatant discarded. The bacterial pellet was suspended in 250 µl chilled resuspension buffer P1 containing RNase A. the bacterial cells were lysed by adding 250 µl lysis buffer P2, the tubes inverted a few times to mix the reaction mixture and incubated at room temperature for 5 minutes. To stop the bacterial lysis, 350 µl of stabilization buffer N3 was added and tubes inverted until the solution was homogenous. The tubes were then spun at 13,000 RPM for 10 min to pellet the precipitated cell debris. The supernatant was carefully added to a filter column and spun at 13,000 RPM for 1 minute and the flow through discarded. 750 µl washing buffer PE was added to filter cartridge and spun at 13000 RPM for 1 min, flow through was discarded. The column was then spun for a further 3 minutes at 10,000 RPM to remove residual ethanol. To elute plasmids, 20-30µl elution buffer EB was added to filter column and incubated at room temperature for 1 minute followed by centrifugation at 13000 RPM for 1 minute. Plasmids were store at -20°C.

2.1.4 Plasmid Maxiprep

For large scale plasmid isolation, the EndoFree plasmid maxi kit (qiagen) was used. For plasmid maxiperp preparation, seed cultures were prepared by inoculating 5 ml of LB media plus antibiotics with a single bacterial colony and incubating at 200 RPM at 37°C for 4-6 hours. 1ml of seed culture was inoculated in 100 ml LB media with antibiotic and incubated at 200 RPM at 37°C overnight.

The next day, cultures were spun at 4000 RPM for 15 minutes and the supernatant removed. The bacterial pellet was re-suspended in 10ml of resuspension buffer P1 and mixed by pipetting. 10 ml of lysis buffer P2 was then added and incubated for 5 minutes at room temperature.

The bacterial lysis was stopped by adding 10 ml of pre-chilled neutralization buffer P3 and inverting the tubes to mix reaction mixture. The lysate was poured into the barrel of QIAfilter cartridge and incubated at room temperature for 10 minutes. The lysate was then collected in to a 50 ml tube by inserting plunger into the cartridge. 2.5 ml buffer ER was added in filtered lysate and mixed by inverting tube 10 times, followed by incubation on ice for 30minutes. The QIAGEN-tip 500 was equilibrated by adding 10 ml QBT buffer. Supernatant containing plasmid was then added to the QIAGEN tip 500 and washed twice by adding 30 ml QC buffer twice, flow through was discarded.

Plasmids were eluted with 15ml buffer QN and flow through containing plasmid was collected in a 50 ml falcon tube. To precipitate plasmids, 10.5 ml isopropanol was added to the supernatant and centrifuged at 4000 RPM for 60 minutes at 4°C. Plasmid pallet was washed with 5ml 70% ethanol and centrifuged at 4000 RPM for 10 minutes. Pallet was now air dried and dissolved in endotoxin free TE buffer to a final concentration of 500 µg/µl and stored at -20°C.

2.1.5 Nanodrop spectrophotometer

Plasmids were quantified using Nano spectrophotometer. To begin quantification, 1 µl deionised water was pipetted on to the sensor to calibrate Nano spectrophotometer followed by blanking using the appropriate buffer. Samples of the plasmids were subsequently pipeted on to sensor to quantify. The 260/280 ratio as ~1.8 was considered as a quality check of plasmid. Nano spectrophotometer sensor was cleaned with deionised water and wiped after quantification.

2.1.6 Plasmid digestion

The restriction digestion of plasmid was performed using 1 µg plasmid combined with 1 µl restriction enzyme and 1x buffer in a total of 50 µl reaction mixture. The reaction was carried out as shown in the Table below. The digested plasmid was resolved in 1.2% agarose gel as standard.

Restriction enzyme	buffer	Incubation time (Min)	Incubation temperature
<i>EcoRI</i> (New England Biolabs)	1x Cutsmart	60	37°C
<i>HindIII</i> (New England Biolabs)	1x NEB buffer 2.1	60	37°C
<i>XmaI</i> (New England Biolabs)	1x Cutsmart	60	37°C

2.1.7 Plasmid ligation

For plasmids ligation, a total of 50 ng digested plasmid was combined with three fold molar excess of insert and the total volume was adjusted up to 10 µl using ddH₂O. This mixture of then mixed to 10 µl of 2X Quick Ligation Reaction Buffer and 1 µl of Quick T4 DNA Ligase (New England Biolabs). The reaction mixture was centrifuged briefly and incubated at room temperature for 5 minutes followed by transformation into Top10 cells.

2.2 Molecular biology techniques

2.2.1 Polymerase chain reaction (PCR)

All PCR reactions were carried out using phusion polymerase master mix (New England Biolabs). 1.25 µl each forward and reverse primer (10µm) and 100 ng template DNA was mixed with 12.5 µl phusion master mix and water was added up to 25 µl per reaction mixture.

The following standard thermocycler program was used for all reactions, with annealing temperature range between 58-62°C depending on primers (Supplementary data).

98°C	1 minute	1X
98°C	10 Seconds	35X
Annealing temperature- 58-62°C	15 Seconds	
72°C	20 Seconds / kb	
72°C	5 Minutes	1X
4°C	Hold	

2.2.2 Isolation/ collection of pure DNA after PCR- For downstream reactions i.e. cloning or sequencing, the PCR was purified using one of following techniques.

2.2.2.1 Isolation of DNA from gel band

DNA was isolated from gel band using Purelink quick gel extraction kit (Thermo Fisher Scientific). DNA fragments of interest was excised from agarose gel in to a 1.5 ml tubes and 600 µl of gel solubilisation buffer L3 was added. The tubes were then incubated at 50°C for 10 minutes with a frequent inversion every 3 minutes. When the gel was completely dissolved it was loaded on to Quick gel extraction columns and centrifuged at 12,000 RPM for 1 minute. The flow-through was discarded and 500 µl wash buffer W3 was added to the column. The column was then centrifuged at 12,000 RPM for 1 minute and the flow through discarded before the column was centrifuged a second time to completely remove all traces of ethanol. To elute the DNA, 20 µl of elution buffer E5 was added to the column and incubated for 1 minute at room temperature. The column was centrifuged at 12,000 RPM and the purified DNA collected stored at -20°C.

2.2.2.2 Cleaning PCR reaction

ChargeSwitch PCR clean-up kit (Thermo Fisher Scientific) was used as standard to clean-up PCR reactions. 25 µl PCR reaction mixtures were mixed with 25 µl purification buffer N5. 10 µl of pre-vortexed ChargeSwitch magnetic

beads were added and gently mixed by pipetting to prevent the formation of bubbles. The mixture was then incubated for 1 minute at room temperature and placed in MagnaRack station for 1 minute. Once the magnetic beads had formed a pellet, the supernatant was carefully removed and discarded. The magnetic beads were then washed with 150 µl wash buffer W12 gently to prevent bubbles. The tube was again placed in the MagnaRack for 1 minute until the beads formed a pellet. The supernatant was carefully removed and discarded. To elute DNA from magnetic beads, 20 µl Elution buffer E5 was added to beads and gently mixed by pipetting. The tube was incubated for 1 minute at room temperature and placed in MagnaRack station until beads formed a pellet. Supernatant containing the purified DNA was collected and stored at -20°C.

2.2.3 Gel electrophoresis

Agarose gel electrophoresis was used to analyse PCR reactions and plasmids digested with restriction enzymes. Tris-Acetate EDTA (TAE) buffer and appropriate amount agarose powder (Thermo Scientific) was used to make agarose gel. Gels were run in 1X TAE buffer using DNA ladder mix (Thermo Fisher scientific) in first well for reference.

2.2.4 Gel documentation

Gel doc was used to capture images of resolved PCR reactions on agarose gel. All images were stored as hard as well as soft copies.

2.2.5 DNA Sequencing

The sequencing of cleaned up PCR products and plasmids was carried out by Edinburgh genomics using Sanger sequencing. 15-20 ng PCR product or 300-400 ng double stranded plasmid were mixed with 3.2 µg forward/reverse primer in a total 6 µl reaction. The results were analysed using laser gene software (DNA star).

2.2.6 Luciferase assay

The reporter assay was carried out using the Promega dual luciferase kit, HC11 cells were cultured in a 6-well plate and transfected with luciferase expressing pGL3B vector, 24h post transfection culture media was removed and cells were rinsed with PBS and cells were lysed with 500 µl passive lysis buffer (1x) for 15 minutes at room temperature. Subsequently 100 µl luciferase assay buffer (LARII) was added in each well and firefly luciferase activity was analysed using a luminometer. For assessing Renilla luciferase activity 100 µl of the Stop and Glo reagent was added in each well and measured using a luminometer.

2.3 Cell culture methods

2.3.1 Mouse embryonic fibroblast (MEF) cell culture

MEF cells were isolated from 13.5 day old mouse embryos at The Roslin Institute and stored at -150°C at P0 stage. To culture, cells were thawed at room temperature and seeded in a media containing high glucose Dulbecco's Modified Eagle Medium (DMEM) (Gibco) with 10% Fetal bovine serum (Gibco) and penicillin streptomycin antibiotics (Thermo Fisher Scientific). MEF cells were cultured in an incubator at 37°C with 5% CO₂.

2.3.2 HC11 (mouse mammary epithelial) Cell culture

HC11 cells were acquired from Dr Andreas Kolb, The University of Aberdeen, UK. These cells were cultured in a media prepared with RPMI-1640 medium with L-Glutamine and sodium bicarbonate (Gibco), 10% Fetal Bovine Serum (Gibco) and penicillin streptomycin antibiotics (Thermo Fisher Scientific). These cells were cultured in an incubator at 37°C with 5% CO₂.

2.3.3 Cryopreservation of cultured cells

To freeze ~80% confluent MEF and HC11 cells growing in a T-25 cm² flask, the culture medium was removed from the cells and rinsed with PBS.

Cells were then dissociated using 2ml of TrypLE™ Express (Invitrogen) and incubated at 37°C for 3-5 minutes. 10ml pre-warmed medium was added to the cell mixture and transferred into a 50ml falcon tube and centrifuged at 500 RPM for 5 minutes. The supernatant was removed and the cell pellet was re-suspended in 6ml of freezing medium (appendix 3). 1ml of the mixture was then aliquoted into 1.5ml tubes, and first stored at -80°C overnight followed by transferring to a -150°C freezer for long-term storage.

2.3.4 Recovery of cryopreserved cells

Cryopreserved MEF and HC11 cells were thawed in a 37°C water bath, and then added to 5 ml pre-warmed culture medium and centrifuged for 5 minutes at 500 RPM. The cell pellets were then re-suspended in appropriate volume of medium for the culture flask and grown as standard in 37°C incubator with 5 % CO₂.

2.3.5 Transfection by electroporation

CRISPRs were transfected into MEF cells using neon electroporation apparatus (Thermo Fisher Scientific). For each transfection, 5 x 10⁵ MEF cells were suspended in resuspension buffer R, 2-4 µg of CRISPR plasmid was gently added and mixed to the cells. Neon apparatus was setup by adding 3ml electrolyte buffer E to the neon tube. Buffer containing cells and plasmid was aspirated into neon tip without making air bubbles and inserted in neon tube. The electric pulse was started according to pre-optimised parameter, transfected cells were then quickly suspended in pre-warmed culture medium in a 6 well culture plate.

Cells were checked for transfection efficiency and cell viability at 24 and 48 hours post transfection, after 48 hours, cells harvested and DNA was isolated.

2.3.6 Transfection by lipofection

HC11 cells were transfected using lipofectamine 3000. Cells were seeded in a 6 well plate two days prior to transfection. On the day of transfection, media was removed from ~80% confluent HC11 cells and replaced with 3ml Opti-mem (Gibco) medium. For transfection, 7.5 μ l lipofectamine 3000 was added to 117.5 μ l Opti-mem. 2-4 μ g Plasmid was suspended in 4-8 μ l of P3000 reagent and Opti-mem was added up to 125 μ l. Diluted plasmid and lipofectamine were mixed with each other by gentle vortexing. The reaction mixture was then incubated at room temperature for 15 minutes and added to cells drop by drop. Cells were checked for transfection efficiency and cell viability at 36 hours post transfection, at that point cells were harvested and DNA was isolated.

2.3.7 Fluorescent activated cell sorting (FACS)

GFP positive MEF and HC11 cells were collected through FACS, either as single cells or total GFP cells pool. Single cells were collected and grown in 96 well plates whereas pooled cells were collected in an Eppendorf tube and grown in 6 or 24 well plates.

2.3.8 Isolation of DNA from animal cells

DNA from cultured cells was isolated using DNeasy Blood and Tissue kit (Qiagen). Culture medium from cells were removed and cells were washed with PBS. Cells were dissociated with TrypLE Express enzyme 1X (Thermo fisher Scientific) for 3-4 minutes and centrifuged at 500 RPM for 5 minutes at room temperate. Supernatant was removed and cells were suspended in 200 μ l PBS. In cell suspension, 20 μ l proteinase K and 200 μ l lysis buffer AL was added. Cells were mixed thoroughly by vortexing and lysate was pipetted in DNeasy mini spin column. Spin columns were centrifuged at 8000 RPM and flow through was discarded. The spin columns were then washed with buffer AW1 at 8000 RPM for 1 minute and AW2 at 14,000 RPM for 3 minutes one by one by and flow through was discarded.

To elute DNA, spin column was placed in a collection tube and 50 µl buffer AE was added. Spin columns were incubated at room temperature for 1 minute and centrifuged at 8000 RPM for 1 minute. DNA was stored at -20°C.

2.3.9 Isolation of DNA from animal tissues

A total of 200 µl QuickExtract solution (Epicentre) was added to finely chopped mouse ear notch sample or tail tip in a 1.5 ml tube and incubated at 65°C for 15 minutes in a water bath. The mixture was vortexed for 2 minutes and again incubated at 65°C for 15 minutes in a water bath, followed by 2 minutes incubation at 98°C in a heat block. The mixture was vortexed briefly and stored at -20°C.

2.4 Protein biology methods

2.4.1 Western blotting

Isolating proteins from HC11 cells

Proteins from HC11 cells were isolated using RIPA lysis and extraction buffer (Thermo Fisher Scientific). Culture medium was removed (stored for protein isolation) from HC11 cells growing in 6 well plates. Cells were washed twice with cold PBS and lysed using 250 µl RIPA buffer by incubating cells on ice for 5 minutes. The cell lysate was then collected in a 1.5ml tube and centrifuged at 12,000 RPM for 15 minutes to pellet the cell debris. The supernatant was carefully collected and stored at -80°C

Sample preparation

Protein samples for western blotting were prepared using 4X Laemmli Sample Buffer (BioRed). The sample buffer was prepared by adding 10 µl of reducing agent (2- Mercaptoethanol) to 90 µl of laemmli buffer. One part of 4X sample buffer was mixed with 3 parts of protein sample. The samples were heated at 95 for 5 minutes prior to loading in SDS-PAGE gels.

Protein separation by SDS-PAGE

4- 20% Mini PROTEAN precast gels (Bioered) were used to separate proteins samples. The bioered apparatus was setup and filled with 1X running buffer (appendix 3). 10 µl protein sample was loaded in each well with PageRular pre-stained protein ladder (Thermo Fisher Scientific) in the first well for reference. The gel was run at 120v for 90 minutes until all samples were visibly resolved.

Transferring proteins from PAGE gel to nitrocellulose membrane

A nitrocellulose membrane was cut with scissors and dipped in freshly made in 1X transfer buffer, 15 minutes prior to protein transfer. The transfer sandwich was assembled in a transfer cassette by placing nitrocellulose membrane on acrylamide gel and sandwiched with 2 filter papers being careful not to form air bubbles. The transfer tank was filled with 1X transfer buffer and the cassette set up and run at 120v for 120 minutes.

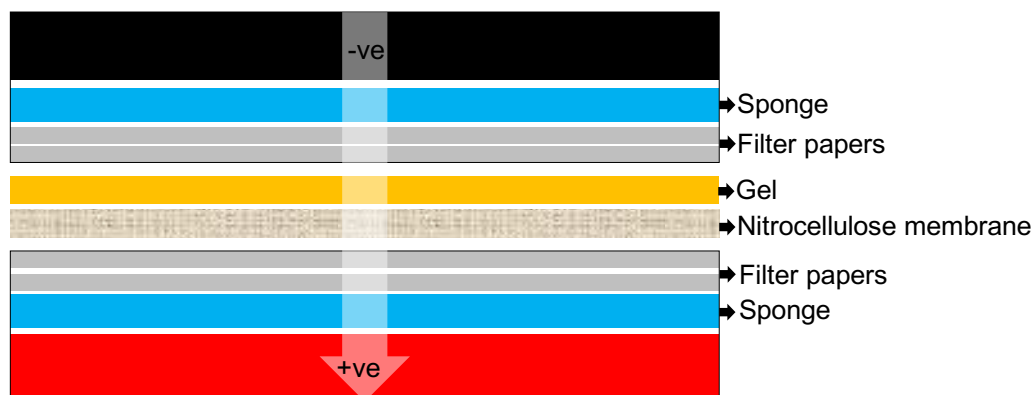


Figure 2. 1: Image of wet transfer set up for western blotting

Primary antibody incubation

Nitrocellulose membrane was washed with PBST buffer and blocked by incubating in TBST with 5% skimmed milk for 60 minutes at room temperature. Antibody solution was prepared by diluting appropriate amount of primary

antibody in PBST buffer with 5% skimmed milk. The membrane was then incubated with 5ml primary antibody solution at 4°C overnight.

Secondary antibody incubation

The membrane incubated with primary antibody was washed three times with TBST before incubating with secondary antibody. HRP conjugated secondary antibody was diluted in TBST with 5% skimmed milk. The membrane was incubated for 1 hour at room temperature with 5ml of secondary antibody solution. After incubation, the membrane was washed with TBST for 30 minutes, changing the TBST every 10 minutes.

2.5 CRISPR constructs

2.5.1 Designing CRISPR guides

CRISPR oligos were designed using an online tool <https://crispr.cos.uni-heidelberg.de>. This online software blasts sequence of interest with chosen organism and finds 20 bp long target sequence depending on presence of promoter adjacent motif (PAM) sequence.

CRISPRs to target mouse genome were designed by providing ~500 bp target sequence to the website. The selected PAM was for *streptococcus pyrogenes* i.e. NGG and the organism was selected as *mus musculus*. The sense and antisense oligos were designed with specific overhang compatible for cloning upstream of the U6 promoter in plasmid *pSL66*. The *pSL66* plasmid (Appendix. 7.4A) contains the backbone of *pX458* plasmid (Addgene, catalogue number 43138) containing a Chen mutation created by Dr. Simon Lillico (Chen et al., 2013).

2.5.2 Cloning guide oligonucleotides into *pSL66* plasmid

Sense and antisense oligos were initially diluted in TE buffer to the concentration of 100µM. Both oligos were annealed using following reaction.

Components	Volume (μl)
sgRNA sense oligo (100 μM)	0.5
sgRNA anti-sense (100 μM)	0.5
10X NEB Buffer 2	0.5
ddH ₂ O	3.5
Total	5

The sense and antisense oligos were annealed in a thermocycler at 37°C for 30 minutes and 95°C for 5 minutes, ramping down to 25°C at 5°C per min. After annealing, 2 μl of annealed oligos were diluted in 98 μl TE buffer. The following reaction mixture was prepared to ligate annealed oligos in *pSL66* plasmid.

Components	Volume (μl)
<i>pSL66</i> (100ng/μl)	0.5
Annealed oligo duplex	0.5
10X Tango buffer	1.0
DTT (10mM)	0.5
ATP (10mM)	0.5
BbsI	0.5
T7 Ligase	0.25
H ₂ O	6.25
Total	10

The above reaction was incubated in a thermocycler using the following program.

37°C	5 minutes	6x
21°C	5 minutes	
4°C	hold	

2.5.3 Plasmid safe nuclease treatment

To remove contaminating DNA and primers, Plasmid Safe ATP dependent DNase (Epicentre) was used, following the following setup.

Components	Volume (μl)
Ligated oligos in <i>pSL66</i> plasmid	10
Plasmid safe nuclease	2
Plasmid safe buffer 10X	2
ATP (10mM)	5
H ₂ O	1
Total	20

The reaction was carried out in a thermocycler at 37°C for 30 minutes, 70°C for 30 minutes.

2.5.4 Colony PCR/ Bacterial PCR

After transforming ligated *pSL66* guide plasmids in to DH5α cells, a colony PCR was carried out using DreamTaq green PCR master mix (Thermo fisher Scientific) on bacterial colonies to confirm successful ligation of guide oligos into the plasmid. The following reaction mixture was prepared and the PCR carried out in a thermocycler.

Components	Volume (μl)
Bacterial Colony	-
DreamTaq green PCR master mix (2X)	12.5
Primer oSL35 (20μM)	0.5
Sense oligo (20μM)	0.5
H ₂ O	11.5
Total	25

95°C	60 Seconds	1X
95°C	30 Seconds	35X
50°C	30 Seconds	
72°C	30 Seconds	
72°C	5 minutes	1X
10°C	Hold	

2.5.5 T7 Endonuclease assay

T7 Endonuclease assays were performed with DNA isolated from CRISPRs transfected MEF and HC11 cells. A 25 μl PCR was carried out

across the target site using Phusion polymerase master mix (NEB). 2 μ l of PCR amplified DNA was resolved in an agarose gel to estimate DNA concentration using DNA ladder mix (Thermo fisher Scientific) as reference. After estimation, 200ng of PCR product was denatured and re-annealed using the following program.

Cycle number	Temperature and time
1.	95°C, 10 minutes
2.	95-85°C, -2°C s ⁻¹
3.	85°C, 1 minute
4.	85-75°C, -3°C s ⁻¹
5.	75°C, 1 minute
6.	75-65°C, -3°C s ⁻¹
7.	65°C, 1 minute
8.	65-55°C, -3°C s ⁻¹
9.	55°C, 1 minute
10.	55-45°C, -3°C s ⁻¹
11.	45°C, 1 minute
12.	45-35°C, -3°C s ⁻¹
13.	35°C, 1 minute
14.	35-25°C, -3°C s ⁻¹
15.	25°C, 1 minute
16.	25-4°C, -3°C s ⁻¹
17.	4°C hold

The re-annealed PCR product was then incubated with 1 μ l T7 endonuclease enzyme at 37°C for 15 minutes, following the incubation, reaction was quickly resolved in 2% agarose gel.

2.5.6 HDR template designing

The ssOligos as homology directed repair (HDR) templates were designed to repair CRISPRs targeted DNA. The 90 and 120 bp HDR templates were designed with 45:45 bp and 90:30 bp homology arms respectively containing the desired changes in base pairs. The PAM site of its respective CRISPR was changed to make it unavailable to be targeted and a restriction site for *EcoRI* or *HindIII* was incorporated for rapid assessment of its integration into the targeted site.

2.6 Tissue Histology

Collection of tissue samples

Pregnant mice injected with ECR edited embryos were euthanized at E18.5 day by cervical dislocation. Mice were disinfected with 70% ethanol before the abdominal skin and wall were cut using scissors to expose the uterus. The uterus was then removed and the embryos were carefully isolated. Embryos were cleaned with PBS and decapitated using a sterile scalpel. The heads from the embryos were stored separately in 15ml Formalin Solution, Neutral Buffered 10% (Sigma aldrich) at room temperature. Tail tips were cut using scissors and the DNA isolated for genotyping.

Tissue Processing

After analysing the genotypes by PCR, relevant mouse embryo heads were processed in a tissue processor (Leica APS300s). The tissues were processed by repeated washing with denatured alcohol and xylene to wash tissues and finally coat tissues in a thin layer of wax as shown in the following table.

Reagent	Incubation time
70% Ethanol	1 Hour
95% Ethanol	1 Hour
Absolute Ethanol	1 Hour
Absolute Ethanol	1 Hour
Absolute Ethanol	1 Hour
Absolute Ethanol	1 Hour
50% Ethanol, 50% Xylene	1 Hour
Absolute Xylene	1 Hour
Absolute Xylene	1 Hour
Wax	1 Hour
Wax	1 Hour
Wax	1 Hour
Wax	1 Hour

Embedding of processed tissues into paraffin wax

Wax blocks of processed tissues were made using Leica Arcadia. A thin layer of melted paraffin wax was poured in an embedding mould to initially support the tissue, when set the moulds were filled with wax and left for setting and drying overnight before Sectioning

Sectioning and staining

Sections of embedded tissue were prepared on a microtome (Leica) using disposable microtome blades. Initially, Slides were placed on a heating rack at 60°C, covered with deionised water. Excess wax from moulds was removed and Sections of the desired thickness (usually 8-12mm) were made. Tissue Sections were then gently placed on the slide covered with warm water. After making appropriate number of Sections, water was carefully removed using a blotting paper. Sections were dried at room temperate overnight. The following day, slides were stained using Haematoxylin and Eosin (H&E) (appendix 3) in a Leica autostainer XL slide stainer using the following program.

Reagent	Time
Xylene	3 min
Xylene	2 min
Xylene	2 min
100% Alcohol	1 min
100% Alcohol	1 min
95% Alcohol	1 min
H ₂ O	1 min 30 sec
Haematoxylin	2 min
H ₂ O	1 min
Buffer	45 sec
H ₂ O	1 min
95% Alcohol	1 min
Eosin	1 min
100% Alcohol	1 min
100% Alcohol	1 min
100% Alcohol	1 min
Xylene	1 min
Xylene	1 min

The slides were mounted with Histomount (Invitrogen) and covered with a glass coverslip and analysed under a bright field microscope.

3 Chapter 3 Targeting the ECR, *Odam* and ECR-*Odam* in mouse

3.1 Introduction

The casein locus in mammals harbours milk protein genes such as α_s1 , α_s2 , β and κ caseins, along with some other non-milk genes such as odontogenic ameloblast associated (*Odam*). In milk, the genes involved in bio-mineralisation are calcium sensitive α_s1 , α_s2 and β casein genes. The presence of these genes in the casein locus suggests they might have evolved from a common ancestor and thus *Odam* is hypothesised as the progenitor of milk protein genes the casein locus (Kawasaki et al., 2011).

The milk caseins were primarily studied by Linderstrom-Lang in 1925, who discovered the presence of calcium-sensitive and calcium insensitive caseins expressed from this locus (Linderstrom-Lang K, 1925). The first casein nomenclature was proposed much later in 1984 by the American Dairy Science Association Committee that classified bovine caseins; α_s1 , α_s2 , β as calcium-sensitive and κ as calcium-insensitive. The presence of these four caseins was later confirmed in the monotremes, marsupials, and the eutherians figure 3.1. (Lefevre et al., 2009).

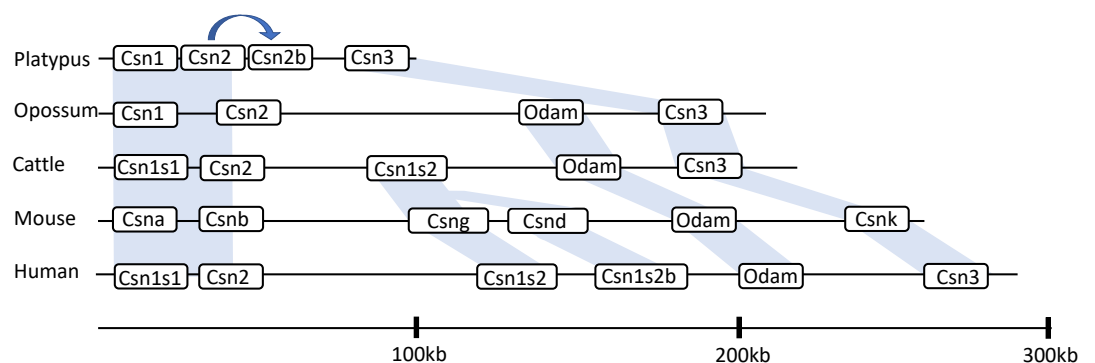


Figure 3. 1: Comparison of the mammalian casein locus: Organisation of the casein locus in platypus (Monotreme), opossum (Marsupial) and cow, mouse, human (Eutherian) genomes (Adapted from (Lefevre et al., 2009).

The casein locus in mammals spans between 250- 350kb and contains various genes, each individually controlled by its promoter. Along with *Odam*, the large intergenic region present between casein genes have shown the presence of some other non-milk genes such as statherin (*STATH*), histatin 1 and 2 (*HTN1*, *HTN2*) and follicular dendritic cell secreted protein (*FDCSP*), mainly expressed in salivary gland and activated T cells of humans respectively (Rijnkels, 2002).

The *Odam* gene, the predicted milk protein genes progenitor, is expressed in tooth associated epithelium and mainly regulates tooth enamel mineralisation in mammals (Lee et al., 2010). This gene is believed to have evolved in the early tetrapods when bio-mineralisation of minerals such as silica, calcium phosphate and calcium carbonate started and triggered bone and teeth formation.

The early tetrapods harboured two genes involved in bio-mineralisation, *Odam* and secreted protein, acidic, cysteine-rich (*SPARC*) or secreted protein, acidic, cysteine-rich like 1 (*SPARCL1*). The *SPARC* /*SPARCL1* gene was involved in the expression of secretory calcium-binding phosphoproteins (SCPPs) which is associated with tissue mineralisation in teeth (Kawasaki et al., 2005).

The Amniotes were perhaps the first animals where initial *Odam* gene shuffling and duplication occurred, and genes such as Secretory calcium-binding phosphoprotein proline-glutamine rich 1 (*SCPPPQ*) and the follicular dendritic cell secreted peptide (*FDCSP*) appeared. The gene duplication of *Odam* created *SCPPPQ1* which, during the succeeding gene duplication, arose an early milk gene called *CSN1/2*. The next significant gene shuffling and duplication of *CSN1/2* in the early synapsids gave rise to calcium sensitive milk genes such as *CSN1S1*, *CSN2* and *CSN1S2*, at the same time, the calcium-insensitive milk gene *CSN3* was originated by *FDCSP*, and the diversion of *SCPPPQ1* and *SPARCL1* genes from this locus occurred (Figure 3. 2A-D) (Jones et al., 1985), (Kawasaki et al., 2011) (Jones et al., 1985). The effects of *Odam* deletion on the development of teeth were analysed in an

Odam knockout mouse and shown to result in detachment in the junctional epithelium in older animals. Its impact in milk casein expression was not recorded (Wazen et al., 2015).

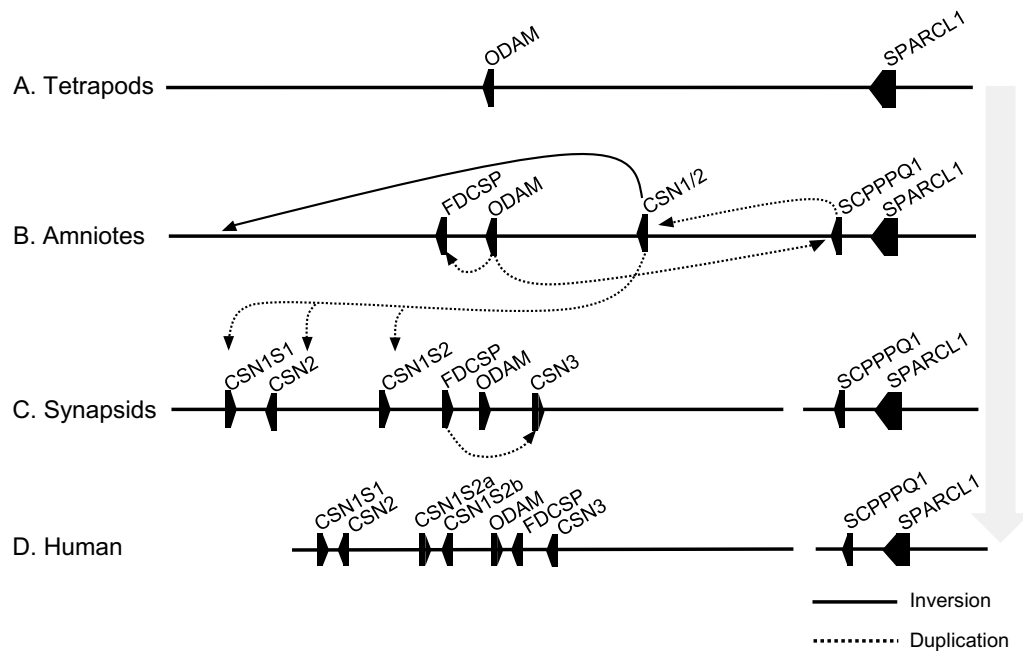


Figure 3. 2: Evolution of the casein locus from the early tetrapod to humans:

A. The presence of *Odam* and *SPARCL1* in the early tetrapods B. The rise of *FDSCP* and the early casein gene *CSN1/2* in the amniotes from *Odam* and *SCPPPQ1* C. Significant gene duplication and inversion of *CSN1/2* and *Odam* and rise of milk genes in the synapsids. D. The casein locus of humans with a well-developed casein locus (Adapted from Kawasaki, Lafont and Sire, 2011).

Identification of a conserved region in casein locus of mammals

A comparison of DNA sequences from multiple species is a powerful way to identify such genetically constrained sequences (Frazer et al., 2003). A comparative analysis of the casein locus of various mammals such as human, cow, mouse, chimpanzee, dog, and rat revealed the presence of a highly conserved region upstream to the *Odam* gene, named as an Evolutionary conserved region or ECR (Kaimala and Kumar, 2015). Figure 3. 3 shows the organisation of the casein locus and the relative location of ECR and *Odam* in mouse, rat, cow and human.

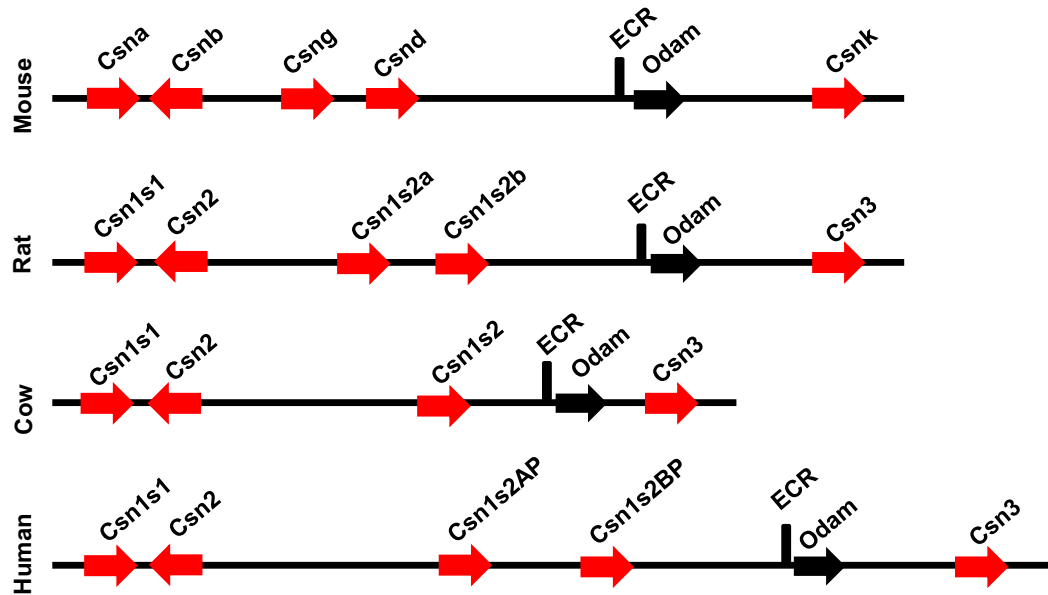


Figure 3. 3: Location of ECR and the *Od* gene in the casein locus of mammals: Comparison of the genetic organisation of casein locus of mouse, rat, cow, and human (Not drawn to scale). The location of the ECR and *Od* is highlighted by a black line and black arrow. Other milk genes in the locus are marked as red arrows, and the direction of the arrow indicates the direction of transcription of the gene. ECR refers to an evolutionarily conserved region (adapted from Kaimala and Kumar, 2015).

The ECR has not been found in non-mammalian vertebrates such as frog, chicken, and zebrafish. The size of ECR varies from 139-147 bp in different mammals (Table 3. 1), with mouse ECR found to be highly similar to the ECR of other mammals with over 85% base pairs identity.

Organism	Length (Prakobphol et al.)	The distance between <i>Odam</i> and ECR (Prakobphol et al.)
Mouse	147	756
Rat	146	52
Human	139	950
Chimpanzee	139	850
Cow	139	836
dog	147	890

Table 3. 1: Comparison of the ECR in different mammals: The mouse ECR was subjected to BLAST analysis against the genomic sequences of various mammalian species in the NCBI server. The ECR length and the distance of ECR from *Odam* are indicated (adapted from Kaimala and Kumar, 2015).

In-silico analysis of ECR showed the presence of over 29 putative transcription factors binding site in this region, some of which are known to regulate casein genes. The significant transcription factor binding sites involved in casein protein regulation found in ECR were CEBP/b, YY1, E4BP4, and Activator protein 1 (AP1). CEBP/b is member of a transcription factor family that regulates growth and differentiation of multiple cell lines (Halterman et al., 2008). YY1 and E4BP4 are found to be actively involved in the activation and repression of various promoters hence controlling gene regulation (Cowell et al., 1992).

The mouse ECR was predicted as a unidirectional repressor based on reporter assays in human cell lines such as HepG2, MCF 7 and HeLa cells which resulted in 2.5 to 6-fold decrease in expression with sense strand of ECR as compared to antisense ECR. Since *Odam* is hypothesised to be the precursor for the milk genes populating the casein locus, we set out to modify both *Odam* and ECR. Three types of knockout mice were envisaged (ECR knockout, *Odam* knockout, and ECR-*Odam* knockout) to be generated using CRISPRs to analyse effects of their deletion in casein genes expression.

3.2 Aims

1. Design and test CRISPRs targeting mouse ECR and *Odam* in mouse embryonic fibroblast cells.
2. Generate ECR, *Odam* and ECR-*Odam* KO mice using CRISPRs.
3. Analyse the impact of ECR and *Odam* loss in the regulation of casein locus and its implications on mammary gland and lactation.

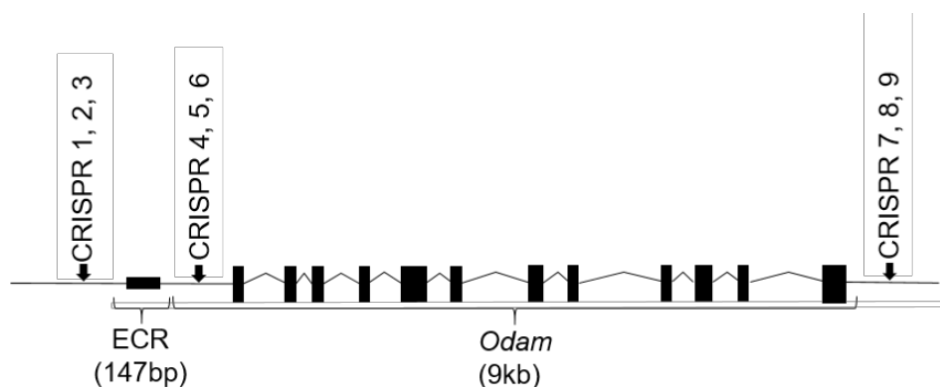
3.3 Results

3.3.1 Designing and validating CRISPRs activity to target ECR, *Odam* and ECR-*Odam* in mouse embryonic fibroblast (MEF) cells

Designing and cloning CRISPRs to target mouse ECR and *Odam*

The mouse ECR and *Odam* sequences were retrieved using Ensemble genome browser and the GRCm38.p6 database. The 147 bp ECR sequence was located upstream to the *Odam* gene based on the PCR primers used by Kaimala *et al*, 2015 to amplify this region. CRISPR guides to target ECR upstream (number 1, 2 and 3) and downstream (number 4, 5, 6) and *Odam* downstream (number 7, 8, 9) were designed using an online tool (<https://crispr.cos.uni-heidelberg.de>) as described in Chapter 2 (Section 2.5.1). The relative position of each set of guides is shown in Figure 3.4A-C.

A.



B.

```

GTGGAACAAAATTAGCCTAGGGGAGTTAGTCATGACCAAGAAAAAGCCAGGAGCG
                                     1
AGCACAGAAAAATGTCTTTGGGGTAATTGAAAGGCATTACTTAGGGTGTGGAAT
      2           3
CCATACACCACAAAGATGATTAGGAAAGGGTTATTAGCTAAGCCATTGTCTAAA
ACCCACTGAATCACTGGTCCCTTTCAATCATGATTGGTCCCTTTTTTTTTTCTAAA
GATGGGTTTTTAAATTTTAAGAGGATCTGAACAGCAGACTGAGAAAGAAAGGGGCA
CAGATCTTGCAAAGAAGTAAGGTAAGATATCAAATCTTTGCTTTGGTAGTTTAAG
                                     6 4
AGCCAGCTCTGATGCACTACCATGGGCACTCTATTGTTTGTCTATGTCTGGGAT
      5

```

C.

```

GATTGGTGTTGACATTAAACTATGTGTCTATTTTTCTCATGTCAATATCTGAT
TATTTTTCCCATGAACAAGGTTCAATTCTTTCAGCTTGTTTGTGACTCATTCTC
CATGAATATCAAATTTGTTTTGTTTTGCTCTGTTTGCCTTTTTTTAAGGCA
                                     7
GGATCTCATACTATAGCCTTGACTATCCTTGAACTACTTAGTAGGTAGGCCAG
                                     8
GCTGGCCTGGTATTCATAGAAATCTACCAGTTCCTGACTCTGAATGTGCTTAA
ACTGTATACCACCACACCTGGTCCTCCAAAATGGTTTTGAAACAGATATTGAC
      9

```

Figure 3. 4: CRISPRs targeting mouse ECR and *Od*am: A. CRISPRs targeting locations at mouse ECR (147 bp) and *Od*am (9 kb) B. Individual CRISPRs targeting mouse ECR upstream (numbers 1, 2 and 3) and downstream (numbers 4, 5 and 6). The mouse ECR is in the blue font C. Individual CRISPRs targeting *Od*am (bold) downstream sequence (numbers 7, 8 and 9). The PAM sites are in the red font.

To construct CRISPR plasmids, the sense and antisense oligos were annealed and cloned into the *pSL66* plasmid followed by transformation in top10 cells (Invitrogen). A colony PCR was performed on four transformed bacterial clones from each of the 9 CRISPRs as described in Chapter 2

(Section 2.5.4). Figure 3.5 exemplifies a colony PCR of four bacterial clones for CRISPR number 1.

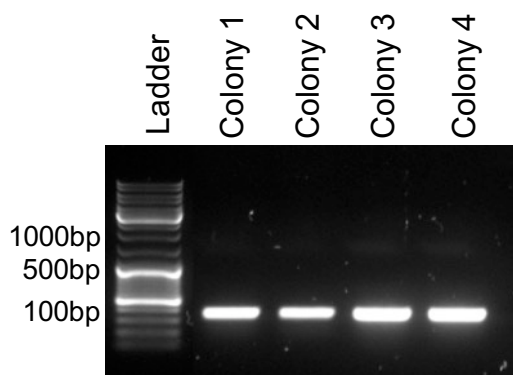


Figure 3. 5: Colony PCR of four bacterial clones of CRISPR number 1 targeting the mouse ECR: The PCR was carried out using CRISPR 1 sense oligo as a forward primer and a reverse primer *oSL35* which binds to the *pSL66* sequence downstream of oligo insert. The presence of 350 bp fragment confirms the successful cloning of CRISPR oligo duplex in the *pSL66* plasmid.

The cloning was confirmed for all 9 CRISPRs using the same colony PCR approach. Plasmids were isolated from all bacterial clones using the QIAprep spin miniprep kit (Qiagen) to verify the presence of CRISPR guides using Sanger sequencing carried out by Edinburgh genomics (Figure 3. 6).

Optimising CRISPRs transfection efficiency in MEF cells

The cutting activity of CRISPRs targeting mouse ECR and *Odam* was tested in MEF cells. The transfection optimisation of MEF cells was performed using 2 µg of *pSL66* plasmid for each transfection into 5×10^5 MEF cells. Various transfection parameters were tested using the neon electroporation system (Invitrogen) as described in chapter 2 (Section 2.3.5). The presence of viable and GFP positive cells and were checked at 24h and 48h post-transfection by visual observation using a fluorescent microscope.

Pulse voltage (v)	Pulse Width (ms)	Pulse Number	Viable cells at 24h post transfection	GFP positive cells at 24h post transfection	Viable cells at 48h post transfection	GFP positive cells at 48h post transfection
1600	10	1	~100%	~10%	~95%	~20%
1600	20	1	~100%	~10%	~95%	~20%
1600	30	1	~100%	~10%	~90%	~25%
1700	10	1	~100%	~10%	~90%	~30%
1700	20	1	~100%	~15%	~90%	~30%
1700	30	1	~100%	~10%	~90%	~30%
1800	10	1	~100%	~15%	~80%	~30%
1800	20	1	~90%	~30%	~70%	~50%
1800	30	1	~80%	~30%	~65%	~50%

Table 3. 2 : Transfection optimisation of mouse embryonic fibroblast cells: Table showing transfection optimisation in MEF cells at 24h and 48h post transfection using neon electroporation system (Invitrogen). The optimisation was concluded by visual microscopic observation of maximum GFP positive and viable cells with a particular transfection parameter.

As shown in Table 3. 2, transfection at 1800 pulse voltage, 20ms pulse width and 1 pulse number using the neon electroporation system (Invitrogen) showed the utmost viable cells (70%) with maximum GFP positive cells (50%) at 48h post transfection. Unless otherwise stated, this parameter was used for all MEF cells transfections using neon electroporation system (Invitrogen).

3.3.2 CRISPR cutting efficiency in MEF cells

CRISPRs targeting mouse ECR and *Odam* were transfected individually into MEF cells using the pre-optimised protocol. The cutting of target site and NHEJ induction by CRISPRs was analysed using T7 endonuclease assay as described in chapter 2 (Section 2.5.5). PCR for CRISPR number 1, 2, 3, 4, 5 and 6 was carried out using different primer pairs 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6 respectively whereas PCR for CRISPR number 7, 8 and 9 was using one primer pair 3.7 only. All primers are listed in chapter 7 (Appendix. 7.1) (Figure 3.7).

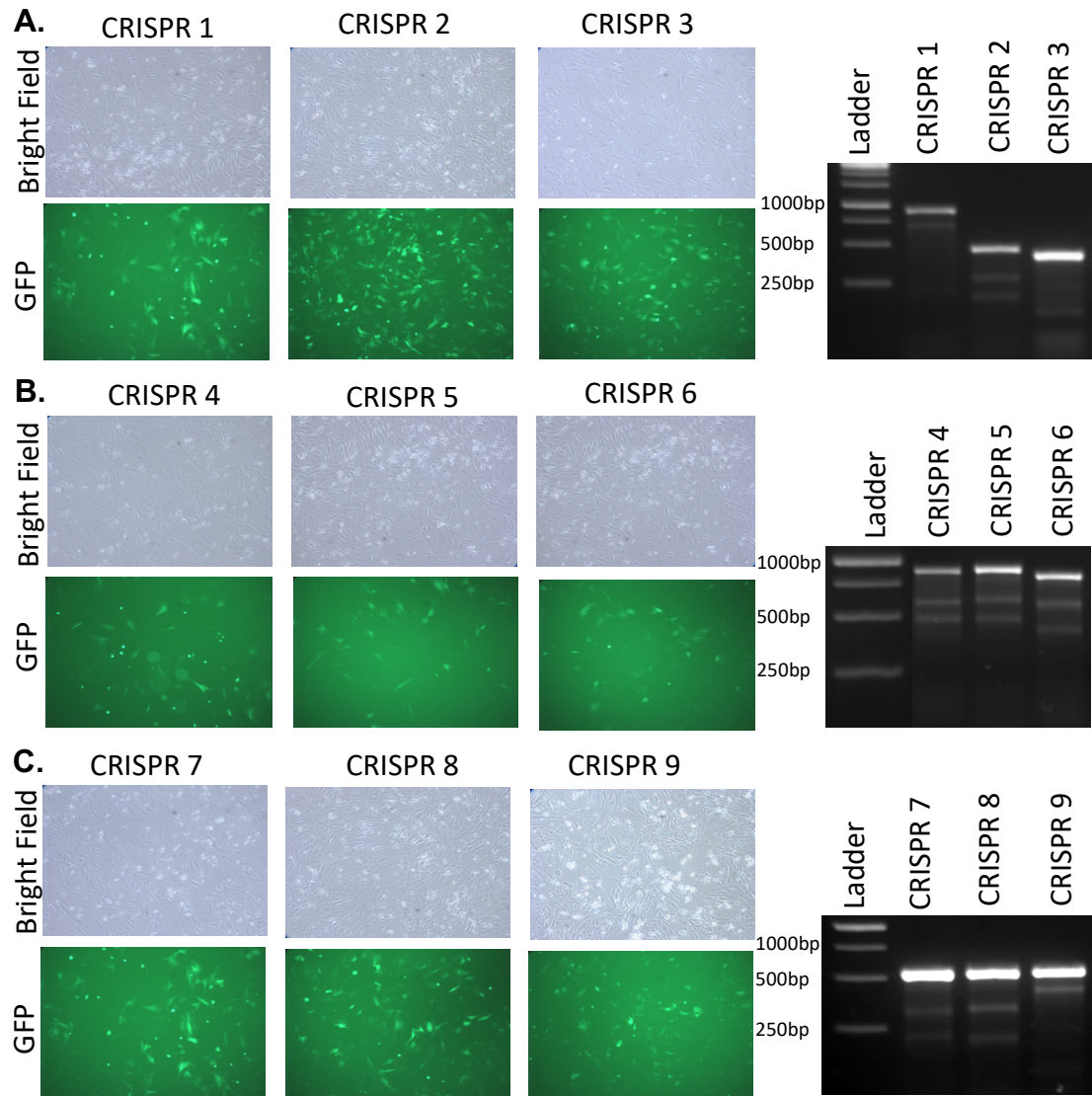


Figure 3. 7: Transfection of CRISPRs targeting mouse *ECR* and *Odam* and accessing NHEJ induction and cutting of the target site by T7 endonuclease assay: A. Transfection of CRISPRs targeting the *ECR* upstream in MEF cells and T7 assay showing wild-type and subsequent fragments from targeted DNA sized 900: 682,209 bp, 469: 262,184 bp and 425: 272,175 bp for CRISPRs number 1, 2, and 3 respectively B. Transfection of CRISPRs targeting the *ECR* downstream in MEF cells and T7 assay showing wild-type and subsequent fragments from targeted DNA sized 929: 525,381 bp, 964: 560,381 bp and 860: 529, 307 bp. for CRISPRs number 4, 5, and 6 respectively C. Transfection of CRISPRs targeting the *Odam* downstream in MEF cells and T7 assay showing wild-type and subsequent fragments from targeted DNA sized 537: 341,200 bp, 537: 314,200 bp and 537: 405,109 bp for CRISPRs number 7, 8, and 9 respectively.

The individual cutting activity of CRISPRs targeting the ECR and *Odsm* was successfully achieved in MEF cells. The collective activity was then tested by preparing CRISPR combinations based on their target site to delete ECR, *Odsm* and ECR-*Odsm* in MEF cells.

Targeting ECR in MEF cells

In an attempt to target the ECR in MEF cells, a total of nine plasmid combinations were tested using CRISPRs targeting ECR upstream (number 1, 2 and 3) and ECR downstream (number 4, 5 and 6). These CRISPRs were co-transfected in quantity of 1+1 µg (targeting ECR upstream and downstream) into MEF cells using the Neon electroporation system (Invitrogen) using the pre-optimised parameter.

Transfected cells were harvested at 48h post-transfection and DNA was isolated from total cell pool using the QuickExtract DNA extraction solution (Lucigen), and PCR was carried out. A primer pair (3.8, Appendix. 7.1) was designed flanking mouse ECR, producing a ~649 bp wild-type fragment, used to detect the combined cutting activity of CRISPRs combinations targeting mouse ECR. A smaller PCR fragment was expected in case of ECR deletion. PCR was carried out at 20s extension time for 35 cycles using the standard conditions described in chapter 2 (Section 2.2.1) and electrophoresed on a 1.2% agarose gel with 1kb DNA ladder (Invitrogen G5711) (figure 3.8).

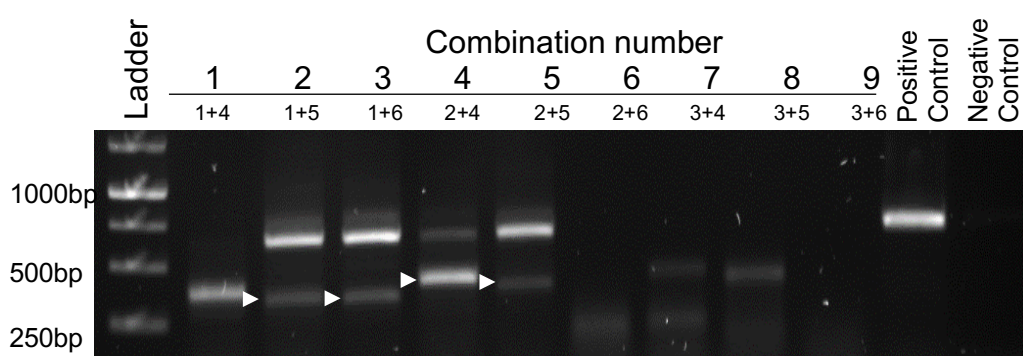


Figure 3. 8: Targeting ECR in MEF cells: The PCR of DNA from ECR targeted MEF cells showing a WT fragment (649 bp) and subsequent smaller fragment in combination numbers 2 (320 bp), 3 (304 bp), 4 (384 bp) and 5 (347 bp) (marked with a white arrow) whereas combinations 1, 7 and 8 showed faint fragments.

However, a wild-type fragment was absent in these combinations. The combinations 6 and 9 showed no fragments. The positive control (PC) was carried out with WT MEF DNA whereas negative control (NC) had no DNA.

The CRISPR combinations 2, 3, 4 and 5 showed a WT and a putative truncated fragment. The DNA from marked PCR fragments (figure 3.8) were isolated using the QIAquick DNA extraction kit (Qiagen) to analyse ECR deletion by Sanger sequencing (Edinburgh genomics). The combination number 1, 6, 7, 8 and 9 were not further analysed.

Verifying ECR deletion with Sanger sequencing

Sanger sequencing was carried out for DNA samples collected from combination number 2, 3, 4 and 5. A representation of the sequencing of CRISPR combination number 4 is shown in Figure 3. 9.

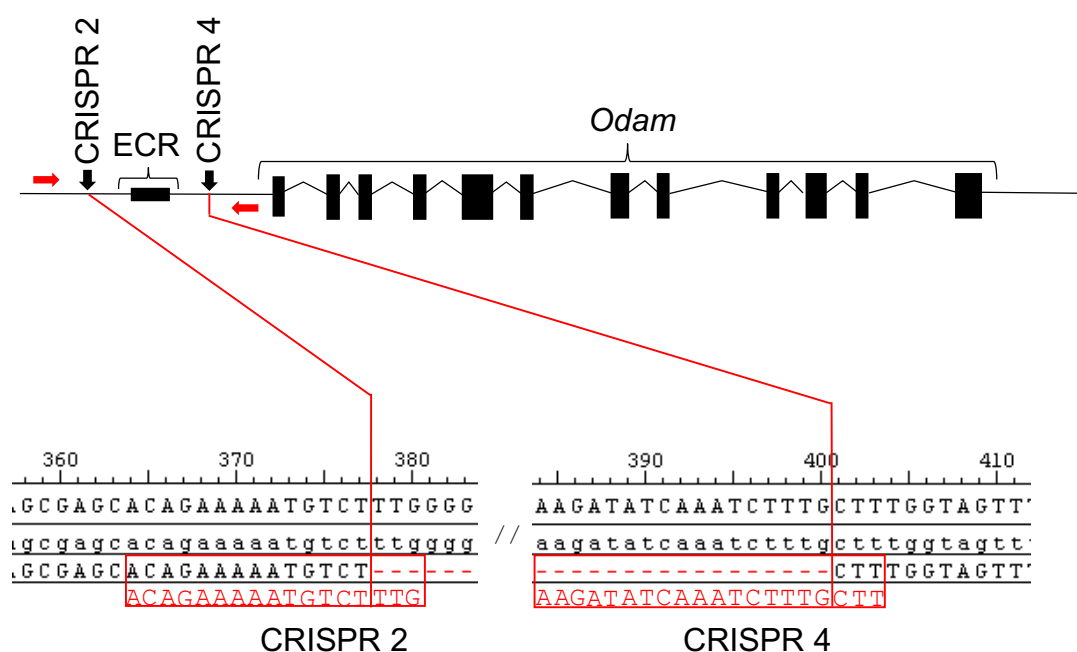


Figure 3. 9: Sequencing of DNA isolated from ECR targeted MEF cells: A Sanger sequencing showing deletion of 147 bp ECR in MEF cells using CRISPR combination number 4 (guides 2 + 4). The guide sequences are aligned underneath, with vertical lines indicating the predicted cut sites and the PAM sequences are underlined for both CRISPRs. The relative location of PCR primer 3.8 is shown by red arrows flanking the targeted region.

The ECR deletion in MEF cells was confirmed for PCR amplicons from remaining three CRISPR combinations (2, 3 and 5) using the same approach.

Targeting *Odam* in MEF cells

CRISPR plasmids targeting ECR downstream (or *Odam* upstream) and *Odam* downstream were co-transfected into MEF cells to delete *Odam*. Similarly, as with the ECR deletion, a total of 9 CRISPRs combinations were prepared using CRISPR set 2 (number 4, 5 and 6) and set 3 (number 7, 8 and 9) and transfected in quantity of 1 µg each in MEF cells using Neon electroporation system (Invitrogen) with pre-optimised conditions.

To detect the *Odam* deletion in MEF cells, a primer pair (3.9, Appendix. 7.1) was designed flanking the 9kb *Odam* gene in the mouse. PCR conditions were designed such that a PCR product was expected only after successful deletion of the *Odam* gene, no product was expected from wild-type DNA. PCR was carried out at 20s extension time for 35 cycles using the standard conditions described in chapter 2 (Section 2.2.1) and electrophoresed on a 1.2% agarose gel with DNA ladder mix (Invitrogen 15628019) (Figure 3.10).

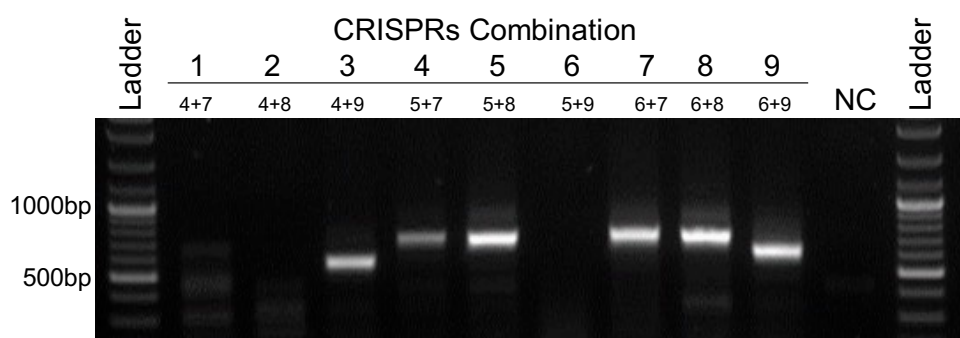


Figure 3. 10: Targeting *Odam* in MEF cells: PCR of DNA from *Odam* targeted MEF cells showing a fragment of the expected size in combination numbers 3 (600 bp), 4 (829 bp), 5 (829 bp), 7 (829 bp), 8 (829 bp) and 9 (630 bp). No full length (9kb) fragment was expected.

The CRISPR combinations 3, 4, 5, 7, 8 and 9 produced a fragment of expected size at 600 bp, 829 bp, 829 bp, 829 bp, 829 bp and 630 bp

respectively. The combination number 1, 2 and 6 showed no fragments and were not further analysed.

Verifying *Odam* deletion with sequencing

The PCR amplicon from combinations 3, 4, 5, 7, 8, and 9 (Figure 3.11) was isolated using a QIAquick DNA extraction kit (Qiagen). A Sanger sequencing (Edinburgh genomics) was carried out for all six samples.

Figure 3.11 exhibits the sequencing results for CRISPR combination number 3, targeting *Odam* in MEF cells.

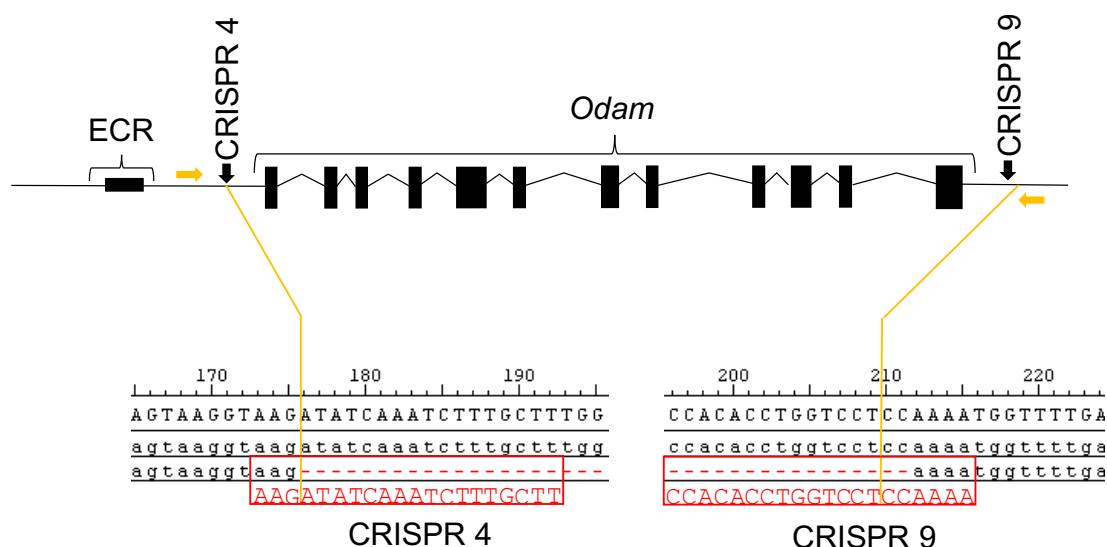


Figure 3. 11: Sequencing of DNA isolated from *Odam* targeted MEF cells: A Sanger sequencing reaction showing *Odam* deletion in MEF cells using CRISPR combination 3 (guides 4 + 9). The sequencing results are aligned with guide sequences showing the NHEJ point and deletion site with underlined PAM sequence for both CRISPRs. The PCR primer 3.9 is shown as blue arrows flanking the targeted *Odam* gene in mouse casein locus.

The deletion of *Odam* in MEF cells using CRISPRs combinations 4, 5, 6, 7 and 8 were similarly analysed by Sanger sequencing of PCR amplicons.

Deletion of ECR and *Odam* in mouse embryonic fibroblast cells

ECR and *Odam* were collectively targeted in MEF cells using CRISPR sets 1 and 3. A total of nine combinations were made using CRISPRs targeting ECR upstream (numbers 1, 2 and 3) and *Odam* downstream (numbers 7, 8 and 9). The plasmid co-transfection in MEF cells was carried out by transfecting 1 µg of each CRISPRs using Neon electroporation system (Invitrogen) and pre-optimised conditions.

CRISPRs were co-transfected into MEF cells, and ECR-*Odam* deletion was analysed using PCR at 20s extension time for 35 cycles using the standard conditions described in chapter 2 (Section 2.2.1) and electrophoresed on a 1.2% agarose gel with 1kb DNA ladder (Invitrogen G5711). PCR was carried out using a primer pair (3.10, Appendix. 7.1) flanking mouse ECR and *Odam*; as with the previous experiments to delete *Odam*, a PCR fragment was expected only when ECR-*Odam* deletion has occurred. No wild-type fragment was expected (Figure 3.12).

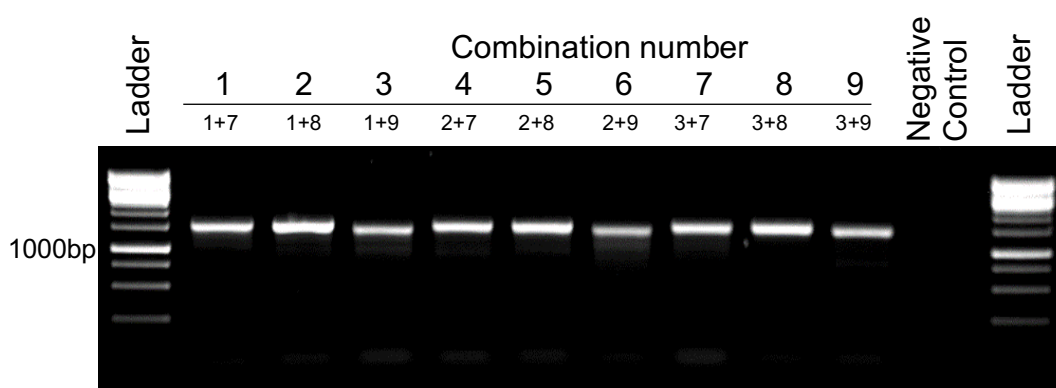


Figure 3. 12: Deletion of ECR- *Odam* in MEF cells: PCR of DNA from ECR-*Odam* targeted MEF cells showing a fragment in combination number 1 (1543 bp), 2 (1543 bp), 3 (1452 bp), 4 (1570 bp), 5 (1570 bp), 6 (1479 bp), 7 (1581 bp), 8 (1581 bp), 9 (1490 bp); no wild type and negative control fragment was expected.

PCR of DNA from ECR-*Odam* targeted MEF cells showed fragments of expected sizes. All PCR fragments were further analysed using Sanger sequencing (Edinburgh genomics).

Verifying ECR-*Od*am deletion with sequencing

DNA was isolated from ECR-*Od*am PCR fragments of all 9 combinations from agarose gel using the QIAquick DNA extraction kit (Qiagen). Sanger sequencing (Edinburgh genomics) was carried out for all nine samples. A representation of a sequencing result for combination number 6 is shown in figure 3.13.

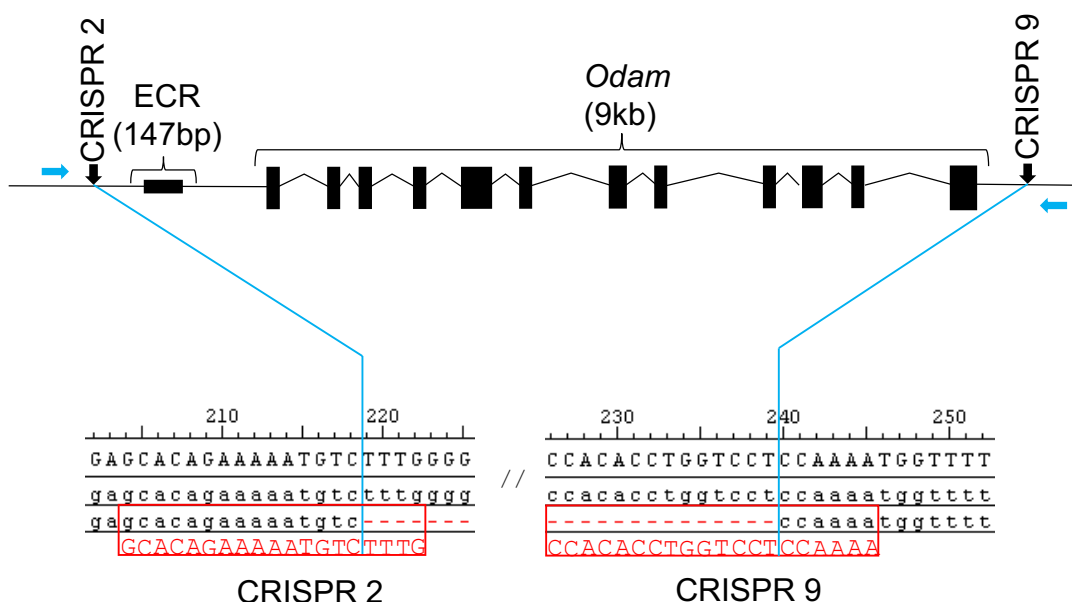


Figure 3. 13: Sequencing of DNA isolated from ECR and *Od*am targeted MEF cells: A Sanger sequencing showing ECR-*Od*am deletion in MEF cells using CRISPRs combination number 6 (number 2+ 9). The sequencing results are compared with guide sequences showing the NHEJ point and deletion site with underlined PAM sequence for both CRISPRs. The orange arrows are PCR primer 3.10 flanking the targeted ECR and *Od*am.

The deletion of ECR and *Od*am in MEF cells for all nine CRISPRs combinations was similarly analysed by Sanger sequencing.

3.3.3 Deleting ECR, *Odam* and ECR-*Odam* in mouse embryos

Deleting ECR and *Odam* (separately and collectively) in the mouse by pronuclear microinjections

The CRISPRs targeting mouse ECR and *Odam* were initially tested in MEF cells, and all the deletions were analysed by PCR and Sanger sequencing. Based on these results, three CRISPR combinations, number 2, 4 and 9, were planned to be simultaneously injected in mouse zygotes by pronuclear microinjections to generate an assortment of ECR, *Odam* and ECR-*Odam* KO mice.

The ECR in mouse is predicted as a controller of the *Odam* gene which is expressed in tooth enamel, although a previous report on *Odam* KO mice showed no adverse effects on health and teeth development of pups, however the mouse ECR has never been deleted so the phenotyping results were unknown. The CRISPR injected zygotes were planned to be checked at the embryonic day (E) 18.5 to ensure that the ECR or ECR-*Odam* targeting was safe to the mouse embryos and the teeth development was regular. All pronuclear microinjections were carried out by the biological research facility (BRF) at the Roslin Institute, The University of Edinburgh.

In an attempt to generate ECR, *Odam* and ECR-*Odam* knockout mice, all three selected CRISPRs, each targeting ECR upstream and downstream and *Odam* downstream (CRISPR plasmid number 2, 4 and 9) were co-injected in mouse zygotes. The microinjection buffer, named as EO: 1, was prepared by diluting CRISPR plasmids number 2, 4 and 9 in low EDTA Tris buffer at a concentration of 5 ng/μl each. The buffer was filtered through a 0.2μm syringe filter and stored at -20°C.

For pronuclear microinjections, the mouse zygotes were collected from donor female B6CBA/F1 mice by superovulation followed by mating. A successful mating was confirmed by the presence of a vaginal plug on the next day and zygotes were collected from the oviduct of donor females. The zygotes

were then placed in M2+ media (Sigma Aldrich) and were injected with 2-5 picolitre of buffer EO: 1 each via injection capillary using a holding capillary for grasp. The recipient mice were prepared by mating with a vasectomised male one day before the pronuclear microinjections. The re-implantation of the injected zygotes was carried out by transferring 20 injected zygotes each into the oviduct of each of 5 pseudo-pregnant recipient mice (Figure 3.14). All pronuclear microinjections were carried out using the same method.

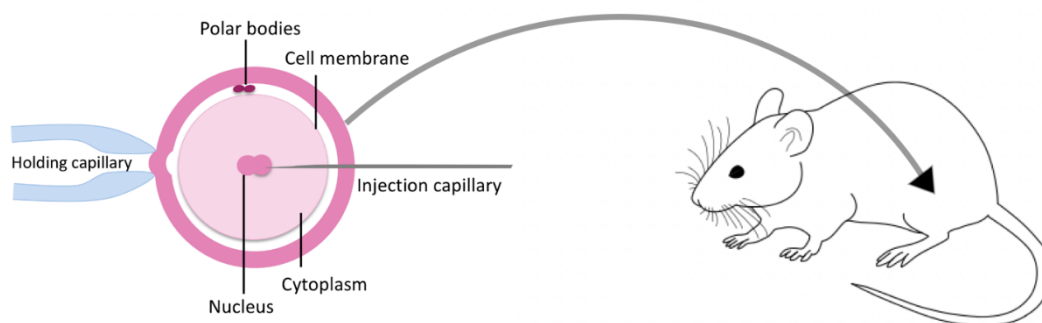


Figure 3. 14: Pronuclear microinjections and transfer of injected zygotes into a recipient mouse.

***In vitro* analysis of EO:1 injected zygotes**

The viability of EO:1 injected zygotes were tested by culturing 10 injected embryos in EmbryoMax KSOM Embryo Culture media (Sigma Aldrich) overnight at 37°C, a two-cell stage was observed in all zygotes on the following day.

Isolation of E18.5 mouse embryos for tooth development analysis

Embryos were isolated at the E18.5 stage by euthanising recipient females by cervical dislocation, no physical difference in all collected embryos was visually observed. The tail tips of embryos were collected for DNA isolation using the QuickExtract DNA extraction solution (Lucigen) as described in chapter 2 (Section 2.3.9) for genotyping. The embryo heads were collected by decapitation and stored in neutral buffered formalin 10% (Sigma Aldrich) at room temperature for tooth analysis after genotyping. The litter size of all recipient mice is shown in table 3.3.

Recipient mouse	Litter size
1	4
2	4
3	5
4	1
5	2
	16

Table 3. 3: Litter size of five recipient females transferred with EO: 1 buffer injected zygotes: A total of 5 recipient mice generated 16 embryos.

The genotyping of each embryo was carried out using 3 PCR reactions with different primer pairs, previously used to check the deletion of ECR, *Odam* or ECR-*Odam* in MEF cells (Figure 3.15). The primer pair 3.8 was flanked mouse ECR and expected to show 649 bp WT and another 384 bp fragment in ECR KO embryos targeted by CRISPR number 2 and 4. The primer pair 3.9 and 3.10 flanked *Odam* and ECR-*Odam* respectively; these primers were expected to show fragments only in *Odam* or ECR-*Odam* KO embryos.

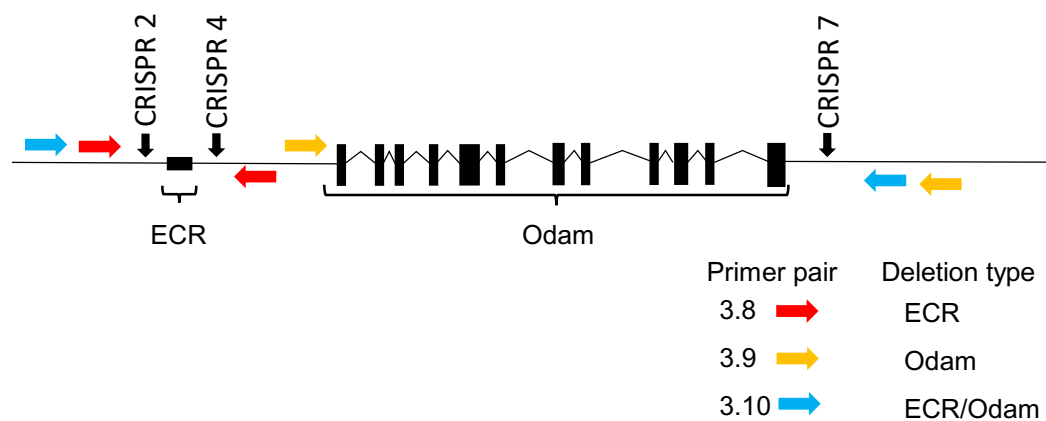


Figure 3. 15: Schematic of primer pairs to genotyping mouse embryos: The three primer pairs were used to analyse ECR, *Odam* and ECR-*Odam* deletion. Primer pair 3.8, flanking the mouse ECR, produces a 649 bp WT fragment or a 384 bp ECR deletion fragment. The primer pairs 3.9 and 3.10 targets *Odam* and ECR-*Odam* respectively and were expected to produce a PCR fragment only in the event of their respective deletions.

The positive control PCR was carried out using primer pairs shown in figure 3.15, along with DNA isolated from ECR, *Odami* and ECR-*Odami* targeted MEF cells with primer pair number 3.8, 3.9 and 3.10 respectively at 20s extension time for 35 cycles using the standard conditions described in chapter 2 (Section 2.2.1). The PCR was electrophoresed on a 1.2% agarose gel (Figure 3.16). All PCR using these three primer pairs were carried out using the same standard conditions.

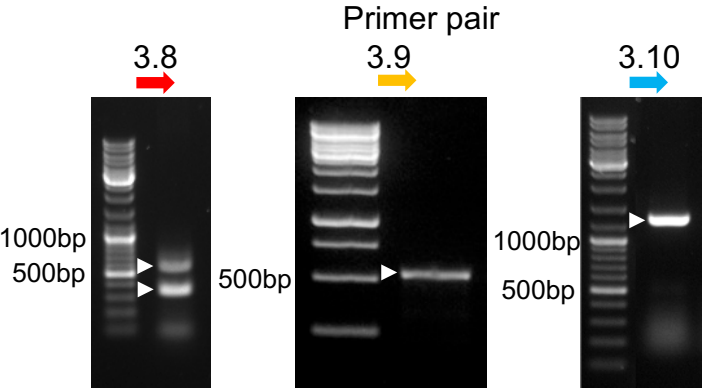


Figure 3. 16: Checking primer pairs for genotyping of ECR and *Odami* targeted mouse pups: The representative PCR carried out with DNA isolated from ECR, *Odami* and ECR-*Odami* targeted MEF cells and along with primer pair 3.8, 3.9 and 3.10 showing fragments of 649:384, 600 and 1479 bp respectively.

Primer pair	Expected band size	Genotype
3.8a 3.8 3.8b	649bp	→ WT
	384bp	→ ECR KO- heterozygous
		→ ECR KO- homozygous
3.9	600bp	<i>Odami</i> KO
3.10	1479bp	ECR- <i>Odami</i> KO

Table 3. 4: Concluding genotype of ECR and *Odami* targeted mouse embryos: The primer pair 3.8 was expected to show a WT fragment of 649 bp (3.8a) and ECR deletion fragment of 384 bp (3.8b). The presence of the only 3.8a was concluded as WT and presence of 3.8b was to be considered as ECR KO homozygous whereas the presence of both, 3.8a and 3.8b, was considered ECR KO heterozygous. The presence of a 600 bp with primer pair 3.9 would be considered as *Odami* KO and presence of a 1479 bp fragment with primer pair 3.10 was considered ECR-*Odami* KO.

The mock PCR shown in Figure 3.16 were always carried out as a positive control for all genotyping PCR.

Genotyping mouse embryos

The genotyping of each mouse embryo was carried out using primer pairs 3.8, 3.9, and 3.10. A representation of PCR of one of the embryo using three primer pairs is shown in figure 3.17. The primer pair 3.8 produced a single 384 bp fragment indicating ECR deletion, whereas primer pair 3.9 and 3.10 showed no fragments indicating no *Odham* or ECR-*Odham* deletion. All genotyping PCR were carried out with positive controls using DNA from ECR, *Odham* and ECR-*Odham* targeted MEF cells, showing fragments of expected sizes.

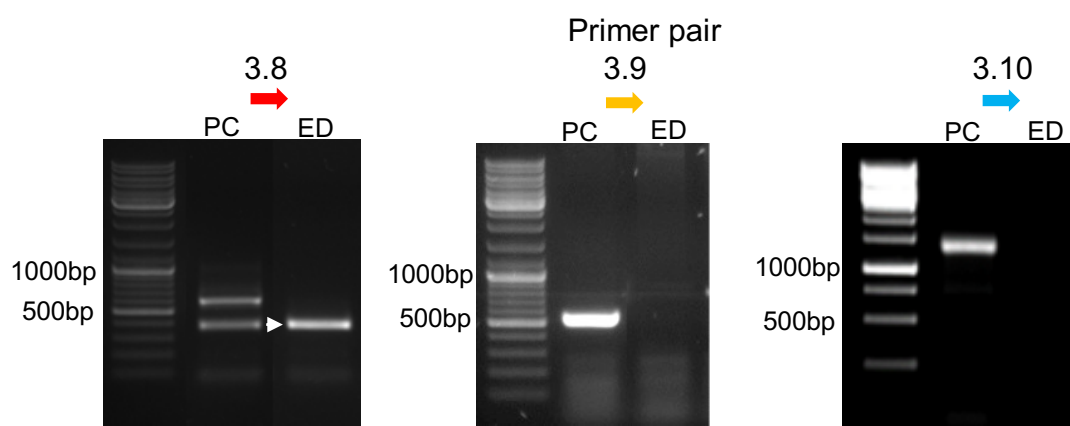


Figure 3. 17: Genotyping an ECR and *Odham* targeted mouse embryo: A total of three PCR were carried out using the three primer pairs to genotype one mouse embryo. The primer pair 3.8 showing the positive control (PC) fragments at 649 and 384 bp whereas DNA isolated from mouse embryo (ED) showing one fragment of 384 bp only (marked) In PCR with primer pair 3.9 and 3.10 the positive control showing fragments of expected sizes (600 and 1479 bp) whereas DNA from targeted mouse embryo showing no fragment.

The PCR in figure 3.17 showed only one 384 bp fragment with primer pair 3.8 hence it was considered as ECR $-/-$. The PCR fragments, wherever observed, were analysed using Sanger sequencing (Edinburgh genomics).

Verifying ECR deletion with sequencing

PCR fragments were purified from the gel using QIAquick DNA extraction kit (Qiagen). A representation of Sanger sequencing for the marked fragment in figure 3.17 is shown in figure 3.18.

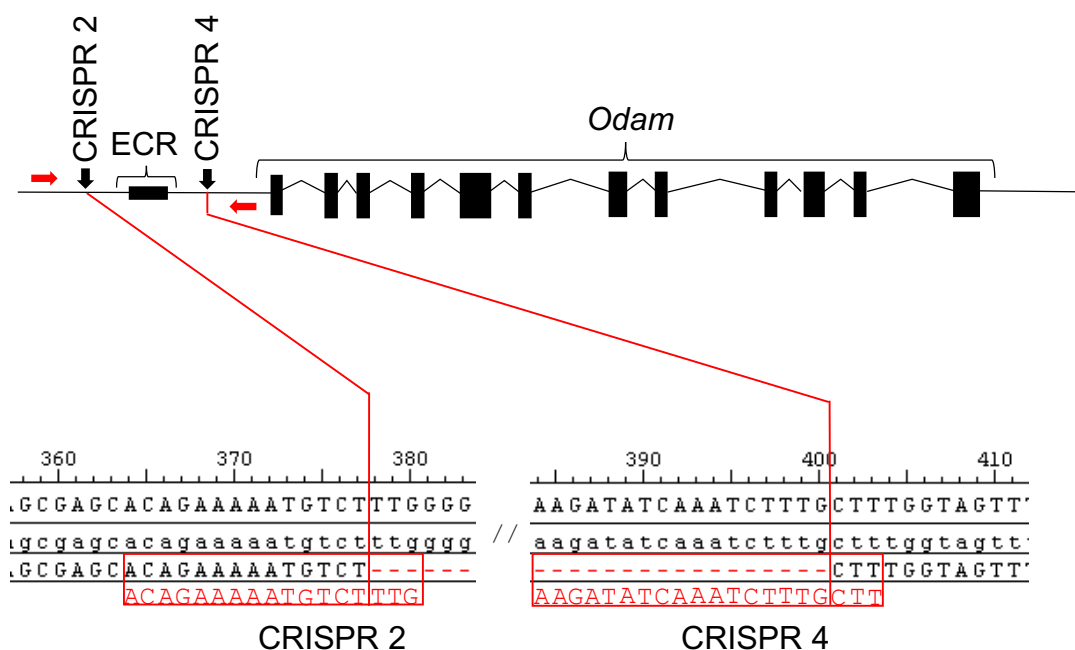


Figure 3. 18: Sequencing of an ECR targeted mouse embryo: A Sanger sequencing reaction showing ECR deletion the embryo using CRISPRs number 2+ 4. The sequencing results are compared with guide sequences containing underlined PAM sequence showing the NHEJ point and deletion site for both CRISPRs. The primer pair 3.8 from is indicated by red arrows.

Genotyping of all 16 embryos was carried out using PCR and Sanger sequencing, showing a total of 5 ECR KO embryos (31.25% KO efficiency). Out of 5 ECR knockouts, two were ECR $-/-$ whereas three were ECR $+/-$, (Table 3.5) no *Odham* knock or ECR-*Odham* knockout was found.

Recipient mouse	Litter size	Number of edited embryos	Genotype
1	4	1 (ECR -/-)	Hom
2	4	1 (ECR +/-)	Het
3	5	2 (ECR -/-, ECR +/-)	Hom & het
4	1	0	-
5	2	1 (ECR -/-)	Hom
Total-	16	5	

Table 3. 5: Genotyping of microinjection buffer EO: 1 injected zygotes:

Genotyping of 16 mouse embryos produced by 5 recipient females resulted in a total of 5 ECR KO embryos. Out of 5 embryos, two were ECR KO heterozygous, and three were ECR KO homozygous.

Analysis of mouse tooth development in ECR edited and WT mouse embryos

Mice embryo heads (E:18) from all ECR edited pups (two ECR -/- and three ECR +/-) were used for tooth development analysis along with three ECR+/+ as controls. All embryo heads were processed in a tissue processor (Leica APS300s) and embedded in molten wax (Leica Arcadia). The embedded heads were sectioned using a microtome (Leica) and dried at room temperature overnight as described in chapter 2 (Section 2.6)

The slides were stained using Haematoxylin and Eosin (H&E) in a Leica autostainer XL slide strainer followed by imaging using a bright field microscope (Figure 3.19).

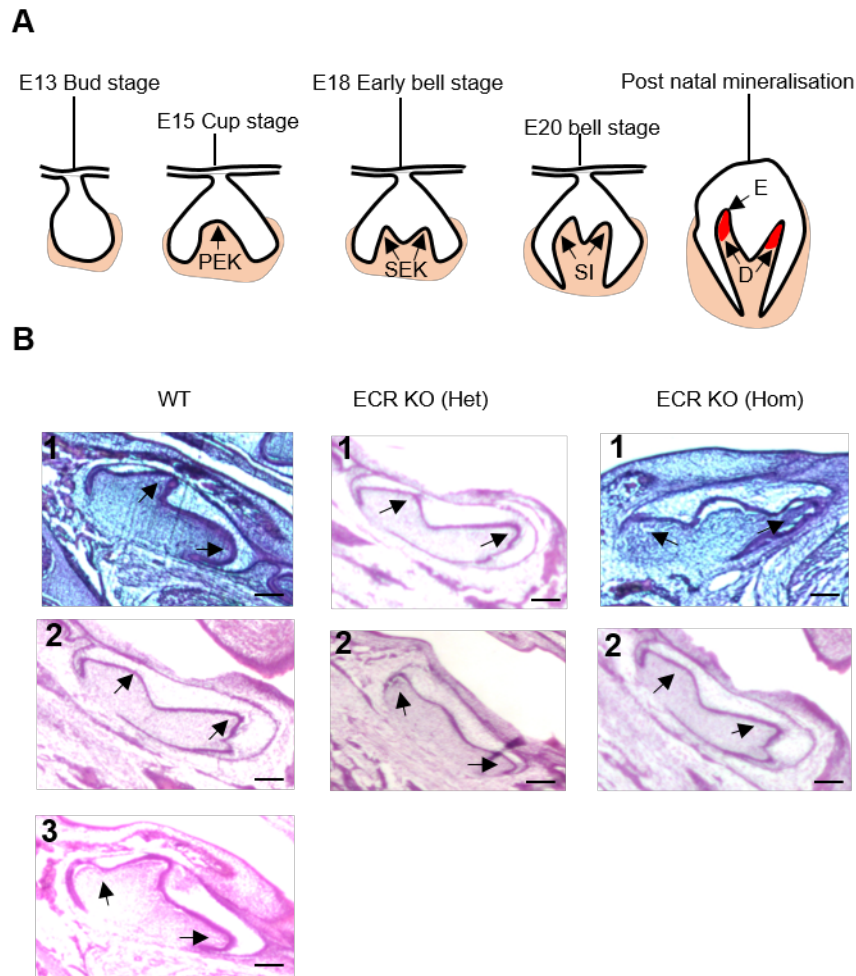


Figure 3. 19: Mouse tooth development analysis: A. The development of mouse tooth at E13 with a bud formation leading to a cup stage at E15 with formation of a primary enamel knot (PEK). The early bell stage appears at E18 with formation of a secondary enamel knot (SEK) which develops into a bell stage with appearance of stratum intermedium (SI) followed by post-natal mineralisation of teeth with developed enamel (E) and dentine (D) formation. B. The histological analysis of wild type, ECR knockout heterozygous and ECR knockout homozygous mice at E18.5 stage. The lower jaw sections of all mice shows a developed secondary enamel knot (marked by arrow). (Scale bar 100 μ m)

The lower jaw of all mice, three WT, two ECO KO het and three ECR KO hom, were analysed tooth development in an early bell stage. The overall

shape of tooth bell structure in all samples appeared same with development of secondary enamel know in all mice.

Since these results showed no apparent effect of ECR deletion on mice embryo growth and the tooth development at the E18.5 stage, the further microinjections were carried out to generate the ECR, *Odam* or ECR-*Odam* KO mice using buffer EO:1 to analyse the effect of their absence on milk proteins regulation and mammary gland.

Microinjections of EO:1 buffer to generate ECR, *Odam* or ECR-*Odam* knockout mice for phenotyping

Pronuclear microinjection of buffer EO:1 (CRISPR number 2+ 4+ 9) to generate ECR, *Odam* or ECR-*Odam* KO mice was carried out as previously. The re-implantation of zygotes was carried out by transferring 20 injected zygotes into the oviduct of each of 6 recipient females. Mean litter size was 2.5, with a total of 15 pups (Table 3.6). These pups were weaned at 22nd day postnatal and were fed with mashed food as a consideration of any potential effects of the desired deletions on teeth development in later developmental stages. Ear notch samples were collected from each pup and DNA was isolated using the QuickExtract DNA extraction solution (Lucigen) as described in chapter 2 (Section 2.3.9) to analyse their genotype.

Recipient mouse	Litter size
1	5
2	3
3	2
4	0
5	3
6	2
Total-	15

Table 3. 6: Litter size for buffer EO:1 injected zygotes: A total of 15 pups were born from 6 recipient females.

Genotyping

The genotyping of all 15 pups was carried out using the three primer pairs, 3.8, 3.9 and 3.10 as previously used. Figure 3.20 shows PCR of two mouse pups along with their positive controls. The positive control PCR was carried out with each primer pair using DNA from targeted MEF cells.

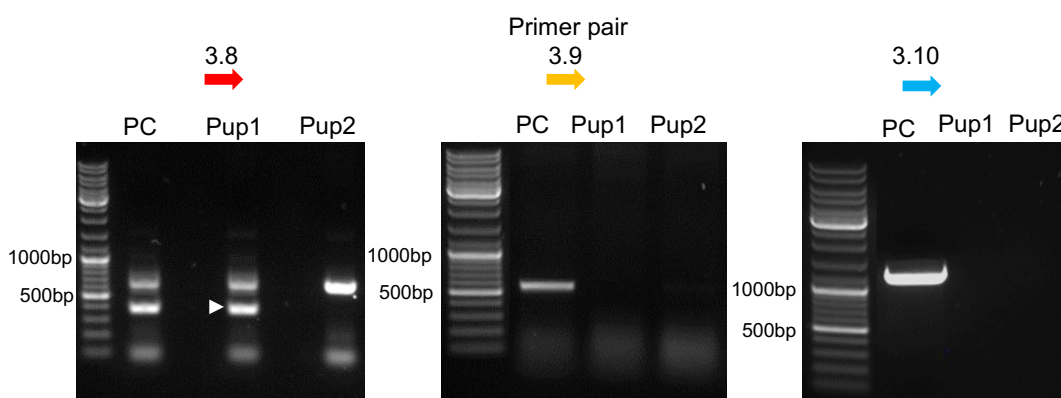


Figure 3. 20: Genotyping of ECR and *Odam* targeted mouse pups: PCR of 2 pups with primer pair 3.8, 3.9 and 3.10 along with a positive control for each primer pair. In PCR with primer pair 3.8, the positive control produced 649 bp WT fragment an additional 384 bp fragment and similar two fragments shown by pup 1 indicating the presence of a truncated allele with ECR removed in pup 1 (marked) whereas pup 2 produced WT fragment only. The PCR of primer 3.9 and 3.10 showed fragments of 600 bp and 1479 bp respectively whereas pup 1 and pup 2 produced no fragment.

The marked PCR fragment was excised from the gel and purified using QIAquick PCR purification kit (Qiagen) and Sanger sequencing (Edinburgh

genomics) was carried out. Table 3.7 shows the final genotyping results acquired through PCR and sequencing.

Recipient mouse	Litter size	Genotyping results	Genotype
1	5	1 (ECR -/-)	Hom
2	3	0	-
3	2	1 (ECR +/-)	Het
4	0	0	-
5	3	1 (ECR +/-)	Het
6	2	0	-
Total-	15	3	

Table 3. 7: Genotyping of microinjection buffer EO: 1 injected zygotes:

Genotyping of total 15 mouse pups produced by 6 recipient females resulted in a total of 3 ECR KO pups with KO efficiency of 20%.

The genotyping results of 15 mouse pups revealed 3 ECR KO pups, but no *Odam* or ECR-*Odam* KO mouse pup was observed.

Pronuclear microinjection carried out using buffer EO:1 containing CRISPR number 2, 4 and 9 resulted in the generation of a total of 31 pups from 11 recipient females. A total of 8 ECR KO pups were obtained, but no *Odam* or ECR-*Odam* pups were generated. In an attempt to generate *Odam* and ECR-*Odam* KO, further microinjections were carried out using another CRISPRs combination targeting ECR upstream and downstream and *Odam* downstream (2, 5 and 8). These CRISPRs were selected based on their targeting activity with each other in MEF cells accessed by PCR and Sanger sequencing.

Preparation of buffer EO:2 and pronuclear microinjections

Microinjection buffer EO:2 was prepared using CRISPRs targeting ECR upstream (number 2) and downstream (number 5) and *Odam* downstream (number 8). All CRISPRs were diluted in low EDTA tris buffer (Invitrogen) at 5 ng/μl each and filtered through a 0.2μm syringe filter.

The pronuclear injections were carried out with zygotes isolated from the oviduct of donor B6CBA/F1 donor mice. Re-implantation of injected zygotes was carried out by transferring a 20 zygotes into the oviduct of each of 5 recipient females (Figure 3.14), resulting in the generation of 13 pups with an average litter size of 2.6 (Table 3.8).

Recipient mouse	Litter size
1	4
2	2
3	2
4	2
5	3
Total-	13

Table 3. 8: Litter size for buffer EO: 2 injected zygotes: Generation of a total of 13 pups from 5 recipient females.

Genotyping

The genotyping of all 13 pups was carried out using the three primer pairs shown in Figure 3.15, with a positive control for each deletion fragment. All PCR fragments were analysed using Sanger sequencing (Edinburgh genomics). Table 3.9 exhibits the genotyping results of 13 mouse pups. PCR and sequencing revealed 4 ECR KO pups; no *Odam* or ECR-*Odam* mouse pups were achieved.

Recipient mouse	Litter size	ECR Edited pups	Genotype
1	4	2 (ECR +/-, ECR +/-)	Het, Het
2	2	1 (ECR -/-)	Hom
3	2	0	-
4	2	0	-
5	3	1 (ECR +/-)	Het
Total-	13	4	

Table 3. 9: Genotyping of microinjection buffer EO:2 injected zygotes: Genotyping of total 13 mouse pups produced by 5 recipient females resulted in a total of 5 ECR KO embryos (30.7%). No *Odam* or ECR-*Odam* pups were found.

A total of 11 recipients transferred with buffer EO:1 injected zygotes and 5 recipient females transferred with buffer EO:2 injected zygotes produced a

total of 44 pups. 12 ECR KO pups were produced with an overall 27% knocking out efficiency. However, no *Odam* or ECR-*Odam* KO pups were produced. Due to time restrictions, no further microinjections were carried out to generate the missing genotypes.

Analysing the NHEJ induction at *Odam* downstream target site in ECR-*Odam* targeted zygotes

The mice zygotes injected with buffer EO: 2 containing CRISPRs targeting ECR upstream, downstream and *Odam* downstream resulted in the generation of ECR KO mice only, suggesting that CRISPR targeting the *Odam* downstream might not have worked efficiently with any of other two CRISPRs.

To analyse the induction of NHEJ by CRISPR number 8, targeting the *Odam* downstream, PCR was carried out using DNA isolated from ECR KO and WT mice and primer pair 3.7 (Figure 3.21).

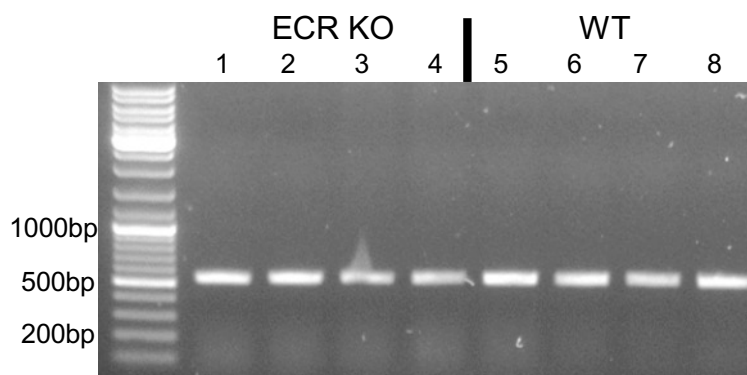


Figure 3. 21: Amplifying the mouse *Odam* downstream target site: PCR carried out with primer flanking the *Odam* downstream target site using DNA isolated from ECR KO and WT mice showing expected sized fragment of 537 bp.

All PCR fragments were excised from the gel and DNA was isolated using QIAquick gel extraction kit (Qiagen) and analysed by Sanger sequencing (Edinburgh genomics) (Figure 3.22).

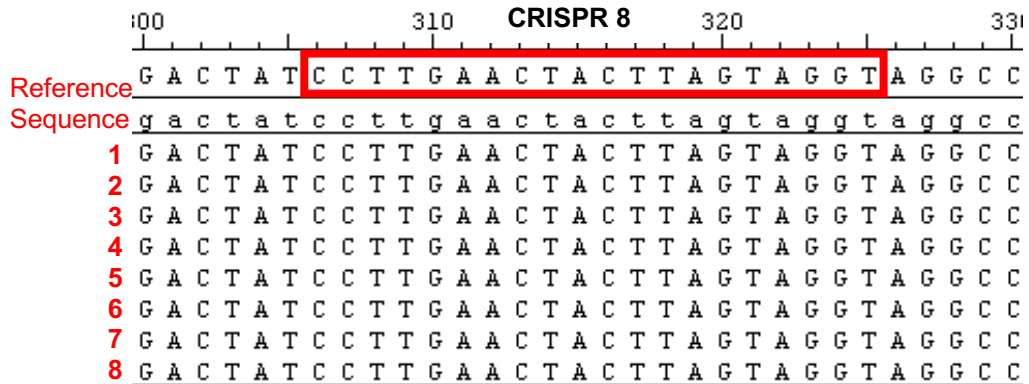


Figure 3. 22: Sequencing of *Odam* downstream target site: A Sanger sequencing showing generation of no NHEJ site on *Odam* downstream in DNA isolated from ECR KO mice and WT litter mates. Sequencing results are aligned with a reference sequence with marked CRISPR number 8.

The Sanger sequence analysis of *Odam* downstream target site in four ECR KO and four WT mouse mice showed no NHEJ induction by CRISPR number 8.

One inference from the overall results of the CRISPR mediated deletions of the ECR, the ODAM gene, and ECR + ODAM suggest that the effectiveness of the deletion process is dependent on the distance of the target sites for the guide RNAs. This could be investigated further through attempting to quantify by measuring the PCR products using quantitative PCR. I have not pursued this aspect further.

Breeding of the ECR KO mice

A colony of ECR knockout mice was established by breeding B6CBA/F2 ECR KO males +/- with C57BL/6J WT females. In order to generate ECR -/- mice, 4 ECR KO +/- males and 4 ECR KO +/- females were transferred to our collaborator's laboratory at Centre of Cellular and Molecular Biology, Hyderabad, India where phenotyping experiments are currently being performed.

3.4 Discussion

Conserved DNA sequences, identified via sequence alignment of multiple species, could potentially have a functional or regulatory role (Frazer et al., 2003). One such conserved sequence in the casein locus of mammals was identified upstream to the dentine gene *Odam* in cow, mouse, rat, human, dog and chimpanzee and referred to as an Evolutionary Conserved Region (ECR). The ECR in the casein locus of mouse was shown to act as a unidirectional repressor and controlling element of the *Odam* gene (Kaimala and Kumar, 2015).

The function of a gene or DNA element can be studied by deleting or inactivating it in a cell line or by generating a KO animal. The ECR is a non-coding element and is expected to suppress the *Odam* gene which is a predicted casein locus progenitor. The deletion of ECR and *Odam* is expected to affect casein locus regulation and lactation and hence generation of ECR, *Odam* and ECR-*Odam* KO mice was envisaged. CRISPRs has been proven highly capable in gene deletion from a few base pairs (Kim and Kim, 2017) up to 5mb (Kato et al., 2017) in the mouse. Multiple CRISPRs were designed to target the ECR upstream and downstream, and the *Odam* downstream and their individual cutting activity was tested in MEF cells. A single gene editor, when cleaving the target site results in the generation of a mutation which is repaired by NHEJ. This mismatch in the targeted DNA can be assessed by an endonuclease enzyme such as T7 E1 and CEL 1 nuclease. In a study comparing the efficiency of both enzymes to detect a mismatch in frog embryos generated by zinc finger nucleases found the T7 E1 to be more sensitive as compared to the CEL1 nuclease (Vouillot et al., 2015). The cutting activity of individual CRISPRs targeting mouse ECR and *Odam* was analysed by T7 E1 assay showed a successful generation of NHEJ at the target site.

The deletions of ECR, *Odam* and ECR-*Odam* were carried out in MEF cells by co-transfecting these CRISPRs and the results were analysed by PCR and Sanger sequencing. To generate the KO mice, three CRISPRs each targeting the ECR upstream and downstream and the *Odam* downstream were

co-injected in mouse zygotes in an attempt to obtain all three types of KOs by pronuclear microinjections. These microinjections were part of a technical staff's training session.

The conserved elements

The most common way of identifying a conserved element is multiple sequence alignment across different species, however, there are several online tools offered by University of California, Santa Cruz (UCSC's <http://genome.ucsc.edu/>) and the National Centre for Biotechnological Information (NCBI's <https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) that readily identifies conserved domains in coding and non-coding nucleotide sequences.

The conserved elements become interesting when they are in proximity to a well-identified gene and contain essential transcription factor binding sites as this could mean their potential role in the regulation of adjacent genes. These elements can also help in understanding the evolution and common ancestry between animals, e.g. the encyclopaedia of DNA elements (Yue et al.) study of human and mouse genome involving their chromatin landscape, transcription factor binding sites and replication domains revealed diversity in regulatory functions and a high similarity in the regulatory elements between these lineages (Yue et al., 2014). In another study, a pairwise comparison of human genome against sea anemone showed a 796 bp long ultra-conserved element and other many short conserved elements (~50 bp) common in human, sponge and fruit fly (Yue et al., 2014). There are several examples establishing the importance of conserved elements and their role as an enhancer or repressor of a gene. Some examples include the deletion of a 500 bp highly conserved element from the Hmx1 locus in the mouse revealed it to be a regulatory element of Hmx1, whose dysregulation led to the congenital ear malformation in mouse (Quina et al., 2012).

Similarly, targeting a conserved element in the ADH gene in *Drosophila* resulted in improvement in the expression of this gene while delaying

development of the fly; hence it was identified as a gene repressor. However in some cases, deleting a conserved element does not show any effect in the regulation of adjacent genes or any changes in the phenotype of targeted animal, one such example is a 144 bp highly conserved sequence found in the DACH1 gene in a total of 11 vertebrates including mouse and human and was predicted as a regulatory element of this gene. The DACH1 gene is expressed in brain, the deletion of this element in the mouse showed no observable effect on brain development (Poulin et al., 2005).

The ECR was identified in mammals using the mouse casein locus sequence as a reference and its alignment against the sequence of human, rat, chimpanzee, cow and dog. A percentage identity plot was generated showing the sequence identity in casein locus of these animals. The ECR was found to be located on upstream of the *Odam* gene containing several transcription factor binding sites known to regulate the casein locus. The *in vitro* activity of ECR was analysed by reporter assay in three different human cell lines and thus predicted as a strong candidate with a regulatory function (Kaimala and Kumar, 2015).

Another evidence suggesting the possibility of ECR having a regulatory function is its location. It is proposed that the *Odam* gene was expressed in the early tetrapods, which gave rise to the amniotes where major shuffling in the *Odam* gene occurred. The amniotes were diverged into the synapsids (mammalian vertebrates) and the sauropsids (non-mammalian vertebrates) approximately 315 million years ago. The ECR in the casein locus has not been found in non-mammalian vertebrates such as frog, chicken and zebrafish indicate that ECR was absent in amniotes and might have appeared only in the synapsids when the *Odam* gene created milk genes.

CRISPR/Cas9 cleavage site: predicted vs. observed

Target specificity is an essential factor involved in the development of precise gene editors. The PAM sequence for Cas9 endonuclease derived from *S. pyogenes* is 5' NGG 3' (Jinek et al., 2012), which cleaves approximately 3-

4 bp upstream to the PAM site. This has been validated several times e.g. The CFTR gene was targeted by CRISPR/Cas9 in human trachea epithelial cells. The deep sequence analysis of NHEJ repair site targeted by a single CRISPR revealed that deletions were majorly centred at 3 bp upstream to the PAM (Hollywood et al., 2016).

In this chapter, the mouse ECR and *Odiam* were targeted using a total of 9 CRISPRs and their combined cutting activity was analysed by T7 E1 assay and Sanger sequencing. The cleavage sites generated by all CRISPRs were analysed to check their precision, e.g. taking 3 CRISPRs into account, each targeting ECR upstream (number 2), ECR downstream (number 4) and *Odiam* downstream (number 9) were injected with each other in MEF cells and mouse zygotes. The Sanger sequencing (data shown in this Chapter) revealed that the majority of deletions had occurred at 3-4 bp upstream to the PAM sequence. However, one exception, CRISPR number 9, produced the cleavage 6 bp upstream to the PAM every time when co-transfected with other CRISPRs. This suggests that gene targeting using the CRISPR/Cas9 is usually, however not always, accurate in the generation of a precise cleavage site. All CRISPRs used in this chapter were designed a few base pairs away from the target site thus a highly accurate targeting was not required.

There are several precise versions of gene editors continuously being developed e.g. SpCas9-HF1 with mutation in N497A, R661A, Q695A, and Q926A residues (Kleinstiver et al., 2016) and a version of Cas9 with an improved specificity, eSpCas9 (Slaymaker et al., 2016). The recently developed adenine base editors allow conversion of AT to GC with an extremely low rate of indels (0.1%) and off-target modifications (Gaudelli et al., 2017).

The generation of KO mice

Since ECR and ECR-*Odiam* are not previously targeted in mouse and their phenotype is unknown; initially, mouse embryos were collected at embryonic day 18.5 to analyse the potential effect of ECR and *Odiam*-ECR

deletions on mouse tooth development. The microinjections resulted in the generation of ECR KO mice only, the histological analysis of ECR KO mice tooth development showed no observable difference when compared to the tooth of WT littermates. The further pronuclear microinjections were carried out to generate KO mice for phenotyping. A total of 21 recipient females, each transferred with 20 injected zygotes, gave birth to a total of 44 pups. The low average litter size, i.e. 2 pups per recipient females was initiated assumed to be a result of the technical staff's micromanipulator handling and practice, however, similar low litter size has been observed in CRISPRs targeted KO mice (Hirose et al., 2017), (Zuo et al., 2017).

The genotyping of 44 pups showed a generation of 12 ECR KO mice with over 27% deletion efficiency whereas no *Odam* and ECR-*Odam* mice were generated. The ECR KO mice were bred and transported to the collaborator's laboratory at Centre of Cellular and Molecular Biology, Hyderabad, India where phenotyping is currently being carried out.

Mouse tooth development analysis at E18.5

The histological analysis of mice tooth at E18.5 revealed the presence of secondary enamel knot in ECR KO (hom and het) and their WT littermate mice. The ECR has been considered as a controlling element of the *Odam* gene, involved in tooth enamel mineralization which initiates mainly in the post-natal stages after disappearance of the secondary enamel knots. Hence, it is unlikely that the deletion of ECR would affect the tooth development at E18.5 in mice. It has been already demonstrated that the deletion of coding exon number 3, 4, 5 and 6 from the *Odam* gene of mice did not result in any adverse effect on tooth development (Wazen et al., 2015).

Simultaneous deletions using CRISPRs

The multiplexed gene targeting was introduced in 2013 by simultaneously targeting multiple loci in human 293FT cells using CRISPRs (Yan et al., 2016). This technique has been developing ever since, e.g. concurrent mutations in five different genes in mouse embryonic stem cells

(Wang et al., 2013) and simultaneous deletions of three genes in pigs to generate a human Parkinson's disease pig model (Wang et al., 2016). In a recent study, the gene targeting in GFP expressing mouse and monkey embryos using multiple CRISPRs was successfully achieved. A total of four CRISPRs, each with a targeting site not more than 200 bp away were co-injected in mouse and monkey embryos resulted in a complete knockout of the GFP gene (Zuo et al., 2017).

In this chapter, we set out to target a conserved DNA element (ECR) and its downstream gene (*Odam*) simultaneously in mice zygotes using CRISPRs. Three types of deletions, the ECR (147 bp), *Odam* (~9.0kb) and the ECR-*Odam* (~9.2kb) were expected using three CRISPRs each targeting the mouse ECR upstream and downstream and *Odam* downstream. The microinjections of these three CRISPRs in mouse zygotes resulted in the generation of ECR KO mice only. The CRISPRs targeting the *Odam* downstream were tested and verified in MEF cells when co-transfected with other CRISPRs, however when co-injected with CRISPRs targeting the ECR upstream and downstream in mouse zygotes, it did not cleave the target site and the Sanger sequencing of *Odam* downstream target site in DNA isolated from ECR KO mice and WT littermates showed no NHEJ induction.

We assume that ECR deletion was achieved due to prioritisation of these adjacent CRISPRs spaced at approximately 150 bp. Their individual cutting efficiency might also have played a role in successful cleavage of the ECR. Since the CRISPR targeting the *Odam* gene is ~9kb away from CRISPRs targeting the ECR upstream and the downstream, it might have not worked efficiently. There are several examples shows that CRISPRs mediated large deletion can result in a very low efficiencies (Birling et al., 2017). Moreover, in another research, simultaneously targeting juxtaposed sites in a loci in mouse zygotes resulted in unintended large and complex deletions (Shin et al., 2017).

By considering the observations acquired in this chapter and the literature available we suspect that in deletions using multiple CRISPRs, their

targeting location in relation to each other may play a crucial role in their efficiency and it could also lead to unintended genotypes. The use of multiple CRISPRs to target one gene appears to be safe whereas the use of multiple CRISPRs to target more than one gene to generate an assortment of genotypes in mouse zygotes should be perhaps avoided.

The utility of ECR KO mice

The ECR KO mice are expected to shed light on the regulation of the *Odam* gene and the casein locus and if the lactation is affected in these mice. The ECR is present in between the cluster of calcium-sensitive (α S1, α S2, β) and calcium-insensitive (k) caseins. The calcium-sensitive protein genes originated from an early milk gene *Csn1/2* are present to the upstream of the ECR whereas calcium-insensitive protein gene derived from follicular dendritic cell secreted protein (FDSCP) is on the downstream of ECR. It would be interesting to analyse the ECR's regulatory effect in the expression of these proteins in the casein locus. Since caseins make ~80% of total milk proteins, a better understanding of the regulation of the casein locus will help in its genetic modification which could be applied to enhance the protein content of milk whether to improve offspring's health, expression of proteins of interest in this locus or dairy production.

4 Chapter 4: Deleting the entire casein locus in mouse

4.1 Introduction

Caseins are ~80% of the total milk proteins, and a key component in the transfer of amino acids, calcium and phosphorus from mothers to suckling Infants. They are a group of acidic, proline-rich, conjugated proteins that contain a phosphate group, vital for the formation of a casein micelle structure (Rijnkels, 2002). These micelles are composed of calcium-sensitive proteins such as α casein (*CSN1S1*, *Csna*), β casein (*CSN2*, *Csnb*) and α s2 (*CSN2S2*, *Csng*, *Csnd*) and a calcium-insensitive K casein (*CSN3* or *Csnk*). All of the genes encoding these proteins are present in a single gene cluster called the casein locus.

Mapping of the casein locus from mammals, such as mouse, human and cattle, have unveiled similarities in the structural organisation and genetic conservation of this genomic region (Rijnkels et al., 1997). The length and the number of genes in the casein locus varies between species. Figure 4. 1 exemplifies the casein locus of cattle, goat, human, mouse and rat.

Along with milk protein genes, the casein locus contains several non-milk genes such as the histatin/statherin family, follicular dendritic cell secreted protein (*FDCSP*) and a tooth enamel protein expressing odontogenic ameloblast associated (*Odam*) gene (Rijnkels, 2002).

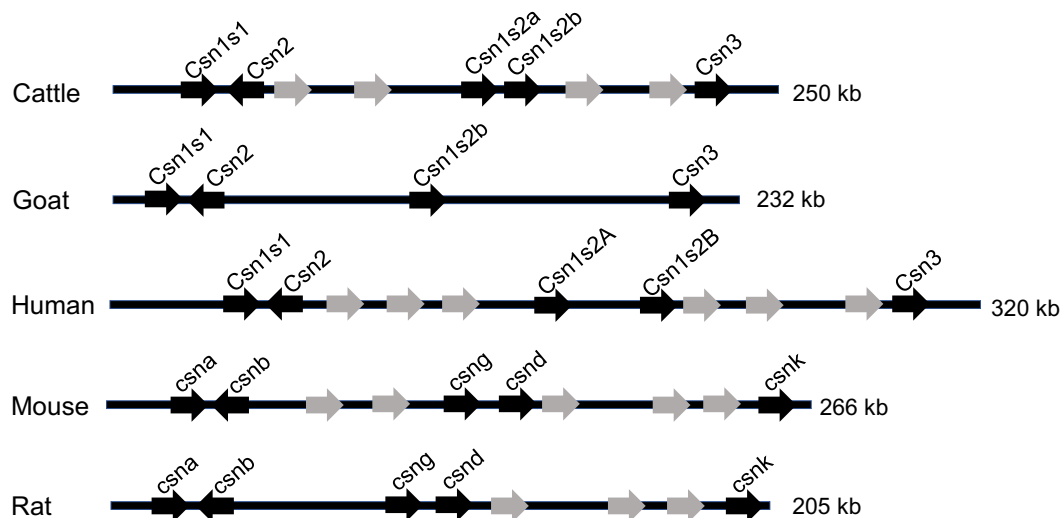


Figure 4. 1: Organisation of the casein locus in mammals: Schematic representation of the casein locus of cattle, goat, human, mouse and rat showing the relative size of the locus and the number and organisation of milk protein genes (not to scale, arrows does not convey the size of the genes). The Non-milk genes in this locus are shown in grey arrows and are not named. (Adapted from Rijnkels, 2002).

In milk, the caseins exist as submicron spherical aggregates called casein micelles (CMs). These are formed by the fusion of the calcium-sensitive (α and β) and calcium insensitive (κ) caseins held together by hydrophobic and electrostatic interactions (Waninge et al., 2005). The inner core of CMs is composed of calcium phosphate clusters and hydrophilic α and β caseins whereas the hydrophobic κ -casein forms an outer layer which provides structural stability to the CMs (Figure 4. 2) (Rijnkels, 2002).

The formation of CMs is a thermodynamically driven process of fusion between hydrophobic and hydrophilic molecules that are affected by pH, temperature and salt concentration. Each CM is composed of approximately $2 \times 10^4 - 1.5 \times 10^5$ independent casein molecules and possesses a polydisperse size property, containing casein molecules of different sizes, with an average diameter of $\sim 150\text{nm}$ and ranging from $40\text{-}300\text{nm}$ (Kruif, 1998).

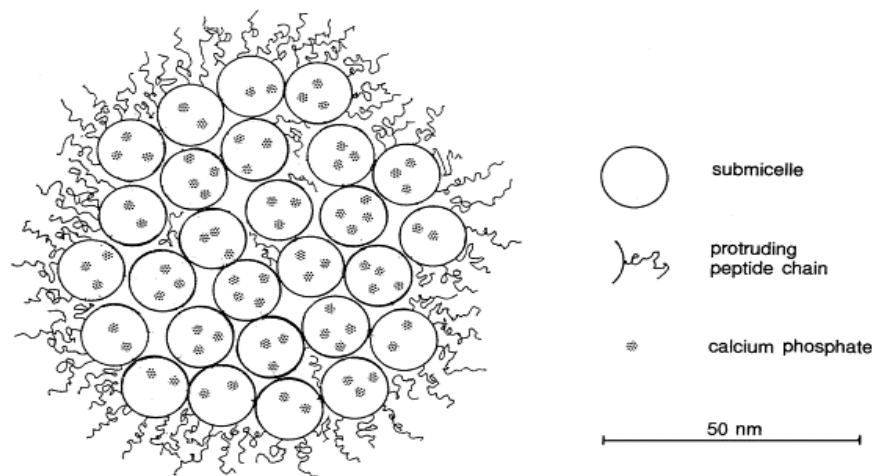


Figure 4. 2: Casein micelles structure: The structure of casein micelles consist of submicelles containing calcium phosphate and peptide chains (Walstra, 1990).

In the past two decades, the casein locus of the mouse has been evaluated by removing its significant components individually. Mice lacking either the α , β or κ -casein have been generated, and the resultant effects on milk composition and suckling infants have been analysed.

The α casein in the mouse was targeted by inactivating its second exon and transitional start codon in embryonic stem (ES) cells. Successfully modified ES cells were used to generate the α casein null (-/-) and heterozygous (+/-) KO mice. The α casein null mice showed a complete absence of α casein in the milk and significant reduction of β and γ (Csng) casein and a marginal reduction in κ casein.

The mRNA expression level of the β , γ and κ casein in the mammary gland of α casein null mice was found to be significantly reduced when compared to the wild type mice. This shows that the absence of α casein on both alleles affects the expression of other genes of this locus. However, no major morphological changes were observed by histological analysis in the lactating mammary gland of the α casein null and heterozygous (+/-) KO mice. The WT and α casein heterozygous (+/-) pups were cross fostered by α casein null (-/-) and WT females respectively. The growth of WT pups feeding on α casein null milk was severely reduced as compared to the α casein

heterozygous (+/-) and WT pups feeding on wild-type milk on days 11 (Figure 4. 3) and 21 of age (Kolb et al., 2011).

Similarly, in naturally occurring α_{S1} loss goat, the rate of secretion of other casein is reduced whereas no major effect on WAP secretion (Chanat et al., 1999) showing that absence of α effects the expression and secretion of other caseins expressed in the locus.



Figure 4. 3: Mice nursed by wild-type and α casein deficient dams: The mice nursed by the α casein deficient dam are smaller in size as compared to the mice nursed by a wild-type dams at 11 days of age (Kolb et al., 2011).

The β casein knock-out mice, generated by gene targeting in mouse embryonic stem cells, resulted in a reduction in the size of casein micelles and the total casein content in milk. The pups feeding on β casein knock-out mice were viable and fertile, and no effects on their growth and development were observed (Kumar et al., 1994).

The role of κ casein in milk production and casein micelle formation was assessed in κ casein deficient mice generated by targeting its first exon. Female mice lacking κ casein showed a blockage of the mammary gland lumina due to de-stabilisation of the casein micelle structure (Figure 4. 4). This resulted in a failure in the secretion of milk from the mammary gland hence

successful pups feeding was not achieved. The northern blot and mRNA expression analysis showed no effect in other caseins, whey acidic protein and alpha lactalbumin gene expression (Shekar et al., 2006).

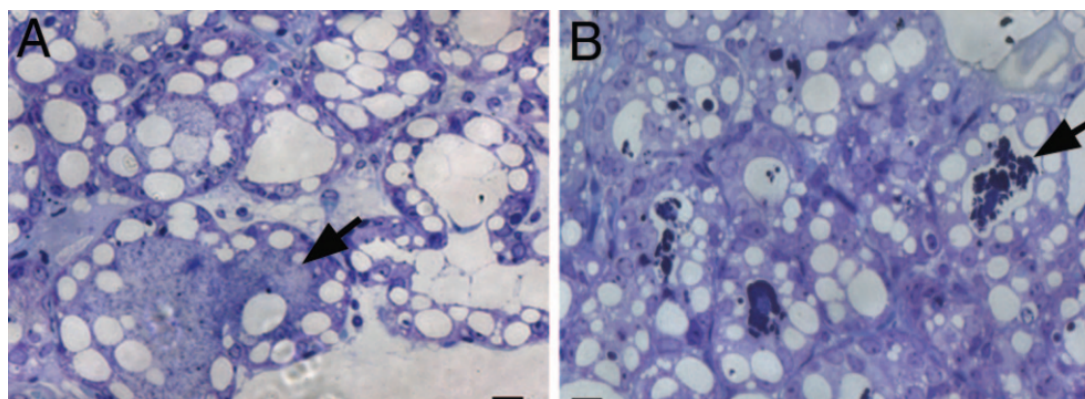


Figure 4. 4: Toluidine blue staining of mammary gland Sections of wild-type and Csnk/ mice at mid-pregnancy (16.5 days post coitum): A. Alveolar lumina (arrow) of the wild type mice without precipitates. B. Precipitated caseins in Csnk/ alveolar lumina. (Scale bars 5µm) (Shekar et al., 2006).

These findings give an insight into the importance of casein gene components in successful lactation and the impact of their absence. The α and κ casein genes have been considered crucial for successful lactation and nurturing of offspring (Kolb, 2011, Shekhar, 2006) whereas β casein does not appear to play an essential role in infant's health (Kumar et al., 1994). Since the collective effect of the entire casein locus deletion on milk production in mice has not been recorded, we set out to delete the whole casein locus in the mouse to assess its impact on lactation and mammary gland. In this chapter, the 266kb long mouse casein locus was targeted by designing CRISPRs on each side of this locus. The individual and collective cutting activity of CRISPRs was tested in mouse embryonic fibroblast cells by PCR and T7 endonuclease assay followed by attempts to target this locus in mouse zygotes.

4.2 Aims:

1. Design and testing CRISPRs flanking the casein locus in mouse.
2. Generate the casein KO mice by direct editing of murine zygotes.
3. Assess the impact of deletion of the casein locus on mammary gland structure and on lactation.

4.3 Results:

4.3.1 Design and validation of CRISPRs targeting the mouse casein locus

Designing and cloning CRISPRs to target the entire casein locus

The mouse casein locus sequence was retrieved using Ensemble genome browser from GRCm38.p6 database. Three CRISPR guides were designed on each side of the mouse casein locus using an online designing tool (<https://crispr.cos.uni-heidelberg.de>) as described in chapter 2 (Section 2.5.1). The first set of CRISPR guides (numbers 10, 11 and 12) was designed upstream to the mouse α -casein gene (*Csna*) and the second set (numbers 13, 14 and 15) was designed downstream to κ -casein gene (*Csnk*) (Figure 4.5A-C).

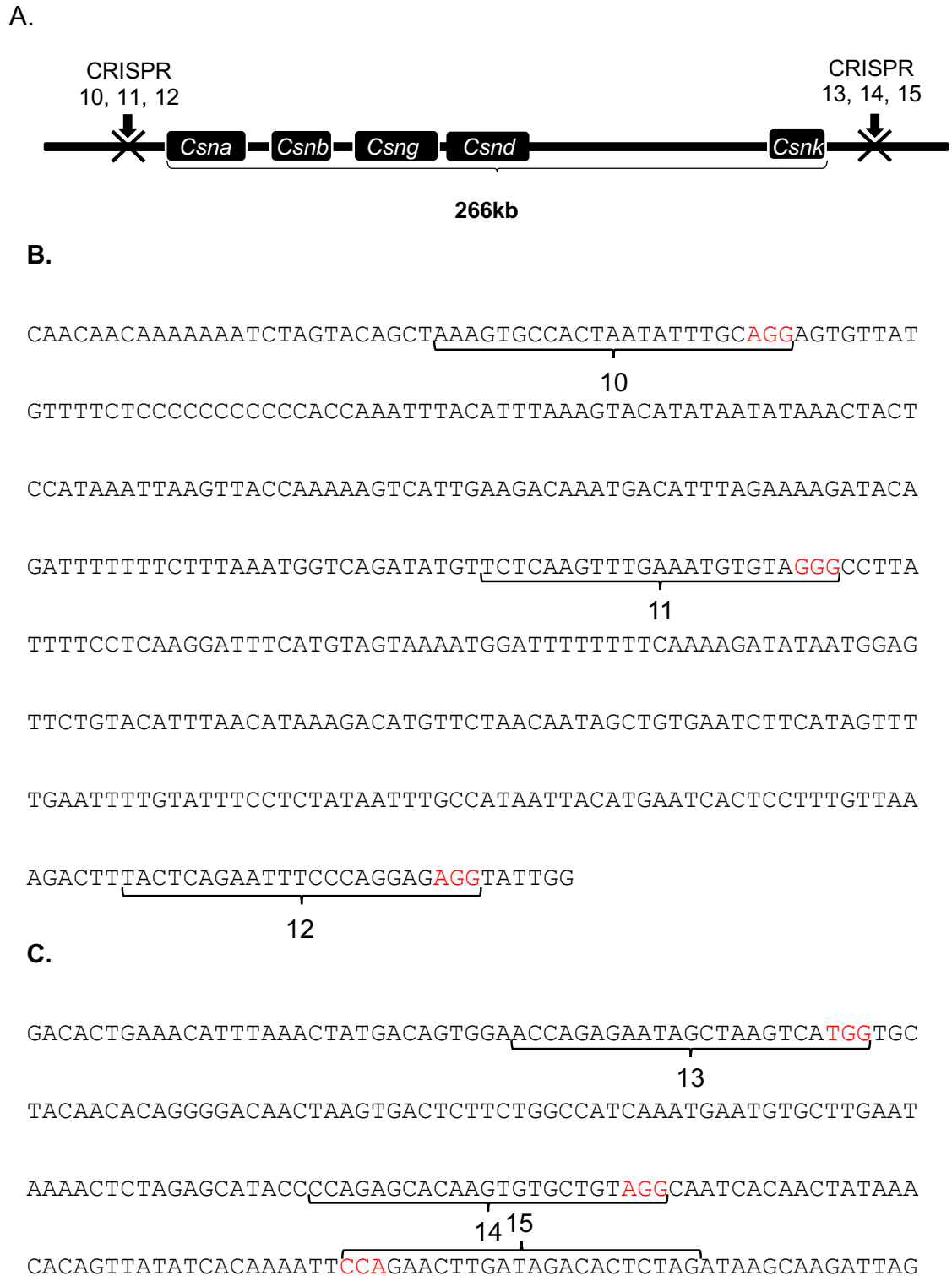


Figure 4. 5: CRISPRs targeting the mouse casein locus: A. sgRNA locations relative to the mouse casein locus. B. Individual sgRNA binding sites targeting

the mouse casein locus upstream (numbers 10, 11 and 12) and C. downstream (numbers 13, 14 and 15). The PAM sequences of all CRISPRs are in red font.

CRISPR plasmids were constructed by annealing sense and antisense oligos and cloning these into the *pSL66* plasmid followed by transformation into Top10 cells (Invitrogen) as described in chapter 2 (Section 2.5.4). A colony PCR was carried out using four transformed bacterial clones from each of the 6 CRISPRs as described in chapter 2 (Section 2.5.4). Figure 4.6 represents a colony PCR of four bacterial clones for CRISPR number 10.

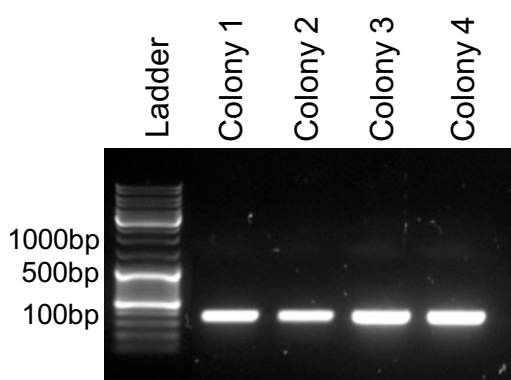


Figure 4. 6: Colony PCR of four bacterial clones of CRISPR number 10 targeting the mouse casein locus upstream: The PCR was carried out using CRISPR 10 sense oligo as the forward and a reserve primer *oSL35* which binds to the *pSL66* sequence downstream to the oligo insert. The presence of a 450 bp fragment confirms the successful cloning of the oligo duplex into the *pSL66* plasmid.

The cloning was confirmed for all 6 CRISPRs using the same approach. The presence of guides in the cloned vector was further confirmed by Sanger sequencing (Edinburgh Genomics) (Figure 4.7).

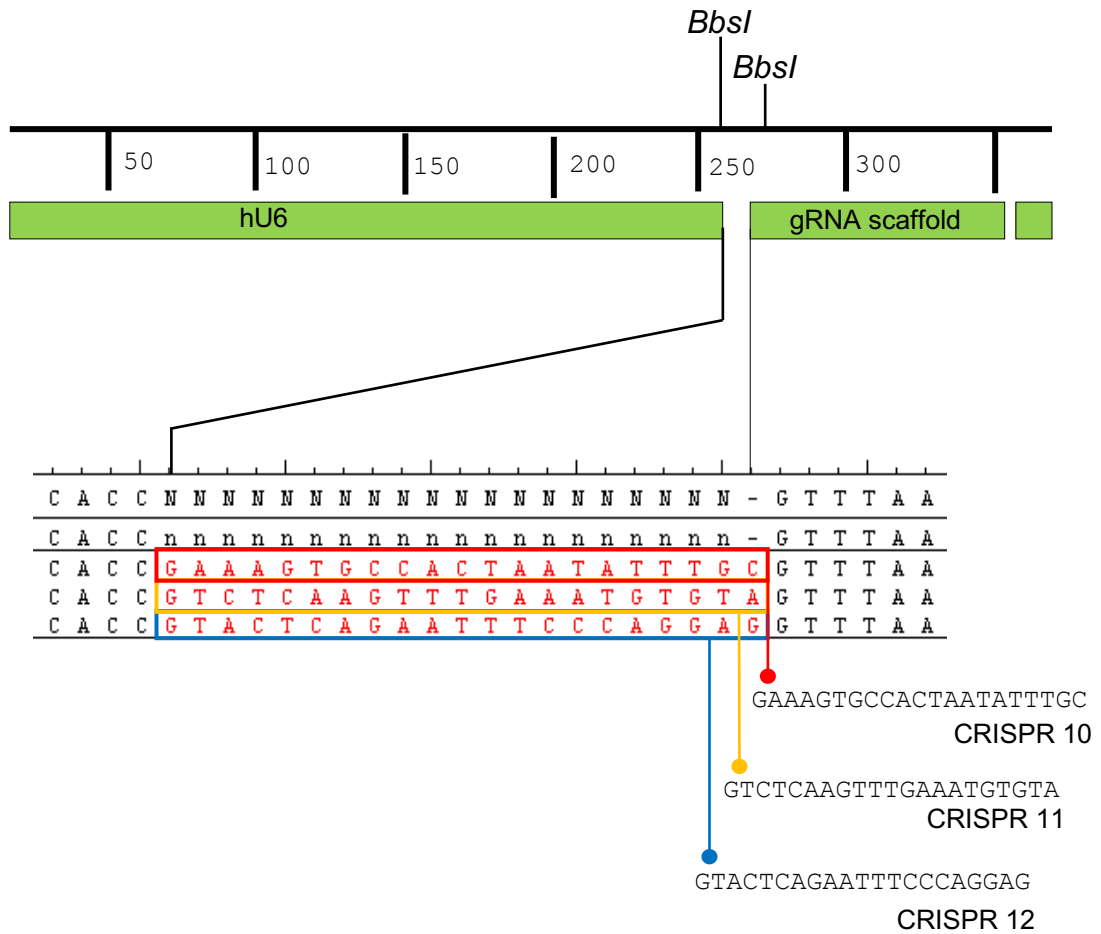


Figure 4. 7: Sequencing of CRISPRs targeting upstream of the mouse casein locus: A Sanger sequencing representing cloning of three CRISPRs number 10, 11 and 12 into the *pSL66* plasmid. The CRISPRs guide sequences are cloned in between the U6 promoter and gRNA scaffold region of *pSL66* plasmid at *BbsI* restriction site.

After confirmation with Sanger sequencing, all CRISPR plasmids were isolated using the endotoxin-free Maxiprep kit (Qiagen) and stored at -20°C.

4.3.2 Assessing CRISPR cutting activity in MEF cells

The cutting activity of CRISPRs targeting the mouse casein locus was analysed by individually transfecting them into MEF cells using pre-optimised parameters (Table 3.2). DNA was isolated from transfected cells using the DNA extraction solution (Lucigen) and PCR was carried out to amplify the targeted region using primer pair 4.1, 4.2 and 4.3 (appendix 7.1) for DNA isolated from MEF cells transfected with CRISPRs number 10, 11 and 12 respectively and primer pair 4.4 was used with DNA isolated from MEF cells individually transfected with CRISPR number 13, 14 and 15. PCR was carried out at 20s extension time for 35 cycles using the standard conditions described in chapter 2 (Section 2.2.1). CRISPR-induced cleavage at the target site is predicted to be resolved predominantly by NHEJ, resulting in INDEL formation which was quantified using T7 endonuclease assay as described in chapter 2 (Section 2.2.5) (Figure 4.8A-B).

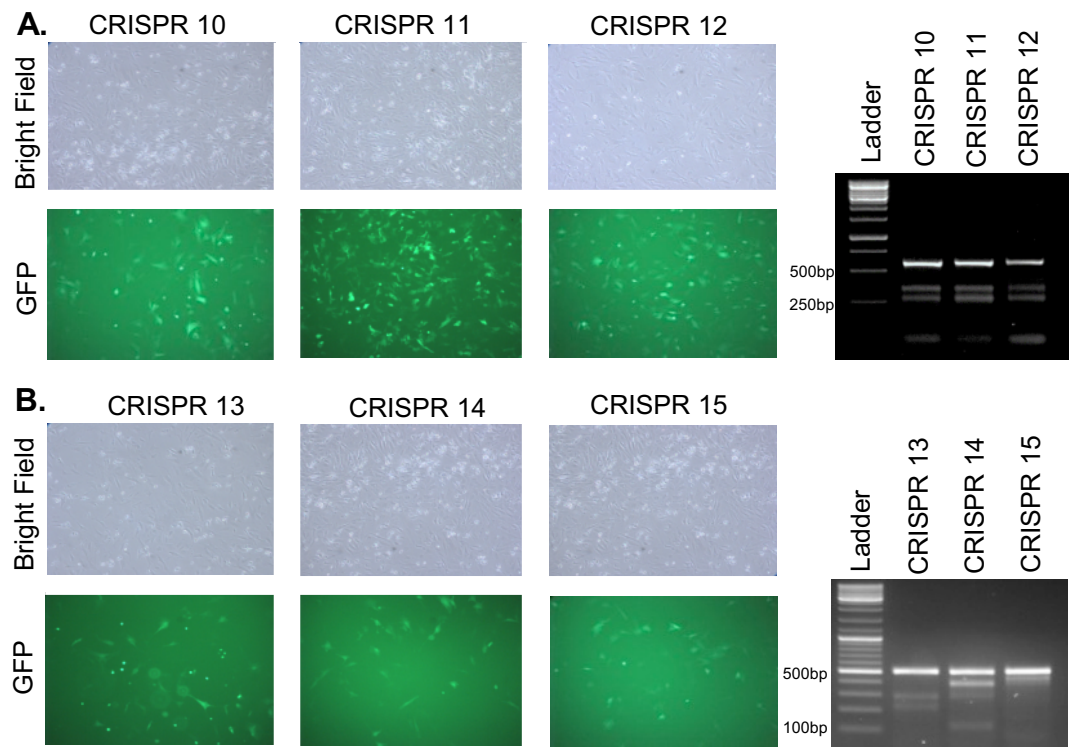


Figure 4. 8: Transfection of CRISPRs targeting the mouse casein locus and assessment of NHEJ induction at the target site by T7 endonuclease assay:

A. Transfection of CRISPRs targeting the casein locus upstream in MEF cells showing the GFP positive cells and the T7 assay showing a 672, 690 and 710 bp wild-type and digested fragments of targeted DNA sized 355: 305, 340: 295 and 375:320 bp for CRISPRs number 10, 11 and 12 respectively B. Transfection of CRISPRs targeting the casein locus downstream in MEF cells showing GFP positive cells and T7 assay showing a 508 bp wild-type and subsequent 270: 218, 371: 117 and 433: 55 bp fragments for CRISPRs number 13, 14 and 15 respectively. An extra unexpected fragment was observed in T7 assay of CRISPR plasmid number 14.

After confirming that each of the designed CRISPRs targeting the casein locus of the mouse was capable of inducing NHEJ formation at their target sites, the ability of these reagents to generate a deletion of the locus when combined in a pairwise manner was assessed. Plasmids encoding CRISPRs targeting each side of casein locus were co-transfected into MEF cells.

Deletion of the casein locus in MEF cells

In an attempt to delete the entire casein locus in MEF cells, a total of nine plasmid combinations were tested using CRISPRs targeting upstream (number 10, 11 and 12) and downstream (number 13, 14 and 15) of the casein locus. The CRISPR combinations were co-transfected in MEF cells in quantity of 1 µg of each plasmid using the Neon electroporation system (Invitrogen) using the pre-optimized conditions (Table 3.2) and as described in chapter 2 (Section 2.3.5).

The transfected cells were harvested at 48h post-transfection, and the DNA was isolated using the QuickExtract DNA extraction solution (Lucigen). A primer pair (primer pair 4.5, Appendix. 7.1) was designed out with the target sites flanking the casein locus (Figure 4.9). PCR conditions, 20s extension time for 35 cycles and standard conditions described in chapter 2 (Section 2.2.1) were designed such that a product was expected only in the event of casein locus deletion. The expected product size for each combination is listed in table 4.1, no product was expected from wild-type DNA.

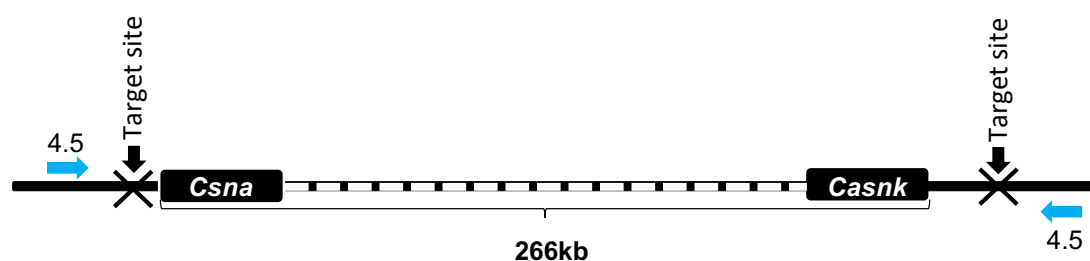


Figure 4. 9: Schematics of primer positions flanking the casein locus and cut sites: The CRISPR target sites are marked, and the blue arrows represent the primer pair 4.5 designed to PCR the casein locus, a PCR fragment is expected only in the absence of entire casein locus. The *Csna* and *Csnk* genes are present at extreme ends of this locus.

Combination number	Expected fragment size
1 (10+13)	296 bp
2 (10+ 14)	195 bp
3 (10+15)	133 bp
4 (11+13)	389 bp
5 (11+14)	288 bp
6 (11+15)	226 bp
7 (12+13)	594 bp
8 (12+14)	493 bp
9 (12+15)	431 bp

Table 4. 1: The combinations of CRISPR plasmids and their expected sized of PCR product after casein locus deletion in MEF cells.

The PCR was carried out using all 9 CRISPRs combinations with primer pair 4.5 (Figure 4. 9) and pre-described PCR conditions and run in a 1.2% agarose gel with DNA ladder mix (Invitrogen) generated fragments of expected sizes (Figure 4. 10).

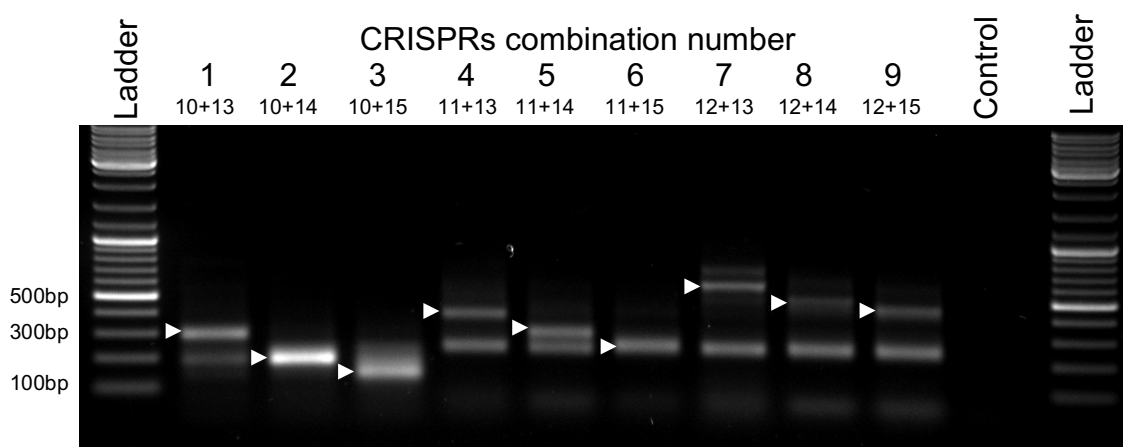


Figure 4. 10: Deletion of the casein locus in MEF cells: The PCR of DNA from casein locus targeted MEF cells, CRISPR combination numbers 2, 3 and 6 showed a single fragment of the expected size (195, 133 and 226 bp respectively) whereas combination numbers 1, 4, 5, 7, 8 and 9 produced spurious products in addition to the predicted fragments of 296, 389, 288, 594, 493 and 431 bp respectively. The control PCR was carried out using WT MEF DNA, no product was anticipated. The expected sized fragments are marked with a white arrow.

All CRISPR combinations targeting the entire casein locus in MEF cells produced PCR fragments of expected sizes (marked on Figure 4.10), which were excised from the gel. DNA was isolated using the QIAquick gel extraction solution (Qiagen). The spurious PCR amplicons were not further analysed.

Verifying casein locus deletion with Sanger sequencing

Sanger sequencing was carried out (Edinburgh genomics) on all purified DNA fragments. The sequencing result of all CRISPR combinations is shown in figure 4.11.

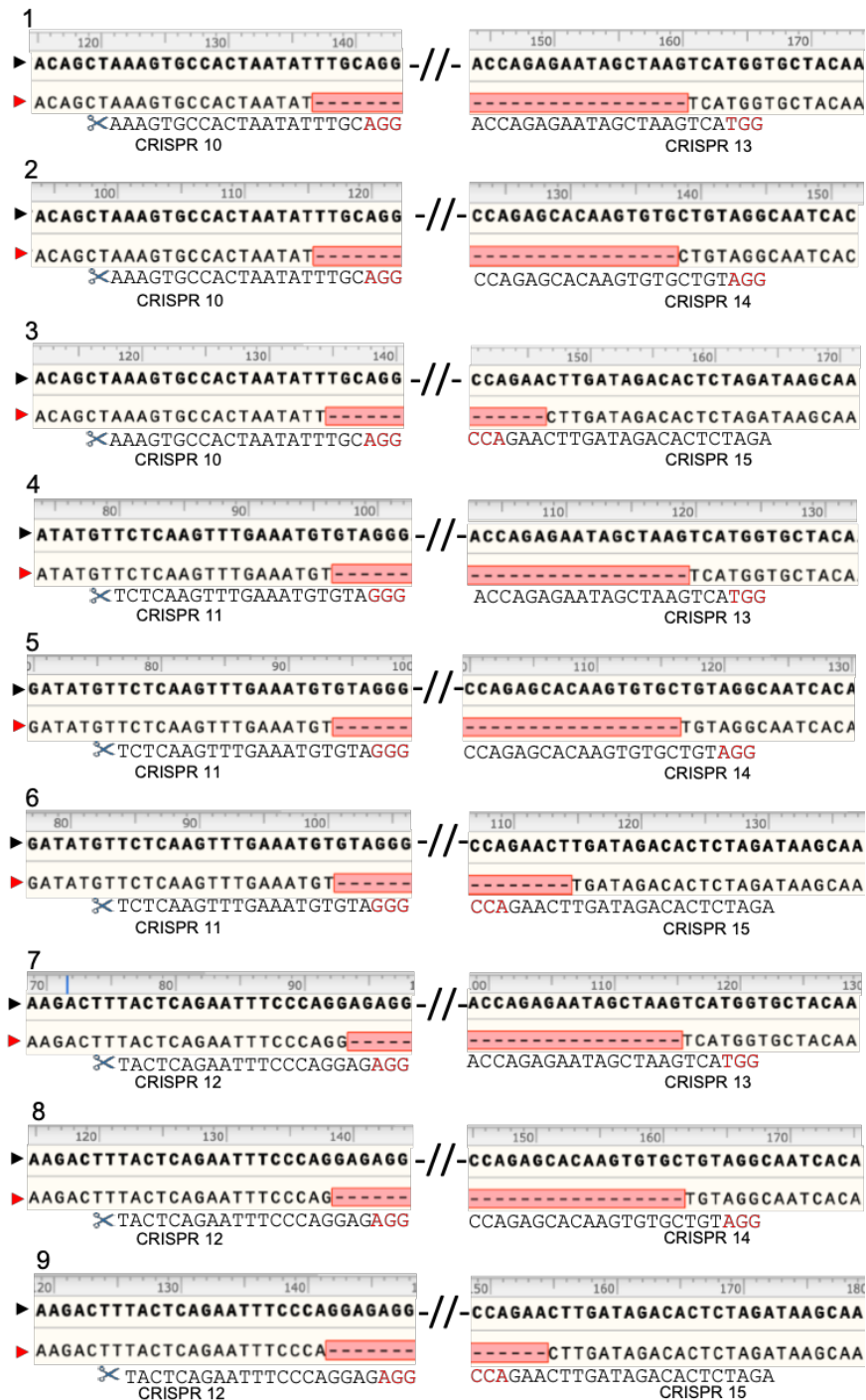


Figure 4. 11: Sequencing of DNA isolated from casein locus targeted MEF cells: Sanger sequencing reactions showing deletion of 266kb mouse casein locus in MEF cells using a total of nine CRISPR combinations. The reference sequences are marked with black arrows and aligned with red arrow marked targeted sequences. The CRISPR guides are shown underneath the targeted sequences are indicated by a blue scissor sign. The entire casein locus has been omitted from the reference sequence is shown by -//- sign.

The casein locus deletion in MEF cells was confirmed for all 9 PCR amplicons marked in figure 4. 10 using Sanger sequencing.

4.3.3 Targeting the entire casein locus in mouse zygotes

CRISPRs targeting the entire casein locus of the mouse were first tested in combinations in MEF cells and deletions were verified by PCR and Sanger sequencing. Since all CRISPR combinations successfully deleted the casein locus in MEF cells and it was not possible to assess which combination was the most efficient based on the PCR assay, hence in the first attempt of targeting casein locus in mouse zygotes was performed using all CRISPR reagents in combination. The pronuclear microinjections were carried out at the biological research facility (BRF) at the Roslin Institute, The University of Edinburgh.

CasDel:1 Buffer preparation and microinjections

In an attempt to target the casein locus in mouse zygotes, a microinjection buffer was prepared by diluting all CRISPRs targeting upstream (number 10, 11 and 12) and downstream (number 13, 14 and 15) of this locus in low EDTA Tris buffer (Invitrogen) at a concentration of 5 ng/μl each and named as CasDel 1. The buffer was filtered through a 0.2μm syringe filter and stored at -20°C.

For pronuclear microinjections, the mouse zygotes were generated in donor B6CBA/F1 female mice by superovulation and mating, which was confirmed by the presence of a vaginal plug on the next day. Zygotes were collected from the oviduct of donor mice and were placed in M2+ media (Sigma Aldrich). Each zygote was injected with 2-5 Pico litre of CasDel: 1 buffer via injection capillary. The female recipient mice were prepared by mating with a vasectomised male one day prior to the pronuclear microinjections. The re-implantation of the injected zygotes was carried out by transferring 20 injected zygotes each in the oviduct of each of 3 pseudo-pregnant recipient mice. All pronuclear microinjections were carried out using the same method.

***In vitro* analysis of injected zygotes**

A total of 12 injected zygotes were cultured in EmbryoMax KSOM Embryo Culture media (Merck Millipore) overnight at 37°C after each microinjection session to assess *in vitro* viability. All embryos were observed in a 2-cell stage on the following day indicating that successful hatching was achieved in all injected zygotes.

PCR screening of pups resulting from zygote microinjection

In order to thoroughly assess the casein locus in pups, three different primer pairs were designed. Primer pair 4.5, flanking the casein locus, was used previously to confirm that guide pairs resulted in deletion events in MEF cells. Another two pairs of primers (number 4.6 and 4.7, appendix 7.1) were designed flanking an internal and downstream cut site respectively. The both primer pairs, number 4.6 and 4.7, were expected to produce a PCR fragment in WT and casein locus heterozygous pups but not in casein locus homozygous pups and thus assessing the complete absence of casein locus in pups (Figure 4. 12). All PCR with primer pair 4.5, 4.6 and 4.7 were carried out at 20s extension time for 35 cycles using standard conditions shown in chapter 2 (Section 2.2.1) and electrophoresed through a 1.2% agarose gel along with DNA ladder (Invitrogen).

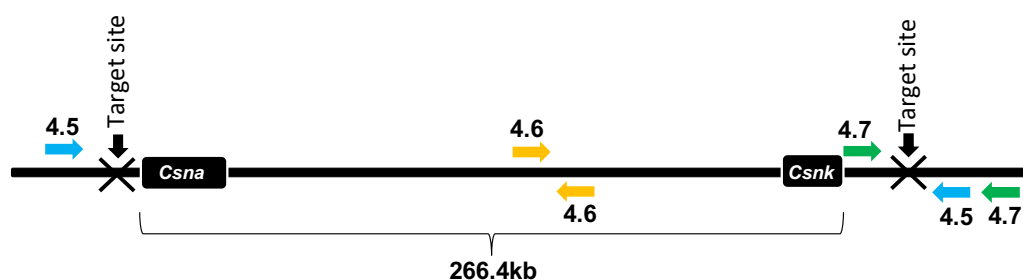


Figure 4. 12: Schematics of primer organisation to assess casein locus deletion: The Primer pair 4.5, spanning the casein locus, is expected to generate PCR fragments ranging from 133- 594 bp in the event of successful deletion. Primer pair 4.6 is situated upstream of the ECR and is expected to produce a 450 bp PCR fragment from the intact locus. Primer pair 4.7 spans the target sites for downstream guides and is anticipated to produce a 510 bp fragment from the intact locus.

The genotype of casein locus targeted pups was concluded based on the presence and absence of a PCR fragment produced by three independent PCRs (Figure 4.13) as shown in table 4.2.

Genotype	Primer pair 4.5 	Primer pair 4.6 	Primer pair 4.7 
Wild-type	✗	✓	✓
Heterozygous	✓	✓	✓
Homozygous	✓	✗	✗

Table 4. 2: Predicting genotype of the casein locus targeted mouse pups:

Assessment of mouse embryo genotypes based on PCR fragments generated by the three primer pairs. Primer pair 4.6 flanks the entire casein locus and is expected to generate a PCR fragment (133- 594 bp) in the event of deletion events. Primer pair 4.6 and 4.7 are expected to generate a PCR fragment (450 and 510 bp) from intact loci.

A total of three representative PCR were carried out using primer pair 4.5, 4.6 and 4.7. PCR with primers pair 4.5 was carried out with DNA isolated from MEF cells targeted with CRISPR number 11 and 15 whereas PCR with primer pair 4.6 and 4.7 was carried out with WT MEF DNA. All PCR were carried out using pre-described conditions and electrophoresed through a 1.2% agarose gel with DNA ladder mix (Invitrogen) (Figure 4.13).

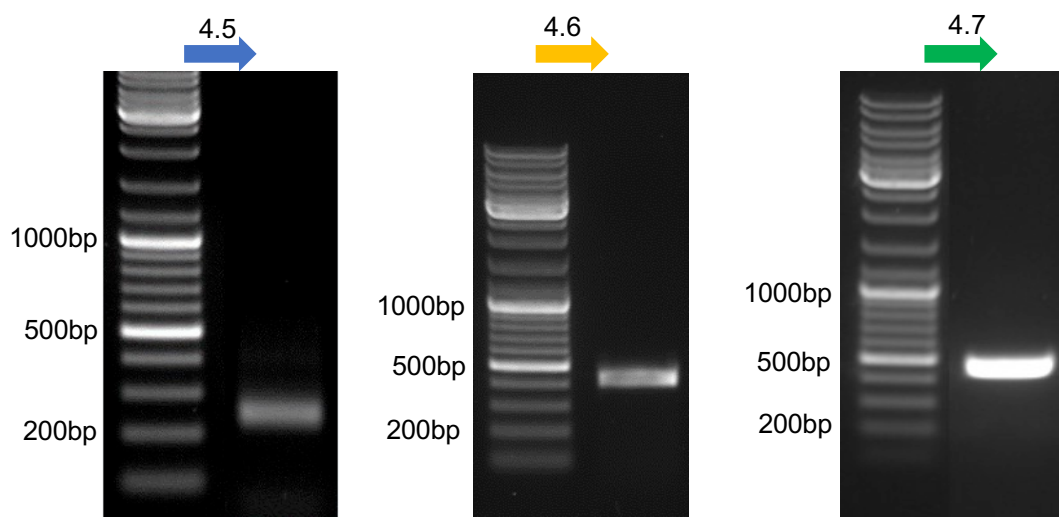


Figure 4. 13: PCR to evaluate casein deletion using the three primer pairs:

A representation of PCR with primer pair 4.5, 4.6 and 4.7. The primer pair 4.5, 4.6 and 4.7 showing a fragment of 226 bp, 450 bp and 510 bp respectively as expected.

Genotyping results

The pronuclear microinjections of buffer CasDel1 with three recipient females resulted in the generation of a total 7 pups. The ear-notch samples were collected from mice pups at the time of weaning, DNA was isolated from ear notch samples using QuickExtract DNA extraction solution (Lucigen) as described in chapter 2 (Section 2.3.9) to assess their genotype

Three independent PCRs were carried out to genotype the 7 pups as described above. PCR fragment generated by primer 4.5 was considered crucial to positively identify homozygous or heterozygous deletion events. Figure 4. 15 exemplifies the genotyping of seven mouse pups (number 1-7), The primer pair 4.5 was expected to generate a fragment ranging from 195-594 bp as shown in Table 4.1, however, no PCR fragment was detected. The primer pairs 4.6 and 4.7 produced fragments of the expected size of 450 bp and 510 bp respectively. The genotyping PCR with primer pair 4.5 was always carried out with a positive control using DNA from casein locus targeted MEF cells (Figure 4. 14).

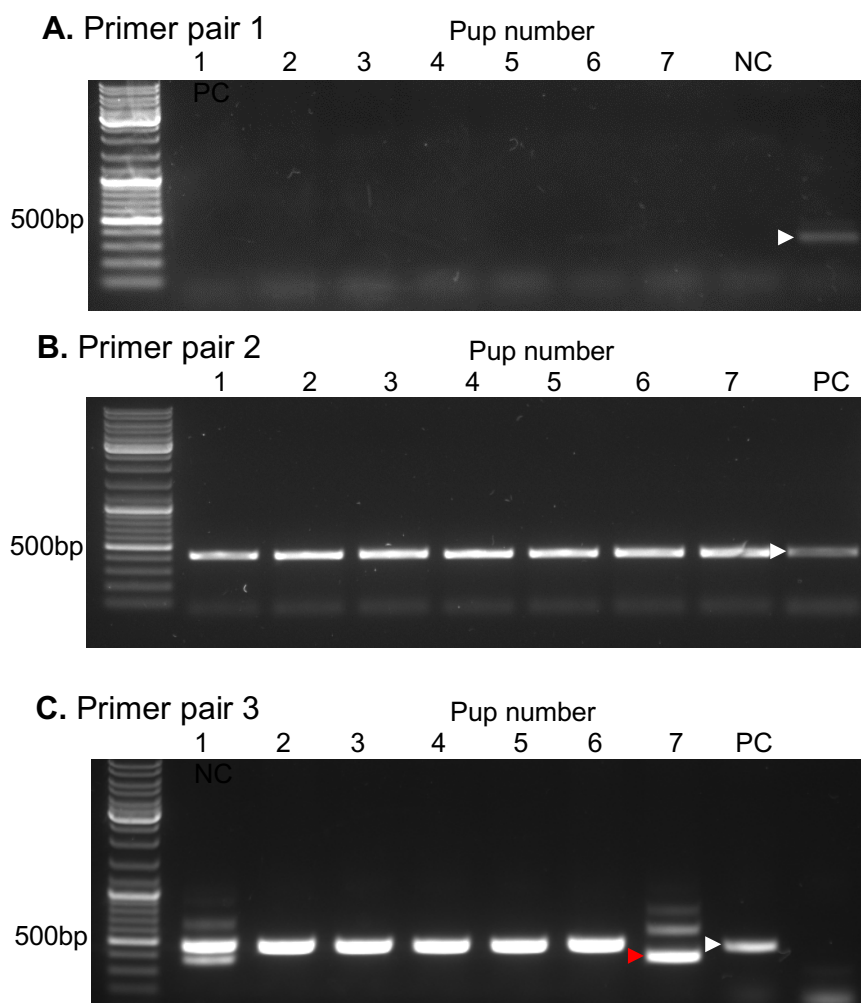
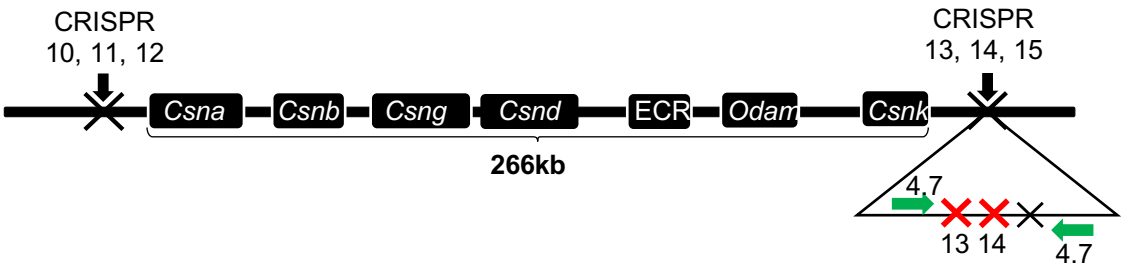


Figure 4. 14: PCR screening of casein locus targeted mouse pups: PCR of DNA isolated from 7 pups resulting from zygotes injection. A. PCR screening with primer pair 4.5 showing no fragments in DNA isolated from 7 pups. The positive control (PC) showing a 266 bp fragment (marked with a white arrow). B. PCR with primer pair 4.6 showing a 450 bp fragment in DNA isolated from all pups and positive control (marked with a white arrow). C. PCR with primer pair 4.7 showing fragment of 510 bp in all samples and positive control (marked with a white arrow) except in pup number 7 (marked with a red arrow). The negative control (NC) was carried out with no DNA.

All seven mouse pups were genotyped using the same approach, and all seven failed to generate the crucial primer pair 4.5 amplified PCR fragment. However, one exception was detected in one of the pups (pup7) where primer pair 4.7 produced a smaller than the expected sized 510 bp fragment (Figure

4.15C). DNA was isolated from the marked PCR fragment using the QIAquick gel extraction kit (Qiagen), and Sanger sequencing (Edinburgh genomics) was carried out. The resulting sequence was aligned with sequence from the WT locus (reference) using a multiple sequence alignment tool, clustal omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (Figure 4.15).

A.



B.

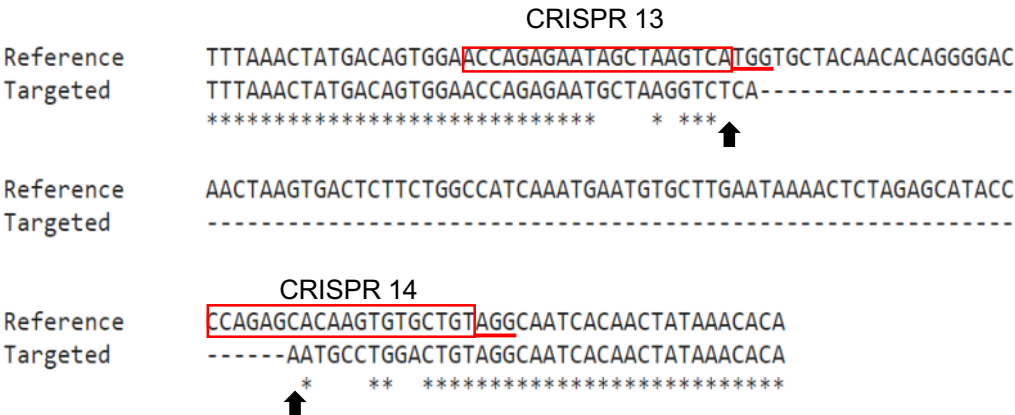


Figure 4. 15: Sequencing of a casein locus targeted mouse pup and the analysis of deletion: A. The schematics of genotyping of pup7 using primer pair 4.7, the injection of all six CRISPRs targeting both sides of mouse casein locus resulted in the deletion of a small region on the downstream of this locus. The targeting CRISPRs number 13 and 14 (highlighted). The relative location of primer pair 4.7 is shown by green arrows. B. The alignment of the WT reference sequence and PCR amplicon underneath showing the cleavage site (marked with a black arrow) of CRISPRs number 13 and 14. The CRISPR sequences are in boxes with their PAM sequences underlined.

The genotyping of pup number 7 using the primer pair 4.7 revealed deletion of a 121 bp DNA fragment consistent with cutting by CRISPRs 13 and 14 with some INDEL formation at the target site during break repair. Based on the PCR acquired from primer pair 4.5, 4.6 and 4.7 with all seven pups, none of them was considered a casein locus KO. The number of recipient mice, their litter size and genotyping results are presented in table 4.3.

Recipient mouse number	Litter size	Casein KO pups
1.	2	0
2.	2	0
3.	3	0
Total-	7	0

Table 4. 3: Genotyping of the casein locus targeted mouse pups: Number of recipient mice with their litter size and genotyping results.

The PCR and Sanger sequencing results obtained by genotyping of 7 pups were discussed with Dr Peter Hohenstein (currently at the Leiden University Medical Centre, Netherland), who has used CRISPR to generate numerous mutations in mice. At his suggestion it was decided to not use all six CRISPRs simultaneously to target casein locus in zygotes as it could lead to unintended small deletions between the flanking guides or inversion of targeted genomic region along with some other gene targeting complications.

Subsequent microinjections were carried out using a buffer with only two CRISPRs, one targeting upstream and one downstream of the mouse casein locus. CRISPR combination number 2 (10+ 14) and combination number 3 (10+ 15) were chosen for further microinjections based on PCR and Sanger sequencing analysed in MEF cells.

CasDel:2 buffer preparation and microinjections

The second buffer (CasDel:2) to target casein locus in mouse zygotes was prepared using CRISPR number 10 and 14 at a concentration of 5ng/ul each in low EDTA Tris buffer (Invitrogen) and filtered through a 0.2µm syringe

filter. Mice zygotes were isolated from the oviduct of donor B6CBA/F1 mice, and pronuclear injections were performed with CasDel 2 buffer using the same approach as shown in Figure 4.12. A total of 20 injected zygotes were transferred in the oviduct of each of seven recipient females each.

Genotyping mouse pups

A total of seven recipient mice resulted in the generation of 22 pups. The genotyping was initially carried out with DNA isolated from ear notch samples and primer pair 4.5 only. A fragment of 195 bp was expected in the event of successful casein locus deletion by CRISPR number 10 and 14. However, no fragment of expected size was detected in PCR with DNA samples of 22 pups.

Figure 4.16 shows PCR of DNA isolated from 22 pups along with its positive control carried out with DNA isolated from casein locus targeted MEF cells using the pre-described conditions and resolved in a 1.2% agarose gel. The PCR with primer pairs 4.6 and 4.7 was to be carried out only when amplification with primer pair 4.5 indicated generation of a deletion allele.

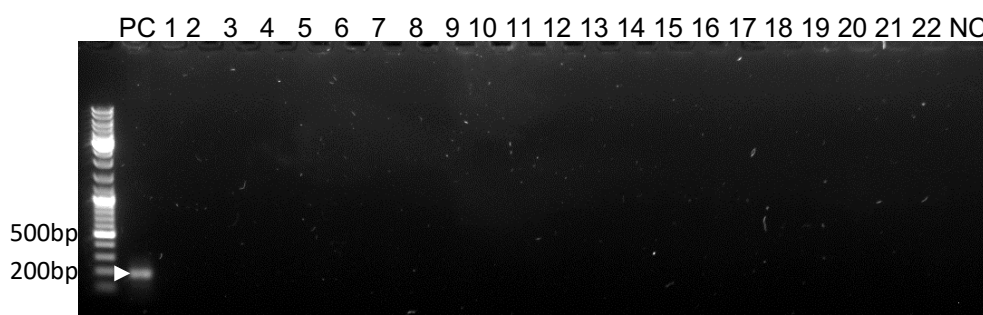


Figure 4. 16: Genotyping with casein locus targeted mouse pups: In PCR with primer pair 4.5, DNA isolated from 22 pups showed no fragment of expected the size whereas the positive control (PC) resulted in a 195 bp fragment. Negative control (NC) was run with no DNA.

Based on the results acquired from PCR of DNA isolated from 22 pups with primer pair 4.5, none of them was concluded to be have a casein KO allele (Table 4.4).

Recipient mouse number	Litter size	Casein KO pups
1	4	0
2	1	0
3	4	0
4	2	0
5	4	0
6	5	0
7	2	0
Total-	22	0

Table 4. 4: Litter size and genotyping of the casein locus targeted mouse

pups: Number of recipient mice with their litter size and the genotyping results.

As CRISPR combination 2 (10+ 14) did not produce casein KO mice, CRISPR combination 5 (11+ 15) was now used to target this locus in mouse zygotes.

CasDel:3 buffer preparation and microinjections

The third buffer (CasDel:3) to delete casein locus in the mouse was prepared using CRISPR number 11 and 15 at a concentration of 5ng/ul each in low EDTA Tris buffer (Invitrogen) and filtered through a 0.2µM syringe filter. Mice zygotes were isolated from the oviduct of donor B6CBA/F1 mice, and pronuclear injections were performed using CasDel 3 buffer. A total of 20 injected zygotes were transferred in five recipient females each.

Genotyping mouse pups

The five recipient females resulted in the birth of a total of 12 pups. Their genotyping was performed using primer pair 4.5 using pre-described PCR conditions, a PCR fragment of 226 bp was expected in casein locus deletion with CRISPR number 11 and 15. The genotyping PCR was carried out with primer pair 4.5 along with DNA isolated from ear notch samples of 12 pups along with a positive control. The PCR was resolved in 1.2% agarose gel and run with DNA ladder mix (Invitrogen) (Figure 4.17).

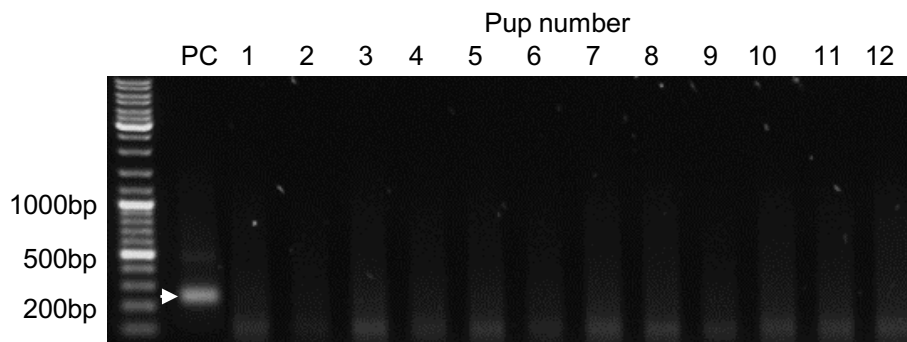


Figure 4. 17: Genotyping of the casein locus targeted mouse pups: PCR with primer pair 4.5 on the positive control (PC) produced a fragment of 226 bp whereas DNA isolated from 12 pups showing no fragment. Negative control (NC) was run with no DNA.

The PCR with primer pair 4.5 was carried out similarly with all 12 pups and the presence of a 226 bp fragment was considered crucial to detect a successful deletion. However no fragment of the expected size was observed from any of the pups (Table 4.5).

Recipient mouse number	Litter size	Casein KO pups
1	3	0
2	2	0
3	3	0
4	2	0
5	2	0
Total-	12	0

Table 4. 5: Litter size and genotyping of the casein locus targeted mouse pups: Number of recipient mice with their litter size and the genotyping results.

To generate the casein KO mice, a total of 15 microinjection sessions were carried out using microinjection buffers with different compositions (buffer CasDel 1, 2 and 3). A total number of 41 pups were born with average litter size 2.7. The initial genotyping PCR was carried out using the primer pair 4.5, however, casein KO pups were not obtained.

4.3.4 Culturing mouse zygotes injected with CRISPRs targeting the entire casein locus

The casein locus of mouse contains long intergenic regions along with milk protein genes like α , β and κ -casein. It also possesses many non-milk genes that play roles out-with milk production hence it was possible that targeted zygotes did not survive because of essential nature of one of the genes situated in this locus. To explore any potential role of casein locus in growth and survival of mouse zygotes, the *in vitro* culturing of mouse zygotes injected with CRISPRs combination 2 (CasDel 2- CRISPR 10 + 14) was performed. The mouse zygotes were planned to be cultured till the blastocyst stage as this can be usually achieved in an incubator at 5% O₂ and 5% CO₂ at 38.5°C.

Mouse zygotes isolation and microinjections

A total of 4 Donor B6CBAF1 mice were super ovulated and mated, and zygotes were collected from the oviduct as described in the previous Section. A total of 45 zygotes were divided into three equal groups of 15. Control zygotes were not injected. Blank zygotes were injected with low EDTA tris buffer only. Test zygotes were injected with CasDel 2 buffer. All zygotes were incubated at 5% O₂ and 5% CO₂ at 38.5°C for 5 days. We reasoned that these three categories would allow us to evaluate any potential effect of CRISPRs injection and casein locus targeting in mouse zygotes. The control zygotes were to reveal if pronuclear microinjections were safe under normal circumstances and the blank could be compared with this to assess if pronuclear microinjections were affecting the zygotes survival. The test zygotes allowed us to analyse if targeting the casein locus had any effect on zygotes.

Culturing control, blank and test mouse zygotes

Each group of 15 mouse zygotes was cultured in a five-well plate containing 500 µl KSOM embryo culture media (Sigma Aldrich) supplemented with 10% rat serum (BioSera) at 38.5°C and 5% O₂ and CO₂ in an incubator for 72 hours. The zygotes were covered with a layer of mineral oil to prevent evaporation.

Analysing mouse zygotes 72 hours post microinjections

Cultured mouse zygotes were visualised using a microscope with a preheated platform at 38.5°C. The number and growth stages are shown in table 4.6.

	Control		Blank		Test	
	EB	LB	EB	LB	EB	LB
No. of zygotes	7	2	5	0	6	1

Table 4. 6: Analysing cultured mouse zygotes at 72h post microinjections:

In the control zygotes, 7 were found at EB stage and 2 at the late blastocyst (LB) stage. The blank zygotes, resulted in 5 at early blastocyst (EB) stage, and no zygotes were observed at LB state. In test zygotes injected with buffer CasDel 2, six were in EB stage whereas 1 was in LB stage. All the remaining zygotes were in a two-cell stage.

Analysing zygotes at 96 hours post microinjections

Mouse zygotes (control, blank and test) visualised at 96 hours post microinjections were expected in an early rosette stage however all zygotes were observed in early or late blastocyst stage. All zygotes were lost while harvesting into individual wells of a 96 well plate. Hence, the experiment was discontinued at this stage. Since the experiment did not provide the expected outcome, a careful repetition was required to assess the effects of casein locus deletion on the mouse zygotes. However, due to time and funding constraints

this was not possible.

4.4 Discussion

The casein locus of mammals spans hundreds of kb depending on the species and contains genes encoding many of the major milk proteins, along with some non-milk genes. The organisation of the casein locus in mammals shows robust conservation throughout evolution (Bouniol et al., 1993), however, the existence of several non-milk genes makes it an interesting locus to study (Rijnkels, 2002). The conventional gene knockout studies have revealed the presence of essential α and κ casein (Kolb et al., 2011), (Shekar et al., 2006) and β casein with no major deficit in development of infant drinking the β casein null milk (Kumar et al., 1994). However, the collective effect of total casein locus absence on lactation and the mammary gland is unknown. In an attempt to remove the entire casein locus in mice, CRISPRs were designed and tested in MEF cells. CRISPRs have been proven highly capable of generating large genomic deletions in mouse (Zhang et al., 2015), (Kato et al., 2017). The casein locus deletion was successfully achieved in MEF cells using nine combinations of CRISPRs targeting each side of this locus were assessed by PCR and Sanger sequencing. Several attempts were then made to generate casein KO mice using pronuclear microinjections. However, the casein KO mice were not obtained.

These results encouraged us to assess any possible effect of casein locus deletion on mouse zygotes; hence whole zygote culture was carried out. While the one attempt of zygote culture failed and remained inconclusive, it should probably be the primary step to create the casein KO mice in the future. The zygote culture is not only likely to help in finding essential elements in casein locus involved in embryonic development, it can also help in evaluating the efficiency of large genomic deletions in mouse using CRISPRs.

Generation of large deletions

The deletion of large DNA fragments was primarily achieved by targeting 981kb DNA in zebrafish using TALENs (Zu et al., 2013). CRISPRs

have been proven equally efficient for the successful deletion of large fragments; 24.4mb in mouse and rat (Birling et al., 2017) 105kb in rabbit (Song et al., 2016) and up to 1mb in human embryonic kidney 293FT cells (He et al., 2015). The deletion efficiency of large fragments has usually been observed between 1-10%. The entire 266kb casein locus of the mouse was targeted for deletion by designing CRISPRs on each side of this locus. The deletion in MEF cells was analysed using a primer pair flanking the targeted region, with a PCR fragment expected only in the case of a casein locus deletion. The CRISPRs cleavage site was usually observed as expected to be 3 bp upstream to the PAM sequence, however, the deletion frequency of CRISPRs combinations was not estimated.

The generation of ECR KO mice had been successfully achieved by pronuclear microinjections described in chapter 3. The microinjections to generate casein KO mice were carried out by the same technical staff using the same method. Initially, a combination of six CRISPRs with three targeting upstream and three downstream of the casein locus were injected in mouse zygotes. While deletion of the entire locus was not detected, sequencing of an unexpected PCR product revealed a deletion event between two guides at the 3' end of the locus. This deletion event might have resulted in destruction of the CRISPRs binding sites making it impossible to be recut due to the modification at the original target site which was verified by Sanger sequencing analysis of the site revealing unintended INDELS near the targeting site. There are evidence that the use of multiple CRISPRs may result in unintended deletions and insertions (Shin et al., 2017) hence, the use of only two CRISPRs one targeting the upstream and one targeting the downstream of casein locus was planned. A total of 20 injected zygotes were transferred in a total of 12 recipient mice each resulted in the generation of a total of 34 pups, with no casein KO pup.

We assume there could be three reasons for not obtaining the casein KO pups, the first is that the CRISPRs mediated large deletion have previously shown efficiencies between 1-10% hence it is possible that the casein locus

deletion was not achieved in 34 pups generated by injecting combination of two CRISPRs in zygotes. The second reason is the presence of essential elements in casein locus required for zygote development. In that case, the successfully targeted zygote may retard during its development. Another reason could be that the cutting frequency of CRISPRs combinations targeting the casein locus in MEF cells was not estimated. There are a few ways which could have been used to quantify the CRISPRs cutting activity to choose the optimum guide combination to increase the probability of success. The digital PCR analysis using serial dilution of DNA isolation from CRISPRs co-transfected cells in different concentration. The deletion frequency is then estimated based on the lowest DNA concentration which results in generation of a PCR product. Another way to quantify the CRISPRs activity is using an online tool 'Tracking of INDELS by Decomposition or TIDE' (<https://tide.deskgen.com/>) which estimates deletion frequency based on Sanger sequencing of WT control DNA and the targeted DNA sequence.

Analysing the casein locus in mouse

To understand the landscape of mouse casein locus and to find if any essential gene/element could be involved in embryonic development, the non-milk gene expressed by this locus were reviewed. The non-milk genes expressed in the casein locus of the mouse are the *Odam*, FCSDP and histatin/Statherin gene family.

The *Odam* gene is expressed in tooth associated epithelia (Moffatt et al., 2008) whose function was assessed in the *Odam* KO mice which showed its role in maintaining attachment of junctional epithelium in older mice, however its effect on milk proteins expression was not recorded (Wazen et al., 2015). The function of FDCSP gene was assessed in the transgenic mice expressing this gene in lymphoid tissues showed a significant decrease in IgA level (Al-Alwan et al., 2007) whereas an increase in IgA level was observed in the FDCSP deficient mice (Al-Alwan et al., 2007). The association of FDCSP with IgA nephropathy (IgAN), deposition of IgA in kidneys of affected patients, was checked in the tonsils of IgAN patients

resulted in decreased FDCSP expression and increased IgA level shows involvement of FDCSP in IgA production (Hou et al., 2014). The human casein locus contains a family of three highly identical genes i.e. histatin 1 (HTN1) histatin 2 (HTN2) and statherin (STATH) expressed in the salivary gland. The human HTN1 shows 89% and 77%-81% overall and coding sequence identity with HTN2 and STATH genes respectively and hence the HTN/STATH protein family has been proposed to have a common ancestor and might have evolved with milk proteins in the casein locus. These gene have been located in casein locus of mammals including mouse, monkey and cattle and were initially associated with oral health defence (Sabatini et al., 1993), (Mardh et al., 2002). The histidine 1 peptide in human saliva have been connected with cell-cell adhesion and angiogenesis (van Dijk et al., 2015), (van Dijk et al., 2017) whereas statherin have shown to play a role in prevention of calcium phosphate degradation caused by bacteria, aiding the tooth enamel health (Johansson et al., 2000), (Helmerhorst et al., 2010).

The individual gene knock out study of α , β , κ caseins, *Odam* and FDCSP exhibit that they may not be associated with embryonic development in mouse whereas the histatin/Statherin gene family does not appear to express in embryos. The milk proteins (α , β and κ caseins) were initially proposed to express in mammary gland only, however, in an interesting study the human α casein (*Csn1s1*) mRNA was identified in monocytes and T cells by enzyme linked immunosorbent assay (Vordenbaumen et al., 2011) and was later shown to differentiate monocytes in to macrophage like cells in *in vitro* conditions (Vordenbaumen et al., 2013). This indicates that proteins expressed by the casein locus has regulatory function in immune system apart from their nutritional properties. The hematopoietic stem cells are progenitor cells that differentiate in to blood cells involved in innate immunity. In mouse the haematopoiesis starts at embryonic day (E) 10.5 in thymus, liver, bone marrow and spleen and differentiate in to the blood cells after birth (Landreth, 2002), suggests that targeting the casein locus in mouse should not be lethal for zygote development however this could only be verified by

in vitro culture of casein locus deleted mouse zygotes.

Mouse zygote culture

In early zygote, the first phase of gene expression is dependent on macromolecules derived from the maternal oocytes which prompt the next phase where embryonic gene expression starts as a result of fertilisation (Davidson 1986) (Rutledge, 1997). In mouse, the transition of gene expression from the first phase to second occurs at 2 cell stage (Bolton and Braude, 1984) hence it is unlikely that removal of a gene would have any effect in the first 48 hours of zygote development.

The whole zygote culture is a technique of growing mouse zygotes in *in-vitro* conditions to analyse organogenesis at different stages of development. This technique can also be applied to check the effects of specific gene deletion on early growth of the mouse zygote. The mouse zygotes injected with CRISPRs targeting the entire casein locus were cultured to assess any potential effect of casein locus deletion on zygotes development until the blastocyst stage, as this could be achieved in a standard low oxygen incubator. A total of 45 zygotes were cultured into three categories, the control, blank and test. The microinjections of mouse zygotes were carried out at The Roslin Institute and were transported in a portable incubator at 38.5°C to the embryology suite at large animal research and imaging facility at the Dryden farm, The University of Edinburgh. The zygotes were then cultured in an incubator at 5%O₂ and 5%CO₂ and 38.5°C for four days.

The microscopic observation after 72 hours revealed a total of 21 (46%) alive zygotes whereas, at 96-hour observation, all zygotes were expected in the early rosette stage however all zygotes were observed in an underdeveloped stage. All zygotes were lost in an attempt to harvest each zygote in to a single well of 96well plate

The mouse zygote has been cultured successfully i.e. for two weeks using enriched culture medium in a roller culture apparatus (rotating ~30 RPM)

within an incubator (BTC, Engineering Cambridge, UK) (Piliszek et al., 2011) at 5%CO₂ and 38.5°C with 5%O₂ at embryonic day 0.5 increased to 95%O₂ on embryonic day 10.5. We suspect that due to unavailability of this apparatus and transporting zygotes after microinjections might have affected their survival and the lack of practice with mouse zygote handling affected their harvesting nevertheless, optimisation of the mouse zygote culture using a roller culture apparatus and better handling of zygotes could improve the results.

The utility of casein KO mice

A casein KO mouse, if successfully generated, will be helpful in understanding the collective role of this locus in mammals. The individual deletion of α and κ casein established their essential roles in growth of the infant and successful lactation. Hence the casein KO mice is unlikely to nurture its pups, which can be accomplished by using a surrogate female.

The caseins constitute ~80% of total milk proteins, deleting them is expected to affect the expression level of other milk components and would be interesting to analyse if lactation were to be sustained and if total casein loss affects the mammary gland morphology. The deletion of caseins will result in a huge decrease in milk protein and calcium which the mammary gland could compensate by improved secretion of other components such as whey proteins, fats and lactose. If the milk lactose is upregulated it will open new doors for research as enhanced milk lactose has been connected with improved litter number and body weight in pigs and mouse.

The whey acidic protein (WAP) are absent in milk of human and cattle whereas it is present in mouse milk and its loss has been connected with malnourishment in pups feeding on WAP loss milk (Triplett et al., 2005). The loss of total casein may increase the WAP secretion in mouse milk compensating the loss of major milk proteins. During pregnancy the majority of calcium consumed by mother is directed toward the embryos via placenta (Widdowson, 1979) whereas during lactation calcium is directed toward the mammary glands. A low calcium intake during pregnancy results in low blood

pressure in mother and reciprocal high blood pressure in offspring (Roberts and D'Abarno, 1996). It will be interesting to analyse how the high amounts of calcium is handled by the mammary gland of casein KO mice supplied with *ad libitum* diet.

A casein KO mice could also be used as a model organism in redesigning a tailored casein locus. The caseins are present in micelle structure in milk whose size ranges from 40-300nm in diameter with an ability to penetrate into the cell membrane therefore they have been considered as an alternative for biological delivery of therapeutic drugs with poor cellular uptake (Morris et al., 2000), (Mandal et al., 2017). The arrangement of casein micelle reveals the encapsulation of α and β casein in a hydrophobic core, stabilised by κ casein. This property has been used to encapsulate vitamins such as Vitamins D2 (Menendez-Aguirre et al., 2014) and B9 (JiangYi, 2015) and D3 (Levinson et al., 2016) by mechanical methods. Studies have revealed the structure of casein micelles to be porous and sponge-like (Bouchoux et al., 2010) which can be used to absorb pharmaceutical and nutritional substances. The redesigning of the casein locus may help in understanding more about the structure and stability of casein micelles. Moreover, gene editing techniques could reveal opportunities of *in vivo* encapsulation of micelles with the substance of interest.

The β casein, when deleted in mice showed no effect on the health and fertility of suckling infants. However, this protein has been connected with milk stability (Clark 1987), an essential consideration for global dairy industries. Out of a total of 12 variants of β casein, the A1 and A2 are mostly expressed in cattle (m Sodhi, 2008). The major difference in β casein variants is in their conformational disparity due to a single nucleotide polymorphism in amino acid histidine (CAT) and proline (CCT) at position 67 in A1 and A2 respectively (Kaminski et al., 2007) The A1 variant has been considered a gastrointestinal allergen whereas no such reactions have been associated with A2 variant (Jianqin et al., 2016), (Brooke-Taylor et al., 2017). The gene editing techniques could help in reversing the mutation in β casein and could also be used in

exploring its role in improving milk's shelf life.

5 Chapter 5: Modifying the alpha-lactalbumin promoter

5.1 Introduction

Lactose is a sugar that is exclusive to milk and provides energy to infant mammals. The production of lactose in milk is controlled by alpha-lactalbumin which forms the regulatory subunit of lactose synthase. In mammary epithelial cells, alpha-Lactalbumin and B 1,4-galactosyltransferase form a catalytic component that produces lactose in the golgi complex (Kuhn et al., 1980).

Alpha-lactalbumin in the mouse is encoded by a gene called *alba*. Activation of the *alba* promoter is dependent on the differentiation of mammary epithelial cells by lactogenic hormones such as prolactin, epidermal growth factor (EGF) and insulin (Morrison and Cutler, 2009). Prolactin is a luteotropin and lactogenic hormone secreted by the pituitary gland in mammals. In 1928, Stricker and Grueter showed that injection of crude pituitary extract from lactating rabbits into castrated pseudo-pregnant animals resulted in the secretion of milk, thus establishing a role for the pituitary in lactation (Young, 1938). Isolated and purified from pituitary extract, prolactin was named by Riddle and his colleagues in 1933 (Riddle O, 1933). It is one is the primary hormones involved in the differentiation of mammary epithelial cells and transcriptional regulation of various milk protein genes during pregnancy (Prunet et al., 1985) whereas insulin and EGF support the proliferation of mammary epithelial cells (Oka et al., 1974).

Signal transducer and activator of transcription factor 5 (STAT5) and induction of alpha-lactalbumin

Signal transducer and activator of transcription (STAT) is a family of seven regulatory proteins that are involved in numerous physiological processes such as cell differentiation, transcription activation and apoptosis. STAT5 is a member of this protein family involved in haematopoietic, immune and lactogenic hormone's cytokines regulation (Rochman et al., 2009). These proteins bind to the promoter of the target gene at their consensus motif and

activate transcription via signalling pathways (Harrison, 2012). STAT5 is an important mediator that binds to promoters of milk genes in response to differentiation of mammary epithelial cells by lactogenic hormones (Liu et al., 1996).

STAT5 was initially identified as a mammary gland factor or milk protein binding factor due to its involvement in mammary differentiation and milk protein production (Schmitt-Ney et al., 1991), (Watson et al., 1991). STAT5 proteins have two highly homologous variants called STAT5a and STAT5b, which share 91% identity in human (Hou et al., 1995). The binding of STAT5 proteins in the *lalba* promoter occurs at a unique consensus binding site called the gamma interferon activation site (GAS). The GAS is a palindromic sequence that acts as a binding site for dimerised and/or tetramerised STAT5 protein. It was identified as an IFN gamma response element (Horvath et al., 1995) and within that, the consensus binding site for the STAT protein family was identified as 5' TTC 3'.

Initially, the core binding sequence was identified only for STAT1 and STAT6. STAT1 showed higher affinity towards a 3 bp spacing TTC-N₃-GAA whereas STAT6 had higher affinity for a 4 bp spacing TTC-N₄-GAA (Schindler and Darnell, 1995). Increasing the spacing to more than 11 bp was shown to be large enough to allow differentiation between the two GAS motifs. It has also been demonstrated that an incomplete tetramer binding site is sufficient to anchor a STAT5 tetramer however its cellular activity was not analysed (Liu et al., 1995). Subsequently, TTC-N₃-GAA was discovered to be the consensus binding site for STAT5 dimer proteins, for binding of the STAT5 tetramer, a 6 bp spacer between two GAS motifs was found to be the most efficient (Soldaini et al., 2000). The mouse STAT5 transcription factor binding site (TFBS) was identified as 5' TTC-N₃-GAA 3' (Malewski, 1998).

The role of STAT5 protein variants was individually analysed in STAT5a and STAT5b knockout mice generated by targeted embryonic stem cells. The STAT5a deficient mice showed no terminal differentiation of mammary alveolar cells, resulting in the disruption of lactation (Liu et al., 1997) whereas

the absence of STAT5b in mice resulted in reduced body growth of adult STAT5b knockout mice when compared to the wild type mice (Udy et al., 1997). Subsequently, the STAT5a/b knockout mice showed reduced growth and body weight in mutant mice at 12 weeks of age when compared to the wild-type litter mates. The ovarian development and the differentiation of terminal mammary buds were highly affected in mutant mature virgin mice, although no response was observed in the cytokine regulation of erythropoietin hormone. The conditional gene inactivation of STAT5a in the different phases of mammary gland morphogenesis such as cell proliferation, differentiation and maintenance in mice resulted in the absence of secretory alveoli and premature cell death in the mammary gland (Cui et al., 2004).

The *in vivo* expression of STAT proteins was analysed in Saanen goats by injecting recombinant bovine growth hormone which resulted in an increased level of not only STAT5 but also STAT1 and STAT3 in mammary epithelial cells. However, northern blot analysis of prolactin receptor and k-casein mRNA showed no effect in their expression levels (Boutinaud and Jammes, 2004). In recent research, the effect of STAT5 dimer and tetramer in natural killer (NK) cell regulation was assessed in the STAT5 tetramer deficient mice showing that NK cells develop in the presence STAT5 dimer however their maturation is highly compromised when STAT5 tetramer is absent (Li et al., 2017a).

Similarly to STAT5, alpha-lactalbumin deficiency in mice shows its importance in lactation. The alpha-lactalbumin KO mice produce a very viscous fluid in the mammary gland which the pups were unable to suckle (Stinnakre et al., 1994). In a similar experiment, lactation was disrupted by deletion of alpha-lactalbumin in mice but was restored by replacing with human alpha-lactalbumin. The pups did not thrive while feeding on the milk from alpha-lactalbumin knockout mice but survived when human alpha-lactalbumin was expressed in the milk (Stacey et al., 1995), highlighting the importance of alpha-lactalbumin for successful lactation and the survival of offspring.

In the early stages of infant development, lactose not only provides energy to infants but also encourages them to feed more. The increased expression of lactose in milk has been associated with higher survival rates and increased weight in pigs and mice. For e.g. the effect of increased milk lactose in suckling infants was assessed by generating transgenic pigs expressing bovine alpha-lactalbumin. The transgenic pigs produced as much as 50% more lactose compared to control pigs. However, the concentration of bovine alpha-lactalbumin in transgenic pigs decreased with the progression of lactation. Some of the reasons for this decrease were assumed to be the use of shorter bovine alpha-lactalbumin construct which might have lost some essential regulatory elements or this gene might have different controlling mechanism. The effect of milk with increased lactose on suckling piglets was not recorded (Bleck et al., 1998). In a similar research including mice and pigs expressing the bovine alpha-lactalbumin resulted in 4% increased growth of the mice pups on day 10 of lactation (Boston et al., 2001) whereas piglets consuming the transgenic milk had higher weaning weight as compared to control piglets feeding on the WT milk (Noble et al., 2002)

In an attempt to maximise the expression of Beta lactalbumin, extra STAT5 binding sites were incorporated in the Beta lactalbumin (BLG) promoter of the mouse. A 195 bp sheep BLG gene fragment containing three STAT5 dimer binding sites was incorporated in the BLG promoter in mouse mammary epithelial HC11 cells and in transgenic mice. A total of three duplicative iterations of this 195 bp fragment were analysed individually; two head-to-tail fragments (280F), two fragments with an 88 bp spacer (370F) and two fragments with a 179 bp spacer (460F). Northern blot analysis of BLG mRNA expression from all three duplications showed identical expression levels when compared to BLG expression of WT HC11 cells. The *in vivo* analysis of sheep BLG gene duplication, 280F, 370F and 460F, with multiple STAT5 dimer binding site in BLG gene of mice showed low transgene copy number and low BLG expression. (Pena and Whitelaw, 2005).

In a similar attempt, we set out to analyse the STAT5 binding site in the *lalba* gene of mouse, to seek an approach to upregulate the expression of *lalba* gene promoter in mouse mammary epithelial HC11 cells.

5.2 Aims

1. Comparative analysis of mouse *lalba* promoter with other milk producing livestock animals and analysing the STAT5 transcription factor binding sites of various mammals.
2. Assess approaches to enhance alpha-lactalbumin expression by editing STAT5 binding sites in *lalba* promoter of mouse mammary epithelial (HC11) cells.

5.3 Results:

5.3.1 Analysing the *lalba* promoter in mammals

Locating and aligning the *lalba* promoter in mammals

The *lalba* promoters of the mouse, human and livestock animals such as pig, sheep and cow, were analysed to determine the sequence identity among these mammals. The *lalba* gene sequences containing 1500 bp upstream from the transcription start site (TSS) were retrieved from the Ensemble genome browser as shown in (Table 5. 1A). These sequences were aligned using the Clustal Omega multiple sequence alignment tool (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). The *lalba* promoter from these mammals showed more than 50% identity to one another. The *lalba* promoter sequence of sheep and cow showed the maximum identity with over 92.91% whereas the lowest identity was found between the mouse and pig *lalba* promoters (51.92%) (Table 5. 1B).

A.

Species	Chromosome number	Start	End
Cattle	5	31,346,638	31,348,138
Human	12c	48,569,660	48,571,160
Mouse	15	98,482,311	98,483,811
Pig	5	82,538,714	82,540,214
Sheep	3	137,214,968	137,216,468

B.

	Mouse	Pig	Sheep	Cattle
Human	57.75	59.86	60.63	58.32
Mouse		51.92	53.79	52.39
Pig			62.59	60.59
Sheep				92.91

Table 5. 1: The *lalba* promoter sequence retrieval and its alignment among selected mammals: A. Table showing the chromosome number and the start and end positions for the 1500 bp sequence upstream of the *lalba* TSS. B. The sequence identities (in %) found when comparing the *lalba* UTR promoter sequences of selected mammals.

Identification of core STAT5 dimer motif in the mouse genome

The core binding site for STAT5 TF was identified in the entire mouse genome using the Jasper database (<http://jaspar.genereg.net/matrix/MA0519.1/>) by selecting the organism and TF as *Mus musculus* and STAT5 respectively. The STAT5 motif was found to be based on a gamma interferon activation motif, i.e. a palindromic 5'-TTCN₃GAA-3' sequence with a spacer of 3 bp for a dimer STAT5a and STAT5b (Figure 5. 1).

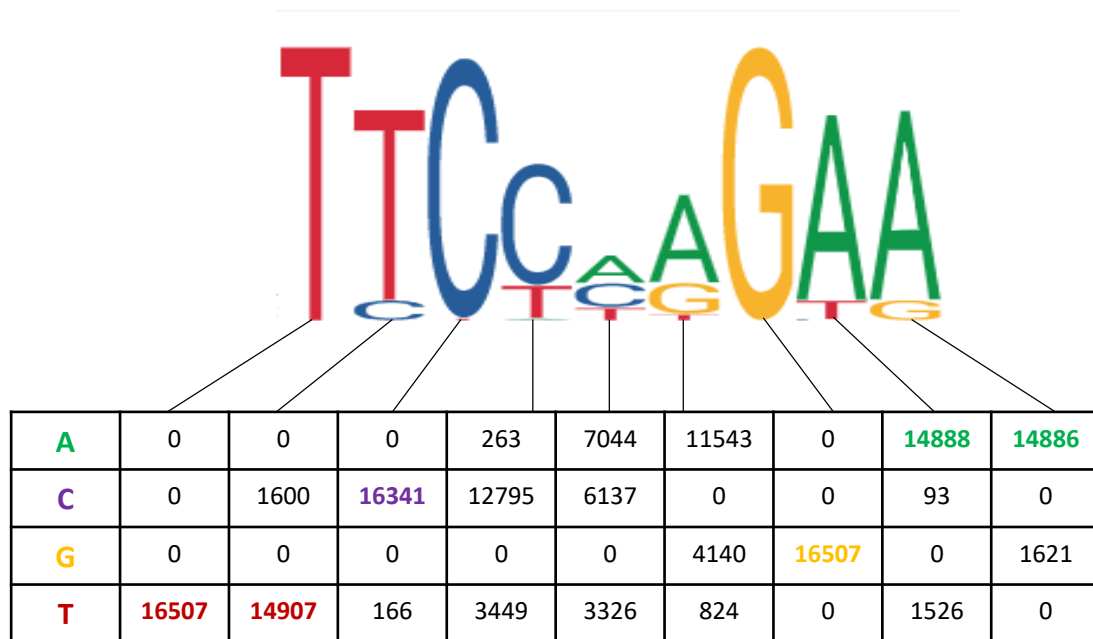


Figure 5. 1: Maximum frequency matrix of the mouse STAT5 dimer binding site throughout the genome: A maximum frequency matrix of the STAT5 binding site in the mouse acquired from the Jasper database. The numbers represent the frequency of a specific base pair appearing and established the consensus sequence as the palindrome 5'-TTCN₃GAA-3' in the mouse genome (acquired from Jasper Database).

The mouse STAT5 dimer binding motif detailed above was used as a template to identify the binding site(s) within the *lalba* promoter of mouse using an online TF searching tool (ContraV2 <http://bioit2.irc.uhent.be/contra/v3/#/step/1>). This was carried out by selecting the 1500 bp promoter region of mouse *lalba* gene which revealed a list of TFBS in the selected region. The STAT5 TF binding site was selected from a list of TFs and searched with the maximum stringency to minimize false positive, resulting in the detection of one potential STAT5 dimer binding site with the core sequence of 5' TTC-N₃-GAA 3' (Figure 5.1).

Identifying STAT5 binding sites in the *lalba* gene of other mammals

ContraV2 was also used to identify the STAT5 binding sites in several other mammals using the same methods as used for mouse. A comparison was carried out between the STAT5 dimer binding sites and the lactose

percentage in milk in an attempt to identify a potential relationship in selected mammals (Table 5.2).

Organism	STAT5 Sequence	Lactose in milk (%)
Rhesus Monkey	TGCTTCCTAGAAACCA	7.0
Donkey	TGCTTCCTAGAAACCA	6.9
Human	TGCTTCCCAAAACCA	6.8
Horse	TGCTTCCTAGAAACCA	6.1
Dolphin	TGCTTCCTAGAAACCA	5.9
Sheep	CAATTCCTAGAAAGTC	5.3
Pig	CAATTCCTAGAAAGTC	4.9
Cow	GAATTCCTAGAAATC	4.7
Goat	TGCTTCCTAGAAACCA	4.6
Wild Yak	TGCTTCCTAGAAACCA	4.6
Dog	TGCTTCCTTCAAGAA	3.7
Domestic Cat	TGCTTCCTAGAAACCA	3.4
Elephant	TGCTTCCTAGAACTG	3.4
Rabbit	CAATTCCTAGAACTC	3.4
Mouse	CAATTCCTAGAAAGTC	3.0
House Mouse	TGCTTCCTGAAAGCA	2.9
Kangaroo Rat	TGGTTCCTAAACTA	2.9
Weddell Seal	TGCTTCCTAGAAACA	2.6
Sperm Whale	TGCTTCCTAGAAACCA	1.8

Table 5. 2: The STAT5 dimer binding site and lactose concentration of milk in mammals: A Table showing the sequence of the STAT5 dimer binding sites in the *lalba* promoter and the lactose concentration of milk in selected mammals. Identical bases within the STAT5 base pair triplets are shown in red and dissimilar bases shown in blue.

All mammals analysed appear to contain only one STAT5 dimer binding site in the *lalba* promoter. However, no relationship between the STAT5 binding site and milk lactose concentration was observed.

Searching for STAT5 tetramer binding site in the *lalba* promoter of mammals

The STAT5 TF can bind to a dimer and/or a tetramer binding motif. The previously located STAT5 dimer binding site in the *lalba* promoter of all mammals shown in Table 5.2 was further analysed by manual visualisation to identify a potential tetramer binding site. An incomplete STAT5 tetramer binding site with a 6 bp spacer was found as a putative binding sequence in several animals listed in Table 5. 3.

Organism	STAT5 Sequence
Cat	TTCCTAGAA CCAAATCTACAAGAA
Cow	TTCCTGGAAGCAGCCCTAAAAGAA
Dolphin	TTCCTAGAA CCAACACTACAAGAA
Elephant	TTCCTAGAA CTGACCCTACAAGAA
Goat	TTCCTAGAA CCAACACTACCAGAA
Horse	TTCCTAGAA CCAACCTACAAGAA
Human	TTCCTGGAAGCAGCCCTAAAAGAA
Mouse	TTCCTGGAAGCAGCCCTAAAAGAA
Pig	TTCCTGGAAGCAGCCCTAAAAGAA
Sheep	TTCCTGGAAGCAGCCCTAAAAGAA

Table 5. 3: Alignment of putative STAT5 tetramer binding sites in the *lalba* promoter of mammals: A list of the STAT5 tetramer binding sites identified in selected mammals. Identical bases are shown in red and dissimilar bases are shown in blue.

A putative STAT5 tetramer binding site was found to be identical in all mammals except in horse. The editing of WT STAT5 binding site in the *lalba* promoter of mouse was proposed to be a model for application in livestock. In an attempt to assess the effect of a complete STAT5 tetramer binding site on the mouse *lalba* promoter activity and ultimately on alpha-lactalbumin expression, we set out to edit the wild-type (WT) STAT5 binding site. Two types of STAT5 editing were envisaged, one where a complete STAT5 tetramer binding site (5'-TTC-N₃-GAA-N₆-TTC-N₃-GAA-3') was introduced into

the mouse *lalba* promoter, and one where the endogenous STAT5 binding site was removed.

5.3.2 Amplifying the mouse *lalba* promoter

In order to create reporter constructs containing the mouse *lalba* promoter, the promoter region was initially amplified using DNA isolated from mouse mammary epithelial HC11 cells. These cells were cultured and maintained as described in chapter 2 (Section 2.3.2). DNA was isolated from HC11 cells using the DNeasy blood and tissue kit (Qiagen). A primer pair (5.1, listed in appendix 7.1) flanking the mouse *lalba* promoter and the first coding exon was designed based on the sequence retrieved from Ensemble genome browser and the GRCm38.p6 database using primer blast online tool (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) (Figure 5. 2A,B).

to the transcription start site B. The 1544 bp mouse *lalba* promoter sequence acquired from Ensembl database, the STAT5 binding site is bold and underlined fonts and primer pair 5.1F and 5.1R is shown in red. The transcription start site is marked with a black arrow.

1544 bp of the mouse *lalba* promoter was amplified by PCR using DNA isolated from HC11 cells and primer pair 5.1 at 20s extension time for 35 cycles using the standard conditions described in chapter 2 (Section 2.2.1). The PCR was electrophoresed on a 1.2% agarose gel with 1kb DNA ladder (Invitrogen G5711) (Figure 5. 3).

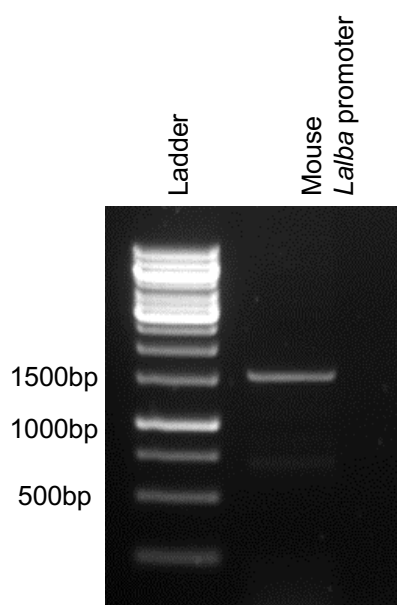


Figure 5. 3: PCR amplification of mouse *lalba* promoter: The PCR was carried out using DNA from HC11 cells as template for primer pair 5.1. Following electrophoresis a fragment approximating the expected 1544 bp was observed.

The PCR fragment (marked) was excised from the gel and DNA was isolated using a QIAquick gel extraction kit (Qiagen). The presence of STAT5 binding site in the WT *lalba* promoter was confirmed using Sanger sequencing (Edinburgh genomics).

Verifying the WT STAT5 binding site in the mouse *lalba* promoter

The sequencing results were compared to the *lalba* promoter sequence acquired from the Ensembl database (reference sequence) to locate the WT STAT5 binding site (Figure 5. 4), which was found to be identical to the published sequence.

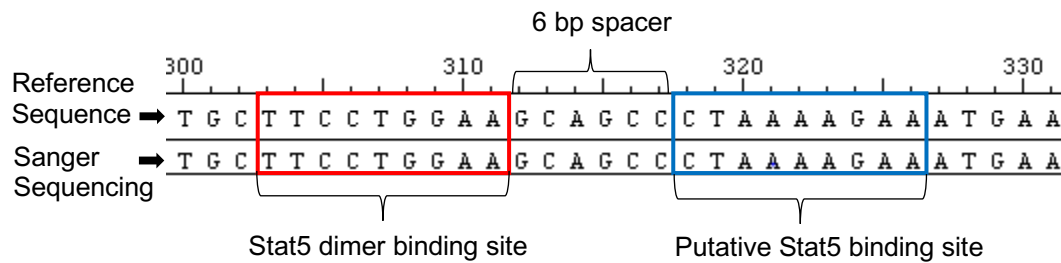


Figure 5. 4: Sequencing of STAT5 binding site in the *lalba* promoter of mouse: Sanger sequencing exemplifying the presence of WT STAT5 binding site in the *lalba* promoter. The STAT5 dimer binding site is shown in the red box and the putative STAT5 binding site is the blue box with 6 bp spacer. The resulting sequence is aligned with a reference sequence retrieved from the Ensemble database.

Differentiation optimisation of HC11 cells

The HC11 cells, when differentiated with lactogenic hormones, activate the binding of STAT5 TF to their consensus motif in a gene's promoter (Morrison and Cutler 2009). To assess the effect of differentiation on the *lalba* gene and alpha-lactalbumin expression in HC11 cells, different combinations of lactogenic hormones were trialled and the resultant alpha-lactalbumin expression was assessed by western blotting.

HC11 cells were differentiated using combinations of insulin (Sigma Aldrich), epidermal growth factor and prolactin (Sigma Aldrich) for eight days. On day 0, 1×10^5 HC11 cells were plated in six-well plates and were cultured for the next three days at 37°C in RPMI medium. Culture medium was replaced on days 3 and 5 to include lactogenic hormones as shown in Figure 5. 5.

HC11 cells growing in well number 1 were cultured as control in RPMI media only. Cells growing in well number 2, 3, 4 and 5 were initially differentiated using RPMI media supplemented with insulin at different concentrations. Cells in well number 2 were grown in RPMI- insulin media for 8 days whereas culture media in well number 3, 4 and 5 was replaced with RPMI- insulin media supplemented with prolactin and epidermal growth factor at different concentrations on day 5th.

Similarly, HC11 cells growing in well number 6, 7, 8 and 9 were initially differentiated by adding RPMI media supplemented with prolactin on day 3rd. Cells in well number 6 were cultured in RPMI- prolactin media from day 3rd to day 8th. On day 5th culture media in well number 7 and 9 was replaced with RPMI- prolactin media supplemented with higher prolactin concentration and epidermal growth factor respectively. In well number 8, media was switched back to RPMI on day 5th of differentiation. Assessment of alpha-lactalbumin expression was carried out by western blotting as described in chapter 2 (Section 2.4).

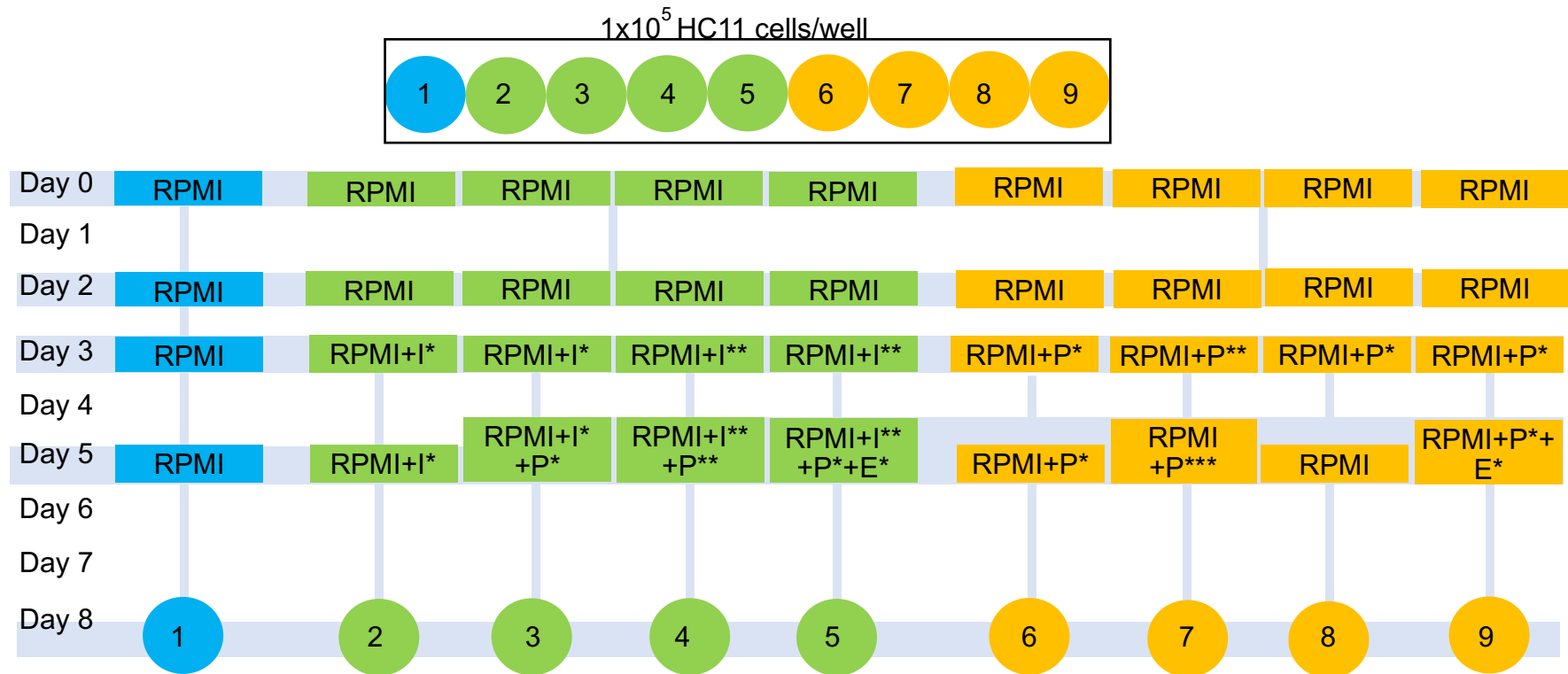


Figure 5. 5: Media composition used during the differentiation of HC11 cells: The undifferentiated cells (well number 1) were grown in RPMI for 8 days before harvesting. Cells in well numbers 2, 3, 4 and 5 were initially grown in RPMI medium, replaced with RPMI supplemented with insulin at different concentrations on the 3rd day. The culture media was again replaced on 5th day in well number 3 and 4 with RPMI + insulin and prolactin whereas well number 5 was replaced with media composed of RPMI + insulin, prolactin and epidermal growth factor. Similarly, in well numbers 6, 7, 8, and 9, media was replaced on the 3rd day of differentiation with RPMI supplemented with prolactin, and on the 5th day with different concentrations of prolactin and epidermal growth factor. Cells from all 9 well were harvested in the 8th day.

- I- Insulin
- P-Prolactin
- E- Epidermal growth factor
- * 5ng/ml
- ** 10ng/ml
- *** 15ng/ml

Quantitative analysis of alpha-lactalbumin and β casein in HC11 cells

Total proteins were isolated from the cells harvested from all nine wells (Figure 5.5) using RIPA lysis and extraction buffer (Invitrogen) as described in chapter 2 (Section 2.4) and quantified by Bradford assay. Western blotting was carried out using 10 μ g of each of the nine protein samples electrophoresed through a 4–20% Mini-PROTEAN Precast Protein Gel (Biorad) with a BLUEstain protein ladder (Abcam). The expression level of alpha-lactalbumin was analysed using an anti-alpha-lactalbumin antibody (SourceBio; catalogue number LS-C302325-100) whereas β casein expression was analysed using β casein antibody (Santacruz; catalogue number SC 166684) (Figure 5. 6 A,B)

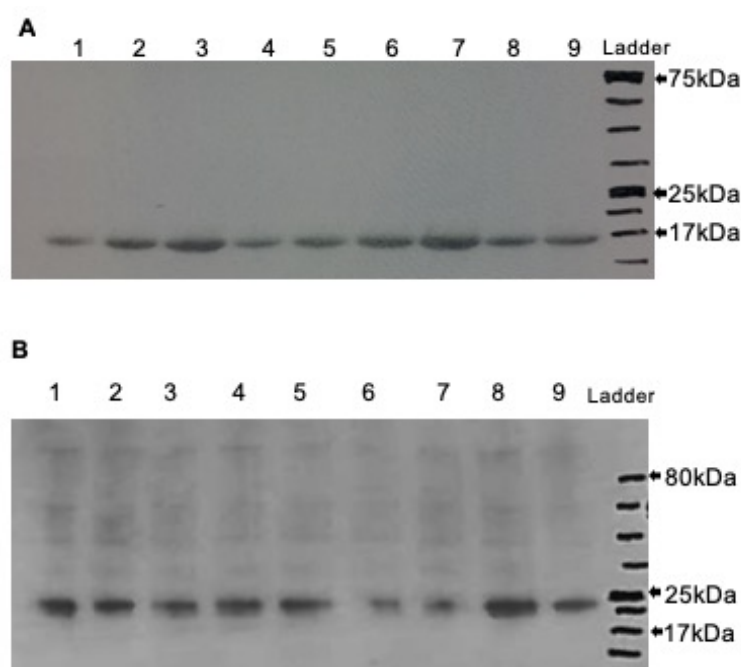
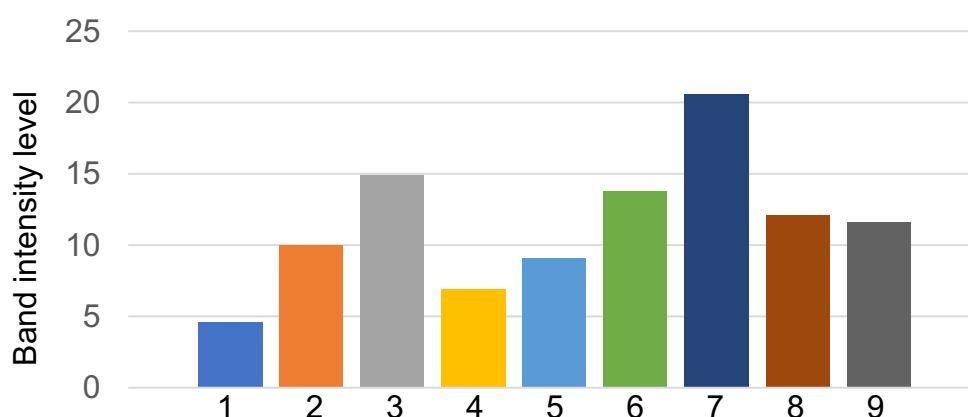


Figure 5. 6: Western blot analysis of mouse alpha-lactalbumin and β casein:

A. Expression of mouse alpha-lactalbumin: The expression of alpha-lactalbumin (17kDa) in differentiated and undifferentiated HC11 cells. The lane 1 showing alpha-lactalbumin expression in undifferentiated cells whereas lane 2-5 and 6-9 showing alpha-lactalbumin expression in HC11 cells primarily differentiated using Insulin and prolactin respectively. B. Expression of mouse β casein: The lane 1 showing β casein (24kDa) expression in undifferentiated cells whereas lane 2-5 and 6-9 showing β casein expression in HC11 cells primarily differentiated using Insulin and prolactin respectively.

The western blot image (figure 5.6 A,B) revealed varying levels of alpha-lactalbumin and β casein expression. However, only the alpha-lactalbumin band intensities were quantitated using the gel analysis software Fiji (<https://fiji.sc/>), to quantify the bands in gel image, all bands were boxed in one single box using the rectangular selection tool. The plot lanes were visualised using gel analyse tool resulted in generation of individual peaks representing intensity of individual bands (Graph 5. 1).



Graph 5. 1: Quantitation of alpha-lactalbumin bands detected by western blot: The measurement of bands using Fiji software acquired from undifferentiated and differentiated HC11 cells. The lane 1 is showing alpha-lactalbumin expression level in undifferentiated HC11 cells. Lane number 2-5 and 6-9 showing alpha-lactalbumin expression level in HC11 cells primarily differentiated using Insulin and prolactin respectively where lane number 7 is showing the maximum expression level.

The analysis of western blot bands using Fiji software revealed the lowest expression of alpha-lactalbumin in undifferentiated HC11 cells (sample number 1) whereas sample number 7, differentiated with a high concentration of prolactin (15 ng/ml), showed the highest *lalba* expression. Based on these results all future HC11 cell differentiation was carried out using RPMI medium supplemented with 15 ng/ml prolactin.

Creating a PCR template of WT *lalba* promoter

WT *lalba* promoter, amplified using HC11 cell DNA and verified by Sanger sequencing (Figure 5. 4) was cloned in a zero blunt vector (Invitrogen) as described in chapter 2 (Section 2.1.2), in order to use it as the template to

amplify the LacP and to create promoters with altered STAT5 binding site. The WT promoter cloned in a zero blunt vector, referred as zbLac, was analysed for successful cloning by restriction digestion with *EcoRI* at 37°C for one hour. The reaction was resolved in a 1.2% agarose gel and run with DNA ladder mix (Invitrogen) resulted in the separation of a 1544bp WT promoter fragment (Figure 5. 7).

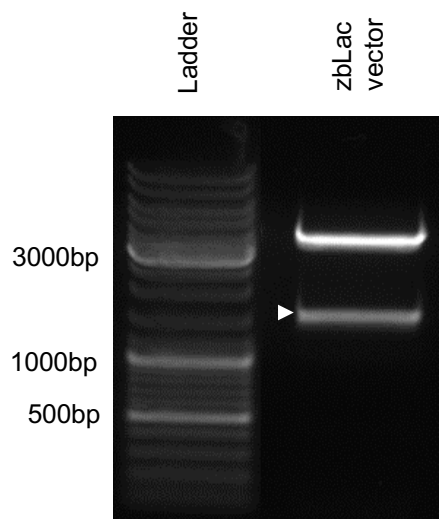


Figure 5. 7: Gel image of digested zbLac vector: Restriction digestion of zero blunt plasmid containing mouse *lalba* WT promoter insert using *EcoRI* showing 1544 bp *lalba* promoter fragment (marked) and 3519 bp digested zero blunt vector.

The zbLac containing WT mouse *lalba* promoter was now used as a template to generate WT and altered promoters namely LacP, LacT and LacD as shown in figure 5.8.

5.3.3 Creating mouse *lalba* promoters with altered STAT5 binding site using fusion PCR for reporter assay

The mouse WT *lalba* promoter editing was carried out using fusion PCR to introduce changes to generate a tetramer binding site (LacT) and another promoter with no STAT5 binding site (LacD) to check the effect of STAT5 binding site removal on the *lalba* promoter's expression. The activity of both altered promoters (LacT and LacD) were compared with mouse WT *lalba* promoter (LacP). All three promoters, LacT, LacD and LacP were amplified using a modified primer pair 5.1F- 5.1R (appendix 7.1) each containing a

restriction site overhangs for *HindIII* and *XmaI* enzymes respectively. These overhangs were specifically chosen to facilitate cloning of all promoter fragments, amplified using these primers, into a firefly luciferase expressing vector *pGL3B* (acquired from Dr Robert Dalziel group, The Roslin Institute). The analysis of all three promoters cloned in the *pGL3B* plasmid was carried out in differentiated and undifferentiated HC11 cells. The creation of WT mouse *alba* promoter with restriction site overhangs and generation of altered mouse *alba* promoter LacT and LacD is shown in figure 5.8.

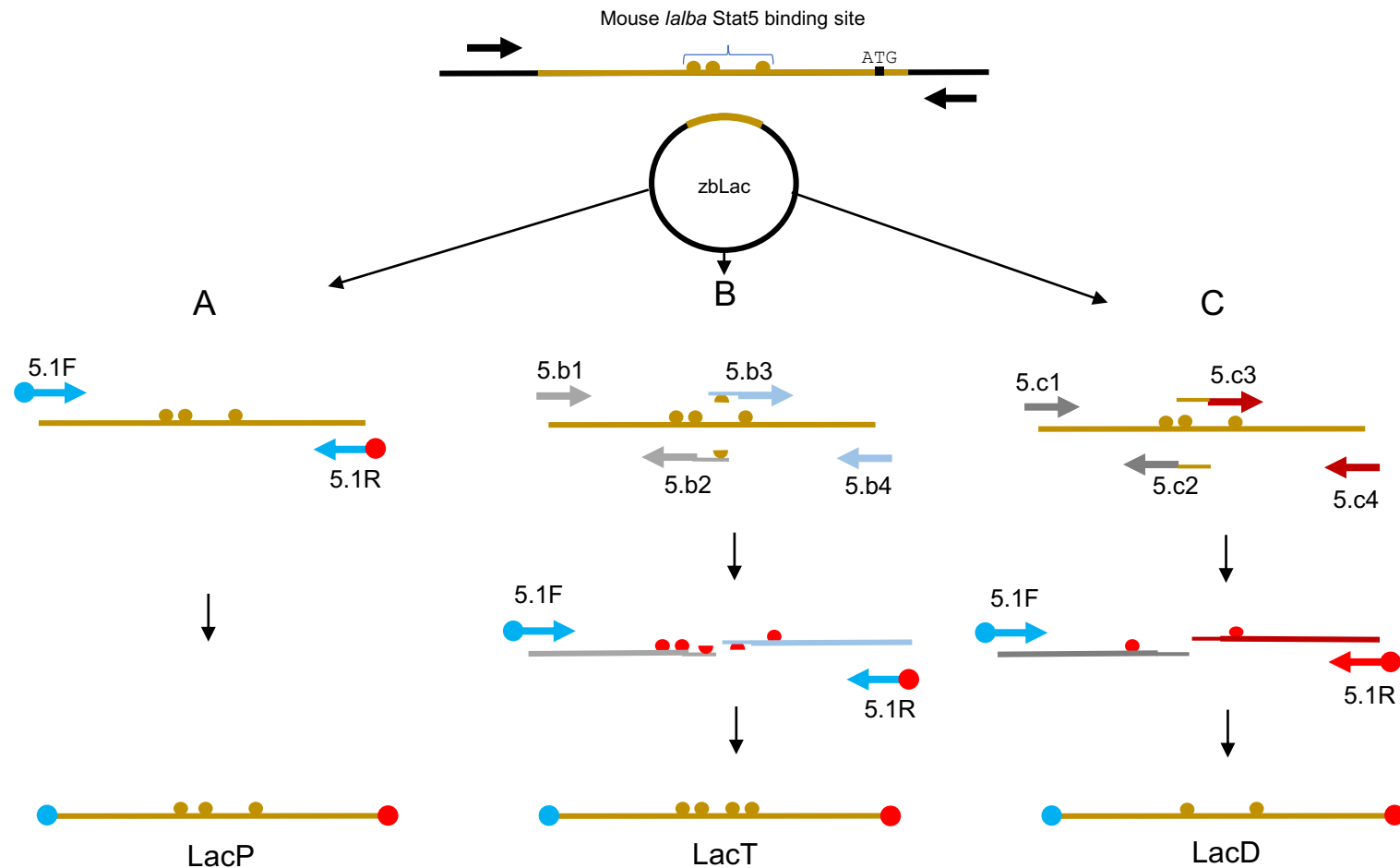


Figure 5. 8 : Creating the WT and altered mouse *lalba* promoters with *HindIII* and *XmaI* restriction overhangs: A. Amplifying WT *lalba* promoter (LacP) using zblac as the PCR template with primers 5.1F-5.1R with *HindIII* and *XmaI* restriction overhangs. B. Fusion PCR to create mouse *lalba* promoter with a tetramer STAT5 binding site (LacT) using primers 5.b1-5.b2 and 5.b3- 5.b4 and creating restriction overhangs using primers 5.1F and 5.1R. C. Fusion PCR to create mouse *lalba* promoter with no STAT5 binding site (LacP) using primers 5.c1-5.c2 and 5.c3- 5.c4 and creating restriction overhangs using primers 5.1F and 5.1R.

Amplifying the LacP promoter

To amplify the LacP promoter containing restriction overhangs, a PCR was carried out using 100ng zbLac as DNA template with primer pair 5.1F-5.1R at 20s extension time for 35 cycles using the standard PCR conditions as shown in chapter 2 (Section 2.2.1). PCR reaction was electrophoresed through a 1.2% agarose gel with a 1kb DNA ladder mix (Invitrogen) produced a predicted ~1544bp fragment (Figure 5. 9).

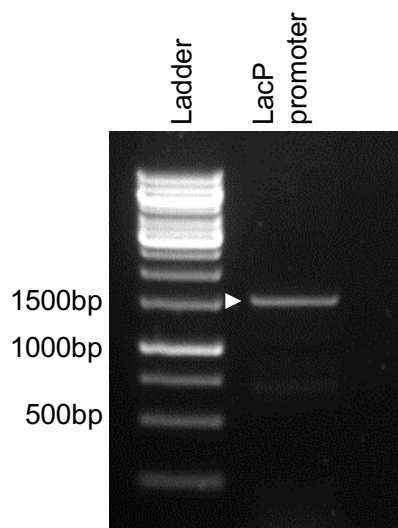


Figure 5. 9: PCR to generate the LacP promoter fragment: PCR of mouse LacP promoter (marked) with primer 5.1F and 5.1R using zbLac as the template DNA.

The marked fragment was excised from agarose gel, and DNA was extracted using the QIAquick gel extraction kit (Qiagen). The LacP promoter was then cloned in a zero blunt vector (Invitrogen) as described in Chapter 2 (Section 2.1.2). The cloned LacP was now referred to as zbP and stored at -20°C.

Creating the LacT promoter with STAT5 tetramer binding site

The LacT promoter fragment was created as shown in figure 5.10A. The first round of PCR was carried out using two primer pairs 5.b1- 5.b2 and 5.b3- 5.b4 (appendix 1), where primer Pb2 and Pb3 contain complimentary overhangs with a STAT5 tetramer binding site. A total of two PCRs were carried out using each of the primer pair and 100ng zbLac as template DNA at 20s extension time for 35 cycles using the standard PCR conditions described in chapter 2 (Section 2.2.1). The PCRs were resolved in a 1.2% agarose gel alongside DNA ladder mix (Invitrogen) and showed ~1200bp and ~700bp fragments (Figure 5.10B) larger than expected sized 1050bp and 650bp fragments, supposedly resulted from using PCR primers 5.b2 and 5.b3 with large overhangs. The modification of WT STAT5 binding site into a tetramer binding site is shown in figure 5.10C.

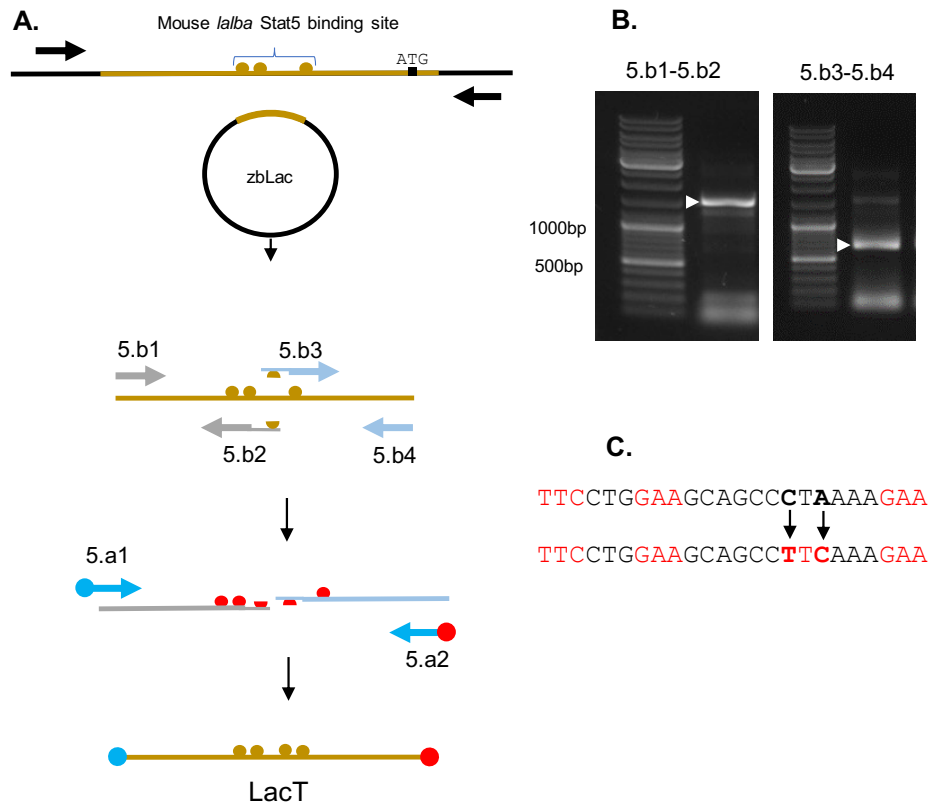


Figure 5. 10: Creating the LacT promoter fragments: A. The schematics of generation of LacT promoter using zbLac as template DNA and fusion PCR using primers 5.b1- 5.b2 and 5.b3 -5.b4 and further incorporating *HindIII* and *XmaI* restriction site using primers 5.1F and 5.1R. B. PCR was resolved in agarose gel is showing ~1200bp and ~700bp fragments (marked) generated by 5.b1-5.b2 and 5.b3- 5.b4 primers with altered STAT5 tetramer binding site overhangs. C. The conversion into a STAT5 tetramer binding site.

The PCR fragments marked in figure 5.10B were excised, and DNA was isolated using the QIAquick gel extraction kit (Qiagen). The additional unexpected fragments were not analysed. A fusion PCR was carried out using 100ng of each of the DNA fragments as the template and the primer pair 5.1F and 5.1R at 30s extension time for 35 cycles using the standard PCR reaction shown in chapter 2 (Section 2.2.1). The PCR was electrophoresed through a 1.2% agarose gel with a DNA ladder mix (Invitrogen) showed a 1544bp fragment (Figure 5. 11).

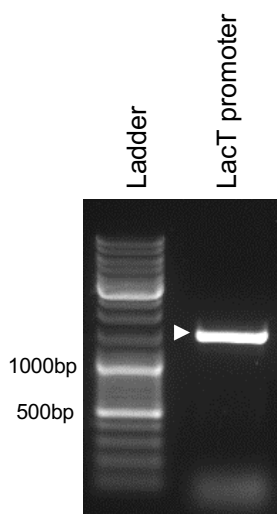


Figure 5. 11: Gel image of Fusion PCR creating the LacT promoter fragment: The PCR to amplify LacT promoter was carried out using 5.b1-5.b2 and 5.b3-5.b4 amplified DNA and 5.1F-5.1R primer pair resulted in the generation of a 1544 bp fragment (marked).

The marked fragment was excised from the agarose gel, and DNA was isolated using the QIAquick gel extraction kit (Qiagen). The isolated DNA fragment was cloned in a zero blunt vector (Invitrogen) as described in Chapter 2 (Section 2.1.2). The cloned fragment, referred as zbT was stored at -20°C. The same method was used to create another altered promoter with no STAT5 binding site (LacD)

Creating LacD promoter with no STAT5 binding site

The LacD promoter without a STAT5 binding site was generated by fusion PCR as shown in (Figure 5.12A). A total of two primer pairs namely 5.c1- 5.c2 and 5.c3- 5.c4 were designed where primers 5.c2 and 5.c3 contained complimentary overhangs with no STAT5 binding site. The conversion of GAA into CTG was chosen to disrupt the 9 bp STAT5 dimer binding motif palindrome which is a prerequisite for a STAT5 TF binding (5.12C)

Initially two PCRs were carried out using zbLac as template DNA and 5.c1- 5.c2 and 5.c3- 5.c4 primer pairs at 20s extension time for 35 cycles using the standard PCR reaction shown in chapter 2 (Section 2.2.1). The PCR was run in a 1.2% agarose gel with a DNA ladder mix (Invitrogen) showed 1250 bp and ~800 bp fragments (5.12B) larger than expected 1110 bp and 690 bp fragments due to primers 5.c2 and 5.c3 containing large overhangs.

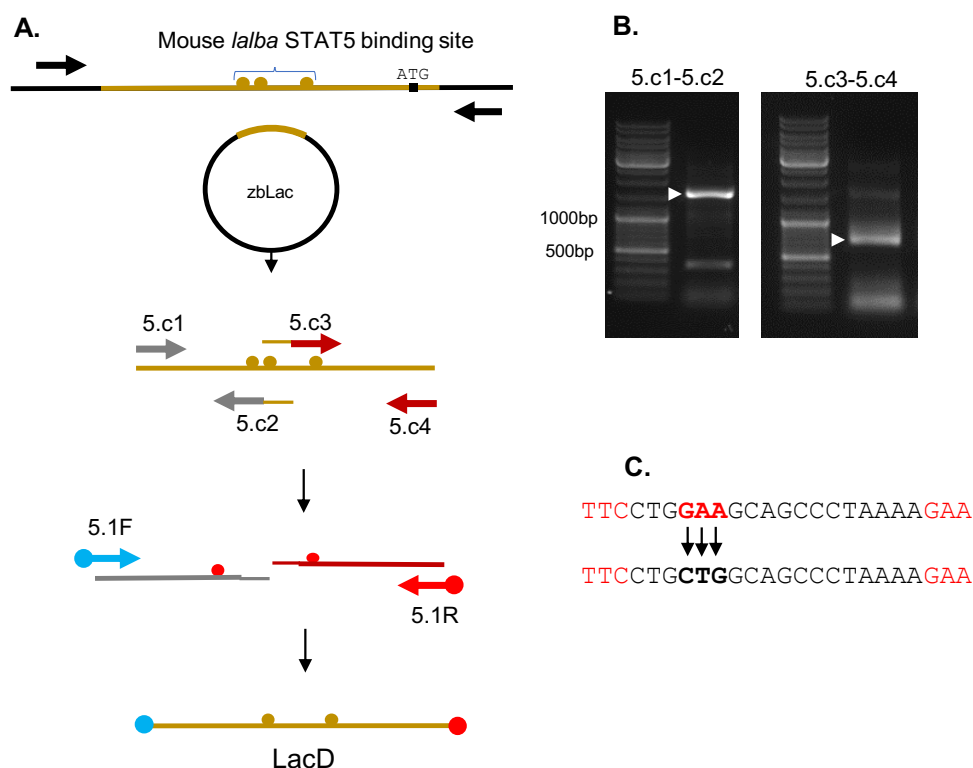


Figure 5. 12: Creating the LacD promoter fragments. A. The schematics of generation of LacD promoter using zbLac as template DNA and fusion PCR using primers 5.c1- 5.c2 and 5.c3- 5.c4 and incorporating *HindIII* and *XmaI* restriction site using primers 5.1F and 5.1R. B. PCR showing ~1250 bp and ~800 bp fragments (marked) using 5.c1- 5.c2 and 5.c3-5.c4 primers with altered STAT5 binding site overhangs. C. The disruption of WT STAT5 binding site and generation of a LacD site.

The marked PCR fragments were excised from the agarose gel, and DNA was isolated using the QIAquick gel extraction kit (Qiagen). A fusion PCR was carried out using both DNA as the template DNA and the primer pair 5.1F and 5.1R with 30s extension time for 35 cycles using the standard PCR reaction shown in chapter 2 (Section 2.2.1). The PCR was run in a 1.2% agarose gel with a DNA ladder mix (Invitrogen) showed 1544 bp product (Figure 5. 13).

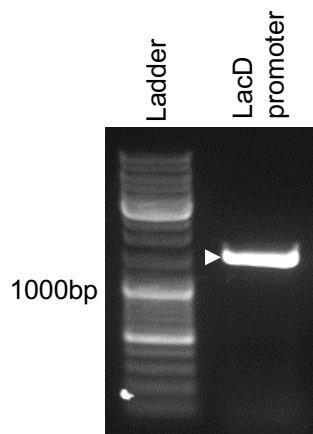


Figure 5. 13: Gel image of Fusion PCR creating the LacD promoter fragment: The PCR to amplify LacD promoter was carried out using 5.c1-5.c2 and 5.c3-5.c4 amplified DNA and 5.1F- 5.1R primer pair resulted in the generation of a 1544 bp fragment (marked).

The marked fragment was excised from the gel and DNA was isolated from agarose gel using QIAquick gel extraction kit (Qiagen). The isolated DNA was cloned in a zero blunt vector (zbD) as described in chapter 2 (Section 2.1.2) and stored at -20°C.

The restriction digestion of zbP, zbT and zbD vectors

The LacP, LacT and LacD promoter fragments cloned in a zero blunt vector (zbP, zbT and zbD) were extracted by restriction digestion of the 1 µg plasmid with *HindIII* (New England Biolabs) and *XmaI* (New England Biolabs) for 1 hour at 37°C. The restriction digest was resolved in a 1.2% agarose gel and run with a DNA ladder mix (Invitrogen) (Figure 5. 14).

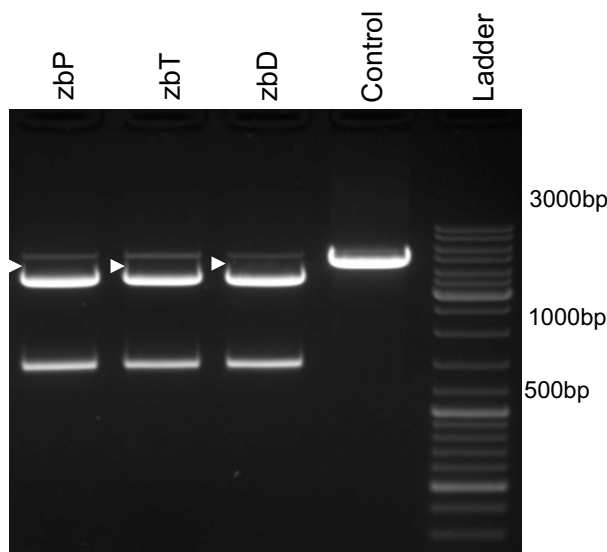


Figure 5. 14: Gel image of digested zbP, zbT and zbD vectors: Restriction digestion of LacP, LacT and LacD plasmids from zpP, zbT and zbD vectors using *HindIII* and *XmaI* showing 1544 bp promoter fragments (marked) and digested zero blunt vector (3519 bp). The control is an undigested cloned zero blunt plasmid.

All three fragments, now referred as LacP, LacT and LacD were excised from agarose gel, and DNA was isolated using the QIAquick gel extraction kit (Qiagen).

Restriction digestion of *pGL3B* plasmid and ligation of LacP, LacT and LacD

1 μ g *pGL3B* plasmid (appendix 7.4B) was digested using *HindIII* (New England Biolabs) and *XmaI* (New England Biolabs) at 37°C for one hour and the reaction was resolved in a 1.2% agarose gel (Figure 5.15).

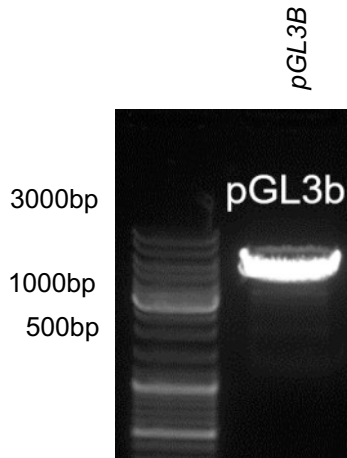


Figure 5. 15: Gel image of digested *pGL3B* plasmid: Restriction digestion of *pGL3B* plasmid (4798 bp) using *HindIII* and *XmaI* enzymes. The digested 20 bp fragment is not visible in the gel image.

The digested *pGL3B* vector was isolated from agarose gel using the QIAquick gel extraction kit (Qiagen) and quantified using nanodrop spectrophotometer. The ligation of LacP, LacT and LacD into digested *pGL3B* vector was carried out using quick ligase buffer (New England Biolabs) as described in chapter 2 (Section 2.1.7) and the transformation of ligated vector into DH5α competent cells was carried out as described in chapter 2 (Section 2.1.2). On the next day, single bacterial clones were inoculated in 5ml LB media supplemented with carbenicillin and incubated in a shaking incubator at 37°C, 200RPM overnight. The plasmid isolation was carried out using the QIAprep spin miniprep kit (Qiagen) as described in chapter 2 (Section 2.1.3). The reaction was confirmed by digesting all plasmids using *HindIII* (New England Biolabs) and *XmaI* (New England Biolabs) restriction enzymes at 37°C for one hour; the digest was resolved in 1.2% agarose gel and run with a DNA ladder mix (Invitrogen) (Figure 5. 16).

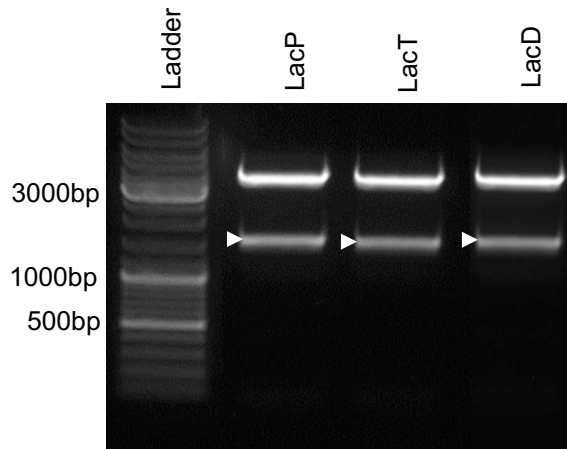


Figure 5. 16: Gel image of digested LacP, LacT and LacD vectors: Restriction digestion of LacP, LacT and LacD vectors (4797 bp) using *HindIII* and *XmaI* restriction enzyme. The digested promoter fragments are marked with white arrows.

Sequencing LacP, LacT and LacD vectors

The promoter fragments digested from LacP, LacT and LacD vectors were excised from agarose gel and DNA was isolated using the QIAquick gel extraction kit (Qiagen). To confirm the WT and altered STAT5 binding site in promoter fragments of LacP, LacT and LacD plasmids, the Sanger sequencing (Edinburgh genomics) was carried out using fragment DNA and a primer pair flanking the WT and altered STAT5 binding site before transfection in HC11 cells for luciferase assay (Figure 5. 17).

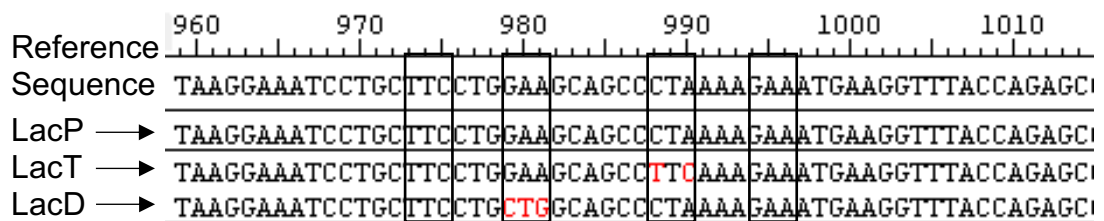


Figure 5. 17: Sequencing of STAT5 binding site in LacP, LacT and LacD vectors: Sanger sequencing showing LacP vector containing a WT STAT5 binding site, The LacT comprising a STAT5 tetramer binding site and LacD with the disrupted STAT5 binding site. The sequencing results are aligned with the WT reference sequence.

The successful cloning of the WT vector LacP (7.4C) and altered LacT and LacD (appendix 7.4D and E) vectors was confirmed using Sanger sequencing using a primer pair flanking the STAT5 binding site in the mouse *lalba* promoter. Prior to the transfection of these promoters in HC11 cells for

luciferase assay, a transfection optimisation with *pGL3B* plasmid was carried out to assess the best transfection condition in HC11 cells using Lipofectamine 3000 (Invitrogen).

Transfection optimisation of HC11 cells

The transfection conditions of HC11 cells were optimised using Lipofectamine 3000 (Invitrogen) as described in chapter 2 (Section 2.3.6). The reaction mixtures were prepared as shown in Table 5. 4 transfecting 2 µg of the plasmid *pGL3B* with two different Lipofectamine reagent volumes, 3.75 µl and 7.5 µl, each containing 4 µl P3000 reagent diluted in Opti-mem to a final volume of 250 µl.

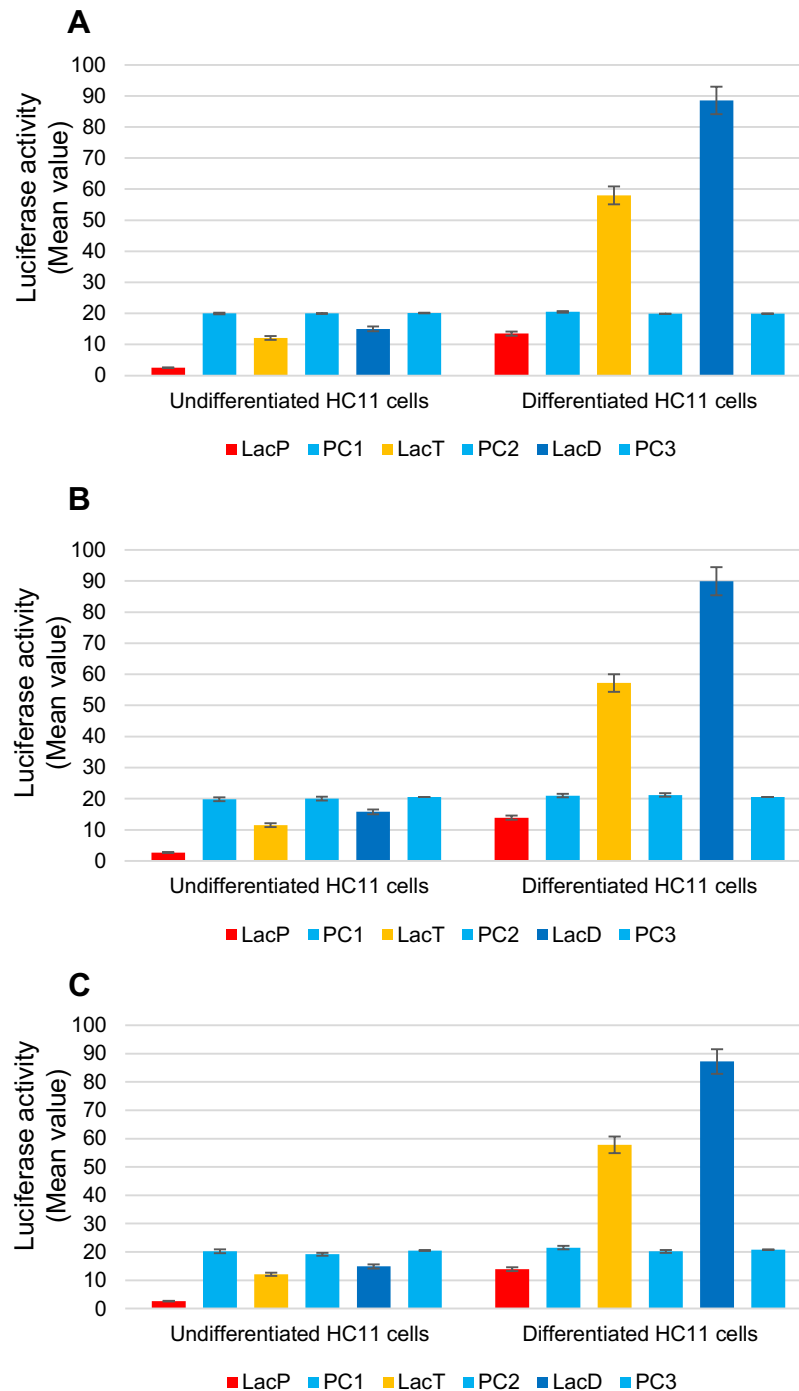
The reaction mixture was added to 5×10^5 HC11 cells growing a six-well plate. The transfected cells were visually observed using a florescent microscope at 24, 36 and 48 hours post transfection. The highest number of GFP positive cells (~30%) with minimum cell death were visually observed at 36h post-transfection with 7.5 µl Lipofectamine 3000. All subsequent HC11 cell transfections, unless otherwise stated, were carried out using 7.5 µl of Lipofectamine 3000 reagent.

Lipofectamine reagent	Plasmid (<i>pGL3B</i>)	P3000 reagent	Opti-mem medium	GFP positive cells (%)			Viable cells (%)		
				24h	36h	48h	24h	36h	48h
3.75 µl	2 µg	4 µl	240.25 µl	~10%	~20%	~20%	~100%	~95%	~95%
7.5 µl	2 µg	4 µl	236.25 µl	~10%	~30%	~30%	~100%	~95%	~95%

Table 5. 4: Transfection optimisation of HC11 cells: HC11 cells were transfected with two different volumes of Lipofectamine 3000 reagent. The cells were visually observed at 24, 36 and 48 hours post transfection to view the presence of GFP positive and viable cells. The maximum GFP positive cells with utmost viable cells were obtained with 7.5 µl Lipofectamine 3000 reagent at 36 and 48h post transfection.

5.3.4 Assessing expression from the altered *lalba* promoter in HC11 cells using luciferase assay

The expression level assessment of all three promoters (LacP, LacT and LacD cloned in the *pGL3B* plasmid) was carried out in HC11 cells using dual luciferase assay. On day one, 4×10^5 HC11 cells were plated in each well of 3x six-well plates and maintained for the next 72 hours. The undifferentiated HC11 cells were grown in RPMI media supplemented with 10% FBS and PenStrep. Differentiation was carried out with RPMI media + 10% FBS and PenStrep enriched with 15 ng/ml prolactin. On the 4th day, transfection was carried out with 2 μ g of each vector (LacP, LacT and LacD) along with 100ng of a transfection control vector, the renilla luciferase expressing *pRL-TK* plasmid using Lipofectamine 3000 (Invitrogen) according to the pre-optimised method. A negative control transfection was carried out with no plasmid using Lipofectamine 3000 (Invitrogen). Cells were harvested 24 hours post-transfection and the activity of firefly and renilla luciferases was checked sequentially using a luminometer as described in chapter 2 (Section 2.2.7) using the Dual Reporter Luciferase assay kit (Promega). The transfections of LacP, LacT and LacD was carried out a total of three times in differentiated and undifferentiated HC11 cells (Graph 5.2A, B, C).



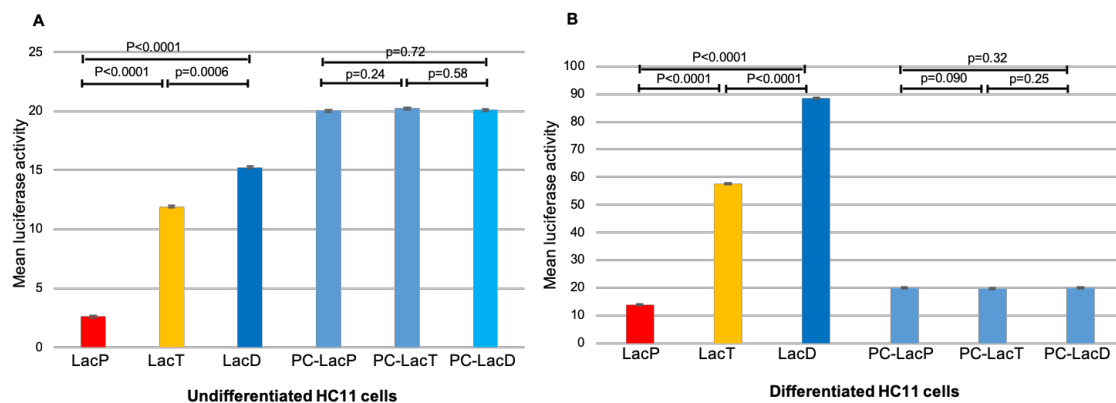
Graph 5. 2: Luciferase assays of undifferentiated and differentiated HC11 cells transfected with LacP, LacT and LacD vectors: A total of three luciferase assays, carried out at different time points (A, B and C), showing the mean value of luciferase activity in differentiated and undifferentiated HC11 cells with LacP, LacT and LacD plasmids with their positive control (PC) mean value shown beside. All three plasmids are showing a considerable high expression level in HC11 cells differentiated using prolactin. The mean value of luciferase activity with all plasmids is shown with standard error bars.

In all three luciferase assays, the promoter containing a tetramer binding site (LacT) showed a similar ~3 fold higher expression as compared to WT promoter (LacP) in both differentiated and undifferentiated cells. It had been anticipated that the promoter with no STAT5 binding site (LacD) would show significantly reduced activity compared to either the WT LacP or the tetramer LacT. However, the luciferase assay of LacD showed five-fold higher expression as compared to LacP in and 0.5 fold higher expression than LacT in undifferentiated cells, whereas in differentiated cells, LacD showed about nine-fold higher expression than LacP and two-fold higher than LacT. The expression of positive control plasmid *pRL-tk* was constant in differentiated and undifferentiated HC11 cells in all three assays.

To further analyse these results shown in graph 5.2A, B and C, an analysis of variation (ANOVA) was carried out using an inbuilt data analysis tool in the Microsoft office excel spreadsheet. Graph 5.3A represents the mean luciferase activity and indicating the level of significance (p value) between LacP, LacT and LacD plasmids and their positive controls in undifferentiated HC11 cells. The p values between LacP-LacD, LacP-LacT and LacT-LacD were found to be $p < 0.0001$, $p < 0.0001$ and $p = 0.0006$ respectively which is much lower than the significance level of $p = 0.05$ showing a highly significant difference in their expression level whereas their F statistical value was found to be greater than F critical value. On the contrary, the p values for positive controls of LacP-LacD, LacP-LacT and LacT-LacD were $p = 0.72$, $p = 0.24$ and $p = 0.56$ respectively showing insignificant difference in expression of control plasmid renilla with their F value lower than F critical value.

Similarly, the mean luciferase activity of LacP, LacT and LacD in differentiated HC11 cells showed the p values between LacP-LacD, LacP-LacT and LacT-LacD as low as $p < 0.0001$ in all the three comparisons exhibiting a highly significant difference in their luciferase activity with their F statistical value greater than F critical value hence the null hypothesis was rejected whereas their positive controls showed p values $p = 0.32$, $p = 0.080$ and $p = 0.25$ for LacP-LacD, LacP-LacT and LacT-LacD respectively. This data shows that the presence of different STAT5 binding sites in Lac promoters

resulted in significant differences in luciferase activity of the plasmids where as their renilla expressing positive controls without the Lac promoter and STAT5 binding site showed a constant luciferase activity.

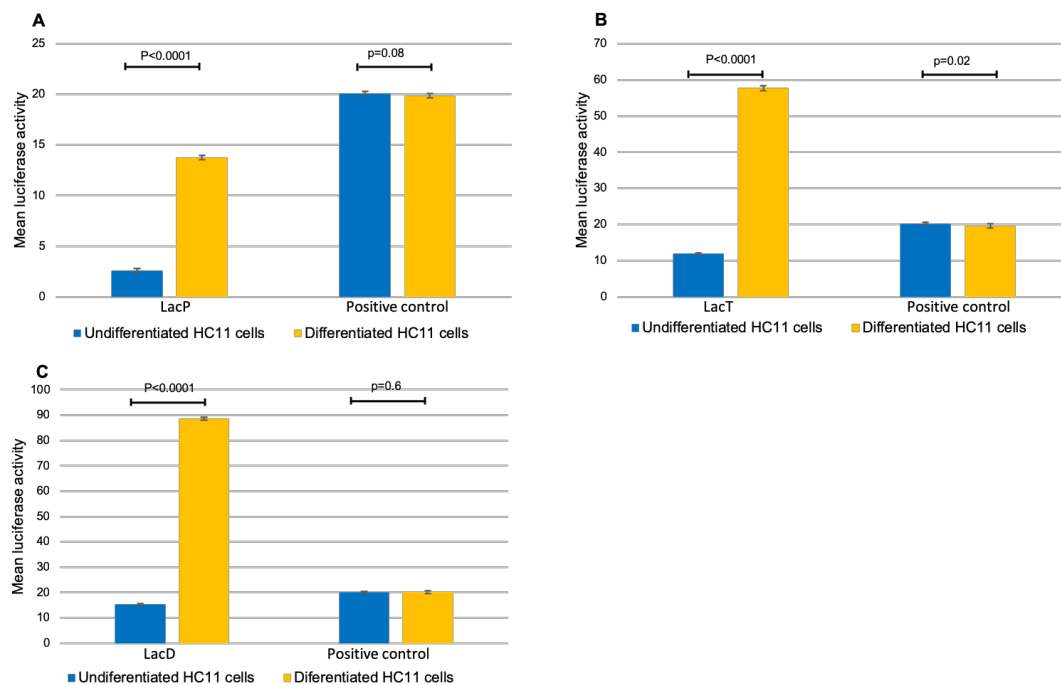


Graph 5. 3: A comparison in the mean luciferase activity and level of significance between Lac plasmids and their controls in undifferentiated and differentiated HC11 cells: A. In undifferentiated HC11 cells the p value between LacP-LacD, LacP-LacT and LacT-LacD showing a highly significant results $p < 0.0001$, $p < 0.0001$ and $p < 0.0006$ respectively whereas their controls shows insignificant $p > 0.05$ p values in all three comparisons B. Similarly, in differentiated HC11 cells the p value between LacP-LacD, LacP-LacT and LacT-LacD showing a highly significant $p < 0.0001$ differences in their expression whereas their controls show insignificant $p > 0.05$ p values. The mean luciferase activity of all plasmids and controls is shown with standard error bars.

To analyse the effect of HC11 differentiation on luciferase activity of LacP, LacT and LacD plasmids. The mean luciferase activity of all plasmids and their positive controls were analysed individually by ANOVA using an inbuilt data analysis tool in the Microsoft office excel spreadsheet. Graph 5.4A, B and C shows a comparison in the mean luciferase activity between LacP, LacT and LacD and their positive controls in undifferentiated and differentiated HC11 cells.

The difference in the p values between all three plasmids in undifferentiated and differentiated HC11 cells was found to be highly significant i.e. $p < 0.0001$ whereas their positive control showed constant renilla expression with insignificant ($p > 0.05$) p values. This analysis depicts that the

addition of prolactin in HC11 cells resulted in their differentiation and increased the luciferase activity of all three plasmids as compared to their activity in undifferentiated HC11 cells.



Graph 5. 4: A comparison in the mean luciferase activity and the p values of LacP, LacT and LacD and their positive controls in undifferentiated and differentiated HC11 cells: The p value of the luciferase activity of LacP, LacT and LacD in undifferentiated and differentiated HC11 cells showing a highly significant ($p<0.0001$) values for all plasmids whereas their positive controls are showing similar level of mean luciferase expression levels with $p=0.08$, $p=0.02$ and $p=0.6$ respectively. The mean luciferase activity of all plasmids and controls is shown with standard error bars.

The overall results acquired in graphs 5.2, 5.3 and 5.4 shows that the luciferase activity of Lac promoter is affected by presence/ absence and oligomerisation of STAT5 binding site (dimer/ tetramer) and the luciferase expression level for all three plasmids (LacP, LacT and LacD) increases when transfected in HC11 cells differentiated using prolactin, never the less, the plasmid containing no STAT5 binding site (LacD) showed the maximum luciferase expression level in differentiated HC11 cells.

5.3.5 Modifying the WT STAT5 binding site of mouse *lalba* promoter in HC11 cells using CRISPR/Cas9 and homology-directed repair (HDR)

In the previous Section, we altered the STAT5 binding site in the *lalba* promoter, and its assessed expression using a reporter assay in mouse mammary epithelial cells. We observed altered expression levels that warranted further investigation. In order to modify the STAT5 binding site of *lalba* in the DNA of the HC11 cells themselves we decided to utilise a CRISPR/Cas9 editing and homology directed repair (HDR) approach

To target the WT STAT5 binding site in HC11 cells, a total of three sgRNAs (Lac1, Lac2 and Lac3) were designed as described in chapter 2 (Section 2.5.1) (Figure 5. 18: C). The guides Lac1 and Lac3 were designed to create a STAT5 tetramer binding site in HC11 cells *lalba* promoter using an HDR template transfected with these CRISPRs separately. Similarly, CRISPR number Lac1 and Lac2 were to create the HC11 cells *lalba* promoter with no STAT5 binding site, LacD, by co-transfection of these CRISPRs with a HDR template.

```

GTTAAGGAAATCCTGCTTCCTGGAAGCAGCCCTA*AAAGAAATGAAGGTTTACCAGAG
                               Lac1
GTTAAGGAAATCCTGCTTCCTGGAAGCAGCCCTA*AAAGAAATGAAGGTTTACCAGAG
                               Lac2
GTTAAGGAAATCCTGCTTCCTGGAAGCAGCCCTA*AAAGAAATGAAGGTTTACCAGAG
                               Lac3

```

Figure 5. 18: CRISPRs targeting location on WT STAT5 binding site in mouse *lalba* promoter: The WT STAT5 binding site in mouse *lalba* promoter is shown in red. The targeting location of CRISPR number Lac1, Lac2 and Lac3, (underlined in black) with their PAM sites (underlined in red) and their predicted cut site is indicated by asterisk sign.

To construct CRISPR plasmids, the sense and antisense oligos were annealed and cloned into the *pSL66* plasmid followed by transformation in top10 cells (Invitrogen). A colony PCR was performed on four transformed bacterial clones from each of the 3 CRISPRs. Figure 5. 19: C exemplifies a colony PCR of four bacterial clones for Lac1.

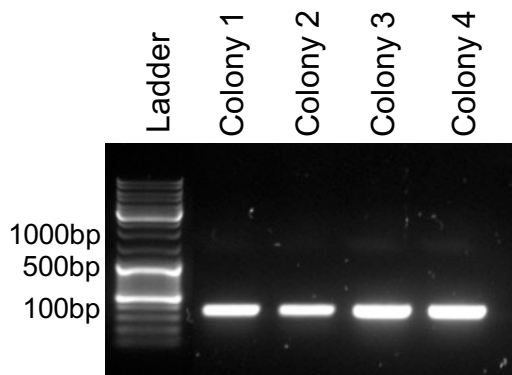


Figure 5. 19: Colony PCR of four bacterial clones containing Lac1 CRISPR oligo cloned in the *pSL66* plasmid: The PCR was carried out using Lac1 sense oligo as a forward primer and a reverse primer *oSL35* which binds to the *pSL66* sequence downstream of oligo insert. The presence of ~400bp fragment confirms the successful cloning of CRISPR oligo duplex in the *pSL66* plasmid.

The cloning was confirmed for CRISPRs Lac1, Lac2 and Lac3 using the similar colony PCR approach. Plasmids were isolated from all bacterial clones using the QIAquick plasmid miniprep kit (Qiagen) and the presence of CRISPR guides was confirmed using the Sanger sequencing (Edinburgh genomics) A representation of CRISPRs guides cloned in *pSL66* plasmid is shown in figure 5.20.

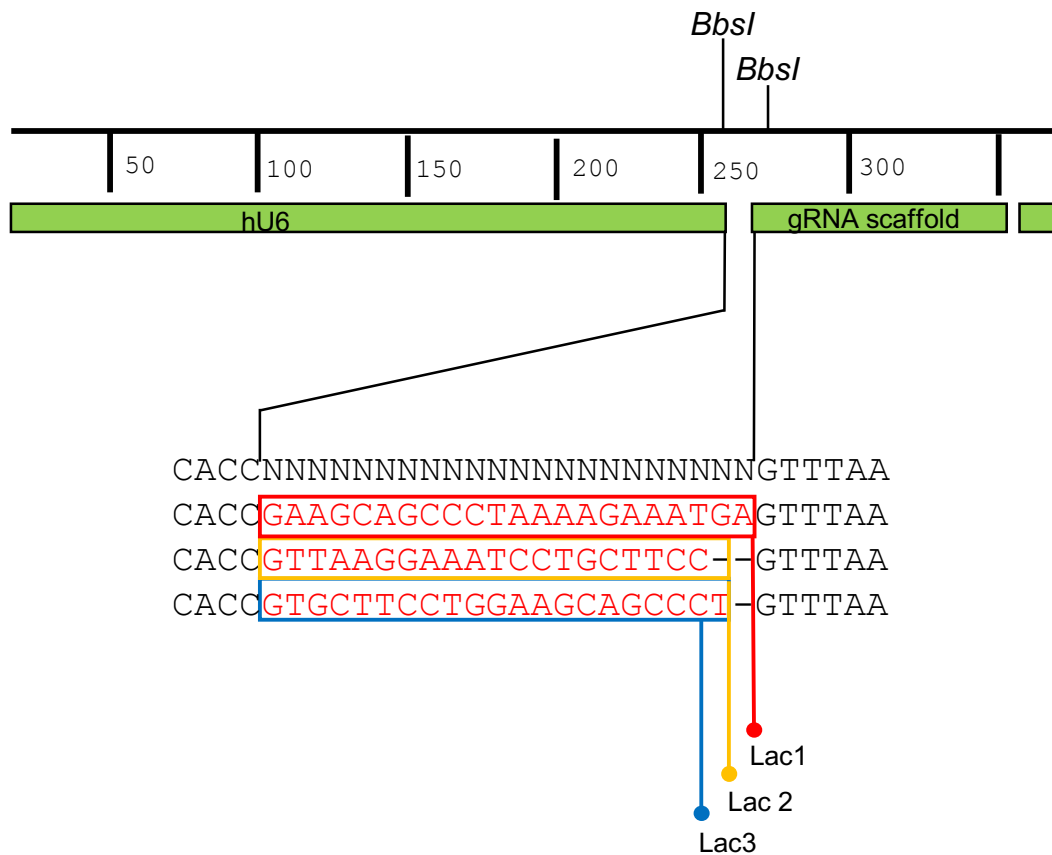


Figure 5. 20: CRISPRs number Lac1, Lac2 and Lac3 cloned in the *pSL66* plasmid: Image representing 3 different oligo pairs of CRISPR Lac1, Lac2 and Lac3 cloned at the *BbsI* restriction sites between the U6 promoter and gRNA scaffold in the *pSL66* plasmid.

After confirmation with Sanger sequencing, all CRISPR plasmids were isolated using the endotoxin-free maxiprep kit (Qiagen) as described in chapter 2 (Section 2.1.4) and stored at -20°C. All CRISPRs (Lac1, Lac2 and Lac3) were transfected into HC11 cells using Lipofectamine3000 (Invitrogen) with pre-optimised transfection conditions in this chapter resulted in visualisation of ~30% GFP positive cells at 36h post transfection. DNA was isolated from the transfected HC11 cells using the Quick Extract DNA extraction solution (Epicentre) and PCR was carried out using a primer pair flanking the targeted region (primer pair 5.2) at 20s extension time and 35 cycles using the standard PCR conditions shown in chapter 2 (Section 2.2.1). The NHEJ induction and

cutting of target site by CRISPRs were analysed using T7 endonuclease assay as described in chapter 2 (Section 2.5.5) (Figure 5.21).

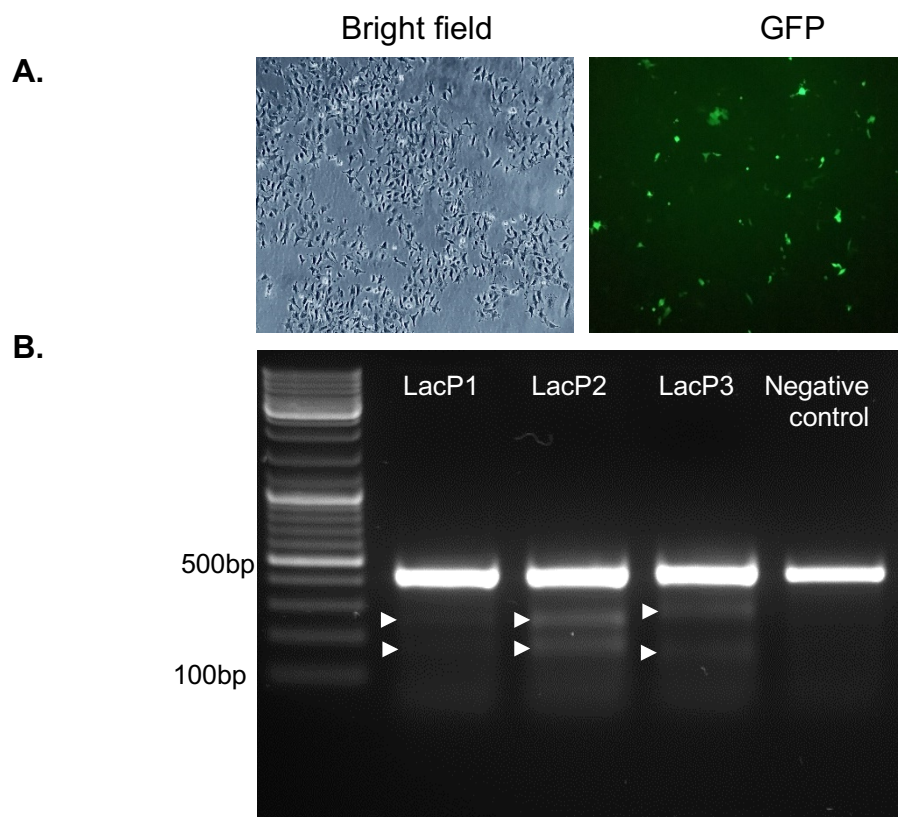


Figure 5. 21: Transfection of CRISPRs targeting the WT STAT5 binding site in mouse *lalba* promoter and assessment of NHEJ induction: A. Images of bright field and GFP positive HC11 cells transfected with CRISPRs 36h post transfection B. T7 endonuclease assay showing a WT fragment of 422 bp and endonuclease-digested fragments for LacP1 (250 bp, 172 bp), LacP2 (240, 182 bp) and LacP3 (280, 142 bp). WT DNA showing no cutting.

TIDE analysis of CRISPR targeting the WT STAT5 binding site

TIDE (<https://tide.deskgen.com/>, Tracking of Indels by Decomposition) is an online tool for quantitative analysis of insertion and deletion (indels) created using CRISPRs by comparing experimental and control sample chromatograms.

A 25 μ l PCR was carried out using DNA from all three CRISPRs transfected cell pools and WT DNA as a control with primer pair 5.2 at 20 s

extension time and 35 cycles using the standard PCR conditions shown in chapter 2 (Section 2.2.1). The PCR reaction was cleaned using the ChargeSwitch PCR clean-up kit (Thermo Fisher Scientific) as described in chapter 2 (Section 2.3.3) and quantified using Nanodrop spectrophotometer. 2 µl of each cleaned PCR was resolved in an agarose gel to detect the presence of expected sized fragments and a Sanger sequencing reaction (Edinburgh genomics) was carried out using the remaining PCR reaction. The chromatograms of all three CRISPRs were subjected to TIDE analysis tool with WT sequencing as a control. Table 5. 5 shows the indel frequency of Lac1, Lac2 and Lac3.

CRISPR name	Indel frequency (in %)
Lac1	14.0
Lac2	19.3
Lac3	37.5

Table 5. 5: TIDE analysis of CRISPRs Lac1 and Lac2 and Lac3: The CRISPRs targeting mouse STAT5 binding site, Lac3 showing the highest indel frequency with 37.5% followed by Lac2 and Lac1 with indel frequency of 19.3% and 14% respectively.

The maximum cutting activity was found in CRISPR Lac3 followed by Lac2 and Lac1. Hence Lac3 and Lac2 were given preference for WT STAT5 binding site editing with HDR templates.

Transfecting CRISPR plasmid and HDR templates in HC11 cells

In order to generate CRISPR-HDR edited HC11 cell lines, two HDR templates containing same modifications as LacT and LacD, hereby referred to as HDRT and HDRD were designed as described in chapter 2 (Section 2.5.6).

Both HDR templates were 90 bp in length with 45 bp each homology arm and named as HDRT1 and HDRD1 (appendix 2). A *HindIII* restriction site was incorporated in both HDR templates for a rapid assessment of their integration. The co-transfection of CRISPR Lac3- HDRT1 and Lac2- HDRD1

was carried out separately using the pre-optimised protocol for HC11 cells as shown in Figure 5.22. The transfection was performed with different CRISPRs and HDR templates concentrations, i.e. 1 μ g, 2 μ g and 100 μ M, 200 μ M respectively to optimise the HDR efficiency. The sequence of all HDR oligonucleotides is presented in appendix 2.

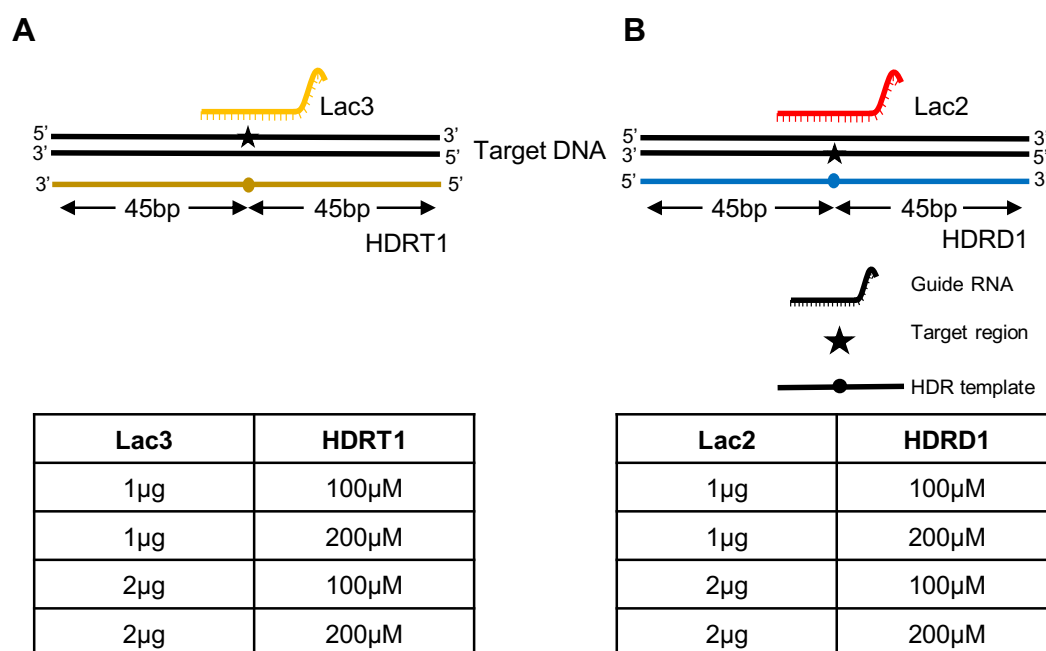


Figure 5. 22: Co-transfection of CRISPRs and HDR templates into HC11 cells: A. co-transfection of CRISPR Lac3 (1 and 2 μ g) and HDRT1 (100 and 200 μ M) in HC11 cells B. Co-transfection of CRISPR Lac2 (1 and 2 μ g) and HDRD1 (100 and 200 μ M) in HC11 cells.

Assessing HDR template integration

HC11 cells transfected with Lac3-HDRT1 and Lac2-HDRD1 were collected at 36h post-transfection by removing culture medium and rinsing with PBS. The cells were dissociated using TrypLE Express (Invitrogen), and GFP positive cells were collected by fluorescence-activated cell sorting (FACS) and plated at 1 cell per well into 5 x 96 well plates containing 100 μ l conditioned RPMI medium in each well. The remaining GFP positive cell pool from both transfections was collected in a 1.5 ml Eppendorf tube and cultured separately

in a single well of a 24 well plate containing 1.5 ml conditioned RPMI medium for the next 48 hours.

DNA was isolated from the cultured cell pool using Quick Extract DNA extraction solution (Epicentre). Two 25 µl PCRs were carried out with DNA isolated from HC11 cells transfected with HDRT1 and HDRD1 with primer pair 5.2 flanking the targeted region at 20s extension time for 35 cycles using the standard conditions described in chapter 2 (Section 2.2.1). The PCR was quantified by resolving 2 µl reaction in a 1.2% agarose gel and 200ng DNA was calculated as described in chapter 2 (Section 2.1.5). A restriction digestion was carried out using 200ng DNA with pre-tested *HindIII* (New England Biolabs) at 37°C for 1 hour as described in chapter (Section 2.1.6) followed by resolving in 1.2% agarose gel (Figure 5.23).

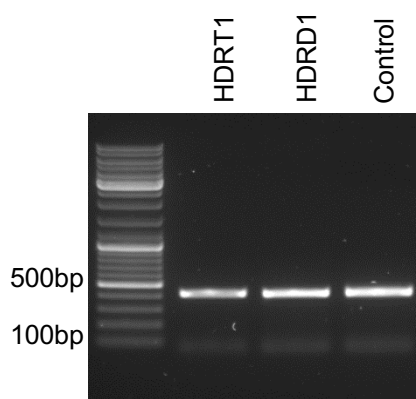


Figure 5. 23: Restriction digestion of DNA isolated from CRISPRs and HDR template transfected HC11 cells: Restriction digestion of PCR product amplified using Lac3-HDRD1 and Lac2 HDRT1 transfected HC11 cells DNA. The control digestion was carried out with WT HC11 DNA. No *HindIII* digested fragments were observed in DNA from CRISPR- HDR transfected cells; all three digestions are showing only one WT 430bp fragment.

The restriction digestion showed no *HindIII* digested fragment suggesting that no HDR template integration at the CRISPRs targeted site occurred.

Transfecting HC11 cells with HDRT2 and HDRD2

Due to the failure to detect HDR events associated with HDRD1 and HDRT1, new HDR templates named as HDRT2 and HDRD2 (appendix 2) were designed to generate LacT and LacD edited HC11 cells. The size of these HDR templates was 120 bp with 90:30 bp homology arms plus a *HindIII*

restriction site. Two different HC11 cells transfections were carried out, one with CRISPR Lac3- HDRT2 template and another with Lac2- HDRD2 template using the same plasmid and template ranges as in the previous experiment and detailed in figure 5.24.

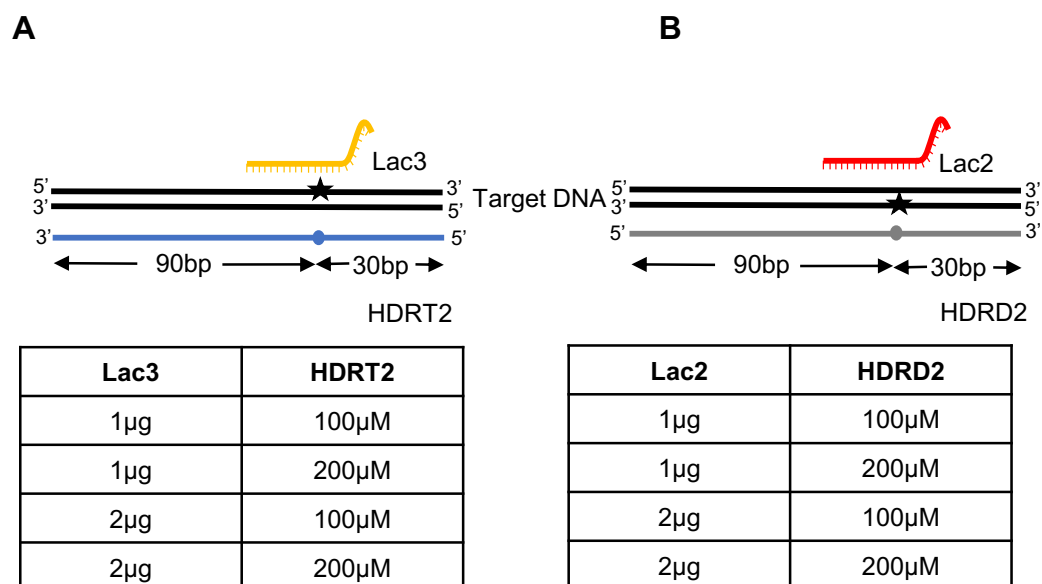


Figure 5. 24: Co-transfection of CRISPRs and HDR template into HC11 cells:

A. co-transfection of CRISPR Lac3 (1 and 2 µg) and HDRT2 (100 and 200uM) in HC11 cells B. co-transfection of CRISPR Lac2 (1 and 2 µg) and HDRD2 (100 and 200uM) in HC11 cells.

HC11 cells transfected with Lac3- HDRT2 and Lac2- HDRD2 were collected at 36h post-transfection as described in the previous Section. The GFP positive cells were plated at one cell per well using FACS into 5 x 96 well plates containing 100 µl conditioned RPMI medium in each well. The remaining GFP positive cells were collected in a 1.5ml Eppendorf tube and transferred to a 24 well plate containing 1.5ml conditioned RPMI media. DNA was isolated from cells growing in a 24 well plate using the Quick Extract DNA extraction solution (Epicentre). Two 25 µl PCRs were carried out with DNA isolated from HC11 cells transfected with CRISPRs and HDRT2 and HDRD2 templates with primer pair 5.2 using the standard PCR conditions (Section 2.2.1). The PCR was quantified by resolving 2 µl reaction in a 1.2% agarose gel and 200 ng

DNA was calculated. The restriction digestion was carried out using 200 ng DNA and *HindIII* (New England Biolabs) at 37°C for 1 hour followed by resolving in 1.2% agarose gel. However, no restriction digestion fragment was detected (Figure 5.25).

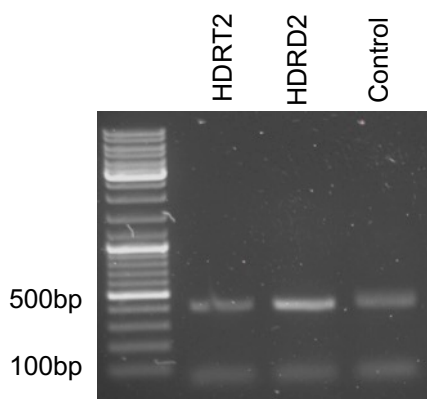


Figure 5. 25: Restriction digestion of DNA isolated from CRISPRs and HDR template transfected HC11 cells: Restriction digestion of PCR product HDRT2 and HDRD2 target region and control PCR carried out with WT HC11 DNA. All three digestions are showing only one WT 430bp fragment. No *HindIII* digested fragments were observed in DNA from HDR transfected cells.

Transfecting HC11 cells with HDRT3 and HDRD3

Two new HDR templates named HDRT3 and HDRD3 (appendix 2) were designed to create a STAT5 tetramer binding site (HDRT) and disrupt the WT STAT5 binding site (HDRD) in HC11 cells respectively. The size of the HDR template was 90 bp with 45:45 bp homology arms with *EcoRI* restriction site. Two different co-transfections were carried out, using Lac1- HDRT3 template and another with Lac1- HDRD3 template at different concentrations as shown in figure 5.26 using prescribed conditions.

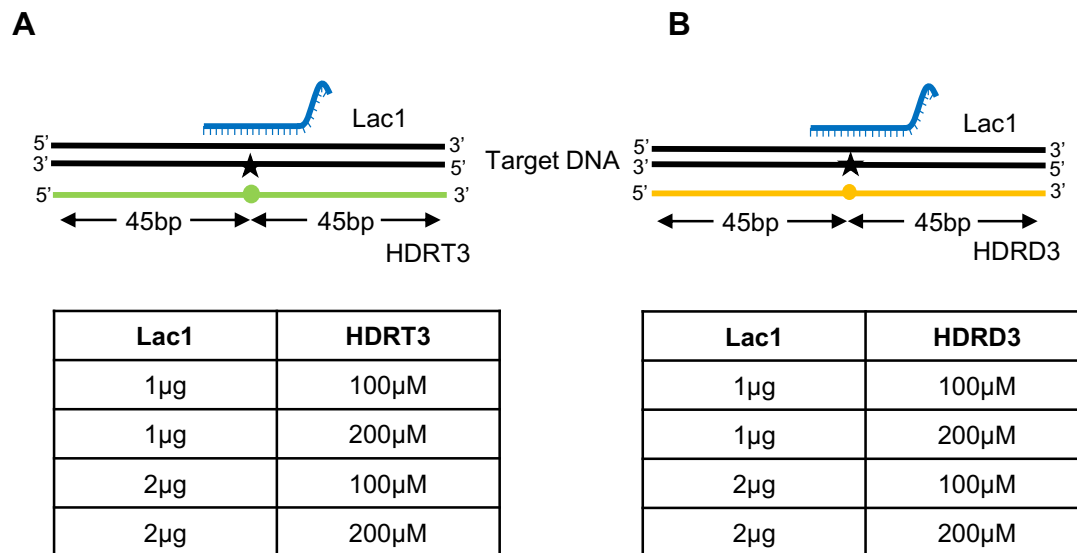


Figure 5. 26: Co-transfection combinations of CRISPR plasmids and HDR templates into HC11 cells: A. Co-transfection of CRISPR Lac1 (1 and 2 µg) and HDRT3 (100 and 200 µM) in HC11 cells B. Co-transfection of CRISPR Lac1 (1 and 2 µg) and HDRD3 (100 and 200 uM) in HC11 cells.

Assessing HDR template integration

The assessment of successful HDR integration was carried out using the same protocol as used in the previous Section. DNA was isolated from GFP positive cells using Quick Extract DNA extraction solution (Epicentre). Two 25 µl PCR were carried out with DNA isolated from HC11 cells transfected with HDRT3 and HDRD3 with a primer pair 5.2 using the standard PCR conditions (Section 2.2.1). The PCR was quantified by resolving 2 µl reaction in a 1.2% agarose gel and 200ng DNA was calculated. The restriction digestion was carried out using 200ng DNA with *EcoRI* (Invitrogen) at 37°C for 1 hour followed by resolving in a 1.2% agarose gel. However, no restriction digestion fragment was detected (Figure 5.27).

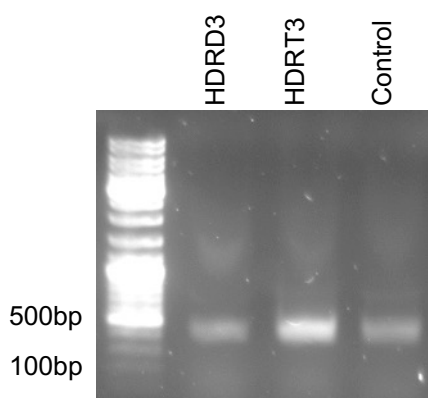


Figure 5. 27: Restriction digestion of DNA isolated from CRISPRs and HDR template transfected HC11 cells: Restriction digestion of PCR product HDRT3 and HDRD3 target region and control PCR carried out with WT HC11 DNA all three digestions are showing only one WT 430bp fragment with no *EcoRI* digested fragments in DNA from HDRD3 and HDRDT transfected cells.

In summary, editing of mouse *lalba* promoter of HC11 cells was attempted using CRISPRs and HDR templates of different sizes and homology arms and transfections at various concentrations. However no successful HDR template integration was achieved. No further experiments were carried out after this point due to time restrictions.

5.4 Discussion

Transcription factors (TF) play a key role in various cellular activities such as expression and repression of essential genes by regulating transcription. Their binding to a transcription factor binding site (TFBS) on DNA is an important primary step which influences transcription (Latchman, 1993). However, a TF can either bind directly to the target DNA or work as a mediator (Thomas Warner, 2009). It is believed that the most convenient way to search for TFBS in multiple sequences is by aligning a regulatory genomic region of the gene of interest in different species which could reveal identical motifs (Djordjevic et al., 2003). One such TF, STAT5, was recognised as mammary gland factor due to the presence of its TFBS as 5' TTC-N₃-GAA 3' in milk proteins such as α , β , γ casein and beta lactalbumin gene promoters in rat, cattle and sheep (Streuli et al., 1995) later, the same STAT5 binding site was identified in the *lalba* promoter of mouse (Malewski, 1998). The *lalba* gene encodes alpha lactalbumin, a milk lactose regulator, which makes a balanced source of glucose and galactose for infant mammals required for energy and brain development (Erbsloh et al., 1958). The increased milk lactose has been

associated with higher weaning weight in pigs and mouse (Bleck et al., 1998), (Noble et al., 2002).

In this chapter the STAT5 TFBS in the *lalba* promoter of 29 mammals was analysed and the expression of mouse *lalba* promoters containing the altered WT STAT5 binding site was assessed in the mouse mammary epithelial HC11 cells using reporter assay. Simultaneously, attempts were made to create similar alterations in the *lalba* promoter of HC11 cells using CRISPRs/ HDR approach.

The STAT5 binding site in *lalba* promoter of mammals

In this chapter, the *lalba* promoters of various mammals were aligned using Clustal Omega, revealing that the 1500 bp 5' to the TSS of sheep and cattle had 92% sequence identity, and that this region of the *lalba* promoter of more diverse species (human, mouse, pig, sheep and cattle) was at least 50% identical at the genomic level.

The STAT5 binding site in mouse and other selected mammals was identified using online transcription factor searching tools such as Jasper and ConvraV2 databases. The STAT5 dimer binding site, i.e. a palindrome 5' TTC-N₃-GGA 3' was found in the *lalba* promoter of a total of 19 mammals including human, aquatic mammals, primates, ruminants and rodents and only one STAT5 TFBC was located in the *lalba* promoter of all mammals.

The lactose concentration in the milk of these 19 mammals varied from 1.8% sperm whale to 7.0% in rhesus monkey and showed no relationship with a number of the sequence of STAT5 TFBS. The STAT5 dimer binding sites of all mammals were visually analysed to locate a palindromic STAT5 tetramer binding site with a distance of 3-6 bp spacer between two STAT5 binding sites, since it has been suggested to be the most efficient tetramer binding motif (Soldaini et al., 2000). In a total of nine mammals including human, mouse, cow, pig, sheep, goat, dolphin, cat and elephant, an incomplete STAT5 tetramer TFBS was found as 5' TTC-N₃-GAA-N₆-CTA-N₃-GAA 3'.

In the past, in an attempt to enhance the BLG gene expression in mouse, a TFBS cluster from sheep BLG gene promoter was expressed as an enhancer element in the proximal promoter of the mouse BLG gene. The *in vitro* and *in vivo* analyses of transgene enhancer in HC11 cells and transgenic mice showed no improvement in the mouse BLG gene expression (James et al., 2000). Similarly, the expression of multiple STAT5 binding sites in the BLG promoter could not elevate the transcriptional activation of this gene in the transgenic mice (Pena and Whitelaw, 2005) hence, in an attempt to achieve enhanced *lalba* expression, the incomplete STAT5 tetramer binding site of mouse was envisaged to edit from 5' TTC-N₃-GAA-N₆-CTA-N₃-GAA 3' into 5' TTC-N₃-GAA-N₆-TTC N₃-GAA 3' making it a palindrome motif based on gamma interferon site (LacT). Another *lalba* promoter with no STAT5 TFBS (LacD) was also analysed.

Analysing the WT and altered mouse *lalba* promoter expression using reporter assay

The dual luciferase reporter assay has been proven a powerful tool to analyse expression level from gene promoters in human (Li et al., 2015) and mouse cells (Xu et al., 2013), (Solberg and Krauss, 2013). The mouse WT (LacP) and altered *lalba* promoters (LacT and LacD), created by fusion PCR, were cloned into a vector expressing firefly luciferase *pGL3B* to compare the luciferase expression level in transfected HC11 cells.

The luciferase assay of all three promoters (LacP, LacT and LacD) in differentiated and undifferentiated HC11 cells showed over 3 fold higher expression of LacT as compared to the LacP promoter, however the expression level of LacD promoter was found to be slightly higher than LacT in undifferentiated cells whereas in differentiated cells the expression of LacD was nearly double when compared to LacT promoter. The similar results were obtained in three luciferase assays carried out at different time points.

These unexpected results encouraged us to reanalyse the sequence of the LacD promoter for the creation of a novel TFBS which might have resulted

from replacement of three base pairs of the WT STAT5 dimer binding site. The modified sequence was pasted into the search box of a TFBS search tool (<http://www.ifti.org/cgi-bin/ifti/Tfsitescan.pl>), however no new TFBS site was detected. Thus it was assumed that the disrupting WT STAT5 binding site with specific bases might have resulted in the unintentional creation of a putative binding motif or an enhancer element which resulted in higher promoter expression in differentiated HC11 cells. However another possibility is that the generation of a no stat5 binding promoter (LacD) might have resulted in inactivation of a silencing DNA element and thus increased its expression.

Stat5 binding site: a limiting factor?

The reporter assay of LacP promoter resulted in lower luciferase expression in differentiated and undifferentiated HC11 cells when compared to the luciferase expression mediated by LacT promoter. This suggests that the WT STAT5 dimer binding site (LacP) in *lalba* promoter of mouse could be negative regulator and its removal might have eliminated its suppression.

The oligomerisation of a TFBS has been associated with up or down regulation of different genes. For e.g. Special AT-rich sequence-binding protein 1 (SATB1) is a TF, involved in regulation of several genes associated T-cells proliferation (J.D. Alvarez, 2000) by binding as a dimer or a tetramer (long J, 2012). The Interleukin-2 (IL-2) gene promoter contains a SATB1 tetramer TFBS, the luciferase reporter assay of IL-2 promoter with a disrupted SATB1 tetramer binding site showed reduced luciferase expression in HEK293T cells when compared to the WT promoter. Subsequently, Hela cell lines over expressing SATB1 tetramer and SATB1 dimer proteins were generated and protein expression was analysed using western blotting. The RT-PCR analysis of gp91phox and IL-2 gene expression in these cells showed reduced expression of these genes in Hela cells expressing SATB1 dimer (Zheng et al., 2017). Similarly the role of STAT5 oligomerisation was analysed in the natural killer cells where maturation of these cells was highly affected in mice lacking STAT5 tetramer (Lin et al., 2017).

The complete removal of STAT5 TFBS (LacD) showed the maximum luciferase expression in differentiated and undifferentiated HC11 cells compared to the luciferase expression of LacP and LacT promoter indicates that the WT STAT5 TFBS in the mouse *lalba* promoter perhaps regulates the expression in a very controlled manner since lactose is responsible for water penetration in mammary alveoli (Palmer et al., 2006) whose expression level could directly affect lactation process and milk composition. The effect of glucose on milk composition in dairy cattle was analysed by injecting glucose intravenously resulted in increased milk quantity with high water and lactose content and decreased protein and fat (Curtis et al., 2014). The presence of a STAT5 dimer binding site in the promoter of other milk protein genes such as α , β and κ casein and BLG in various mammals (Streuli et al., 1995) strengthen this speculation since a balanced expression of all milk components is essential for successful lactation and a balance diet to the infant.

The both hypothesis involving the role of a STAT5 tetramer TFBS and absence of the STAT5 TFBS in regulation of mouse *lalba* gene could be analysed by creating these alterations using CRISPRs/Cas9 and homology directed repair in HC11 cells.

Editing the STAT5 binding site in *lalba* promoter of HC11 cells using CRISPRs-HDR

It has previously been shown that CRISPR/Cas9 can be used to remove a cluster of TFBS including GATA-1, TAL1 and KLF1 TFs from the human β -globin gene in mouse erythroleukaemia cells expressing the human β -globin locus (Kim and Kim, 2017), and that such targeted mutations can be repaired using homology-directed repair by supplying a DNA template (Aird et al., 2018). To further investigate the results acquired from reporter assay, the expression of mouse *lalba* promoter containing the similar edits such as LacT and LacD was planned to analyse in HC11 by editing WT STAT5 binding site using CRISPRs and HDR.

The WT STAT5 binding site of mouse *lalba* promoter was targeted by designing CRISPR guides and cloned into a *Cas9* expressing *pSL66* plasmid to generate LacT and LacD type altered promoters. The cutting efficiency of CRISPR plasmids to create indels was analysed by TIDE assay and the most efficient plasmids were transfected with HDR template to edit the promoter and create LacT or LacD variants. A total of three different HDR templates, varying in length and homology arms, were co-transfected into cells alongside appropriate CRISPR plasmid for each type of promoter editing. All HDR templates were designed with an edited PAM site to make them unavailable for *Cas9* targeting and were incorporated with a restriction site to analyze their integration in cells. The HC11 transfection was carried out with varied CRISPRs and HDR template concentration and the HDR of different length and homology arms, however, no successful HDR event was observed based on restriction digestion of DNA from transfected cells. It has been believed that the DNA repair using a HDR template and the HDR efficiency could be directly affected by the cell type and the transfection efficiency (Zhang et al., 2017) hence optimisations of HC11 cells are perhaps required to obtain successful HDR integration.

The new methods to improve HDR efficiency are currently being discovered, e.g. A 30 fold increase in HDR efficiency was achieved by combining single-stranded HDR template to *Cas9* gRNA via HUH endonuclease (Aird et al., 2018), which contains a catalytic motif to break or join ssDNA (Chandler et al., 2013). The HDR efficiency in iPSCs was improved by 20 -40% by cold shock treatment of CRISPR/*Cas9*- HDR transfected cells at 32°C for 24- 48 hours (Guo et al., 2018). The use of piggyback (PB) transposons in combination with CRISPRs to generate gene replacements is also gaining popularity (Chen et al., 2015), (Li et al., 2017b). The PB transposons are mobile DNA elements initially used for transgenesis in mammalian cells in 2005 (Ding et al., 2005). The human iPS cells were stably transfected with a PB transposon construct containing a *Cas9* transgene. The targeting of Tafaffin (TAZ) gene in the X chromosome of these cells using CRISPR/*Cas9* showed 50% improvement in the indel frequency whereas the

repair of targeted site using HDR template in these cells showed 10-20% improvement when compared to the control WT cells targeted using the same constructs (Wang et al., 2017).

The results obtained in this chapter show that modification of the STAT5 binding site of the *lalba* promoter in a reporter assay results in altered expression in both undifferentiated and hormone induced HC11 cells. The higher expression level of the LacD promoter as compared to LacP and LacT in reporter assay was unexpected yet encouraging and could be further validated by creating similar base pair changes in the mouse *lalba* promoter. Although a deep understanding of all TFBS in the promoter of interest is required before editing, creating or improving TFBS and gene expression, without disturbing other TFs binding site and creating an unintentional TFBS or a controlling element.

6 Chapter: 6 General discussion

The transgenic animals

The emergence of transgenesis in the early 1980s expanded the applications of animals in different research areas beyond traditional usage. Over the years, development of sophisticated gene editing tools such as ZFN and TALENs improved the efficiency of gene editing, however, the discovery of CRISPRs in 2013 turned out to be game changer providing a cutting edge technique to generate transgenic animals, promising a footprint-free and precise editing by allowing gene targeting at various lengths (He et al., 2015).

CRISPRs is now widely used to generate tailored animals in various research areas to enhance their biomedical applications and to attain food security. In biomedical research, transgenic animals are created as a model for human disease. The similarity between the human and pig genome including the presence of several conserved regions (Hart et al., 2007) plus similarities in physiology and anatomy has resulted in the pig becoming preferred as a model animal for human diseases. The early transgenic pig models were created by using the traditional genetic engineering technique, e.g. a mutant rhodopsin transgene was transferred into pig embryos by pronuclear microinjection to generate transgenic pigs as a model for human retinitis pigmentosa disease (Petters et al., 1997) however, lately, CRISPRs is used to efficiently produce such disease models with precise editing. The first intron of pig huntingtin (HTT) gene was targeted in pig fetal fibroblast cells using CRISPRs/Cas9, and a 150 bp long first exon of mutant human huntingtin (HTT) gene was knocked into the targeted region using homologous recombination. These edited cells were then used to generate a transgenic pig line via somatic nuclear transfer serving as a model for human HTT disease (Yan et al., 2018).

Similarly, CRISPRs mediated gene editing has opened many possibilities in xenotransplantation research to generate humanised organs for transplantation. One such human organogenesis platform in pig was

demonstrated by targeting the first exon of *PDX1* gene in pig embryos using CRISPRs to disable pancreatogenesis. The edited embryos were expected to develop human pancreas if combined with human pluripotent stem cells (Wu et al., 2017).

In agriculture, gene editors are applied to reduce the function of a gene to understand the milk genetics and to produce allergen-free milk, e.g. the complete removal of beta-lactoglobulin from milk in goat using CRISPRs (Zhou et al., 2017) and in cattle using TALENs (Wei et al., 2018). The use of gene editors to gain a function is applied in overexpressing milk components or to produce a protein of interest in milk, e.g. the expression of the human alpha-lactalbumin gene in goat using TALENs (Zhu et al., 2016). However, it remains crucial to understand the gene regulation of milk components such as caseins, fat and lactose for precise milk editing.

Altering milk proteins in the mouse genome using CRISPRs/Cas9

A systematic milk protein gene editing could be highly beneficial not only for suckling offspring but also in dairy industries. The zero hunger vision of the European Union aims to achieve global food security by the year 2030, and the dairy has been considered as one of the crucial contributing factor (zero hunger challenge). Several Asian and African countries are majorly dependent on livestock derived food to feed the increasing population projected to grow substantially by the year 2050 (Ritchie et al., 2018). Economically, the global revenue of dairy was over \$550 billion in the year 2017, which generated employment mainly in rural areas worldwide. The consumption of dairy products has been connected with a higher intake of nutrients and minerals (O'Neil et al., 2009), several clinical trials of dairy consumption has shown improved bone health in children and teenagers (Merrilees et al., 2000), (Lau et al., 2004) and in elderly (Thorpe et al., 2008), (Bonjour et al., 2013).

In mammals such as the mouse, human and livestock, an evolutionary conserved region (ECR) was identified in the casein locus, upstream to the

Odam gene, predicted as a controlling element and unidirectional repressor (Kaimala and Kumar, 2015). Some previous research has shown that the milk genes in casein locus such as α , β , γ , δ and κ casein (in mouse) are individually controlled by their promoters (Rijnkels et al., 2003) however the deletion of α casein in mouse resulted in a significant reduction in expression of β , γ caseins in milk lacking the α casein (Kolb et al., 2011). This indicates that the regulation of genes in this locus is perhaps interconnected. The hypothesis of the evolution of all calcium sensitive casein genes (α , β , γ and δ caseins) from an early casein gene *CSN1/2* (Kawasaki et al., 2011) strengthen this speculation. To investigate the function of ECR we envisaged the deletion of ECR, *Odam* and ECR-*Odam* separately in the mouse by injecting CRISPR plasmids targeting each component into zygotes via pronuclear microinjections and transfer of injected zygotes into the oviduct of pseudo-pregnant recipient mice, although, however this resulted in the generation of ECR knockout mice only. The phenotyping of these mice is currently being carried out, and the results are expected to shed light on the significance of ECR and its role in the regulation of *Odam* and other milk proteins.

Previously, in similar research, a highly conserved enhancer element was identified in the first intron of the CD4 gene of murine thymoma cells and was referred as Novel Cis Element (NCE) (Sarafova et al., 2013). The deletion of ~300 bp long NCE in these cells using CRISPRs and non-homologous end joining resulted in the reduction of CD4 gene expression. The NCE was then considered as an essential element for upregulation of CD4 gene during differentiation of thymoma cells (Sarafova et al., 2018).

Over the years, several attempts have been made to enrich milk which includes the removal of beta lactalbumin in cattle to generate more humanised milk lacking an allergic component (Jabed et al., 2012) and the expression of human lysozyme and lactoferrin in goat and cattle milk to improve the innate immunity of suckling infants (Cooper et al., 2015). Similarly, milk with improved caseins content was produced by overexpression of β and κ casein in cattle which resulted in a noticeable increase in sialic acid content (Brophy et al.,

2003) (Laible et al., 2016), which has been considered an essential element for brain development (Wang, 2012).

Improved livestock milk positively affects offspring's health and is crucial for pork, beef and meat industries which majorly rely on the generation of healthy and fast-growing offspring to meet the cost of production. An improved milk lactose has been connected with higher litter size at the time of weaning in pigs and mice (Bleck et al., 1998), (Wheeler et al., 2001), (Noble et al., 2002). Milk lactose is regulated by alpha-lactalbumin which is encoded by the *lalba* gene in the mouse. The regulation of this gene is controlled by transcription factors such as stat5 which bind to a motif in the promoter region. In an attempt to better understand the expression of the *lalba* gene, its promoter from human, mouse and livestock animals was analysed which revealed the presence of an incomplete tetramer stat5 binding site in all promoters. In an effort to elevate the expression of alpha-lactalbumin, we attempted to modify the *lalba* promoter of mouse HC11 cells. To do this, we created two mouse *lalba* promoter constructs, a promoter with a complete tetramer stat5 binding site (LacT) and another promoter with a disrupted/modified stat5 binding site (LacD) using fusion PCR. These constructs (LacT and LacD) and the WT control construct (LacP) were cloned in a firefly luciferase expressing *pGL3-B* plasmid. The reporter assay of the altered promoters (LacT and LacD) and a comparison with the wild-type promoter (LacP) as a control in HC11 cells showed the maximum luciferase activity in promoter with the disrupted stat5 binding site (LacD). The elevated expression of LacD promoter suggested that the presence of a dimer or tetramer STAT5 binding site in the mouse *lalba* promoter could be acting as a limiting factor affecting the expression of this gene.

To verify these results at the cellular level, similar edits (LacT and LacD) were produced in HC11 cells using CRISPRs mediated homology-directed repair (HDR). Previously, CRISPRs has been successfully used to mutate the binding motif of erythroid-specific transcription factors in the β -globin locus of human melanoma cells (Kim and Kim, 2017). To alter the wildtype STAT5

binding site (LacP) into the LacT and LacD type, CRISPRs targeting the native STAT5 binding site and HDR templates of different size (90 and 120 bp) and arm lengths (45:45 bp and 90:30 bp) were transfected in HC11 cells at various concentrations. Nevertheless, no successful HDR event was detected. In the case of CRISPRs-HDR mediated gene replacement, the HDR efficiency plays an important role where the binding of a repair template to the CRISPRs targeted site often require cell line based optimisations. Several new research reports describing enhanced HDR efficiency are being published lately, for e.g. HDR efficiency was improved in HEK293T cells by fusing 65 and 81 nucleotide long single-stranded HDR templates to a sgRNA-Cas9 complex via covalent linkage using SNAP-tag (Savic et al., 2018). Further research on improved HDR efficiency in HC11 cells and *alba* promoter with an altered stat5 binding site, similarly as LacT and LacD, may give insight on new opportunities to enhance *alba* expression and ultimately the milk lactose.

The global dairy industries are majorly dependent on the concentration of specific milk components, for example, the cheese production is affected by milk casein content. The caseins make about ~80% of total milk protein, and the effect of the absence of its three major components, α , β and κ caseins has been individually analysed in the mouse by (Kolb et al., 2011), (Kumar et al., 1994, Shekar et al., 2006) respectively. However, the effect of the entire casein locus deletion on lactation is unknown. We successfully targeted the entire casein locus (266.5kb) in the mouse embryonic fibroblast cells using CRISPRs; however, the subsequent attempts to generate KO mice by pronuclear microinjections of CRISPRs in mouse zygotes did not produce a casein KO mice. We suspect that perhaps CRISPRs targeting the casein locus in the mouse embryos did not work efficiency due to the large size of this locus or the presence of several non-milk genes in this locus affected the development of targeted mouse embryo. To investigate both the speculations, mouse zygotes were injected with CRISPRs targeting the casein locus, these zygotes were culture *in-vitro* for four days; however the results of this culture remained inconclusive because of losing zygotes during harvesting and could not be repeated due to time constraints.

The deletion of a significant component from milk, i.e. caseins, may provide an opportunity to other milk components to express at maximum extent. The casein KO mice could potentially play a role as a model organism to redesign casein locus in livestock to express proteins of interest in this locus and may help in identifying genes with a potential role other than in milk production. The casein proteins in milk are present as a micelle structure where α and β caseins form an inner core, and the κ casein is present on the outer surface, stabilising the micelle's structure. The casein micelles possess an ability to entrap non-polar molecules such as lipids, antimicrobials, antioxidants and vitamins; thus, it has been considered as a potential candidate for biological delivery of these water-insoluble or partially soluble compounds. The successful encapsulation of limonene, lycopene, lutein, and omega-3 fatty in casein micelles has already been achieved (Gursoy and Benita, 2004). The redesigning of the casein locus may exhibit the scope of *in vivo* encapsulation of desired molecules in casein micelles in livestock milk.

Some current requirements of global dairy industries are the generation of high yielding and low emission cattle and production of healthier milk with improved shelf life. The casein locus of mammals could play a crucial role in improving milk properties. E.g. there are two genetic variants of β casein reported in cattle, A1 and A2. The milk with A1 or A1/A2 variant has been associated with gastrointestinal intolerance, and allergic reactions on the other hand milk with the A2 variant is considered healthier (Jianqin et al., 2016), (He et al., 2017). The research on β casein variants and their association with milk allergy could be carried out using gene editors in an attempt to produce humanised milk.

The emission of greenhouse gas, methane, by cattle is one of the major concern not only for dairy industries but also for environmental biologists. Grazing cattle can produce 250-500 litre methane per day losing ~6% of its consumed energy (Johnson and Johnson, 1995). Recently the role of rumen methanogenic archaea in the production of methane as a result of food

digestion was identified. The analysis of a total of 3970 genes in these bacteria revealed the presence of 20 methane producing genes (Roehe et al., 2016), which could be modified using gene editors to generate altered gut microbiota producing less or no methane. However, modified bacteria would be outcompeted by the cattle rumen flora, in that case, an alternative approach to alter the functioning of cattle digestive tract could be considered.

Transgenics in biotechnology

A major utility of milk editing is the production of recombinant proteins for therapeutic usage which started in the late 1980s with the production of human tissue plasminogen activator in mouse (Gordon et al., 1992) and expression of human clotting factor IX and $\alpha 1$ antitrypsin gene in sheep. The global biotechnology market involving recombinant protein production has been expanding ever since and is expected to reach \$600 billion by the year 2020. There are several production systems such as bacteria, yeast, mammalian cells and transgenic animals that can be used to produce recombinant proteins. However, each system has its own advantages and disadvantages. The benefits of using an animal as a bioreactor over bacterial and mammalian cell culture-based methods to produce recombinant proteins have been discussed several times in the past (Clark et al., 1989), (Rudolph, 1999). The yeast and bacterial systems lack the ability of post-transcriptional modification and glycosylation of mammalian proteins, whereas the use of cell culture system generally results in the relatively limited output (Houdebine, 2018). The use of animals as bioreactor appear as one of the most mature and efficient alternatives available where recombinant proteins can be produced in blood or milk while deciphering their genetics. The expression of the foreign protein in blood results in lower production level and could affect animal's health whereas expression of recombinant protein in milk is much safer with high output (Houdebine, 2009) e.g. transgenic mice expressing human insulin produced over 8.1 g/l insulin in milk (Qian et al., 2014). In some cases, bioactive therapeutic proteins could affect the animal which can be overcome by using chicken as a bioreactor to produce recombinant proteins in eggs.

Recently, human cytokines such as interferon $\alpha 2a$ and CSF1 were produced in chicken eggs with the recovery of 15 and 100mg/ml of egg white respectively (Herron et al., 2018).

For transgenesis in mammals, an animal can be selected based on its annual yield and offspring number to express proteins of interest in milk, e.g. rabbit, goat and cow have yearly milk yield of 15, 800 and 8,000 litres respectively (Houdebine, 2009) however, they vary significantly in their gestation period and maintenance cost. Among all animals, goat and rabbit are primarily being applied for expression of recombinant proteins due to their small size and ease in management. Some of the recombinant proteins expressed in transgenic goat milk are human plasminogen activator (Denman et al., 1991), Antithrombin III (Edmunds et al., 1998a), lysozyme (Maga et al., 2006) and human factor IX (Amiri Yekta et al., 2013) whereas recombinant human protein C (Dragin et al., 2005) and human plasminogen activator (He et al., 2018) has been successfully expressed in higher concentrations in rabbit milk. The β casein promoter of goat is majorly used as a regulatory element for the expression of foreign proteins. Buffalo is an emerging interest for producing foreign proteins in milk, recently, the β casein promoter was identified as a key component to express a protein of interest in higher levels (Ganguli et al., 2015). Since 2011, the FDA has approved over 59 therapeutic proteins which include monoclonal antibody, growth and coagulation factors, enzymes and hormones being developed by different production technologies including transgenic animals (Lagasse et al., 2017). The two commercially available recombinant proteins expressed in transgenic animals are an anticoagulant protein, antithrombonin (ATryn), produced in goat milk and the human C1 inhibitor (ruconest) used for the treatment of angioedema, swelling of skin tissues, is produced in rabbit milk.

Genetically modified animals for food security

Gene editing has been applied in animals to improve their growth rate, resistance to specific disease and to gain new characteristics. The research on the only genetically modified animal available currently in the market, a

salmon, started in the late 1980s. A chimeric gene was created by incorporating chinook salmon growth hormone gene into the anti-freeze protein promoter of Atlantic salmon. The gene construct was injected into fertilised Atlantic salmon eggs resulted in the generation of the transgenic salmon line where 40% of the transgenic salmon generated by injected eggs expressed the chimeric gene. The transgenic salmon grew faster than non-transgenic salmon in the first year with a six-fold increase in the body growth and were termed as AquAdvantage Salmon (Du et al., 1992). In the subsequent years, several scientific trials regarding the effects of this Salmon on human consumption and the environment were carried out and resulted were presented to the food and drug administration (FDA) in the USA. Later in 2015, AquAdvantage Salmon became the first genetically engineered animal available the market for human consumption. A similar research in pig to improve its muscle growth was carried out by expressing bovine growth hormone gene in the pig. The early development of this transgenic pig showed improvements in body weight gain. However, the long term effect of high growth hormone in blood plasma showed several undesirable phenotypes (Pursel et al., 1989), (Pursel et al., 1990). In the past few years, an inhibitor of the muscle growth, myostatin, was targeted in sheep and cattle embryos using TALENs (Proudfoot et al., 2015) and CRISPRs (Crispo et al., 2015) aiming to improve the growth and development of livestock. In 2018, the research group of Alison Van Eenennaam at the University of California, Davis, USA knocked-in *SRY* gene in bovine cells using CRISPRs (unpublished). The aim of this project to generate a bull that will only produce male progeny and is supposed to aid the beef industry.

The developments of gene editing tools open enormous possibilities of a better future ensuring food safety, improved biomedical and biotechnological outcomes; however, such research very often debated regarding the containment of transgenic animals, which if escape, can alter the natural ecosystem and the safety on human health if GM animals are consumed regularly. Many segments in public have expressed concerns over the use of

gene editing tools in making genetically modified livestock especially the time and cost-effective CRISPRs.

The very recent use of CRISPRs in human embryos to generate HIV resistant babies in China faced widespread criticism about the responsible use of this technique. Since there are many anti-GMO activist groups worldwide opposing the use of gene modification in food sources, it is necessary that scientists and governments work together to enhance the awareness and demonstrate the safety of this technique and GMO products. Nevertheless, a responsible use of gene editing in animals must be corroborated by scientist worldwide and there must be regulations revised by governments regularly to ensure accountability.

7 Chapter 7 Appendices

7.1 List of PCR primers

Primer name	Forward	Reverse
3.1	CGGAAACCGTAAGAACTGTG	TCAGTGGGTGTTTCAGATCCTCTT
3.2	TCCACCTCATCTTACCTCAACCTA	ATTCAGTGGGTGTTTCAGATCCTCTT
3.3	TTGTCAGCAACCATGCTTTGAATA	TGATTTCAGTGGGTGTTTCAGATCCTC
3.4	TCCACCTCATCTTACCTCAACC	GTGGCAGGTGATTGTTAGGCT
3.5	CCACCTCATCTTACCTCAACCT	GGGAAGAGCTGTTTGCTTGA
3.6	GTCAGCAACCATGCTTTGAA	TGACAGAAGTGAATGCCCTGT
3.7	AACTGGGCAGCATTCAAGT	CAACTGTAGAACCTAGAACTCCA
3.8	CGGAAACCGTAAGAACTGTGA	CCCCTACCTGGAGTTTAACTGAC
3.9	TCGGAAACCGTAAGAACTGTG	TGTTGAGCAGCTGGGGTTTA
3.10	CCACTGAATCACTGGTCCCT	TCACCCTAAACAAACACGCA
4.1	TCTCAACATGTACTATTCCT	AGAATTTCCCAGGAGAGGTA
4.2	TTGGAGTCTAACTTCTTGAG	AAATAGTCAAGGAGCAACCT
4.3	TGCCACTAATATTTGCAGGA	TCAAGGAGCAACCTCCTCAG
4.4	ATGGGATGGTTGTGGAGGT	AAAGTTCATGGCCACACACA
4.5	TTTCCAAGAAAGGACAAATGGC	AAAGTTCATGGCCACACACA

4.6	CAGCAACCATGCTTTGAATA	CACAGATCTTGCAAAGAAGT
4.7	ATGGGATGGTTGTGGAGGT	AAAGTTCATGGCCACACACA
5.1	GGGCTTTCCTGGACTTTGTC	AAAGGCCAGACAATGGGAGAC
5.a1- b.a2	GCTAAGACAGCCCGGGGTGGCTTGGTTAGGATGGA	GGCAGGGCAGGAGAAGTAATTTCGAAGCTAAGACAG
5.b1- 5.b2	TGAGTGTAGTACTCCATGACAGA	TTGAAGGCTGCTTCCAGGAAGCAGGATTTCTTAACCT
5.b3- 5.b4	ACCGAGACCATTTGGAAGTAAAGAACTTCCGACGAAGG	CCTGTCTCCCATTTGTCTGGC
5.c1- 5.c2	AATGAGCACAGGCAGGAAGGA	TTTAGGGCTGCCAGCAGGAAGCAGGATTTCTTAACCT
5.c3- 5.c4	ACCGAGACCATTTGGAAGTAAAGAAAATCCCGACGGTCG	TCAATTACTTCTCCTGCCCTGCC
5.2	ATCCCTTTTCCCAAGGCTCTC	GCCAAAATAGAAACCCTGAGCC

7.2 List of HDR templates

HDRT1	ATTGTCTCTGAAGCCTCCTTTGTTCCAAAAGTAGGTTAAGGAAATCAAGCTTCCTGGAAGCAGCC TTCAAAGAAATGAAGGTTTACCAGA
HDRT2	TCCTTTGTTCCAAAAGTAGGTTAAGGAAATCAAGCTTCCTGGAAGCAGCCTTCAAAGAAATGAAG GTTTACCAGAGCCAAGTGAGAAGCTGGGTCATGTGTGGAATTATGTGGGAAGAAA
HDRT3	ATCCAATTCCTTTAGGACGAAGGACCTTCGTCGGAAGTTTCTTTACTTAAGAATGGTCTCGGTTC ACTCTTCGACCCAGTACACACCTTAATA
HDRD1	TCGGAGGAAACAAGGTTTTTCATCCAATTCCTTTAGGACGAAGGACGACCGTCGGGATTTTCTTTA CTTCGAAATGGTCTCGGTTCACTCTTCG
HDRD2	ACATCGGTTGACACAAAAGGTCAGTGGAGGGAACATAACAGAGACTTCGGAGGAAACAAGGTTTT CATCCAATTCCTTTAGGACGAAGGACGACCGTCGGGATTTTCTTTACTTCGAAATGGT
HDRD3	ATCCAATTCCTTTAGGACGAAGGACGACCGTCGGGATTTTCTTTACTTAAGAATGGTCTCGGTTC ACTCTTCGACCCAGTACACACCTTAATA

7.3 Buffers

Bacterial growth

- Luria Bertani (LB media): 10g Tryptone: 10g, NaCl, 10g, Yeast extract: 5g, ddH₂O: up to 1 litre.
- Luria Bertani agar (LB agar): 10g Tryptone: 10g, NaCl, 10g, Yeast extract: 5g, Agar: 20g, ddH₂O: up to 1 litre.

Animal cells

- Freezing media for MEF cells: Dulbecco's Modified Eagle Medium (DMEM) (Gibco) with 10% Fetal bovine serum (Gibco) and penicillin streptomycin antibiotics (Thermo Fisher Scientific), 10% DMSO (Sigma).
- Freezing media for HC11 cells: RPMI-1640 medium with L-Glutamine and sodium bicarbonate (Gibco), 10% Fetal Bovine Serum (Gibco) and penicillin streptomycin antibiotics (Thermo Fisher Scientific), 10% DMSO (Sigma).

Western blotting

- RIPA lysis and extraction buffer, Sigma (1x): 25mM Tris•HCl pH 7.6, 150mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS, stored at 4°C supplemented with 1x Halt™ Phosphatase Inhibitor Cocktail (Invitrogen) immediately before use.
- Protein sample buffer, Laemmli buffer, BioRad (2x): 65.8 mM Tris-HCl, pH 6.8, 2.1% SDS, 26.3% (w/v) glycerol, 0.01% bromophenol blue, stored at

room temperature was added with 50ul 2- mercaptoethanol per 950ul buffer. Diluted with equal part of protein sample.

- Western running buffer (10x): Tris-base: 30g Glycine: 144g, SDS: 10g, ddH₂O: 1 L. stored at room temperature and diluted with ddH₂O to a final concentration of 1x before use.
- Western transfer buffer (1x): 25mM Tris, 190mM Glycine, 0.1% SDS, prepared fresh in ddH₂O immediately before use.
- Phosphate buffer saline (PBS, 1x): NaCl: 8g, KCl: 0.2g, Na₂HPO₄: 1.44g, KH₂PO₄: 0.24g, ddH₂O: up to 1 litre, pH 7.4, stored at room temperature.
- Phosphate buffer saline tween (PBST, 1x): PBS buffer 1x: 1000ml, Tween 20: 2ml, stored at room temperature.
- Western blocking buffer: 5% skimmed milk powder in PBST

Gel electrophoresis

- TAE buffer (1x): Tris: 40mM, Acetic acid: 20mM, EDTA: 1mM, pH 8.0

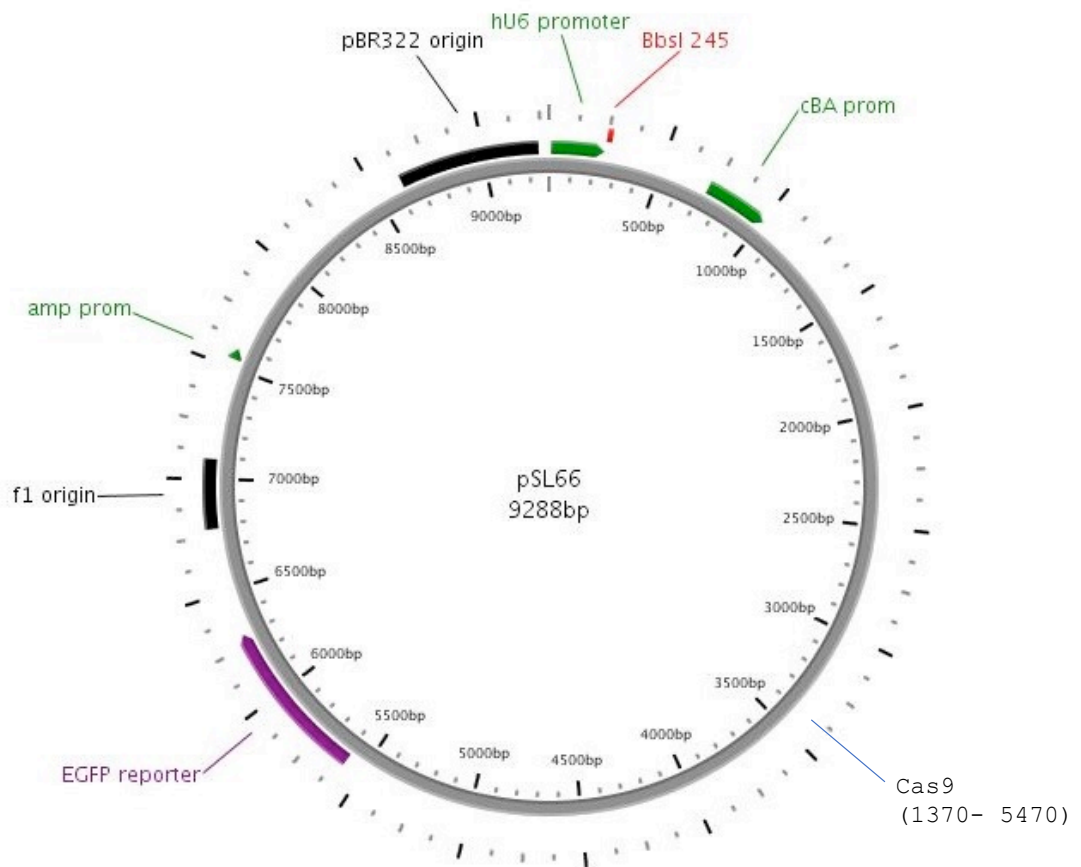
Histology

Eosin Stock and working solution: For Stock solution (1 litre); 10 g Eosin Y Cl 45380 (Sigma), 800 ml Ethanol, 2.5 ml 1 % Phoxine B (Cl number 45410, Acid Red 92), was mixed in deionised water up to 1 litre. For working solution- 100 ml stock solution, 200 ml 70% Ethanol, 1.5ml Glacial acetic acid were mixed.

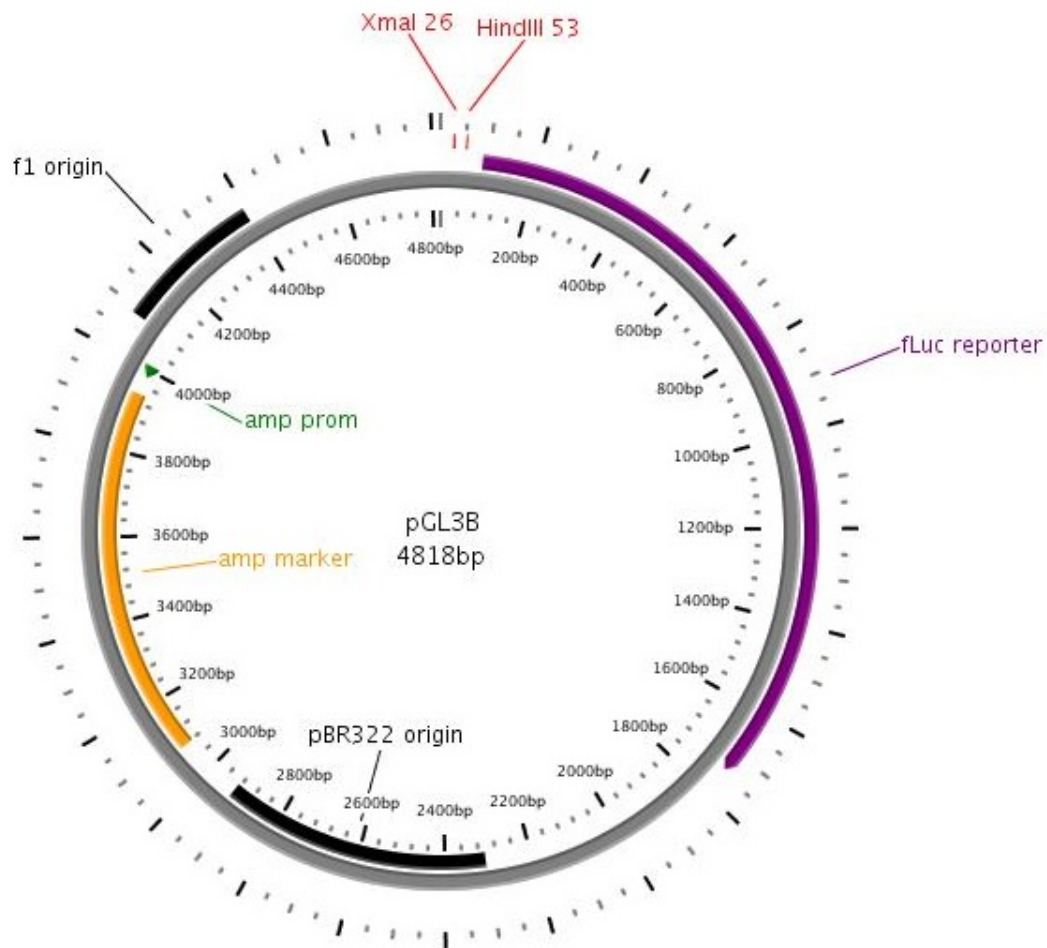
Haematoxylin: Harris haematoxylin (Leica Biosystems)

7.4 Plasmids and vectors

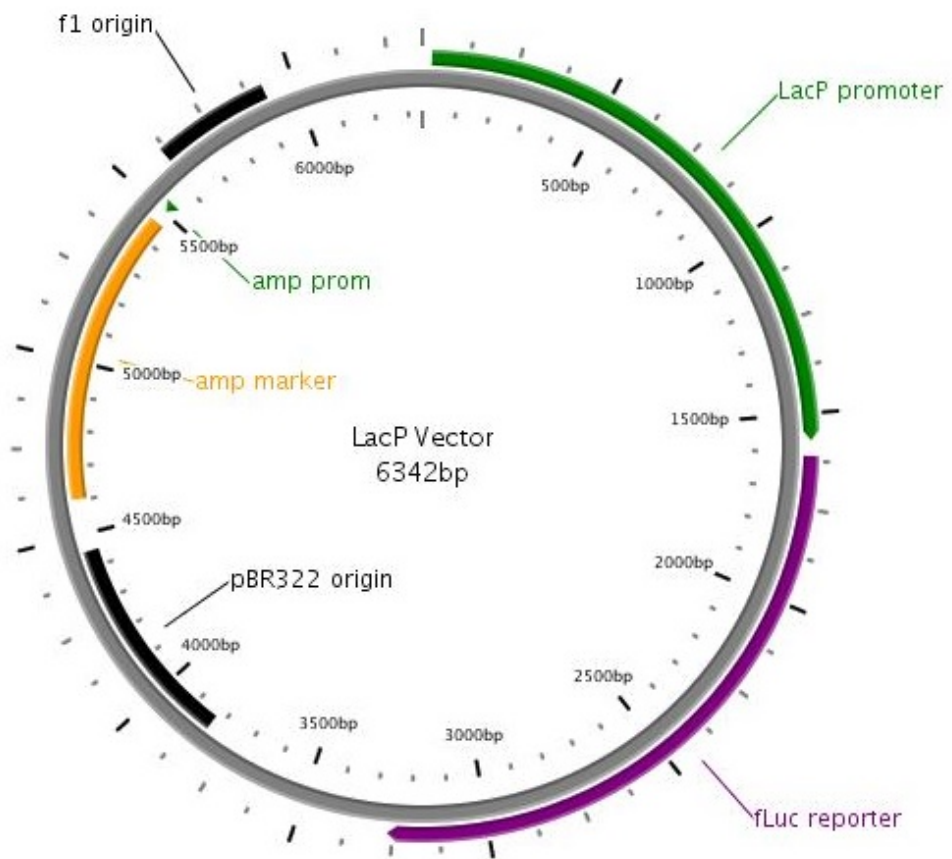
7.4A- Plasmid *pSL66* map



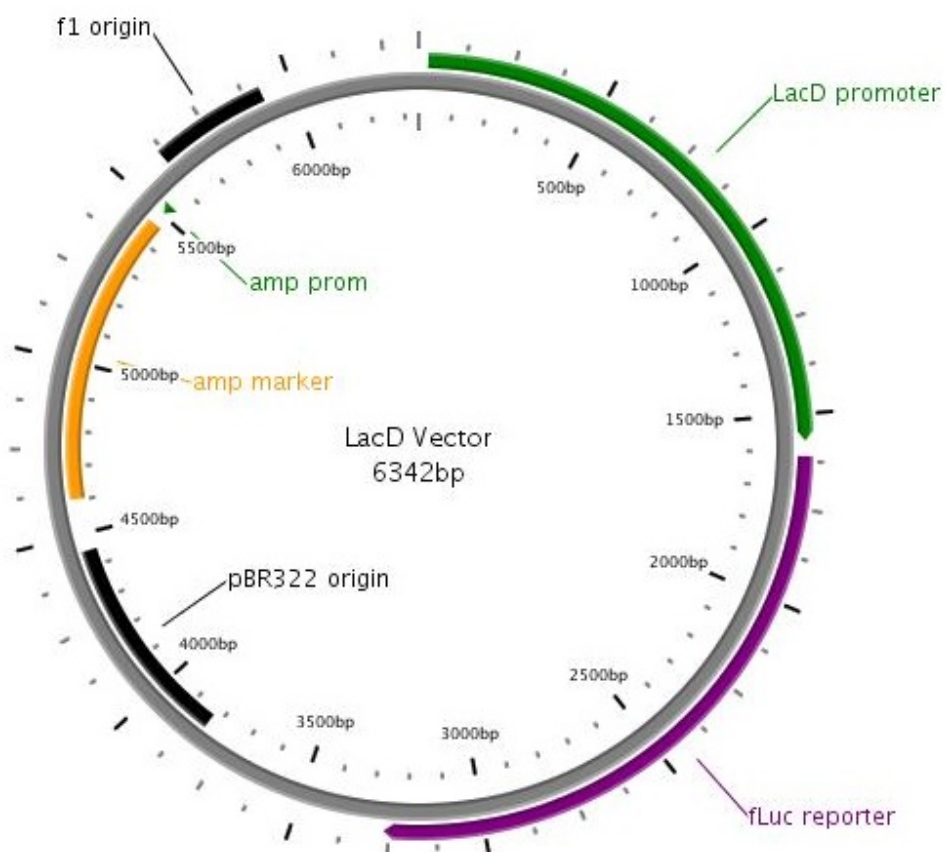
7.4B- Plasmid *pGL3B* map



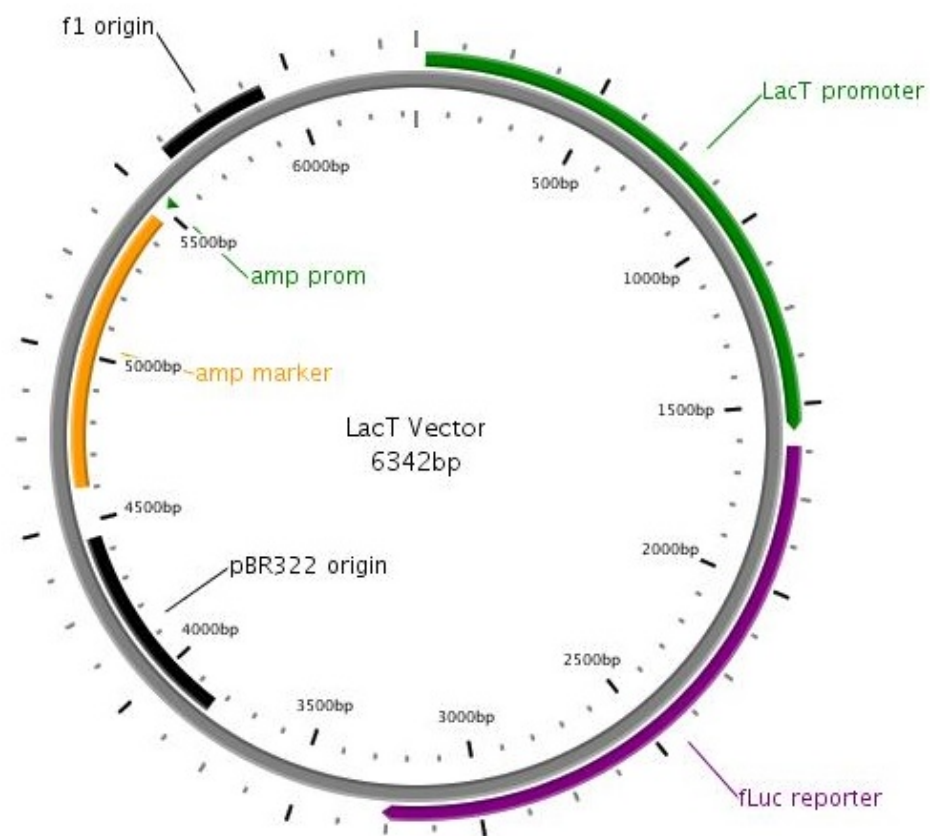
7.4C- Vector LacP map



7.4D- Vector LacT map



7.4E- Vector LacT map



7.5 Reagents

Reagent	Company	Catalogue number
1 KB plus DNA ladder	Thermo fisher scientific	10787018
ChargeSwitch PCR Clean-Up kit	Thermo fisher scientific	CS12000
DNeasy Blood & Tissue Kit	Qiagen	69504
DreamTaq Green PCR Master Mix (2X)	Thermo fisher scientific	K1081
Dual Luciferase Reporter Assay system	Promega	E1910
<i>EcoRI</i>	New England biolabs	R0101S
GeneRuler DNA Ladder Mix	Thermo fisher scientific	SM0331
<i>HindIII</i>	New England biolabs	R0104S
Phusion® High-Fidelity DNA Polymerase	New England biolabs	M0530S
Plasmid-Safe ATP-Dependent DNase	Epicentre	E3101K
PureLink Quick Gel Extraction Kit	Thermo fisher scientific	K210012
QIAGEN Plasmid <i>Plus</i> Midi Kit	Qiagen	12943
QIAprep SpiMiniprep Kit	Qiagen	27104

QuickExtract DNA Extraction Solution	Epicentre	QE09050
RIPA Lysis and Extraction Buffer	Thermo fisher scientific	89901
T4 DNA Ligase	New England biolabs	R0202S
T7 Endonuclease I	New England biolabs	M0320S
<i>Xma</i> I	New England biolabs	R0180S
RIPA Lysis and Extraction Buffer	Thermo fisher scientific	89901
T4 DNA Ligase	New England biolabs	R0202S

8 Chapter 8- Bibliography

AIRD, E. J., LOVENDAHL, K. N., ST MARTIN, A., HARRIS, R. S. & GORDON, W. R. 2018. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun Biol*, 1, 54.

AL-ALWAN, M., DU, Q., HOU, S., NASHED, B., FAN, Y., YANG, X. & MARSHALL, A. J. 2007. Follicular dendritic cell secreted protein (FDC-SP) regulates germinal center and antibody responses. *J Immunol*, 178, 7859-67.

AMIRI YEKTA, A., DALMAN, A., EFTEKHARI-YAZDI, P., SANATI, M. H., SHAHVERDI, A. H., FAKHERI, R., VAZIRINASAB, H., DANESHZADEH, M. T., VOJGANI, M., ZOMORODIPOUR, A., FATEMI, N., VAHABI, Z., MIRSHAHVALADI, S., ATAIEI, F., BAHRAMINEJAD, E., MASOUDI, N., REZAZADEH VALOJERDI, M. & GOURABI, H. 2013. Production of transgenic goats expressing human coagulation factor IX in the mammary glands after nuclear transfer using transfected fetal fibroblast cells. *Transgenic Res*, 22, 131-42.

AMOROSO, E. C. 1959. Comparative anatomy of the placenta. *Ann N Y Acad Sci*, 75, 855-72.

ANTHWAL, N., JOSHI, L. & TUCKER, A. S. 2013. Evolution of the mammalian middle ear and jaw: adaptations and novel structures. *J Anat*, 222, 147-60.

ARCHIBALD, J. D., AVERIANOV, A. O. & EKDALE, E. G. 2001. Late Cretaceous relatives of rabbits, rodents, and other extant eutherian mammals. *Nature*, 414, 62-5.

BAKER, E. N., BAKER, H. M. & KIDD, R. D. 2002. Lactoferrin and transferrin: functional variations on a common structural framework. *Biochem Cell Biol*, 80, 27-34.

BALLARD, O. & MORROW, A. L. 2013. Human milk composition: nutrients and bioactive factors. *Pediatr Clin North Am*, 60, 49-74.

BARBANO, D. M. 2017. A 100-Year Review: The production of fluid (market) milk. *J Dairy Sci*, 100, 9894-9902.

BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315, 1709-12.

BARRINGTON, G. M. & PARISH, S. M. 2001. Bovine neonatal immunology. *Vet Clin North Am Food Anim Pract*, 17, 463-76.

BELLA, R., KAMINSKI, R., MANCUSO, P., YOUNG, W. B., CHEN, C., SARIYER, R., FISCHER, T., AMINI, S., FERRANTE, P., JACOBSON, J. M., KASHANCHI, F. & KHALILI, K. 2018. Removal of HIV DNA by CRISPR from Patient Blood Engrafts in Humanized Mice. *Mol Ther Nucleic Acids*, 12, 275-282.

BEUMER, K. J., TRAUTMAN, J. K., CHRISTIAN, M., DAHLEM, T. J., LAKE, C. M., HAWLEY, R. S., GRUNWALD, D. J., VOYTAS, D. F. & CARROLL, D. 2013. Comparing zinc finger nucleases and transcription activator-like effector nucleases for gene targeting in *Drosophila*. *G3 (Bethesda)*, 3, 1717-25.

BIGGERS, J. D. 1987. Pioneering Mammalian Embryo Culture *The Mammalian Preimplantation Embryo*.

BIRLING, M. C., SCHAEFFER, L., ANDRE, P., LINDNER, L., MARECHAL, D., AYADI, A., SORG, T., PAVLOVIC, G. & HERAULT, Y. 2017. Efficient and rapid

generation of large genomic variants in rats and mice using CRISMERE. *Sci Rep*, 7, 43331.

BLACKBURN, D. G. 1992. Convergent Evolution of Viviparity, Matrotrophy, and Specializations for Fetal Nutrition in Reptiles and Other Vertebrates. *American Zoologist*, 32, 313-321.

BLECK, G. T., WHITE, B. R., MILLER, D. J. & WHEELER, M. B. 1998. Production of bovine alpha-lactalbumin in the milk of transgenic pigs. *J Anim Sci*, 76, 3072-8.

BOCH, J. & BONAS, U. 2010. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol*, 48, 419-36.

BOLOTIN, A., QUINQUIS, B., SOROKIN, A. & EHRLICH, S. D. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 151, 2551-61.

BOLTON, V. N. & BRAUDE, P. R. 1984. Preparation of human spermatozoa for in vitro fertilization by isopycnic centrifugation on self-generating density gradients. *Arch Androl*, 13, 167-76.

BONJOUR, J. P., KRAENZLIN, M., LEVASSEUR, R., WARREN, M. & WHITING, S. 2013. Dairy in adulthood: from foods to nutrient interactions on bone and skeletal muscle health. *J Am Coll Nutr*, 32, 251-63.

BOSTON, W. S., BLECK, G. T., CONROY, J. C., WHEELER, M. B. & MILLER, D. J. 2001. Short communication: effects of increased expression of alpha-lactalbumin in transgenic mice on milk yield and pup growth. *J Dairy Sci*, 84, 620-2.

BOUCHOUX, A., GESAN-GUIZIOU, G., PEREZ, J. & CABANE, B. 2010. How to squeeze a sponge: casein micelles under osmotic stress, a SAXS study. *Biophys J*, 99, 3754-62.

BOUNIOL, C., PRINTZ, C. & MERCIER, J. C. 1993. Bovine alpha s2-casein D is generated by exon VIII skipping. *Gene*, 128, 289-93.

BOUTINAUD, M. & JAMMES, H. 2004. Growth hormone increases Stat5 and Stat1 expression in lactating goat mammary gland: a specific effect compared to milking frequency. *Domest Anim Endocrinol*, 27, 363-78.

BROCK, J. H. 1980. Lactoferrin in human milk: its role in iron absorption and protection against enteric infection in the newborn infant. *Arch Dis Child*, 55, 417-21.

BROOKE-TAYLOR, S., DWYER, K., WOODFORD, K. & KOST, N. 2017. Systematic Review of the Gastrointestinal Effects of A1 Compared with A2 beta-Casein. *Adv Nutr*, 8, 739-748.

BROPHY, B., SMOLENSKI, G., WHEELER, T., WELLS, D., L'HUILLIER, P. & LAIBLE, G. 2003. Cloned transgenic cattle produce milk with higher levels of beta-casein and kappa-casein. *Nat Biotechnol*, 21, 157-62.

CARLSON, D. F., TAN, W., LILICO, S. G., STVERAKOVA, D., PROUDFOOT, C., CHRISTIAN, M., VOYTAS, D. F., LONG, C. R., WHITELOW, C. B. & FAHRENKRUG, S. C. 2012. Efficient TALEN-mediated gene knockout in livestock. *Proc Natl Acad Sci U S A*, 109, 17382-7.

CARNEIRO, I. S., MENEZES, J. N. R., MAIA, J. A., MIRANDA, A. M., OLIVEIRA, V. B. S., MURRAY, J. D., MAGA, E. A., BERTOLINI, M. & BERTOLINI, L. R. 2018. Milk from transgenic goat expressing human lysozyme for recovery and treatment of gastrointestinal pathogens. *Eur J Pharm Sci*, 112, 79-86.

CHANAT, E., MARTIN, P. & OLLIVIER-BOUSQUET, M. 1999. Alpha(S1)-casein is required for the efficient transport of beta- and kappa-casein from the endoplasmic reticulum to the Golgi apparatus of mammary epithelial cells. *J Cell Sci*, 112 (Pt 19), 3399-412.

CHANDLER, M., DE LA CRUZ, F., DYDA, F., HICKMAN, A. B., MONCALIAN, G. & TON-HOANG, B. 2013. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat Rev Microbiol*, 11, 525-38.

CHAPMAN, R. S., LOURENCO, P. C., TONNER, E., FLINT, D. J., SELBERT, S., TAKEDA, K., AKIRA, S., CLARKE, A. R. & WATSON, C. J. 1999. Suppression of epithelial apoptosis and delayed mammary gland involution in mice with a conditional knockout of Stat3. *Genes Dev*, 13, 2604-16.

CHEN, B., GILBERT, L. A., CIMINI, B. A., SCHNITZBAUER, J., ZHANG, W., LI, G. W., PARK, J., BLACKBURN, E. H., WEISSMAN, J. S., QI, L. S. & HUANG, B. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 155, 1479-91.

CHEN, F., ROSIENE, J., CHE, A., BECKER, A. & LOTURCO, J. 2015. Tracking and transforming neocortical progenitors by CRISPR/Cas9 gene targeting and piggyBac transposase lineage labeling. *Development*, 142, 3601-11.

CLARK, A. J., ALI, S., ARCHIBALD, A. L., BESSOS, H., BROWN, P., HARRIS, S., MCCLENAGHAN, M., PROWSE, C., SIMONS, J. P., WHITELOW, C. B. & ET AL. 1989. The molecular manipulation of milk composition. *Genome*, 31, 950-5.

CLEMENS, J. A. L. Z. K.-J. W. A. 1981. Mesozoic mammals, the first two thirds of mammalian history. *Journal of Mammalogy*, 62, 443-445.

COCKRILL, W. R. 1981. The water buffalo: a review. *Br Vet J*, 137, 8-16.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339, 819-23.

COOPER, C. A., GARAS KLOBAS, L. C., MAGA, E. A. & MURRAY, J. D. 2013. Consuming transgenic goats' milk containing the antimicrobial protein lysozyme helps resolve diarrhea in young pigs. *PLoS One*, 8, e58409.

COOPER, C. A., MAGA, E. A. & MURRAY, J. D. 2014. Consumption of transgenic milk containing the antimicrobials lactoferrin and lysozyme separately and in conjunction by 6-week-old pigs improves intestinal and systemic health. *J Dairy Res*, 81, 30-7.

COOPER, C. A., MAGA, E. A. & MURRAY, J. D. 2015. Production of human lactoferrin and lysozyme in the milk of transgenic dairy animals: past, present, and future. *Transgenic Res*, 24, 605-14.

COWELL, I. G., SKINNER, A. & HURST, H. C. 1992. Transcriptional repression by a novel member of the bZIP family of transcription factors. *Mol Cell Biol*, 12, 3070-7.

COWIN, P. & WYSOLMERSKI, J. 2010. Molecular mechanisms guiding embryonic mammary gland development. *Cold Spring Harb Perspect Biol*, 2, a003251.

CRISPO, M., MULET, A. P., TESSON, L., BARRERA, N., CUADRO, F., DOS SANTOS-NETO, P. C., NGUYEN, T. H., CRENEGUY, A., BRUSSELLE, L., ANEGON, I. & MENCHACA, A. 2015. Efficient Generation of Myostatin Knock-Out Sheep Using CRISPR/Cas9 Technology and Microinjection into Zygotes. *PLoS One*, 10, e0136690.

CUI, Y., RIEDLINGER, G., MIYOSHI, K., TANG, W., LI, C., DENG, C. X., ROBINSON, G. W. & HENNIGHAUSEN, L. 2004. Inactivation of Stat5 in mouse

mammary epithelium during pregnancy reveals distinct functions in cell proliferation, survival, and differentiation. *Mol Cell Biol*, 24, 8037-47.

CURRIER, R. W. & WIDNESS, J. A. 2018. A Brief History of Milk Hygiene and Its Impact on Infant Mortality from 1875 to 1925 and Implications for Today: A Review. *J Food Prot*, 81, 1713-1722.

CURTIS, R. V., KIM, J. J., BAJRAMAJ, D. L., DOELMAN, J., OSBORNE, V. R. & CANT, J. P. 2014. Decline in mammary translational capacity during intravenous glucose infusion into lactating dairy cows. *J Dairy Sci*, 97, 430-8.

D'ARCO, F., TALENTI, G., LAKSHMANAN, R., STEPHENSON, K., SIDDIQUI, A. & CARNEY, O. 2017. Do Measurements of Inner Ear Structures Help in the Diagnosis of Inner Ear Malformations? A Review of Literature. *Otol Neurotol*, 38, e384-e392.

DEAKIN, J. E., DELBRIDGE, M. L., KOINA, E., HARLEY, N., ALSOP, A. E., WANG, C., PATEL, V. S. & GRAVES, J. A. 2013. Reconstruction of the ancestral marsupial karyotype from comparative gene maps. *BMC Evol Biol*, 13, 258.

DELTCHEVA, E., CHYLINSKI, K., SHARMA, C. M., GONZALES, K., CHAO, Y., PIRZADA, Z. A., ECKERT, M. R., VOGEL, J. & CHARPENTIER, E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471, 602-7.

DEMMER, J., STASIUK, S. J., GRIGOR, M. R., SIMPSON, K. J. & NICHOLAS, K. R. 2001. Differential expression of the whey acidic protein gene during lactation in the brushtail possum (*Trichosurus vulpecula*). *Biochim Biophys Acta*, 1522, 187-94.

DENMAN, J., HAYES, M., O'DAY, C., EDMUNDS, T., BARTLETT, C., HIRANI, S., EBERT, K. M., GORDON, K. & MCPHERSON, J. M. 1991. Transgenic expression of a variant of human tissue-type plasminogen activator in goat milk:

purification and characterization of the recombinant enzyme. *Biotechnology (N Y)*, 9, 839-43.

DEVINOY, E., THEPOT, D., STINNAKRE, M. G., FONTAINE, M. L., GRABOWSKI, H., PUISSANT, C., PAVIRANI, A. & HOUDEBINE, L. M. 1994. High level production of human growth hormone in the milk of transgenic mice: the upstream region of the rabbit whey acidic protein (WAP) gene targets transgene expression to the mammary gland. *Transgenic Res*, 3, 79-89.

DHOUAILLY, D. 2009. A new scenario for the evolutionary origin of hair, feather, and avian scales. *J Anat*, 214, 587-606.

DIAMOND, J. 1997. *Guns, Germs and Steel*.

DING, S., WU, X., LI, G., HAN, M., ZHUANG, Y. & XU, T. 2005. Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell*, 122, 473-83.

DJORDJEVIC, M., SENGUPTA, A. M. & SHRAIMAN, B. I. 2003. A biophysical approach to transcription factor binding site discovery. *Genome Res*, 13, 2381-90.

DRAGIN, S., CHRASTINOVA, L., MAKAREVICH, A. & CHRENEK, P. 2005. Production of recombinant human protein C in the milk of transgenic rabbits from the F3 generation. *Folia Biol (Krakow)*, 53, 129-32.

DRISCOLL, C. A., CLUTTON-BROCK, J., KITCHENER, A. C. & O'BRIEN, S. J. 2009. The Taming of the cat. Genetic and archaeological findings hint that wildcats became housecats earlier--and in a different place--than previously thought. *Sci Am*, 300, 68-75.

DU, S. J., GONG, Z. Y., FLETCHER, G. L., SHEARS, M. A., KING, M. J., IDLER, D. R. & HEW, C. L. 1992. Growth enhancement in transgenic Atlantic salmon by

the use of an "all fish" chimeric growth hormone gene construct. *Biotechnology (N Y)*, 10, 176-81.

EDMUNDS, T., VAN PATTEN, S. M., POLLOCK, J., HANSON, E., BERNASCONI, R., HIGGINS, E., MANAVALAN, P., ZIOMEK, C., MEADE, H., MCPHERSON, J. M. & COLE, E. S. 1998a. Transgenically produced human antithrombin: structural and functional comparison to human plasma-derived antithrombin. *Blood*, 91, 4561-71.

EDMUNDS, T., VAN PATTEN, S. M., POLLOCK, J., HANSON, E., BERNASCONI, R., HIGGINS, E., MANAVALAN, P., ZIOMEK, C., MEADE, H., MCPHERSON, J. M. & COLE, E. S. 1998b. Transgenically produced human antithrombin: Structural and functional comparison to human plasma-derived antithrombin. *Blood*, 91, 4561-4571.

EID, A. & MAHFOUZ, M. M. 2016. Genome editing: the road of CRISPR/Cas9 from bench to clinic. *Experimental and Molecular Medicine*, 48.

ERBSLOH, F., BERNSMEIER, A. & HILLESHEIM, H. 1958. [The glucose consumption of the brain & its dependence on the liver]. *Arch Psychiatr Nervenkr Z Gesamte Neurol Psychiatr*, 196, 611-26.

FAO 2013. Milk and dairy products in human nutrition. *In*: ORGANIZATION, F. A. A. (ed.). United States.

FENG, W., DAI, Y., MOU, L., COOPER, D. K., SHI, D. & CAI, Z. 2015. The potential of the combination of CRISPR/Cas9 and pluripotent stem cells to provide human organs from chimaeric pigs. *Int J Mol Sci*, 16, 6545-56.

FRANKE, W. W., LUDER, M. R., KARTENBECK, J., ZERBAN, H. & KEENAN, T. W. 1976. Involvement of vesicle coat material in casein secretion and surface regeneration. *J Cell Biol*, 69, 173-95.

FRAZER, K. A., ELNITSKI, L., CHURCH, D. M., DUBCHAK, I. & HARDISON, R. C. 2003. Cross-species sequence comparisons: a review of methods and available resources. *Genome Res*, 13, 1-12.

GALL, H., KALVERAM, C. M., SICK, H. & STERRY, W. 1996. Allergy to the heat-labile proteins alpha-lactalbumin and beta-lactoglobulin in mare's milk. *J Allergy Clin Immunol*, 97, 1304-7.

GALTON, F. 1865. Hereditary talent and character. *Macmillan's Magazine*.

GANGULI, N., GANGULI, N., USMANI, A. & MAJUMDAR, S. S. 2015. Isolation and functional characterization of buffalo (*Bubalus bubalis*) beta-casein promoter for driving mammary epithelial cell-specific gene expression. *J Biotechnol*, 198, 53-9.

GANTZ, V. M. & BIER, E. 2015. Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science*, 348, 442-4.

GARNEAU, J. E., DUPUIS, M. E., VILLION, M., ROMERO, D. A., BARRANGOU, R., BOYAVAL, P., FREMAUX, C., HORVATH, P., MAGADAN, A. H. & MOINEAU, S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468, 67-71.

GAUDELLI, N. M., KOMOR, A. C., REES, H. A., PACKER, M. S., BADRAN, A. H., BRYSON, D. I. & LIU, D. R. 2017. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature*, 551, 464-471.

GERMAN, J. B. & DILLARD, C. J. 2006. Composition, structure and absorption of milk lipids: a source of energy, fat-soluble nutrients and bioactive molecules. *Crit Rev Food Sci Nutr*, 46, 57-92.

GERSBACH, C. A., GAJ, T. & BARBAS, C. F., 3RD 2014. Synthetic zinc finger proteins: the advent of targeted gene regulation and genome modification technologies. *Acc Chem Res*, 47, 2309-18.

GJOREVSKI, N. & NELSON, C. M. 2011. Integrated morphodynamic signalling of the mammary gland. *Nat Rev Mol Cell Biol*, 12, 581-93.

GORDON, K., LEE, E., VITALE, J. A., SMITH, A. E., WESTPHAL, H. & HENNIGHAUSEN, L. 1992. Production of human tissue plasminogen activator in transgenic mouse milk. 1987. *Biotechnology*, 24, 425-8.

GRINEVICH, V., DESARMENIEN, M. G., CHINI, B., TAUBER, M. & MUSCATELLI, F. 2014. Ontogenesis of oxytocin pathways in the mammalian brain: late maturation and psychosocial disorders. *Front Neuroanat*, 8, 164.

GRUTZNER, F., NIXON, B. & JONES, R. C. 2008. Reproductive biology in egg-laying mammals. *Sex Dev*, 2, 115-27.

GU, K., YANG, B., TIAN, D., WU, L., WANG, D., SREEKALA, C., YANG, F., CHU, Z., WANG, G. L., WHITE, F. F. & YIN, Z. 2005. R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature*, 435, 1122-5.

GUO, J., WU, T., ANDERSON, J., KANE, B. F., JOHNSON, D. G., GORELICK, R. J., HENDERSON, L. E. & LEVIN, J. G. 2000. Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *J Virol*, 74, 8980-8.

GUO, Q., MINTIER, G., MA-EDMONDS, M., STORTON, D., WANG, X., XIAO, X., KIENZLE, B., ZHAO, D. & FEDER, J. N. 2018. 'Cold shock' increases the frequency of homology directed repair gene editing in induced pluripotent stem cells. *Sci Rep*, 8, 2080.

GURSOY, R. N. & BENITA, S. 2004. Self-emulsifying drug delivery systems (SEDDS) for improved oral delivery of lipophilic drugs. *Biomed Pharmacother*, 58, 173-82.

HALL, N. L., BUCHANAN, D. S., ANDERSON, V. L., ILSE, B. R., CARLIN, K. R. & BERG, E. P. 2011. Working chute behavior of feedlot cattle can be an indication of cattle temperament and beef carcass composition and quality. *Meat Sci*, 89, 52-7.

HALTERMAN, M. W., DE JESUS, C., REMPE, D. A., SCHOR, N. F. & FEDEROFF, H. J. 2008. Loss of c/EBP-beta activity promotes the adaptive to apoptotic switch in hypoxic cortical neurons. *Mol Cell Neurosci*, 38, 125-37.

HAN, Y., RIWANTO, M., GO, M. L. & EE, P. L. 2008. Modulation of breast cancer resistance protein (BCRP/ABCG2) by non-basic chalcone analogues. *Eur J Pharm Sci*, 35, 30-41.

HANAS, J. S., HAZUDA, D. J., BOGENHAGEN, D. F., WU, F. Y. & WU, C. W. 1983. Xenopus transcription factor A requires zinc for binding to the 5 S RNA gene. *J Biol Chem*, 258, 14120-5.

HARRISON, D. A. 2012. The Jak/STAT pathway. *Cold Spring Harb Perspect Biol*, 4.

HART, E. A., CACCAMO, M., HARROW, J. L., HUMPHRAY, S. J., GILBERT, J. G., TREVANION, S., HUBBARD, T., ROGERS, J. & ROTHSCHILD, M. F. 2007. Lessons learned from the initial sequencing of the pig genome: comparative analysis of an 8 Mb region of pig chromosome 17. *Genome Biol*, 8, R168.

HE, M., SUN, J., JIANG, Z. Q. & YANG, Y. X. 2017. Effects of cow's milk beta-casein variants on symptoms of milk intolerance in Chinese adults: a multicentre, randomised controlled study. *Nutr J*, 16, 72.

HE, Z., JIANG, L., ZHANG, T., ZHOU, M., WU, D., YUAN, T., YUAN, Y. & CHENG, Y. 2018. Efficient increase of the novel recombinant human plasminogen activator expression level and stability through the use of homozygote transgenic rabbits. *Int J Mol Med*, 42, 2269-2275.

HE, Z., PROUDFOOT, C., MILEHAM, A. J., MCLAREN, D. G., WHITELOW, C. B. & LILLICO, S. G. 2015. Highly efficient targeted chromosome deletions using CRISPR/Cas9. *Biotechnol Bioeng*, 112, 1060-4.

HE, Z., PROUDFOOT, C., WHITELOW, C. B. & LILLICO, S. G. 2016. Comparison of CRISPR/Cas9 and TALENs on editing an integrated EGFP gene in the genome of HEK293FT cells. *Springerplus*, 5, 814.

HEBERLEIN, M. T., TURNER, D. C., RANGE, F. & VIRANYI, Z. 2016. A comparison between wolves, *Canis lupus*, and dogs, *Canis familiaris*, in showing behaviour towards humans. *Anim Behav*, 122, 59-66.

HEINRICHS, M. & DOMES, G. 2008. Neuropeptides and social behaviour: effects of oxytocin and vasopressin in humans. *Prog Brain Res*, 170, 337-50.

HELMERHORST, E. J., TRABOULSI, G., SALIH, E. & OPPENHEIM, F. G. 2010. Mass spectrometric identification of key proteolytic cleavage sites in statherin affecting mineral homeostasis and bacterial binding domains. *J Proteome Res*, 9, 5413-21.

HENNESSY, M. B., KAISER, S., TIEDTKE, T. & SACHSER, N. 2015. Stability and change: Stress responses and the shaping of behavioral phenotypes over the life span. *Front Zool*, 12 Suppl 1, S18.

HENNIGHAUSEN, L. G. & SIPPEL, A. E. 1982. Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins. *Nucleic Acids Res*, 10, 2677-84.

HERBECK YE, G. R. 2017. Neuropeptides as facilitators of domestication. *Cell & tissue research*, 1-13.

HERRON, L. R., PRIDANS, C., TURNBULL, M. L., SMITH, N., LILLICO, S., SHERMAN, A., GILHOOLEY, H. J., WEAR, M., KURIAN, D., PAPADAKOS, G., DIGARD, P., HUME, D. A., GILL, A. C. & SANG, H. M. 2018. A chicken bioreactor for efficient production of functional cytokines. *BMC Biotechnol*, 18, 82.

HIROSE, M., HASEGAWA, A., MOCHIDA, K., MATOBA, S., HATANAKA, Y., INOUE, K., GOTO, T., KANEDA, H., YAMADA, I., FURUSE, T., ABE, K., UENOYAMA, Y., TSUKAMURA, H., WAKANA, S., HONDA, A. & OGURA, A. 2017. CRISPR/Cas9-mediated genome editing in wild-derived mice: generation of tamed wild-derived strains by mutation of the a (nonagouti) gene. *Sci Rep*, 7, 42476.

HOCKEMEYER, D., WANG, H., KIANI, S., LAI, C. S., GAO, Q., CASSADY, J. P., COST, G. J., ZHANG, L., SANTIAGO, Y., MILLER, J. C., ZEITLER, B., CHERONE, J. M., MENG, X., HINKLEY, S. J., REBAR, E. J., GREGORY, P. D., URNOV, F. D. & JAENISCH, R. 2011. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol*, 29, 731-4.

HOLLYWOOD, J. A., LEE, C. M., SCALLAN, M. F. & HARRISON, P. T. 2016. Analysis of gene repair tracts from Cas9/gRNA double-stranded breaks in the human CFTR gene. *Sci Rep*, 6, 32230.

HORVATH, C. M., WEN, Z. & DARNELL, J. E., JR. 1995. A STAT protein domain that determines DNA sequence recognition suggests a novel DNA-binding domain. *Genes Dev*, 9, 984-94.

HOU, J., SCHINDLER, U., HENZEL, W. J., WONG, S. C. & MCKNIGHT, S. L. 1995. Identification and purification of human Stat proteins activated in response to interleukin-2. *Immunity*, 2, 321-9.

HOU, S., LANDEGO, I., JAYACHANDRAN, N., MILLER, A., GIBSON, I. W., AMBROSE, C. & MARSHALL, A. J. 2014. Follicular dendritic cell secreted protein FDC-SP controls IgA production. *Mucosal Immunol*, 7, 948-57.

HOUEBINE, L. M. 2009. Production of pharmaceutical proteins by transgenic animals. *Comp Immunol Microbiol Infect Dis*, 32, 107-21.

HOUEBINE, L. M. 2018. Production of pharmaceutical proteins by transgenic animals. *Rev Sci Tech*, 37, 131-139.

HUGHES, R. L. 1993. Monotreme development with particular reference to the extraembryonic membranes. *J Exp Zool*, 266, 480-94.

ISHINO, Y., SHINAGAWA, H., MAKINO, K., AMEMURA, M. & NAKATA, A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol*, 169, 5429-33.

JABED, A., WAGNER, S., MCCracken, J., WELLS, D. N. & LAIBLE, G. 2012. Targeted microRNA expression in dairy cattle directs production of beta-lactoglobulin-free, high-casein milk. *Proc Natl Acad Sci U S A*, 109, 16811-6.

JACOBS, L. L., WINKLER, D. A. & MURRY, P. A. 1989. Modern mammal origins: evolutionary grades in the Early Cretaceous of North America. *Proc Natl Acad Sci U S A*, 86, 4992-5.

JAENISCH, R. & MINTZ, B. 1974. Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. *Proc Natl Acad Sci U S A*, 71, 1250-4.

JAMES, R. M., NEIL, C., WEBSTER, J., ROOS, S., CLARK, A. J. & WHITELAW, C. B. 2000. Multiple copies of beta-lactoglobulin promoter do not function as LCR. *Biochem Biophys Res Commun*, 272, 284-9.

JAMIESON, A. C., KIM, S. H. & WELLS, J. A. 1994. In vitro selection of zinc fingers with altered DNA-binding specificity. *Biochemistry*, 33, 5689-95.

JANNE, J., ALHONEN, L. & LEINONEN, P. 1991. Polyamines: from molecular biology to clinical applications. *Ann Med*, 23, 241-59.

JANSEN, R., EMBDEN, J. D., GAASTRA, W. & SCHOULS, L. M. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*, 43, 1565-75.

JIANQIN, S., LEIMING, X., LU, X., YELLAND, G. W., NI, J. & CLARKE, A. J. 2016. Effects of milk containing only A2 beta casein versus milk containing both A1 and A2 beta casein proteins on gastrointestinal physiology, symptoms of discomfort, and cognitive behavior of people with self-reported intolerance to traditional cows' milk. *Nutr J*, 15, 35.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-21.

JINEK, M., JIANG, F., TAYLOR, D. W., STERNBERG, S. H., KAYA, E., MA, E., ANDERS, C., HAUER, M., ZHOU, K., LIN, S., KAPLAN, M., IAVARONE, A. T., CHARPENTIER, E., NOGALES, E. & DOUDNA, J. A. 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, 343, 1247997.

JOHANSSON, I., BRATT, P., HAY, D. I., SCHLUCKEBIER, S. & STROMBERG, N. 2000. Adhesion of *Candida albicans*, but not *Candida krusei*, to salivary statherin and mimicking host molecules. *Oral Microbiol Immunol*, 15, 112-8.

JOHNSON, K. A. & JOHNSON, D. E. 1995. Methane emissions from cattle. *J Anim Sci*, 73, 2483-92.

- JONES, W. K., YU-LEE, L. Y., CLIFT, S. M., BROWN, T. L. & ROSEN, J. M. 1985. The rat casein multigene family. Fine structure and evolution of the beta-casein gene. *J Biol Chem*, 260, 7042-50.
- KAIMALA, S. & KUMAR, S. 2015. An evolutionarily conserved non-coding element in casein locus acts as transcriptional repressor. *Gene*, 554, 75-80.
- KAMINSKI, S., CIESLINSKA, A. & KOSTYRA, E. 2007. Polymorphism of bovine beta-casein and its potential effect on human health. *J Appl Genet*, 48, 189-98.
- KATO, T., HARA, S., GOTO, Y., OGAWA, Y., OKAYASU, H., KUBOTA, S., TAMANO, M., TERAOKA, M. & TAKADA, S. 2017. Creation of mutant mice with megabase-sized deletions containing custom-designed breakpoints by means of the CRISPR/Cas9 system. *Sci Rep*, 7, 59.
- KAWASAKI, K., LAFONT, A. G. & SIRE, J. Y. 2011. The evolution of milk casein genes from tooth genes before the origin of mammals. *Mol Biol Evol*, 28, 2053-61.
- KAWASAKI, K., SUZUKI, T. & WEISS, K. M. 2005. Phenogenetic drift in evolution: the changing genetic basis of vertebrate teeth. *Proc Natl Acad Sci U S A*, 102, 18063-8.
- KIM, Y. G., CHA, J. & CHANDRASEGARAN, S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A*, 93, 1156-60.
- KIM, Y. W. & KIM, A. 2017. Deletion of transcription factor binding motifs using the CRISPR/spCas9 system in the beta-globin LCR. *Biosci Rep*.
- KISHNANI, P. S., STEINER, R. D., BALI, D., BERGER, K., BYRNE, B. J., CASE, L. E., CROWLEY, J. F., DOWNS, S., HOWELL, R. R., KRAVITZ, R. M., MACKEY, J., MARSDEN, D., MARTINS, A. M., MILLINGTON, D. S., NICOLINO, M.,

- O'GRADY, G., PATTERSON, M. C., RAPOPORT, D. M., SLONIM, A., SPENCER, C. T., TIFFT, C. J. & WATSON, M. S. 2006. Pompe disease diagnosis and management guideline. *Genet Med*, 8, 267-88.
- KLEINSTIVER, B. P., PATTANAYAK, V., PREW, M. S., TSAI, S. Q., NGUYEN, N. T., ZHENG, Z. & JOUNG, J. K. 2016. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 529, 490-5.
- KOLB, A. F., HUBER, R. C., LILLICO, S. G., CARLISLE, A., ROBINSON, C. J., NEIL, C., PETRIE, L., SORENSEN, D. B., OLSSON, I. A. & WHITELAW, C. B. 2011. Milk lacking alpha-casein leads to permanent reduction in body size in mice. *PLoS One*, 6, e21775.
- KONTOPIDIS, G., HOLT, C. & SAWYER, L. 2004. Invited review: beta-lactoglobulin: binding properties, structure, and function. *J Dairy Sci*, 87, 785-96.
- KRISHNA, S. S., MAJUMDAR, I. & GRISHIN, N. V. 2003. Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res*, 31, 532-50.
- KRUIF, C. G. D. 1998. Supra-aggregates of Casein Micelles as a Prelude to Coagulation. *Journal of dairy science*, 81, 3019–3028.
- KUES, W. A. & NIEMANN, H. 2011. Advances in farm animal transgenesis. *Prev Vet Med*, 102, 146-56.
- KUHN, N. J., CARRICK, D. T. & WILDE, C. J. 1980. Lactose synthesis: the possibilities of regulation. *J Dairy Sci*, 63, 328-36.
- KUMAR, S., CLARKE, A. R., HOOPER, M. L., HORNE, D. S., LAW, A. J., LEAVER, J., SPRINGBETT, A., STEVENSON, E. & SIMONS, J. P. 1994. Milk composition and lactation of beta-casein-deficient mice. *Proc Natl Acad Sci U S A*, 91, 6138-42.

KUMAR, S., NAGARAJAN, M., SANDHU, J. S., KUMAR, N., BEHL, V. & NISHANTH, G. 2007. Mitochondrial DNA analyses of Indian water buffalo support a distinct genetic origin of river and swamp buffalo. *Anim Genet*, 38, 227-32.

LAGASSE, H. A., ALEXAKI, A., SIMHADRI, V. L., KATAGIRI, N. H., JANKOWSKI, W., SAUNA, Z. E. & KIMCHI-SARFATY, C. 2017. Recent advances in (therapeutic protein) drug development. *F1000Res*, 6, 113.

LAIBLE, G., SMOLENSKI, G., WHEELER, T. & BROPHY, B. 2016. Increased gene dosage for beta- and kappa-casein in transgenic cattle improves milk composition through complex effects. *Sci Rep*, 6, 37607.

LAMB, B. M., MERCER, A. C. & BARBAS, C. F., 3RD 2013. Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. *Nucleic Acids Res*, 41, 9779-85.

LANDRETH, K. S. 2002. Critical windows in development of the rodent immune system. *Hum Exp Toxicol*, 21, 493-8.

LARSON, G., KARLSSON, E. K., PERRI, A., WEBSTER, M. T., HO, S. Y., PETERS, J., STAHL, P. W., PIPER, P. J., LINGAAS, F., FREDHOLM, M., COMSTOCK, K. E., MODIANO, J. F., SCHELLING, C., AGOULNIK, A. I., LEEGWATER, P. A., DOBNEY, K., VIGNE, J. D., VILA, C., ANDERSSON, L. & LINDBLAD-TOH, K. 2012. Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc Natl Acad Sci U S A*, 109, 8878-83.

LATCHMAN, D. S. 1993. Transcription factors: an overview. *Int J Exp Pathol*, 74, 417-22.

LAU, E. M., LYNN, H., CHAN, Y. H., LAU, W. & WOO, J. 2004. Benefits of milk powder supplementation on bone accretion in Chinese children. *Osteoporos Int*, 15, 654-8.

LEE, H. K., LEE, D. S., RYOO, H. M., PARK, J. T., PARK, S. J., BAE, H. S., CHO, M. I. & PARK, J. C. 2010. The odontogenic ameloblast-associated protein (ODAM) cooperates with RUNX2 and modulates enamel mineralization via regulation of MMP-20. *J Cell Biochem*, 111, 755-67.

LEE, M., MEHTA, R. & LUCIA, S. P. 1962. Amino acid composition and nutritive value of goat milk casein. *Proc Soc Exp Biol Med*, 110, 115-8.

LEE, M. S., GIPPERT, G. P., SOMAN, K. V., CASE, D. A. & WRIGHT, P. E. 1989. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science*, 245, 635-7.

LEFEVRE, C. M., SHARP, J. A. & NICHOLAS, K. R. 2009. Characterisation of monotreme caseins reveals lineage-specific expansion of an ancestral casein locus in mammals. *Reprod Fertil Dev*, 21, 1015-27.

LEVINSON, Y., ISH-SHALOM, S., SEGAL, E. & LIVNEY, Y. D. 2016. Bioavailability, rheology and sensory evaluation of fat-free yogurt enriched with VD3 encapsulated in re-assembled casein micelles. *Food Funct*, 7, 1477-82.

LI, P., MITRA, S., SPOLSKI, R., OH, J., LIAO, W., TANG, Z., MO, F., LI, X., WEST, E. E., GROMER, D., LIN, J. X., LIU, C., RUAN, Y. & LEONARD, W. J. 2017a. STAT5-mediated chromatin interactions in superenhancers activate IL-2 highly inducible genes: Functional dissection of the Il2ra gene locus. *Proc Natl Acad Sci U S A*, 114, 12111-12119.

LI, S., WANG, L., HE, X., XIE, Y. & ZHANG, Z. 2015. Identification and analysis of the promoter region of the STGC3 gene. *Arch Med Sci*, 11, 1095-100.

LI, S., ZHANG, A., XUE, H., LI, D. & LIU, Y. 2017b. One-Step piggyBac Transposon-Based CRISPR/Cas9 Activation of Multiple Genes. *Mol Ther Nucleic Acids*, 8, 64-76.

LI, Y., WANG, G. D., WANG, M. S., IRWIN, D. M., WU, D. D. & ZHANG, Y. P. 2014. Domestication of the dog from the wolf was promoted by enhanced excitatory synaptic plasticity: a hypothesis. *Genome Biol Evol*, 6, 3115-21.

LIANG, P., XU, Y., ZHANG, X., DING, C., HUANG, R., ZHANG, Z., LV, J., XIE, X., CHEN, Y., LI, Y., SUN, Y., BAI, Y., SONGYANG, Z., MA, W., ZHOU, C. & HUANG, J. 2015. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein Cell*, 6, 363-372.

LILLEGRAVEN, J. A., THOMPSON, S. D., MCNAB, B. K. & PATTON, J. L. 1987. The Origin of Eutherian Mammals. *Biological Journal of the Linnean Society*, 32, 281-336.

LILLICO, S. G., PROUDFOOT, C., CARLSON, D. F., STVERAKOVA, D., NEIL, C., BLAIN, C., KING, T. J., RITCHIE, W. A., TAN, W., MILEHAM, A. J., MCLAREN, D. G., FAHRENKRUG, S. C. & WHITELAW, C. B. 2013. Live pigs produced from genome edited zygotes. *Sci Rep*, 3, 2847.

LIN, J. X., DU, N., LI, P., KAZEMIAN, M., GEBREGIORGIS, T., SPOLSKI, R. & LEONARD, W. J. 2017. Critical functions for STAT5 tetramers in the maturation and survival of natural killer cells. *Nat Commun*, 8, 1320.

LINDERSTROM-LANG K, K. S. 1925. Studies on casein. I. On the solubility of caseins in hydrochloric acid. *Trav Lab Carls- berg* 16:1 – 47

LIU, X., ROBINSON, G. W., GOUILLEUX, F., GRONER, B. & HENNIGHAUSEN, L. 1995. Cloning and expression of Stat5 and an additional homologue (Stat5b) involved in prolactin signal transduction in mouse mammary tissue. *Proc Natl Acad Sci U S A*, 92, 8831-5.

LIU, X., ROBINSON, G. W. & HENNIGHAUSEN, L. 1996. Activation of Stat5a and Stat5b by tyrosine phosphorylation is tightly linked to mammary gland differentiation. *Mol Endocrinol*, 10, 1496-506.

LIU, X., ROBINSON, G. W., WAGNER, K. U., GARRETT, L., WYNshaw-BORIS, A. & HENNIGHAUSEN, L. 1997. Stat5a is mandatory for adult mammary gland development and lactogenesis. *Genes Dev*, 11, 179-86.

LOVEDAY, S. M., RAO, M. A., CREAMER, L. K. & SINGH, H. 2009. Factors affecting rheological characteristics of fibril gels: the case of beta-lactoglobulin and alpha-lactalbumin. *J Food Sci*, 74, R47-55.

LUO, Z. X. 2007. Transformation and diversification in early mammal evolution. *Nature*, 450, 1011-9.

LYDERSEN, C. & KOVACS, K. M. 1996. Energetics of lactation in harp seals (*Phoca groenlandica*) from the Gulf of St. Lawrence, Canada. *J Comp Physiol B*, 166, 295-304.

MA, H., MARTI-GUTIERREZ, N., PARK, S. W., WU, J., LEE, Y., SUZUKI, K., KOSKI, A., JI, D., HAYAMA, T., AHMED, R., DARBY, H., VAN DYKEN, C., LI, Y., KANG, E., PARK, A. R., KIM, D., KIM, S. T., GONG, J., GU, Y., XU, X., BATTAGLIA, D., KRIEG, S. A., LEE, D. M., WU, D. H., WOLF, D. P., HEITNER, S. B., BELMONTE, J. C. I., AMATO, P., KIM, J. S., KAUL, S. & MITALIPOV, S. 2017. Correction of a pathogenic gene mutation in human embryos. *Nature*, 548, 413-419.

MACIAS, H. & HINCK, L. 2012. Mammary gland development. *Wiley Interdiscip Rev Dev Biol*, 1, 533-57.

MAGA, E. A., CULLOR, J. S., SMITH, W., ANDERSON, G. B. & MURRAY, J. D. 2006. Human lysozyme expressed in the mammary gland of transgenic dairy goats can inhibit the growth of bacteria that cause mastitis and the cold-spoilage of milk. *Foodborne Pathog Dis*, 3, 384-92.

- MAGA, E. A., DESAI, P. T., WEIMER, B. C., DAO, N., KULTZ, D. & MURRAY, J. D. 2012. Consumption of lysozyme-rich milk can alter microbial fecal populations. *Appl Environ Microbiol*, 78, 6153-60.
- MAKAROVA, K. S., GRISHIN, N. V., SHABALINA, S. A., WOLF, Y. I. & KOONIN, E. V. 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*, 1, 7.
- MAKINO, S., SUGIO, A., WHITE, F. & BOGDANOVE, A. J. 2006. Inhibition of resistance gene-mediated defense in rice by *Xanthomonas oryzae* pv. *oryzicola*. *Mol Plant Microbe Interact*, 19, 240-9.
- MALEWSKI, T. 1998. Computer analysis of distribution of putative cis- and trans-regulatory elements in milk protein gene promoters. *Biosystems*, 45, 29-44.
- MANDAL, A., BISHT, R., RUPENTHAL, I. D. & MITRA, A. K. 2017. Polymeric micelles for ocular drug delivery: From structural frameworks to recent preclinical studies. *J Control Release*, 248, 96-116.
- MARDH, C. K., BACKMAN, B., HOLMGREN, G., HU, J. C., SIMMER, J. P. & FORSMAN-SEMB, K. 2002. A nonsense mutation in the enamel gene causes local hypoplastic autosomal dominant amelogenesis imperfecta (AIH2). *Hum Mol Genet*, 11, 1069-74.
- MASSCHALCK, B. & MICHIELS, C. W. 2003. Antimicrobial properties of lysozyme in relation to foodborne vegetative bacteria. *Crit Rev Microbiol*, 29, 191-214.
- MASSON, P. L. & HEREMANS, J. F. 1971. Lactoferrin in milk from different species. *Comp Biochem Physiol B*, 39, 119-29.

MENENDEZ-AGUIRRE, O., KESSLER, A., STUETZ, W., GRUNE, T., WEISS, J. & HINRICHS, J. 2014. Increased loading of vitamin D2 in reassembled casein micelles with temperature-modulated high pressure treatment. *Food Res Int*, 64, 74-80.

MERRILEES, M. J., SMART, E. J., GILCHRIST, N. L., FRAMPTON, C., TURNER, J. G., HOOKE, E., MARCH, R. L. & MAGUIRE, P. 2000. Effects of dairy food supplements on bone mineral density in teenage girls. *Eur J Nutr*, 39, 256-62.

MILLER, J. C., TAN, S., QIAO, G., BARLOW, K. A., WANG, J., XIA, D. F., MENG, X., PASCHON, D. E., LEUNG, E., HINKLEY, S. J., DULAY, G. P., HUA, K. L., ANKOUDINOVA, I., COST, G. J., URNOV, F. D., ZHANG, H. S., HOLMES, M. C., ZHANG, L., GREGORY, P. D. & REBAR, E. J. 2011. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*, 29, 143-8.

MOFFATT, P., SMITH, C. E., ST-ARNAUD, R. & NANJI, A. 2008. Characterization of Apin, a secreted protein highly expressed in tooth-associated epithelia. *J Cell Biochem*, 103, 941-56.

MORRIS, G. A., FOSTER, T. J. & HARDING, S. E. 2000. Further observations on the size, shape, and hydration of casein micelles from novel analytical ultracentrifuge and capillary viscometry approaches. *Biomacromolecules*, 1, 764-7.

MORRISON, B. & CUTLER, M. L. 2009. Mouse Mammary Epithelial Cells form Mammospheres During Lactogenic Differentiation. *J Vis Exp*.

MUELLER, S. O., CLARK, J. A., MYERS, P. H. & KORACH, K. S. 2002. Mammary gland development in adult mice requires epithelial and stromal estrogen receptor alpha. *Endocrinology*, 143, 2357-65.

NADERI, S., REZAEI, H. R., POMPANON, F., BLUM, M. G., NEGRINI, R., NAGHASH, H. R., BALKIZ, O., MASHKOUR, M., GAGGIOTTI, O. E., AJMONE-

MARSAN, P., KENCE, A., VIGNE, J. D. & TABERLET, P. 2008. The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proc Natl Acad Sci U S A*, 105, 17659-64.

NAGARAJAN, M., NIMISHA, K. & KUMAR, S. 2015. Mitochondrial DNA Variability of Domestic River Buffalo (*Bubalus bubalis*) Populations: Genetic Evidence for Domestication of River Buffalo in Indian Subcontinent. *Genome Biol Evol*, 7, 1252-9.

NAGASAWA, M., MITSUI, S., EN, S., OHTANI, N., OHTA, M., SAKUMA, Y., ONAKA, T., MOGI, K. & KIKUSUI, T. 2015. Social evolution. Oxytocin-gaze positive loop and the coevolution of human-dog bonds. *Science*, 348, 333-6.

NANDA, A. S., BRAR, P. S. & PRABHAKAR, S. 2003. Enhancing reproductive performance in dairy buffalo: major constraints and achievements. *Reprod Suppl*, 61, 27-36.

NEMUDRYI, A. A., VALETDINOVA, K. R., MEDVEDEV, S. P. & ZAKIAN, S. M. 2014. TALEN and CRISPR/Cas Genome Editing Systems: Tools of Discovery. *Acta Naturae*, 6, 19-40.

NICHOLAS, K., SHARP, J., WATT, A., WANYONYI, S., CROWLEY, T., GILLESPIE, M. & LEFEVRE, C. 2012. The tammar wallaby: a model system to examine domain-specific delivery of milk protein bioactives. *Semin Cell Dev Biol*, 23, 547-56.

NISHIMASU, H., RAN, F. A., HSU, P. D., KONERMANN, S., SHEHATA, S. I., DOHMAE, N., ISHITANI, R., ZHANG, F. & NUREKI, O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, 156, 935-49.

NOBLE, M. S., RODRIGUEZ-ZAS, S., COOK, J. B., BLECK, G. T., HURLEY, W. L. & WHEELER, M. B. 2002. Lactational performance of first-parity transgenic gilts expressing bovine alpha-lactalbumin in their milk. *J Anim Sci*, 80, 1090-6.

O'MAHONY, P. F. F. T. U.-L. P. H. L. M. J. A. 1998. *Dairy chemistry and biochemistry*, London, Blackie Academic & Professional.

O'NEIL, C. E., NICKLAS, T. A., LIU, Y. & FRANKLIN, F. A. 2009. The impact of dairy product consumption on nutrient adequacy and weight of Head Start mothers. *Public Health Nutr*, 12, 1693-701.

OFTEDAL, O. T. 2002. The mammary gland and its origin during synapsid evolution. *J Mammary Gland Biol Neoplasia*, 7, 225-52.

OFTEDAL, O. T. 2012. The evolution of milk secretion and its ancient origins. *Animal*, 6, 355-68.

OKA, T., PERRY, J. W. & TOPPER, Y. J. 1974. Changes in insulin responsiveness during development of mammary epithelium. *J Cell Biol*, 62, 550-6.

OSBORN, J. W. & HILLMAN, J. 1979. Enamel structure in some therapsids and mesozoic mammals. *Calcif Tissue Int*, 29, 47-61.

PABO, C. O., PEISACH, E. & GRANT, R. A. 2001. Design and selection of novel Cys2His2 zinc finger proteins. *Annu Rev Biochem*, 70, 313-40.

PAINE, I. S. & LEWIS, M. T. 2017. The Terminal End Bud: the Little Engine that Could. *J Mammary Gland Biol Neoplasia*, 22, 93-108.

PALMER, C. A., NEVILLE, M. C., ANDERSON, S. M. & MCMANAMAN, J. L. 2006. Analysis of lactation defects in transgenic mice. *J Mammary Gland Biol Neoplasia*, 11, 269-82.

PALMITER, R. D., HAMMER, R. E. & BRINSTER, R. L. 1985. Expression of Growth Hormone Genes in Transgenic Mice. *Banbury Rep*, 20, 123-132.

- PAVLETICH, N. P. & PABO, C. O. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*, 252, 809-17.
- PENA, R. N. & WHITELOW, C. B. 2005. Duplication of Stat5-binding sites within the beta-lactoglobulin promoter compromises transcription in vivo. *Biochimie*, 87, 523-8.
- PERLMAN, R. L. 2016. Mouse models of human disease: An evolutionary perspective. *Evol Med Public Health*, 2016, 170-6.
- PETERSON, N. A., ANDERSON, B. F., JAMESON, G. B., TWEEDIE, J. W. & BAKER, E. N. 2000. Crystal structure and iron-binding properties of the R210K mutant of the N-lobe of human lactoferrin: implications for iron release from transferrins. *Biochemistry*, 39, 6625-33.
- PETTERS, R. M., ALEXANDER, C. A., WELLS, K. D., COLLINS, E. B., SOMMER, J. R., BLANTON, M. R., ROJAS, G., HAO, Y., FLOWERS, W. L., BANIN, E., CIDECIYAN, A. V., JACOBSON, S. G. & WONG, F. 1997. Genetically engineered large animal model for studying cone photoreceptor survival and degeneration in retinitis pigmentosa. *Nat Biotechnol*, 15, 965-70.
- PICKERING, M. & JONES, J. F. 2002. The diaphragm: two physiological muscles in one. *J Anat*, 201, 305-12.
- PILISZEK, A., KWON, G. S. & HADJANTONAKIS, A. K. 2011. Ex utero culture and live imaging of mouse embryos. *Methods Mol Biol*, 770, 243-57.
- PLATENBURG, G. J., KOOTWIJK, E. P., KOOIMAN, P. M., WOLOSHUK, S. L., NUIJENS, J. H., KRIMPENFORT, P. J., PIEPER, F. R., DE BOER, H. A. & STRIJKER, R. 1994. Expression of human lactoferrin in milk of transgenic mice. *Transgenic Res*, 3, 99-108.

POULIN, F., NOBREGA, M. A., PLAJSER-FRICK, I., HOLT, A., AFZAL, V., RUBIN, E. M. & PENNACCHIO, L. A. 2005. In vivo characterization of a vertebrate ultraconserved enhancer. *Genomics*, 85, 774-81.

PRAKOBPHOL, A., XU, F., HOANG, V. M., LARSSON, T., BERGSTROM, J., JOHANSSON, I., FRANGSMYR, L., HOLMSKOV, U., LEFFLER, H., NILSSON, C., BOREN, T., WRIGHT, J. R., STROMBERG, N. & FISHER, S. J. 2000. Salivary agglutinin, which binds *Streptococcus mutans* and *Helicobacter pylori*, is the lung scavenger receptor cysteine-rich protein gp-340. *J Biol Chem*, 275, 39860-6.

PROUDFOOT, C., CARLSON, D. F., HUDDART, R., LONG, C. R., PRYOR, J. H., KING, T. J., LILLICO, S. G., MILEHAM, A. J., MCLAREN, D. G., WHITELAW, C. B. & FAHRENKRUG, S. C. 2015. Genome edited sheep and cattle. *Transgenic Res*, 24, 147-53.

PRUNET, P., BOEUF, G. & HOUDEBINE, L. M. 1985. Plasma and pituitary prolactin levels in rainbow trout during adaptation to different salinities. *J Exp Zool*, 235, 187-96.

PURSEL, V. G., BOLT, D. J., MILLER, K. F., PINKERT, C. A., HAMMER, R. E., PALMITER, R. D. & BRINSTER, R. L. 1990. Expression and performance in transgenic pigs. *J Reprod Fertil Suppl*, 40, 235-45.

PURSEL, V. G., PINKERT, C. A., MILLER, K. F., BOLT, D. J., CAMPBELL, R. G., PALMITER, R. D., BRINSTER, R. L. & HAMMER, R. E. 1989. Genetic engineering of livestock. *Science*, 244, 1281-8.

QIAN, X., KRAFT, J., NI, Y. & ZHAO, F. Q. 2014. Production of recombinant human proinsulin in the milk of transgenic mice. *Sci Rep*, 4, 6465.

QUINA, L. A., KURAMOTO, T., LUQUETTI, D. V., COX, T. C., SERIKAWA, T. & TURNER, E. E. 2012. Deletion of a conserved regulatory element required for

Hmx1 expression in craniofacial mesenchyme in the dumbo rat: a newly identified cause of congenital ear malformation. *Dis Model Mech*, 5, 812-22.

REBAR, E. J. & PABO, C. O. 1994. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science*, 263, 671-3.

RENFREE, M. B. 2010. Review: Marsupials: placental mammals with a difference. *Placenta*, 31 Suppl, S21-6.

RENFREE, M. B., O, W. S., SHORT, R. V. & SHAW, G. 1996. Sexual differentiation of the urogenital system of the fetal and neonatal tammar wallaby, *Macropus eugenii*. *Anat Embryol (Berl)*, 194, 111-34.

RIDDLE O, B. R., DYKSHORN SW 1933. The preparation, identification and assay of prolactin – a hormone of the anterior pituitary. *American Journal of Physiology*, 105, 191- 216.

RIJNKELS, M. 2002. Multispecies comparison of the casein gene loci and evolution of casein gene family. *J Mammary Gland Biol Neoplasia*, 7, 327-45.

RIJNKELS, M., ELNITSKI, L., MILLER, W. & ROSEN, J. M. 2003. Multispecies comparative analysis of a mammalian-specific genomic domain encoding secretory proteins. *Genomics*, 82, 417-32.

RIJNKELS, M., KOOIMAN, P. M., KRIMPENFORT, P. J., DE BOER, H. A. & PIEPER, F. R. 1995. Expression analysis of the individual bovine beta-, alpha s2- and kappa-casein genes in transgenic mice. *Biochem J*, 311 (Pt 3), 929-37.

RIJNKELS, M., WHEELER, D. A., DE BOER, H. A. & PIEPER, F. R. 1997. Structure and expression of the mouse casein gene locus. *Mamm Genome*, 8, 9-15.

- RITCHIE, H., REAY, D. & HIGGINS, P. 2018. Sustainable food security in India- Domestic production and macronutrient availability. *PLoS One*, 13, e0193766.
- ROBERT, K. A. & BRAUN, S. 2012. Milk composition during lactation suggests a mechanism for male biased allocation of maternal resources in the tammar wallaby (*Macropus eugenii*). *PLoS One*, 7, e51099.
- ROBERTS, J. M. & D'ABARNO, J. 1996. Effects of calcium supplementation on pregnancy-induced hypertension. *JAMA*, 276, 1386-7; author reply 1388.
- ROBINSON, C. V., GROSS, M., EYLES, S. J., EWBANK, J. J., MAYHEW, M., HARTL, F. U., DOBSON, C. M. & RADFORD, S. E. 1994. Conformation of GroEL-bound alpha-lactalbumin probed by mass spectrometry. *Nature*, 372, 646-51.
- ROCHMAN, Y., SPOLSKI, R. & LEONARD, W. J. 2009. New insights into the regulation of T cells by gamma(c) family cytokines. *Nat Rev Immunol*, 9, 480-90.
- ROEHE, R., DEWHURST, R. J., DUTHIE, C. A., ROOKE, J. A., MCKAIN, N., ROSS, D. W., HYSLOP, J. J., WATERHOUSE, A., FREEMAN, T. C., WATSON, M. & WALLACE, R. J. 2016. Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance. *PLoS Genet*, 12, e1005846.
- ROWE, T., RICH, T. H., VICKERS-RICH, P., SPRINGER, M. & WOODBURNE, M. O. 2008. The oldest platypus and its bearing on divergence timing of the platypus and echidna clades. *Proc Natl Acad Sci U S A*, 105, 1238-42.
- ROWZEE, A. M., LUDWIG, D. L. & WOOD, T. L. 2009. Insulin-like growth factor type 1 receptor and insulin receptor isoform expression and signaling in mammary epithelial cells. *Endocrinology*, 150, 3611-9.

- RUDOLPH, M. C., MCMANAMAN, J. L., HUNTER, L., PHANG, T. & NEVILLE, M. C. 2003. Functional development of the mammary gland: use of expression profiling and trajectory clustering to reveal changes in gene expression during pregnancy, lactation, and involution. *J Mammary Gland Biol Neoplasia*, 8, 287-307.
- RUDOLPH, N. S. 1999. Biopharmaceutical production in transgenic livestock. *Trends Biotechnol*, 17, 367-74.
- RUTLEDGE, J. C. 1997. Developmental toxicity induced during early stages of mammalian embryogenesis. *Mutat Res*, 396, 113-27.
- SABATINI, L. M., OTA, T. & AZEN, E. A. 1993. Nucleotide sequence analysis of the human salivary protein genes HIS1 and HIS2, and evolution of the STATH/HIS gene family. *Mol Biol Evol*, 10, 497-511.
- SANDER, J. D., CADE, L., KHAYTER, C., REYON, D., PETERSON, R. T., JOUNG, J. K. & YEH, J. R. 2011. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol*, 29, 697-8.
- SAVIC, N., RINGNALDA, F. C., LINDSAY, H., BERK, C., BARGSTEN, K., LI, Y., NERI, D., ROBINSON, M. D., CIAUDO, C., HALL, J., JINEK, M. & SCHWANK, G. 2018. Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *Elife*, 7.
- SCHINDLER, C. & DARNELL, J. E., JR. 1995. Transcriptional responses to polypeptide ligands: the JAK-STAT pathway. *Annu Rev Biochem*, 64, 621-51.
- SCHMITT-NEY, M., DOPPLER, W., BALL, R. K. & GRONER, B. 1991. Beta-casein gene promoter activity is regulated by the hormone-mediated relief of transcriptional repression and a mammary-gland-specific nuclear factor. *Mol Cell Biol*, 11, 3745-55.

SCHORNACK, S., MINSAVAGE, G. V., STALL, R. E., JONES, J. B. & LAHAYE, T. 2008. Characterization of AvrHah1, a novel AvrBs3-like effector from *Xanthomonas gardneri* with virulence and avirulence activity. *New Phytol*, 179, 546-56.

SCOTT, A., WELDON, S. & TAGGART, C. C. 2011. SLPI and elafin: multifunctional antiproteases of the WFDC family. *Biochem Soc Trans*, 39, 1437-40.

SCOTT, J. B., DAUGHERTY, R. M., JR. & HADDY, F. J. 1967. Effect of severe local hypoxemia on transcapillary water movement in dog forelimb. *Am J Physiol*, 212, 847-51.

SHAMAY, A., SOLINAS, S., PURSEL, V. G., MCKNIGHT, R. A., ALEXANDER, L., BEATTIE, C., HENNIGHAUSEN, L. & WALL, R. J. 1991. Production of the mouse whey acidic protein in transgenic pigs during lactation. *J Anim Sci*, 69, 4552-62.

SHARMAN, G. B. 1970. Reproductive physiology of marsupials. *Science*, 167, 1221-8.

SHARP, J. A., LEFEVRE, C. & NICHOLAS, K. R. 2008. Lack of functional alpha-lactalbumin prevents involution in Cape fur seals and identifies the protein as an apoptotic milk factor in mammary gland involution. *BMC Biol*, 6, 48.

SHARP, J. A., WANYONYI, S., MODEPALLI, V., WATT, A., KURUPPATH, S., HINDS, L. A., KUMAR, A., ABUD, H. E., LEFEVRE, C. & NICHOLAS, K. R. 2017. The tammar wallaby: A marsupial model to examine the timed delivery and role of bioactives in milk. *Gen Comp Endocrinol*, 244, 164-177.

SHEKAR, P. C., GOEL, S., RANI, S. D., SARATHI, D. P., ALEX, J. L., SINGH, S. & KUMAR, S. 2006. kappa-casein-deficient mice fail to lactate. *Proc Natl Acad Sci U S A*, 103, 8000-5.

SHIN, H. Y., WANG, C., LEE, H. K., YOO, K. H., ZENG, X., KUHNS, T., YANG, C. M., MOHR, T., LIU, C. & HENNIGHAUSEN, L. 2017. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat Commun*, 8, 15464.

SILBERSTEIN, G. B. & DANIEL, C. W. 1987. Reversible inhibition of mammary gland growth by transforming growth factor-beta. *Science*, 237, 291-3.

SIMONS, J. P., MCCLENAGHAN, M. & CLARK, A. J. 1987. Alteration of the quality of milk by expression of sheep beta-lactoglobulin in transgenic mice. *Nature*, 328, 530-2.

SLAYMAKER, I. M., GAO, L., ZETSCHE, B., SCOTT, D. A., YAN, W. X. & ZHANG, F. 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science*, 351, 84-8.

SMITH, J., BIBIKOVA, M., WHITBY, F. G., REDDY, A. R., CHANDRASEGARAN, S. & CARROLL, D. 2000. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res*, 28, 3361-9.

SOLBERG, N. & KRAUSS, S. 2013. Luciferase assay to study the activity of a cloned promoter DNA fragment. *Methods Mol Biol*, 977, 65-78.

SOLDAINI, E., JOHN, S., MORO, S., BOLLENBACHER, J., SCHINDLER, U. & LEONARD, W. J. 2000. DNA binding site selection of dimeric and tetrameric Stat5 proteins reveals a large repertoire of divergent tetrameric Stat5a binding sites. *Mol Cell Biol*, 20, 389-401.

SONG, Y., YUAN, L., WANG, Y., CHEN, M., DENG, J., LV, Q., SUI, T., LI, Z. & LAI, L. 2016. Efficient dual sgRNA-directed large gene deletion in rabbit with CRISPR/Cas9 system. *Cell Mol Life Sci*, 73, 2959-68.

STACEY, A., SCHNIEKE, A., KERR, M., SCOTT, A., MCKEE, C., COTTINGHAM, I., BINAS, B., WILDE, C. & COLMAN, A. 1995. Lactation is disrupted by alpha-lactalbumin deficiency and can be restored by human alpha-lactalbumin gene replacement in mice. *Proc Natl Acad Sci U S A*, 92, 2835-9.

STEPHENSON, A., ADAMS, J. W. & VACCAREZZA, M. 2017. The vertebrate heart: an evolutionary perspective. *J Anat*, 231, 787-797.

STINGL, J. 2011. Estrogen and progesterone in normal mammary gland development and in cancer. *Horm Cancer*, 2, 85-90.

STINNAKRE, M. G., VILOTTE, J. L., SOULIER, S. & MERCIER, J. C. 1994. Creation and phenotypic analysis of alpha-lactalbumin-deficient mice. *Proc Natl Acad Sci U S A*, 91, 6544-8.

STREULI, C. H., EDWARDS, G. M., DELCOMMENNE, M., WHITELAW, C. B., BURDON, T. G., SCHINDLER, C. & WATSON, C. J. 1995. Stat5 as a target for regulation by extracellular matrix. *J Biol Chem*, 270, 21639-44.

TEBAS, P., STEIN, D., TANG, W. W., FRANK, I., WANG, S. Q., LEE, G., SPRATT, S. K., SUROSKY, R. T., GIEDLIN, M. A., NICHOL, G., HOLMES, M. C., GREGORY, P. D., ANDO, D. G., KALOS, M., COLLMAN, R. G., BINDER-SCHOLL, G., PLESA, G., HWANG, W. T., LEVINE, B. L. & JUNE, C. H. 2014. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med*, 370, 901-10.

THORPE, M. P., JACOBSON, E. H., LAYMAN, D. K., HE, X., KRIS-ETHERTON, P. M. & EVANS, E. M. 2008. A diet high in protein, dairy, and calcium attenuates bone loss over twelve months of weight loss and maintenance relative to a conventional high-carbohydrate diet in adults. *J Nutr*, 138, 1096-100.

TILL, B. G., JENSEN, M. C., WANG, J., CHEN, E. Y., WOOD, B. L., GREISMAN, H. A., QIAN, X., JAMES, S. E., RAUBITSCHKE, A., FORMAN, S. J., GOPAL, A.

K., PAGEL, J. M., LINDGREN, C. G., GREENBERG, P. D., RIDDELL, S. R. & PRESS, O. W. 2008. Adoptive immunotherapy for indolent non-Hodgkin lymphoma and mantle cell lymphoma using genetically modified autologous CD20-specific T cells. *Blood*, 112, 2261-71.

TRIPLETT, A. A., SAKAMOTO, K., MATULKA, L. A., SHEN, L., SMITH, G. H. & WAGNER, K. U. 2005. Expression of the whey acidic protein (Wap) is necessary for adequate nourishment of the offspring but not functional differentiation of mammary epithelial cells. *Genesis*, 43, 1-11.

TRUT, L. N. 1999. Early Canid Domestication: The Farm-Fox Experiment: Foxes bred for tamability in a 40-year experiment exhibit remarkable transformations that suggest an interplay between behavioral genetics and development. *American Scientist*, 87, 160-169.

TSABOURI, S., DOUROS, K. & PRIFTIS, K. N. 2014. Cow's milk allergenicity. *Endocr Metab Immune Disord Drug Targets*, 14, 16-26.

TUNICK, M. H. 2009. Dairy innovations over the past 100 Years. *J Agric Food Chem*, 57, 8093-7.

TURCK, D. 2013. Cow's milk and goat's milk. *World Rev Nutr Diet*, 108, 56-62.

TYNDALE-BISCOE, C. H. & HEARN, J. P. 1981. Pituitary and ovarian factors associated with seasonal quiescence of the tammar wallaby, *Macropus eugenii*. *J Reprod Fertil*, 63, 225-30.

TYNDALE-BISCOE, C. H., HINDS, L. A. & HORN, C. A. 1988. Fetal role in the control of parturition in the tammar, *Macropus eugenii*. *J Reprod Fertil*, 82, 419-28.

UDY, G. B., TOWERS, R. P., SNELL, R. G., WILKINS, R. J., PARK, S. H., RAM, P. A., WAXMAN, D. J. & DAVEY, H. W. 1997. Requirement of STAT5b for sexual

dimorphism of body growth rates and liver gene expression. *Proc Natl Acad Sci U S A*, 94, 7239-44.

URNOV, F. D., REBAR, E. J., HOLMES, M. C., ZHANG, H. S. & GREGORY, P. D. 2010. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*, 11, 636-46.

VAN DEN HOUT, J. M., KAMPHOVEN, J. H., WINKEL, L. P., ARTS, W. F., DE KLERK, J. B., LOONEN, M. C., VULTO, A. G., CROMME-DIJKHUIS, A., WEISGLAS-KUPERUS, N., HOP, W., VAN HIRTUM, H., VAN DIGGELEN, O. P., BOER, M., KROOS, M. A., VAN DOORN, P. A., VAN DER VOORT, E., SIBBLES, B., VAN CORVEN, E. J., BRAKENHOFF, J. P., VAN HOVE, J., SMEITINK, J. A., DE JONG, G., REUSER, A. J. & VAN DER PLOEG, A. T. 2004. Long-term intravenous treatment of Pompe disease with recombinant human alpha-glucosidase from milk. *Pediatrics*, 113, e448-57.

VAN DIJK, I. A., FERRANDO, M. L., VAN DER WIJK, A. E., HOEBE, R. A., NAZMI, K., DE JONGE, W. J., KRAWCZYK, P. M., BOLSCHER, J. G. M., VEERMAN, E. C. I. & STAP, J. 2017. Human salivary peptide histatin-1 stimulates epithelial and endothelial cell adhesion and barrier function. *FASEB J*, 31, 3922-3933.

VAN DIJK, I. A., NAZMI, K., BOLSCHER, J. G., VEERMAN, E. C. & STAP, J. 2015. Histatin-1, a histidine-rich peptide in human saliva, promotes cell-substrate and cell-cell adhesion. *FASEB J*, 29, 3124-32.

VELTMAAT, J. M., MAILLEUX, A. A., THIERY, J. P. & BELLUSCI, S. 2003. Mouse embryonic mammaryogenesis as a model for the molecular regulation of pattern formation. *Differentiation*, 71, 1-17.

VORDENBAUMEN, S., BRAUKMANN, A., ALTENDORFER, I., BLECK, E., JOSE, J. & SCHNEIDER, M. 2013. Human casein alpha s1 (CSN1S1) skews in vitro differentiation of monocytes towards macrophages. *BMC Immunol*, 14, 46.

VORDENBAUMEN, S., BRAUKMANN, A., PETERMANN, K., SCHARF, A., BLECK, E., VON MIKECZ, A., JOSE, J. & SCHNEIDER, M. 2011. Casein alpha s1 is expressed by human monocytes and upregulates the production of GM-CSF via p38 MAPK. *J Immunol*, 186, 592-601.

VOUILLOT, L., THELIE, A. & POLLET, N. 2015. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda)*, 5, 407-15.

WALL, R. J. 2001. Pronuclear microinjection. *Cloning Stem Cells*, 3, 209-20.

WALSTRA, P. 1990. Casein sub-micelles: do they exist? *International Dairy Journal*, 9, 189-192.

WANG, B. 2012. Molecular mechanism underlying sialic acid as an essential nutrient for brain development and cognition. *Adv Nutr*, 3, 465S-72S.

WANG, C. W., ROOS, P. G., CHANT, N. S., CIANGARU, G., KHAZAIE, F., D, J. M., NADASEN, A., MILLS, S. J., WARNER, R. E., NORBECK, E., BECCHETTI, F. D., JANECKE, J. W. & LISTER, P. M. 1985. $^9\text{Be}(p,p \text{ } \alpha) ^5\text{He}$ cluster knockout reaction with 150 MeV polarized protons. *Phys Rev C Nucl Phys*, 31, 1662-1672.

WANG, D. L., CHANG, H., YANG, J. X., ZHANG, G. X., WANG, Z. G., YU, B., LIAO, X. J., SONG, W. T., HAN, X., SONG, G. M., WANG, J. H. & LI, R. L. 2007. [Analysis on the genetic structure of 8 Asia buffalo populations]. *Yi Chuan*, 29, 1103-9.

WANG, G., YANG, L., GRISHIN, D., RIOS, X., YE, L. Y., HU, Y., LI, K., ZHANG, D., CHURCH, G. M. & PU, W. T. 2017. Efficient, footprint-free human iPSC genome editing by consolidation of Cas9/CRISPR and piggyBac technologies. *Nat Protoc*, 12, 88-103.

WANG, H., YANG, H., SHIVALILA, C. S., DAWLATY, M. M., CHENG, A. W., ZHANG, F. & JAENISCH, R. 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, 153, 910-8.

WANG, X., CAO, C., HUANG, J., YAO, J., HAI, T., ZHENG, Q., WANG, X., ZHANG, H., QIN, G., CHENG, J., WANG, Y., YUAN, Z., ZHOU, Q., WANG, H. & ZHAO, J. 2016. One-step generation of triple gene-targeted pigs using CRISPR/Cas9 system. *Sci Rep*, 6, 20620.

WANINGE, R., WALSTRA, P., BASTIAANS, J., NIEUWENHUIJSE, H., NYLANDER, T., PAULSSON, M. & BERGENSTAHL, B. 2005. Competitive adsorption between beta-casein or beta-lactoglobulin and model milk membrane lipids at oil-water interfaces. *J Agric Food Chem*, 53, 716-24.

WATSON, C. J., GORDON, K. E., ROBERTSON, M. & CLARK, A. J. 1991. Interaction of DNA-binding proteins with a milk protein gene promoter in vitro: identification of a mammary gland-specific factor. *Nucleic Acids Res*, 19, 6603-10.

WAZEN, R. M., MOFFATT, P., PONCE, K. J., KURODA, S., NISHIO, C. & NANJI, A. 2015. Inactivation of the Odontogenic ameloblast-associated gene affects the integrity of the junctional epithelium and gingival healing. *Eur Cell Mater*, 30, 187-99.

WEI, J., WAGNER, S., MACLEAN, P., BROPHY, B., COLE, S., SMOLENSKI, G., CARLSON, D. F., FAHRENKRUG, S. C., WELLS, D. N. & LAIBLE, G. 2018.

Cattle with a precise, zygote-mediated deletion safely eliminate the major milk allergen beta-lactoglobulin. *Sci Rep*, 8, 7661.

WHEELER, M. B., BLECK, G. T. & DONOVAN, S. M. 2001. Transgenic alteration of sow milk to improve piglet growth and health. *Reprod Suppl*, 58, 313-24.

WIDDOWSON, E. M. 1979. Nutritional requirement and its assessment, with special reference to energy, protein and calcium. *Bibl Nutr Dieta*, 148-54.

WILKINS, A. S., WRANGHAM, R. W. & FITCH, W. T. 2014. The "domestication syndrome" in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*, 197, 795-808.

WOOD, A. J., LO, T. W., ZEITLER, B., PICKLE, C. S., RALSTON, E. J., LEE, A. H., AMORA, R., MILLER, J. C., LEUNG, E., MENG, X., ZHANG, L., REBAR, E. J., GREGORY, P. D., URNOV, F. D. & MEYER, B. J. 2011. Targeted genome editing across species using ZFNs and TALENs. *Science*, 333, 307.

WU, J., VILARINO, M., SUZUKI, K., OKAMURA, D., BOGLIOTTI, Y. S., PARK, I., ROWE, J., MCNABB, B., ROSS, P. J. & BELMONTE, J. C. I. 2017. CRISPR-Cas9 mediated one-step disabling of pancreatogenesis in pigs. *Sci Rep*, 7, 10487.

XU, Y. Z., KANAGARATHAM, C., JANCIK, S. & RADZIOCH, D. 2013. Promoter deletion analysis using a dual-luciferase reporter system. *Methods Mol Biol*, 977, 79-93.

YAN, Q., XU, K., XING, J., ZHANG, T., WANG, X., WEI, Z., REN, C., LIU, Z., SHAO, S. & ZHANG, Z. 2016. Multiplex CRISPR/Cas9-based genome engineering enhanced by Drosha-mediated sgRNA-shRNA structure. *Sci Rep*, 6, 38970.

YAN, S., TU, Z., LIU, Z., FAN, N., YANG, H., YANG, S., YANG, W., ZHAO, Y., OUYANG, Z., LAI, C., YANG, H., LI, L., LIU, Q., SHI, H., XU, G., ZHAO, H., WEI,

H., PEI, Z., LI, S., LAI, L. & LI, X. J. 2018. A Huntingtin Knockin Pig Model Recapitulates Features of Selective Neurodegeneration in Huntington's Disease. *Cell*, 173, 989-1002 e13.

YIN, C., ZHANG, T., QU, X., ZHANG, Y., PUTATUNDA, R., XIAO, X., LI, F., XIAO, W., ZHAO, H., DAI, S., QIN, X., MO, X., YOUNG, W. B., KHALILI, K. & HU, W. 2017. In Vivo Excision of HIV-1 Provirus by saCas9 and Multiplex Single-Guide RNAs in Animal Models. *Mol Ther*, 25, 1168-1186.

YOUNG, S. J. F. A. F. G. 1938. The Effect of Anterior Pituitary Extracts on Established Lactation in the Cow. *Proceedings of the Royal Society of London*, 126, 45-76.

YU, S. H., DEEN, K. C., LEE, E., HENNIGHAUSEN, L., SWEET, R. W., ROSENBERG, M. & WESTPHAL, H. 1989. Functional human CD4 protein produced in milk of transgenic mice. *Mol Biol Med*, 6, 255-61.

YUE, F., CHENG, Y., BRESCHI, A., VIERSTRA, J., WU, W., RYBA, T., SANDSTROM, R., MA, Z., DAVIS, C., POPE, B. D., SHEN, Y., PERVOUCHINE, D. D., DJEBALI, S., THURMAN, R. E., KAUL, R., RYNES, E., KIRILUSHA, A., MARINOV, G. K., WILLIAMS, B. A., TROUT, D., AMRHEIN, H., FISHER-AYLOR, K., ANTOSHECHKIN, I., DESALVO, G., SEE, L. H., FASTUCA, M., DRENKOW, J., ZALESKI, C., DOBIN, A., PRIETO, P., LAGARDE, J., BUSSOTTI, G., TANZER, A., DENAS, O., LI, K., BENDER, M. A., ZHANG, M., BYRON, R., GROUDINE, M. T., MCCLEARY, D., PHAM, L., YE, Z., KUANG, S., EDSALL, L., WU, Y. C., RASMUSSEN, M. D., BANSAL, M. S., KELLIS, M., KELLER, C. A., MORRISSEY, C. S., MISHRA, T., JAIN, D., DOGAN, N., HARRIS, R. S., CAYTING, P., KAWLI, T., BOYLE, A. P., EUSKIRCHEN, G., KUNDAJE, A., LIN, S., LIN, Y., JANSEN, C., MALLADI, V. S., CLINE, M. S., ERICKSON, D. T., KIRKUP, V. M., LEARNED, K., SLOAN, C. A., ROSENBLOOM, K. R., LACERDA DE SOUSA, B., BEAL, K., PIGNATELLI, M., FLICEK, P., LIAN, J., KAHVECI, T.,

LEE, D., KENT, W. J., RAMALHO SANTOS, M., HERRERO, J., NOTREDAME, C., JOHNSON, A., VONG, S., LEE, K., BATES, D., NERI, F., DIEGEL, M., CANFIELD, T., SABO, P. J., WILKEN, M. S., REH, T. A., GISTE, E., SHAFER, A., KUTYAVIN, T., HAUGEN, E., DUNN, D., REYNOLDS, A. P., NEPH, S., HUMBERT, R., HANSEN, R. S., DE BRUIJN, M., et al. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature*, 515, 355-64.

ZHANG, J. P., LI, X. L., LI, G. H., CHEN, W., ARAKAKI, C., BOTIMER, G. D., BAYLINK, D., ZHANG, L., WEN, W., FU, Y. W., XU, J., CHUN, N., YUAN, W., CHENG, T. & ZHANG, X. B. 2017. Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol*, 18, 35.

ZHANG, L., JIA, R., PALANGE, N. J., SATHEKA, A. C., TOGO, J., AN, Y., HUMPHREY, M., BAN, L., JI, Y., JIN, H., FENG, X. & ZHENG, Y. 2015. Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. *PLoS One*, 10, e0120396.

ZHAO, J., ZHAO, J., MANGALAM, A. K., CHANNAPPANAVAR, R., FETT, C., MEYERHOLZ, D. K., AGNIHOTHRAM, S., BARIC, R. S., DAVID, C. S. & PERLMAN, S. 2016. Airway Memory CD4(+) T Cells Mediate Protective Immunity against Emerging Respiratory Coronaviruses. *Immunity*, 44, 1379-91.

ZHENG, M., XING, W., LIU, Y., LI, M. & ZHOU, H. 2017. Tetramerization of SATB1 is essential for regulating of gene expression. *Mol Cell Biochem*, 430, 171-178.

ZHOU, W., WAN, Y., GUO, R., DENG, M., DENG, K., WANG, Z., ZHANG, Y. & WANG, F. 2017. Generation of beta-lactoglobulin knock-out goats using CRISPR/Cas9. *PLoS One*, 12, e0186056.

ZHU, H., LIU, J., CUI, C., SONG, Y., GE, H., HU, L., LI, Q., JIN, Y. & ZHANG, Y. 2016. Targeting Human alpha-Lactalbumin Gene Insertion into the Goat beta-Lactoglobulin Locus by TALEN-Mediated Homologous Recombination. *PLoS One*, 11, e0156636.

ZHU, W., ZENG, Q., COLANGELO, C. M., LEWIS, M., SUMMERS, M. F. & SCOTT, R. A. 1996. The N-terminal domain of TFIIB from *Pyrococcus furiosus* forms a zinc ribbon. *Nat Struct Biol*, 3, 122-4.

ZU, Y., TONG, X., WANG, Z., LIU, D., PAN, R., LI, Z., HU, Y., LUO, Z., HUANG, P., WU, Q., ZHU, Z., ZHANG, B. & LIN, S. 2013. TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat Methods*, 10, 329-31.

ZUO, E., HUO, X., YAO, X., HU, X., SUN, Y., YIN, J., HE, B., WANG, X., SHI, L., PING, J., WEI, Y., YING, W., WEI, W., LIU, W., TANG, C., LI, Y., HU, J. & YANG, H. 2017. CRISPR/Cas9-mediated targeted chromosome elimination. *Genome Biol*, 18, 224.