THE ROLE OF DEFENSINS AND C-X-C CHEMOKINES IN MAMMALIAN INNATE IMMUNITY

by

Linda Marie Rehaume

B.Sc., University of Victoria (*with distinction*), 2001

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Microbiology & Immunology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

November 2010

ABSTRACT

In humans, defensins constitute the largest group of host defence peptides that are evolutionarily conserved components of innate immunity. Defensins share many structural and functional characteristics with C-X-C chemokines, including a C-X-C amino acid motif, net positive charge, disulphide bonding, three-dimensional shape and chemokine activity. Deficiencies in α-defensins and C-X-C chemokines have been correlated with susceptibility to infection and chronic inflammatory diseases. However the genetics and diversity of defensins and mechanisms underlying these disorders were not well understood. This thesis comprises three separate but overlapping approaches to address these issues.

The genomic content of murine α-defensins within the reference C57BL/6J strain was characterized. Novel α-defensin (11) and defensin-related cryptdin (3) genes were found, as were gene duplications and differences in genomic content between strains of mice. A next-generation sequencing method was developed for the quantitative analysis of α-defensin and defensin-related cryptdin gene expression.

The α-defensin DEFA1 induced interleukin (IL) 8 and IL10 release from human PBMCs. The mechanism(s) of action of defensins, which appears to involve induction of chemokines and anti-inflammatory cytokines, needs further elucidation *in vivo*. Consequently, novel murine models of inflammation and immunosuppression were developed. The *IL8* and *Il10* genes were separately cloned, behind an intestine-specific promoter, into eukaryotic expression vectors, which were used to transfect murine embryonic stem cells. Correct targeting was confirmed for both constructs and germline transmission achieved for the IL8 mice. Conditional homozygous mice were generated, which, upon breeding with Cre-expressing mice, will express IL8, a C-X-C chemokine, in an intestinal-specific manner. This will enable analyses of effects of chemokine overexpression on intestinal infection, and on peptide efficacy in the resolution of infection. In other studies to address innate immune mechanisms, the transcriptional profiles of patients susceptible to *Salmonella* and mycobacterial infections due to immunodeficiencies in IL12- and interferon-γ-mediated immunity were generated. These data indicated that the chemokines CXCL9 and CXCL10 might mediate immunity to *Mycobacteria* whereas additional defects in TLR4 responses appeared to underlie susceptibility to *Salmonella*.

The data presented here strengthen our understanding of the murine defensin repertoire and provide tools that enable sophisticated systems level studies of *in vivo* function.

PREFACE

A version of Chapter 2 has been published in:

Amid, C.*, Rehaume, L.M.*, Brown, K.L., Gilbert, J.G.R., Dougan G., Hancock, R.E.W. and Harrow J.L. 2009. Manual annotation and analysis of the defensin gene cluster in the C57BL/6J mouse reference genome. *BMC Genomics* 10:606. Authors retain copyright.

* Equal contribution

Manual annotation of the defensin gene cluster in the C57BL/6J mouse reference genome was performed by Clara Amid in the Human and Vertebrate Analysis and Annotation (HAVANA) group at the Wellcome Trust Sanger Institute (WTSI). Subsequent analyses and writing of the manuscript was shared by Linda Rehaume and Clara Amid with overall revisions being performed by our supervisors. The text has been significantly revised in this thesis. James Gilbert originally created Figure 2.1 that was further modified in collaboration with Clara Amid. Figures 2.10 and 2.11 were generated by Clara Amid. The defensin transcriptional and functional experiments were performed by Linda Rehaume, and have not been published.

Ethical approval was obtained for the work in the thesis involving human material. The research in Chapter 2 is covered by The University of British Columbia Clinical Research Ethics Board Certificate H04-70232. The research in Chapter 4 is covered by The Royal Free Hampstead National Health Service Trust Research Ethics Committee Certificate 04/Q0501/119. Approval for the research carried out in Chapter 4 was also obtained from the Wellcome Trust Sanger Institute Human Material and Data Management Committee.

Throughout the thesis, genes and proteins are referred to using approved nomenclature. This means that references to previous studies, where non-approved nomenclature was used, have been changed accordingly; although the non-approved nomenclature is also indicated in brackets for clarity. The distinction between human and mouse gene and protein symbols is made by capitalization. Human symbols contain all upper-case letters and mouse symbols begin with an upper-case letter followed by lower-case letters.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

A33p, A33 probe

BAC, bacterial artificial chromosome

BLAST, basic local alignment search tool

CCL, chemokine (C-C motif) ligand

CCR, chemokine (C-C motif) receptor

CNV, copy number variation

Crp, cryptdin

CRS, cryptdin-related sequence

CXCL, chemokine (C-X-C motif) ligand

CXCR, chemokine (C-X-C motif) receptor

DC, dendritic cell

DEFA, α-defensin

DEFB, β-defensin

Defcr, defensin-related cryptdin sequence

Defcr-rs, defensin-related cryptdin-related sequence

D-PBS, Dulbecco's phosphate buffered saline

ELISA, enzyme linked immunosorbent assay

ELR, glutamic acid-leucine-arginine amino acid motif

ESC, embryonic stem cell

Fabp, fatty acid binding protein

Gpa33, glycoprotein A33 (transmembrane)

GSMID, genome sequencer multiplex identifier

HGNC, Human Genome Organisation Gene Nomenclature Committee

HNP, human neutrophil peptide

IFNG, interferon-γ

IFNGR, interferon-γ receptor

IL, interleukin

IL10R, interleukin 10 receptor

IL12R, interleukin 12 receptor

IRF, interferon regulatory factor

JAK, Janus activating kinase

KI, knock-in

KO, knock-out

LPS, lipopolysaccharide

LRHE, Linda Rehaume human interleukin eight mouse colony

LRIE, Linda Rehaume interleukin eight mouse colony

LRIT, Linda Rehaume murine interleukin ten mouse colony

LRMT, Linda Rehaume murine interleukin ten mouse colony

MCS, multiple cloning site

MGI, mouse genome informatics

MGNC, Mouse Genomic Nomenclature Committee

NCBI, National Center for Biotechnology Information

NET, neutrophil extracellular trap

NF-H_2O, nuclease free water

NF-κB, nuclear factor of kappa-light-chain-enhancer of activated B cells

NK, natural killer

NTM, non-tuberculosis mycobacteria

OTTID, Vega database identifier

PBMC, polymorphonuclear cell

Pfam, Protein family database

RACE, rapid amplification of cDNA ends

RefSeq, NCBI Reference Sequence database

RGNC, Rat Genome Nomenclature Committee

RSF, research support facility

RT, reverse transcription

SFF, standard flowgram format

SOC, super optimal broth with catabolite repression

SPAG, sperm-associated antigen

SPI, *Salmonella* pathogenicity island

STAT, signal transducer and activator of transcription

Th1, Type 1 helper T cell

Th2, Type 2 helper T cell

TLR, Toll-like receptor

TNF, tumor necrosis factor

Treg, regulatory T cell

TTSS, type III secretion system

TYK, tyrosine kinase

UBC, University of British Columbia

VEGA, Vertebrate Genome Annotation genome browser

WT, wild type

WTSI, Wellcome Trust Sanger Institute

ACKNOWLEDGEMENTS

I would like to start by thanking my supervisor, Bob Hancock, for his unwavering support. I appreciated the freedom he allowed me during my research, especially for the opportunity to work in Cambridge UK, but also the guidance he gave me when I needed to refocus and see the larger picture.

I am grateful to Gordon Dougan for the opportunity to work in his lab at the Wellcome Trust Sange Institute in Cambridge UK, as well as David Adams for tissue culture space and helpful discussions. I would like to thank the Human and Vertebrate Analysis and Annotation group, especially Clara Amid for the defensin annotation and the friendship that has developed along the way. Additionally I have appreciated working with Rainer Doffinger and Dinakantha Kumararatne at the Addenbrooke's Hospital Department of Clinical Biochemistry & Immunology in Cambridge, UK. I would especially like to thank the immune deficient patients and their families for their selfless participation in this research.

I would like to thank my committee members, Megan Levings, David Speert and Ninan Abraham, for their feedback on my research, as well as the members of the Hancock, Dougan and Adams' labs for intellectual and technical support. I have enjoyed working with such hardworking, talented and enthusiastic people and am grateful for all I have learned.

I would especially like to thank my parents, Eleanor and Len Rehaume, and sister, Vicki Rehaume, for their steadfast belief in me always. I have leaned on them for emotional support through good and bad times, and on my Mum and Vicki during the years of this degree. My parents gave me the best start in life and to them I will always be grateful. Finally I would like to thank my partner, Greg Baillie, for all of his love and support, especially for 'doing everything' while I was writing my thesis. Thank you.

DEDICATION

This work is dedicated to my Dad. He taught me so many things and was one of the smartest people I know. I was inspired to pursue this degree in the field of immunology because of what he went through. It is my hope that each new piece of information will not only advance our knowledge of how the immune system works, but lead to novel therapies and improve patient health.

I know my Dad would have been proud of me.

1 INTRODUCTION

1.1 Modulators of the Immune System

Cationic host defence peptides are evolutionarily conserved across a wide range of species. They constitute a major component of the defensive repertoire in lower organisms, whereas in higher species they are part of the complex immune system involved in protecting against infection. These peptides were first described as antimicrobial peptides, but it is now apparent they play a multi-functional role in the immune system. There are several classes of peptides, based on their secondary structure, but they all share several key features. They often require proteolytic cleavage for release of the mature, active peptide which contains 12-50 amino acids and is positively charged and amphipathic (1, 2). The secondary structure of these peptides can be α-helical (e.g. cathelicidins), looped or cyclized with one disulphide bond (e.g. bactenecin), β-stranded with two or more disulphide bonds (e.g. defensins), or extended (e.g. indolicidin) (1, 2). The expression of host defence peptides occurs in a wide range of cell types in either a constitutive or an inducible manner. As such, many host defence peptides are commonly found at increased concentrations within the inflammatory milieu in humans. Additionally several human immune deficiencies have decreased host defence peptide expression, which is often correlated with increased susceptibility to infection, implicating them in disease. Whether these peptides are the cause or effect of such conditions has been difficult to ascertain, due to the often redundant nature of their expression.

The immune system has evolved layers of protection against invading pathogens. Often several molecules or pathways perform the same or similar functions. This becomes apparent in loss-of-function assays if the observed phenotype is milder than expected, or no phenotype is observed. The redundancy of host defence peptide expression and apparent function is analogous to that of cytokines, especially for chemokine function, although cytokines have long been regarded as crucial in the functioning of all aspects of the immune system. Within the chemokine family, often more than one member performs the same function, many bind to more than one receptor, and conversely chemokine receptors usually bind more than one chemokine. The receptor-ligand interaction is not strictly a one-to-one relationship. *In vitro* experiments have been important for determination of their common and distinctive functions. Within each of the cytokine and host defence peptide families there are members that possess overlapping functions. Cytokines are responsible for cellular differentiation, mobilization and activation, as well as regulation of homeostasis, tolerance and repair (3). It is becoming apparent that host defence

peptides also play an important role in many of these processes. The interaction between cytokines and host defensin peptides within the inflammatory milieu has not been extensively studied, and whether these molecules display synergy with respect to any of these functions is unknown.

Defensins and chemokines have direct chemotactic activities for immune cells (although much higher concentrations of defensins are needed), and despite their differing primary sequences, defensins and chemokines have related tertiary structures and cationic nature (4). Both contain three β-strands, an amino terminal α-helix and disulphide bridges (4). It has been proposed that the chemotactic activities of human β-defensin (DEFB) 1, DEFB4 (DEFB2) and chemokine (C-C motif) ligand (CCL) 20 (MIP-3α) are due to their binding to the chemokine (C-C motif) receptor, CCR6 (5-7). Similar to Toll-like receptor (TLR) agonists (8), CCL20 binding of dendritic cells (DCs) via CCR6 leads to increased CCR7 expression and reduced CCR6 expression (4). Additionally murine Defb2, Defb3 and Ccl20 (Mip-3α) demonstrate adjuvant antitumor activity (9). Mice vaccinated with lymphoma-specific nonimmunogenic antigens fused with murine Defb2, Defb3 or Ccl20, generated tumour specific antibodies and were protected upon tumour challenge (9). Furthermore, increased survival was observed when mice bearing tumours were administered the fusion constructs (9). This adjuvant activity was dependent on the targeting of murine Defb2, Defb3 or Ccl20 to immature dendritic cells, presumably through the chemotactic receptor, CCR6 (9). Therefore both host defence peptides and chemokines link the innate and adaptive immune systems through a continuum of cell signalling and activation.

1.2 Defensins

Defensins are the largest family of cationic host defense peptides in humans and are produced by many cell types including neutrophils, Paneth cells, epithelial cells and keratinocytes (10). Defensin genes typically have a two-exon structure, however there are exceptions within the α-defensin family, some of which have three exons, while members of the β-defensin family can have between two to four exons, e.g. fish and birds have three-exon β-defensins (11). The differences between α- and β-defensins have been proposed to be a consequence of gene duplication and subsequent divergence selected during evolution (12). Extensive analysis has provided insight into the evolution of mammalian β-defensins (12-14). β-defensins are evolutionarily older than α-defensins, which are thought to have arisen by repeated gene duplication of β-defensins and positive diversifying selection (12, 15). This is especially

apparent for rodent, and in particular murine, α-defensins, which have most likely "lost" one or more of these exons during evolution, and subsequent gene/chromosome duplication events led to their two exon structure. The high similarity of mouse α-defensin genes and subsequent repetitive nature of their chromosomal position lends support to this model. As a rapidly evolving gene family, defensins provide a means through which to study mammalian evolution.

Defensins are produced as 94-95 amino acid prepropeptides, and sequential cleavage of the signal- and pre-peptides releases the 29-34 amino acid mature peptides (Figure 1.1A). Mature defensin peptides are characterized by six canonical cysteines residues at defined positions, and adopt a β-stranded secondary structure composed of three disulphide bonds. The spacing between the cysteine residues and the arrangement of the three disulphide bonds divides the defensins into α- and β-defensin subfamilies (Figure 1.1B) (16-18).

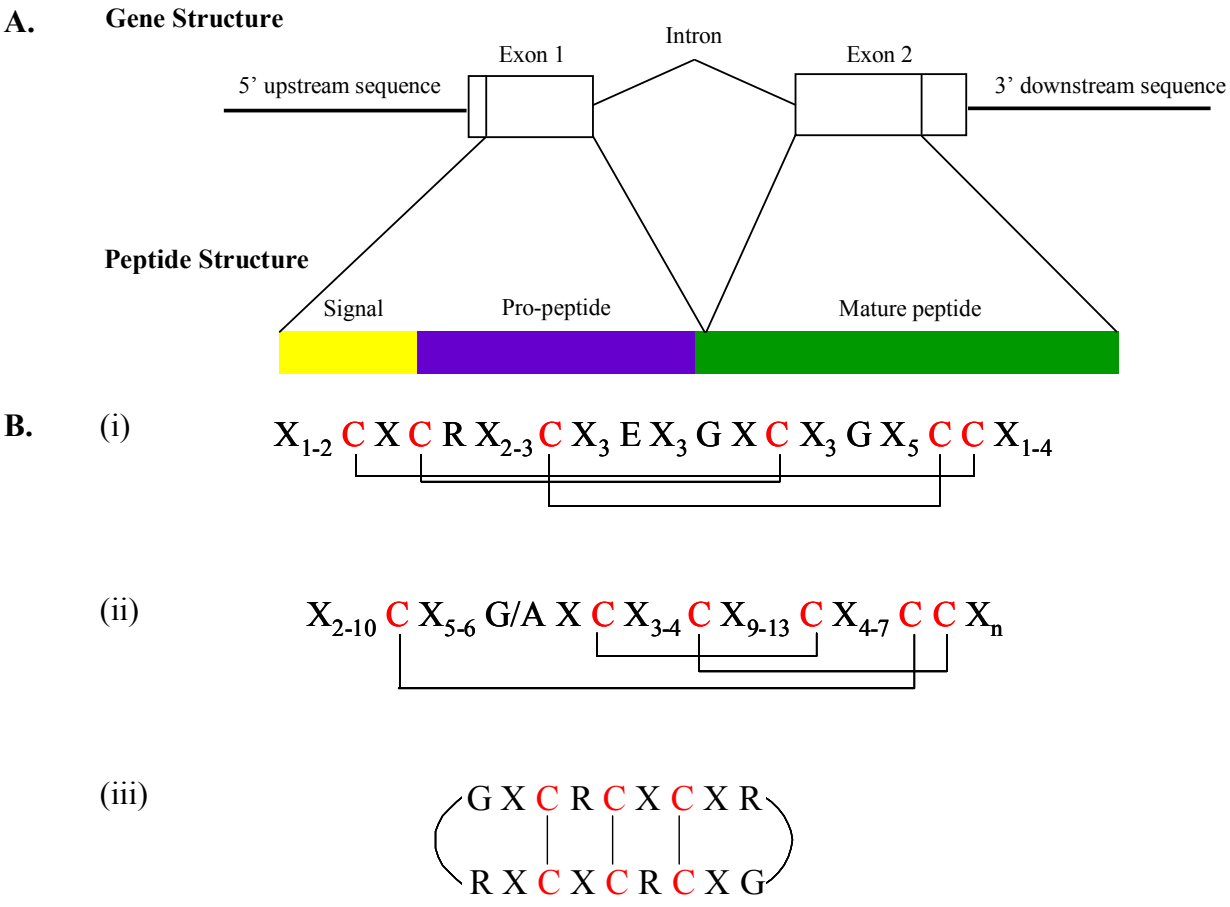


Figure 1.1. Canonical defensin structural characteristics. Schematic of murine α- and CRS-defensin gene and peptide structures (A). The defensin primary sequence, and linkage pattern of the six cysteines residues (B), allowing formation of the three disulphide bonds, divides them into α-defensin (i), β-defensin (ii) and θ-defensin (iii) subfamilies.

β-defensins have a broad tissue expression pattern and are found in most vertebrates and some invertebrate species, whilst α-defensins are specific to certain mammals and are mainly produced by leukocytes of myeloid origin and Paneth cells of the small intestine (16-18). θ-Defensins are believed to be derived by cyclization of α-defensins and seem to be restricted to the leukocytes of Old World monkeys (19). There is debate as to the role of defensins but there is limited evidence to support the hypotheses of their biological function. *In vitro*, under low salt and serum-free conditions, defensins have antimicrobial activity against Gram-positive and Gram-negative bacteria, fungi, parasites, yeast and viruses, however under physiological salt conditions and in the presence of serum, the killing activity of most defensins is abrogated or greatly reduced (2). Under these conditions, DEFB103 (DEFB3) is the only human defensin that retains significant antimicrobial function. Direct killing activity of defensins may be limited to sites at which concentrations are high such as neutrophil phagosomes, intestinal crypts, or at the skin epithelia which has lower concentrations of salts and serum proteins.

The discovery of a mouse Paneth cell α-defensin peptide, termed cryptdin (Crp) due to its expression in the Crypts of Lieberkühn (20), was the first report of defensin expression in a non-myeloid cell lineage (21, 22). The gene coding for cryptdin, defensin-related cryptdin (*Defcr*) was subsequently mapped to mouse Chromosome 8 (23, 24) and since has been discovered to be part of a larger gene family including additional *Defcr* α-defensin and Defcr-related sequence (*Defcr-rs*) genes. *Defcr-rs* were so named due to their sequence similarity and genetic linkage to *Defcr* (21-25). The Defcr-rs genes and peptide products are also referred to as cryptdin-related sequences (CRS). Additional *Defcr*/*Defcr-rs* loci have been discovered in different mouse strains, some of which may be polymorphic or involved in copy number variation (CNV) (23, 24, 26-29). The confusion around gene names, variable copy numbers and polymorphisms has made the study of mouse defensins quite complex.

In humans, α-defensins are most abundant in neutrophils and intestinal Paneth cells (30). The human neutrophil-specific α-defensins, DEFA1, DEFA2 and DEFA3 constitute 30-50% of the protein content in the azurophilic granules (10). *DEFA2* does not exist within the human genome, however the DEFA2 (HNP-2) peptide is presumed to be derived from either DEFA1 or DEFA3 through post-translational cleavage of the N-terminal alanine or aspartic acid residue, respectively (31). Excluding the N-terminal amino acid, the DEFA1 and DEFA3 primary sequences are identical (31). The human intestinal-specific α-defensins *DEFA5* and *DEFA6* are the highest expressed of Paneth cell components (32). Both neutrophil and Paneth cell

stimulation can lead to degranulation and α-defensin release, either into the surrounding tissues or intestinal crypts and lumen, respectively. There are rare human disorders, Chediak Higashi Syndrome and Specific Granule Deficiency, associated with decreased or absent neutrophil α-defensins, however other neutrophil granule components are also deficient which makes it difficult to specifically assign these disorders to defensins (33). Murine neutrophils do not contain α-defensins, therefore Paneth cells are the only source of α-defensins in mice (34). These murine α-defensins comprise 70% of the killing activity of Paneth cells (35). Due to their secondary structure, defensins are relatively resistant to proteolysis and to this extent murine Paneth cell α-defensins have been purified from the lumen of the large intestine, confirming this property *in vivo* (36). This characteristic of α-defensins may result in their effects being propagated throughout the intestinal tract, in addition to within the crypts of the small intestine (36). Reduced DEFA5 and DEFA6 expression has been linked with inflammatory bowel Crohn's disease and, from experiments with transgenic mice expressing DEFA5 in the intestine, is hypothesized to be the result of alterations within the intestinal microbiota, which might contribute to the inflammatory pathology of this disease (32, 37). Interestingly the severity of disease is also correlated with increased IL8 production (32), which may be either a marker of inflammation or a factor in the disease progression.

1.3 Peptide Therapeutics in Models of Inflammation

There is increasing interest in host defence peptides for use as novel therapeutics in the treatment of infectious diseases (38-40). One goal of this work was to develop novel mouse models that would enable elucidation of the *in vivo* mode of action of host defence peptides in intestinal infection under conditions of hyper- and hypo-inflammation. Characterization of the *in vitro* functions of host defence peptides, in particular defensins, on a variety of cell types, including primary cells, has revealed their wide range of functions, including direct chemokine activity, induction of chemokine production and secretion (including IL8), suppression of induced pro-inflammatory responses (in part reflected in increased IL10 production), enhancement of phagocytosis, inhibition of complement, adjuvant activity, co-stimulatory molecule induction, enhancement of cellular growth and stimulation of wound healing (41, 42). Direct antimicrobial activity *in vivo* under physiological conditions remains to be determined (2); however the concentration of defensins upon neutrophil or Paneth cell degranulation would favor high local concentrations and suggest that, under certain conditions, defensins might contribute

to local immunity through direct antimicrobial action. The *in vivo* peptide mechanism of action is harder to determine due to redundancy in peptides, especially in the case of defensins, however the transgenic mice expressing DEFA5 in the intestine were slightly more resistant to oral *Salmonella typhimurium* infection (43). There is interest in generating synthetic derivatives of host defence peptides that have enhanced function(s). Administration of a synthetic peptide, IDR-1, prophylactically or therapeutically to mice improves survival upon infection with both Gram-positive and Gram-negative bacteria (39).

1.4 Interleukin 8

IL8 is a chemokine that demonstrates low basal expression and can be markedly increased in response to various stimuli, e.g. bacteria, viruses, stress factors, cytokines and, pertinent to this thesis, host defence peptides, including defensins and LL-37 (44-47). LL-37 and DEFB1 inhibit lipopolysaccharide (LPS)-induced IL8 production (44, 48). A synthetic peptide IDR-1002 induces robust chemokine, including IL8, production from human polymorphonuclear cells (PBMCs), chemoattracts murine neutrophils and monocytes *in vivo* and protects mice against infection (49). It is therefore important to consider the *in vivo* context of host defence peptides with respect to IL8 production/inhibition.

IL8 is a member of the ELR^+ C-X-C motif family of chemokines, which are primarily selective for neutrophils (50). These chemokines contain the amino acid motif glutamic acid-leucine-arginine (E-L-R) before the first cysteine of the C-X-C motif (50). IL8 chemoattracts and activates neutrophils; activation leads to phagocytosis or degranulation, and can also prepare for the oxidative burst through phosphorylation events needed for nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex activation (51). There are conflicting results as to whether IL8 inhibits neutrophil apoptosis, which may be beneficial for bacterial clearance but could also promote excessive tissue damage (51). Neutrophil production of chemokines is dependent on the inflammatory milieu, as varying combinations of stimuli produce variable chemokines (51). Similarly, neutrophil activation can result from the engagement of the chemokine (C-X-C motif) receptors (CXCR) 1 or CXCR2, or both. However, these receptors, although potentially redundant in function due to overlapping ligand specificity, appear to have unique roles in neutrophil activation (52). The contents released by neutrophils upon degranulation, secreted cytokines and chemokines or cell-cell contact mediate the interaction of neutrophils with endothelial cells, monocytes, dendritic cells and T cells (53). It is now understood that

neutrophils play a more significant role in the immune response through their ability to coordinate the innate and adaptive immune systems. IL8 mediates inflammation through neutrophil activation but also promotes neutrophil release (mobilization) from the bone marrow, which is the primary site of neutrophil precursor differentiation (54). Additionally, IL8 is a potent inducer of angiogenesis (46).

The mouse genome does not contain an Il8-encoding gene although chemokine (C-X-C motif) ligand (Cxcl) 1 (KC) and Cxcl2 (Mip-2β) have both been proposed as functional equivalents due to their ability to chemoattract neutrophils (55). However, the protein alignments, and thus identities, of Cxcl1 and Cxcl2 with IL8 do not reflect true orthology, and they likely have other distinct functions based on their differing receptor binding capabilities (discussed below). Of the rodents with sequenced genomes that are accessible through the Ensembl genome browser, the ones that contain homologues of human IL8 are guinea pig, kangaroo rat, pika, rabbit, squirrel and tree shrew, but homologues are absent from mouse or rat. This indicates that both mouse and rat have likely lost the Il8 gene, rather than species outside of rodents, e.g. primates, having gained the gene.

In humans, IL8 binds two G protein-coupled receptors, CXCR1 and CXCR2. CXCR1 and CXCR2 both bind the chemokines IL8 and CXCL6, whereas CXCR2 also binds CXCL1, CXCL2, CXCL3, CXCL5 and CXCL7 (52). The roles of CXCR1 and CXCR2 are largely thought to be redundant since both bind ELR^+ chemokines. In particular they both bind IL8 with comparable high affinity, although CXCR1 is relatively selective for IL8 compared to CXCR2, which does not display any ligand selectivity (50). However the difference in the ligand-binding repertoires of CXCR1 and CXCR2 indicates that engagement of these two receptors by IL8 may lead to different outcomes. This is in agreement with the expanding role of neutrophils, beyond phagocytosis and cell-mediated killing, although any distinct roles of CXCR1 and CXCR2 need elucidating.

It was initially thought that since mice lack an Il8 gene they would also lack a homologue of human CXCR1; however, mouse Cxcr1 was recently identified. Conversely, the ligand specificity and neutrophil chemoattractant properties of Cxcr2 have been well characterized and orthology between mouse Cxcr2 and human CXCR2 is generally accepted. Cxcr2 binds ELR^+ chemokines, in particular Cxcl1 and Cxcl2, the human homologues CXCL1 and CXCL2, as well as CXCL3 and IL8 (52). The identification and cloning of mouse Cxcr2 was reported by several groups (56-58), but initial characterizations were somewhat conflicting. Cxcr2 bound Cxcl1 and

Cxcl2 with high affinity and IL8 with low affinity. Nevertheless IL8 promoted murine neutrophil chemotaxis, although higher concentrations of IL8 were required to elicit cell numbers equivalent to Cxcl1, *in vitro* (58). The responsiveness of murine neutrophils to IL8 led to speculation that a second IL8 receptor exists in mouse, similar to CXCR1 (58). This was corroborated by Southern blot analysis when a human CXCR2 probe was shown to hybridize to two similarly-sized fragments of murine DNA (58). However Cxcr2-transfected cells bound IL8 in a dose-dependent manner with similar affinity to that of murine neutrophils, and a single band was observed by Southern blot using a conserved CXCR2 probe against DNA from a murine cell line (56). The explanation for the apparent discrepancy of receptor number is likely the use of different Southern blot probes. In the first study, probes derived from either the 5' or 3' end of the human CXCR2 gene both hybridized twice but one of the two bands that were observed with the 5' probe was much weaker compared to similar intensity bands using the 3' probe. It was therefore interpreted that the 3' end of the human CXCR2 gene was more similar to the second receptor than the 5' end (58); alignment of the nucleotide and protein Consensus Coding Sequences (CCDS) of Cxcr1 and Cxcr2 supports this observation (alignment not shown). In the second study, the probe encompassed the internal region of the CXCR2 gene, which apparently only hybridized to Cxcr2; however a Cxcr2 probe hybridized twice to mouse neutrophil mRNA (56), again confirming the presence of another IL8 receptor.

A putative CXCR1 homologue, Cxcr1, was discovered in mice (59-61), although it is still not clear whether this is a true orthologue of CXCR1 due in part to a lack of murine Il8. The first reports of the cloning and tissue expression of Cxcr1 from BALB/c and C57BL/6 mice described its genomic position 30 kb downstream of Cxcr2, through 5' and 3' rapid amplification of cDNA ends (RACE) experiments (59, 60); the gene and protein sequences, as well as its genomic position are preserved in the current murine C57BL/6J reference genome assembly (NCBIM37). More recently, the putative Cxcr1 was also identified through computational searches of mouse expressed sequence tag (EST) data using the CXCR1 cDNA sequence, and subsequently cloned and expressed (61).

Cxcr1 is highly expressed on neutrophils, and in tissues from lung, spleen, caecum, thymus, placenta, testis and stomach, as well as in the mesenteric and peripheral lymph nodes (59-61). Cxcr1 is also expressed in the bone marrow and peripheral blood leukocytes (59, 61), specifically $CD3^+$ T cells and macrophages express both Cxcr1 and Cxcr2, and $CD4^+$ and to a smaller extent $CD8^+$ T cells express Cxcr1; Cxcr2 expression by $CD4^+$ or $CD8^+$ was not tested

(60). LPS-induced inflammation in the lung causes an influx of neutrophils, with increased Cxcr1 but decreased Cxcr2 expression (59). This may be the result of receptor desensitization, since lower concentrations of IL8 are needed for Cxcr2 internalization and Cxcr2 recycling is slower than that of Cxcr1 (52). Similarly, the expression of both Cxcr1 and Cxcr2 is induced in collagen-induced arthritis (61). It has been reported that Cxcr1 does not bind IL8, CXCL1, Cxcl1 or Cxcl2 in contrast to Cxcr2 and CXCR2, which both bind, to some degree, to the aforementioned chemokines (60). However assay conditions may not have fulfilled requirements for Cxcl1 activation by these ligands; Sf9 (*Spodoptera frugiperda*) insect cells were infected with baculoviruses expressing either Cxcr1, Cxcr2 or CXCR2 as well as the heterotrimeric G protein $\alpha_{i2}\beta_1\gamma_3$ (60). Neutrophil chemotaxis mediated by CXCR1 or CXCR2 can be inhibited by pertussis toxin but both receptors may couple with different heterotrimeric G protein subunits for activation (62). This may also be the case for Cxcr1 and Cxcr2. In contrast, Cxcr1 was reported to bind IL8 and CXCL6, analogous to CXCR1, as well as Cxcl6, as determined by binding assays, GTP exchange and chemotaxis of Cxcr1-expressing cells (61). Moreover Cxcr1 did not bind other CXCR2 ligands (61), indicating that this may indeed be the true CXCR1 orthologue. Further experimentation is required to verify this relationship.

1.5 Interleukin 10

IL10 is a major anti-inflammatory cytokine, produced by a variety of leukocyte and non-leukocyte cell types, which include monocytes, macrophages, myeloid DCs, T cells, B cells, mast cells, natural killer (NK) cells, eosinophils and neutrophils, as well as epithelial cells, endothelial cells and keratinocytes, respectively (63-66). IL10 was first described for its ability to polarize T cell responses. Production of IL10 by Th2 cells inhibits cytokine production by Th1 cells, however similar inhibition of Th2 cytokine production was not observed (67). Indeed IL10 inhibits IL12 production, but both CD4+ and CD8+ T cells produce IL10, and in particular the T cell subclasses Th1, Th2, Tr1 and Th17 (64, 66). IL10 suppresses T cells and NK cells by inhibiting cytokine, co-stimulatory molecule and MHC class II production from macrophages and DCs, in addition to upregulating anti-inflammatory molecules (63, 64).

IL10 can be expressed constitutively but also induced in response to a variety of stimuli such as microbial pattern-associated molecular patterns (PAMPs) including superantigens, and cytokines (66, 68). Engagement, by appropriate ligands, of TLR2, TLR4, TLR9, C-type lectin domain family 7 (CLEC7A) (DECTIN1) or CD209 (DC-SIGN) can induce IL10 production

(66). Additionally DEFB2 and IDR-1002 induce IL10 production from human PBMCs and murine bone marrow-derived macrophages, respectively (49, 69). An increase in Il10 *in vivo* upon IDR-1002 administration during infection was not observed (49). The control of IL10 production involves several levels of complexity including type, strength and number of stimuli, activation of enhancing or silencing transcription factors, chromatin remodeling, histone modifications (acetylation and phosphorylation), as well as post-transcriptional regulation (66). The regulation and timing of intestinal IL10 production *in vivo* with respect to either commensal or pathogen stimulation needs further elucidation (66). It is not clear which cells are key IL10 producers in the intestine and whether those that maintain homeostasis are different from those that are involved in the inflammatory response (66). Host defence peptides appear to selectively modulate the immune response, thus it is important to determine their role in immune regulation.

The functional IL10 receptor is composed of two subunits, IL10RA and IL10RB. IL10RA is expressed at low levels on most hematopoietic cells. Expression patterns are cell-type specific and can be up- or down-regulated during activation although IL10RA is constitutively expressed by human and mouse intestinal epithelial cells (63). IL10RB is also expressed by a wide variety of cells, however its expression appears to be constitutive in all of these (63). IL10RA binds IL10 at the surface of the cell and IL10RB transduces the intracellular signal through phosphorylation of Janus activating kinase (JAK) 1 and tyrosine kinase (TYK) 2, which can then activate signal transducer and activator of transcription (STAT) 1, STAT3 or STAT5. STAT3 seems to mediate the majority of IL10-induced responses through transcription of suppressor of cytokine signalling (SOCS) 3 (63).

IL10 has been implicated in a number of human diseases, particularly those involving intracellular pathogens including viruses, autoimmune disorders and sepsis (63). In some cases it is not clear whether IL10 is the cause or a consequence of the disease although high IL10 levels may predispose individuals to systemic lupus erythematosus (63). IL10 protects from endotoxin and death due to sepsis and may be involved in endotoxin tolerance *in vitro* (63). Interestingly certain viruses (EBV, herpes, poxvirus) encode an IL10 homologue, which share high similarity with human IL10 (63), which is likely one mechanism these viruses use to reduce the inflammatory response and aid survival within the host.

1.6 Cytokines in Gastrointestinal Disease

Gastrointestinal diseases pose a significant threat to human health worldwide. These include bacterial and viral infections, due to contaminated water and food, and direct person to person spread (70, 71), as well as autoimmune inflammatory bowel disorders and cancer (72, 73). While the pathology of these diseases is distinct, they share a common underlying etiology, the dysregulation of inflammation. Too much or too little inflammation can be a bad thing and cytokines, including both IL8 and IL10, have been implicated in inflammatory diseases (74, 75).

In humans Shigellosis or bacillary dysentery is caused by several Gram-negative *Shigella* species, *S. flexneri*, *S. dysenteriae*, *S. sonnei* and *S. boydii*. *Shigella* are transmitted via the fecal-oral route and are a serious threat to people in the developing world who do not have access to clean water and proper sanitation. The World Health Organization estimates *Shigella* spp. cause between 80–165 million cases of dysentery and 600,000 deaths annually; approximately 60% of the deaths are for children under five years old (70). There is also a significant burden on developed countries from *Shigella* infections in travellers to developing countries (70).

The virulence of *S. flexneri* is associated with a 140 Mb plasmid because strains lacking this plasmid are non-invasive (76). Gene expression of the human intestinal epithelial cell line, Caco-2, was analyzed by microarray following infection with wild type *S. flexneri* strain M90T or a mutant strain, BS176, lacking the 140 Mb plasmid (77). The gene most highly induced by the invasive strain, but not the mutant non-invasive strain, was *IL8* (~300-fold increase) (77). IL8 protein induction by the invasive strain, but not the non-invasive strain, was confirmed (77).

A vaccine protecting against *Shigella* infection has not been developed, in part because current animal models do not mimic the human disease (78). It has been postulated that the lack of murine Il8 may account for this discrepancy in disease severity because administration of recombinant human IL8 to mice infected with *S. flexneri* induces a disease which better resembles that of the human disease (79). The lack of neutrophil influx in the intestinal tract cannot be explained by differing neutrophil extravasation in human and mouse because a large neutrophil infiltrate is observed in pulmonary *Shigella* infection (80). This difference may be due to a more protective barrier in the intestine due to the barrage of microbes in the gastrointestinal tract compared to that of the lungs. *Shigella* is phagocytosed by resident macrophages and DCs in the follicle following transport through the follicular-associated epithelium via M cells. It then escapes from the vacuole and induces caspase-1 mediated apoptosis of the macrophage or DC (78). Caspase-1 activation increases IL1B production, which leads to neutrophil infiltration,

increased inflammation and increased bacterial invasion (78). IL8 is therefore not the only mediator of neutrophil influx but its absence in mice may be one reason why less neutrophil infiltrate is observed in many mouse models compared to that observed in the human disease, especially those of the gastrointestinal tract.

Therefore it has been hypothesized that IL8 leads to increased neutrophil recruitment and inflammation, and subsequent *Shigella* infection. However this may not be the same for all enteric diseases as neutrophil recruitment could increase clearance of the microorganism. It is unlikely that the same mechanism is exploited by all bacteria similar to the fact that IL1B can be detrimental in terms of *Shigella* infection but is beneficial in others, e.g. *Escherichia coli* (78), and *Candida albicans* (81). *Shigella* is a mucosal pathogen whereas *Salmonella* can cause systemic infection and thus comparison of these two infections will lead to a better understanding of how IL8 mediates disease pathology.

Salmonella enterica serovar Typhi (*S. typhi*) is the causative agent of typhoid fever in humans. An estimated 17 million cases of typhoid fever occur annually and account for 600,000 deaths (70). *S. typhi* is a human pathogen and the disease is modeled in mice by use of *Salmonella enterica* serovar Typhimurium (*S. typhimurium*), which causes the human enteric disease gastroenteritis. This disease can be modeled in mice by administration of antibiotics before infection with *S. typhimurium* (82). There are significant differences in the human diseases caused by *S. typhimurium* and *S. typhi*, in particular patients with *S. typhimurium* have a large intestinal neutrophil infiltrate which is not observed in the disease caused by *S. typhi* (83). This combined with the host-restricted nature of *S. typhi* has led to substantial research comparing these two organisms, both at the genetic and host-interaction levels (82-84). Mice orally infected with *S. typhimurium* develop a disease similar to typhoid fever in humans, as the bacteria invade the intestinal mucosa without neutrophil infiltration or epithelial damage, and then disseminate to the spleen and liver, where granulomas form (82). Mice treated with antibiotics prior to *S. typhimurium* infection develop colitis, which is characterized by a large neutrophil infiltrate consistent with the human disease (82). *S. typhimurium* can also disseminate in this model depending on the genetic background of mice (82). The *Salmonella* pathogenicity islands (SPI) encode virulence factors, which contribute to colonization and invasion of the host (82). *S. typhimurium* contains at least two type III secretion systems (TTSS), SPI-1 and SPI-2 (84). The SPI-1 TTSS is necessary for the translocation of virulence factors into intestinal epithelial cells, thereby *S. typhimurium* promotes its own invasion of these cells (82). Following

successful translocation the SPI-2 TTSS is required for invasion of macrophages, replication and dissemination (82). Investigations of *S. typhimurium* virulence should therefore encompass both epithelial cells and leukocytes. Novel murine models of intestinal inflammation are needed to accurately assess the mechanism by which host defence peptides protect against infection *in vivo*. Additionally there are human immunodeficiencies that increase the susceptibility to *Salmonella* infection. The study of leukocytes from these patients *ex vivo* can be used to complement the *in vivo* murine models of infection, thereby providing a better understanding of the disease.

IL10 is important for immune regulation but during infection, the source of IL10 and the timing of its production are not well understood (64). IL10 production, whether by epithelial cells or intestinal lymphocytes, in the intestinal tract likely plays an important role in the early stages of the immune response, and in general immune surveillance for differentiation between enteric pathogens and commensal organisms. As bacterial-induced inflammation can promote epithelial barrier destruction leading to translocation and successful infection, it is hypothesized that reducing this inflammation through the overexpression of Il10 in the mouse intestinal tract would limit the epithelial damage induced by enteric pathogens and prevent infection and widespread dissemination, thus increasing survival following bacterial challenge. This is consistent with increased survival of mice with reduced lung inflammation following intranasal infection with Il10-secreting *Shigella flexneri* (85). However if the bacteria are able to invade the epithelial barrier in high enough numbers, it is possible that mice will succumb to bacterial infection due to the inhibitory effects of Il10. Novel murine models of immune suppression are also needed to further address the role of host defence peptides in immune homeostasis.

1.7 Project Goals and Hypotheses

This aim of this project was to further elucidate the role of host defence peptides, in particular defensins, in the innate immune system through *in vitro* stimulation of human peripheral blood mononuclear cells. Originally the generation of an α-defensin deficient mouse was proposed in order to study the *in vivo* role of α-defensins within the murine intestinal tract. This highlighted the need for a complete characterization of murine α-defensins within the C57BL/6J reference strain, as well as the development of alternative models with which to study the function of defensins.

The key hypotheses pursued here were:

1. Human α-defensins induce the production of chemokines and cytokines from human peripheral blood mononuclear cells.
2. The discrepancy between genomic and peptide α-defensin repertoire indicates murine α-defensin expression is tightly regulated.
3. Murine models of inflammation and immunosuppression can be used to test the efficacy of host defence peptides against gastrointestinal disease.
4. Cells of patients with immunodeficiencies can be used to dissect the mechanisms of susceptibility to *Salmonella* and mycobacterial infections.

The specific research questions addressed in this thesis are:

1. How extensive and diverse is the alpha defensin locus in mouse and is there evidence in the DNA sequence signatures of how it evolved?
2. Do the different alpha defensin genes represent a repository of silent genes for generation of diversity or are they actually expressed?
3. Since defensins work in part through induction of chemokines and cell recruitment (which are also exceptionally diverse and functionally redundant) are there genetic approaches to permit this property to be explored in infection models?

2 α-DEFENSIN GENE ANNOTATION, QUANTITATIVE TRANSCRIPT EXPRESSION AND FUNCTIONAL ACTIVITY

2.1 Introduction

In humans, α-defensins are most abundant in neutrophils and Paneth cells of the small intestine, however the neutrophils of mice do not contain α-defensins (30, 34), despite the fact that mice have the largest known repertoire of defensin-encoding sequences. Therefore the Paneth cell α-defensins are the only option when considering a murine model with which to study their function. However the fact that murine α-defensins are contained within discrete compartments within the small intestine, i.e. Paneth cell granules, is an advantage for the generation of such a model. Knocking out α-defensins in mice would allow the study of their function in the tolerance of normal flora, upon bacterial infection and during chronic inflammatory diseases. At the beginning of this project, there was increasing evidence of the importance of defensins but their mechanism of action was, and still is, not clear. Additionally no α-defensin deficient mouse models had been generated for the *in vivo* study of defensin function. It was suggested that human neutrophil α-defensins chemoattract immature dendritic cells and T cells *in vitro* (86), and subsequently macrophages and mast cells but with conflicting results for previous observations (87). *In vitro*, in standard low phosphate buffer (10 mM), murine α-defensins have varying antimicrobial activities (88). *In vivo*, β-defensin knock-out mice are more susceptible to certain bacterial infections but it might be difficult to determine the effect of knocking out only one defensin because of the apparent redundancy of these peptides, which in itself if an implication of their importance. *Defb1*, encoding the constitutively expressed β-defensin 1, deficient mice had similar bacterial load to wild type mice but higher numbers of *Staphylococcus* bacteria in the bladder; bacterial clearance from the lung was not altered (89). The redundancy of these peptides is apparent by the delayed, but eventual, ability of Defb1 deficient mice to clear *Haemophilus influenzae* from the lung (90). In support of the immunomodulatory role of defensins, administration DEFA1 (HNP-1) not only decreased the bacterial load following peritoneal administration of *Klebsiella pneumoniae* to mice, but also increased the numbers of macrophages, granulocytes and lymphocytes in the peritoneal cavity, both in a dose-dependent manner (91). The antibacterial activity of DEFA1 was abolished in leukocytopenic mice upon treatment with cyclophosphamide (91). Thus it can be argued that the killing activity of DEFA1 is secondary to the recruitment of effector cells of the immune system.

Paneth cell *DEFA5* transgene knock-in mice are less susceptible to *S. typhimurium* infection and death as compared to wild type mice (43). DEFA5 was knocked-in because it possesses the largest antimicrobial activity of defensins *in vitro*, but this was determined under low salt concentration (10 mM sodium phosphate) (92). The authors argue that the differences seen in the transgenic mice compared to the wild type, with respect to bacterial load and survival, must be due to direct antimicrobial activity of DEFA5 because of the short time frame in which the differences were seen (6-12 hours) (43). However, *in vitro* studies show that in response to defensins, chemotaxis of a variety of cell types (monocytes, mast cells, immature dendritic cells, naïve T cells) can occur in as little as 1-3 hours, and peptides can cause significant increases in cytokine production in 4 hours (86, 93, 94). Recently, β-chemokines have been shown to induce neutrophil chemotaxis in one hour (95), and given the similar tertiary structure and some functions, β-defensins may also be able to chemoattract neutrophils. If other defensins have the ability to chemoattract various immune effector cells, it is possible that DEFA5 could have similar abilities and that these events could occur within the timeframe indicated for its *in vivo* activity. This mouse model and these experiments elegantly show the importance and often underestimated role that defensins play in the proper functioning of the immune system. Direct antimicrobial activity by DEFA5 was not shown definitively, although DEFA5 alters the microbiota composition within the mouse small intestine (96). It is possible that there are also other equally important functions of DEFA5, and indeed other defensins, as compared to direct antimicrobial activity. It is also difficult to determine whether the effects seen are more of an additive effect due to the increased expression of defensins, endogenous plus DEFA5, in the mouse intestine.

Matrix metalloproteinase (MMP) 7 deficient mice were available however, in addition to α-defensin processing, MMP7 is also involved in matrix remodeling (97), wound healing (98) and neutrophil migration (99), all of which could significantly affect bacterial-induced inflammatory responses. The approach of knocking out endogenous α-defensins in mice may be a better indication of their function, in addition to complementing the knock-outs either genetically or by the addition of exogenous peptide. The choice of gene for deletion is complicated owing to the fact that there is not a clear orthologous relationship between human and mouse Paneth cell α-defensins. Whereas humans express two Paneth cell α-defensins, DEFA5 and DEFA6 (100), over twenty α-defensin genes or defensin-related genes have been described in several strains of mouse. Six of these, Crp1-6 that correspond to *Defcr1-6*, have

been purified as peptides from Outbred Swiss mice (101). In this strain, Crp1 is most abundant in adult mice (29), whereas *Defcr6* is most abundant in newborn mice (102). Crp2 and 3 differ in the primary sequence compared to Crp1 at only three amino acids (29). Murine α-defensins are also differentially expressed throughout the small intestine. *Defcr1* and *Defcr5* expression is equivalent in duodenum, jejunum, and ileum, whereas the others show regional differences (103). Taking all of these factors into consideration, the generation of *Defcr1* and *Defcr5* deficient mice was proposed.

In retrospect, for reasons discussed below, these two genes were the least suitable for knock out in the C57BL/6J mouse but serendipitously it drew attention to annotation and assembly problems within the α-defensin region of the C57BL/6J reference genome, as well as differences in α-defensin genomic content between different strains of mice. It also highlighted discrepancies between genome databases, which, until their resolution, prevented any α-defensin deficient mouse generation. A collaboration was established between the Centre for Microbial Disease and Immunity Research at the University of British Columbia (Vancouver, BC, Canada) and the Wellcome Trust Sanger Institute (Hinxton, Cambridge, UK) to investigate the genomic structure of α-defensins, with the ultimate goal of the generation of an α-defensin deficient mouse, with which to study their *in vivo* function.

2.2 Methods and Materials

2.2.1 Genomic analysis and annotation pipeline

Prior to the process of manual annotation, an automated analysis for similarity searches and *ab initio* predictions is run in an extended Ensembl analysis pipeline system (104). All search results are stored in an Ensembl MySQL database. Following genomic sequence masking of interspersed repeats and tandem repeats by RepeatMasker and Tandem repeats finder (105), a WU-BLASTN search against the nucleotide databases is performed. Significant hits are then re-aligned to the unmasked genomic sequence using est2genome (106). The Uniprot protein database is then searched with wuBLASTX. In order to provide prediction of protein domains Genewise (107) is used to align hidden Markov models for Pfam (Protein family database) protein domains against the genomic sequence. Finally, a number of different *ab initio* algorithms are used: Genescan predicts genes (108), tRNAscan predicts tRNA genes (109), and Eponine TSS predicts transcription start sites (110).

After completion of the automated analysis, manual annotation starts using a Perl/Tk based graphical interface, called 'otterlace', to edit annotation data stored in a separate MySQL database system (111). Otterlace provides tools for changing exon coordinates, adding gene names and remarks, assigning genes to different categories or adding genomic features such as polyA sites and signals. The annotation of gene objects requires a visual representation of the genomic region and features such as CpG islands, repeats and polyA sites, gene predictions, evidence to support the annotation of gene structures (EST/cDNA/protein), and all transcript variants created by annotators. This representation is provided by a graphical user interface called ZMap, which was written in C programming language to give the high performance required to display large numbers of features (*personal communication*, R. Storey). An alignment viewer called 'Blixem' allows gapped alignments of nucleotide and protein blast hits to be compared with the genomic sequence (112). Furthermore, a 'Dotplot' tool called 'Dotter' is used to show pair-wise alignments of unmasked sequences, revealing the location of exons that are occasionally missed by the automated blast searches due to their small size and/or match to repeat-masked sequence (112).

All annotation is publicly available in the Vertebrate Genome Annotation (VEGA) browser (113). Definitions of Vega gene and transcript types are also available. Subsequent to the first annotation of the region, the C57BL/6J genomic assembly changed slightly and some of the gene names have been changed. Figure 2.1 represents the current version at the time of publication and writing of this thesis. Vega will be updated accordingly.

2.2.2 Mouse genomic assembly

The manual annotation of the defensin gene cluster region was based on NCBI Build 36 (NCBIM36). However, at the time of writing the new build NCBIM37 was released, and the two assemblies show several differences. The most crucial one is that a new clone AC161189 was added to the new assembly which overlaps partially with and has replaced clone AC140205. Although most genes initially annotated in AC140205 are present in AC161189, there are seven loci missing. One of these codes for β-defensin 33 (*Defb33*) and the remaining ones are pseudogenes. In order to preserve this data we propose that the clone AC140205 should be trimmed from the point where it is unique and be returned to the new assembly. The Genome Reference Consortium (114), which is a collaborative effort between NCBI, WTSI, EMBL-EBI and the Genome Center at Washington University, aim to close remaining gaps in the human and

mouse genomes and remove discrepancies in clones observed by research groups. The issue described here has been submitted to the Genome Reference Consortium.

2.2.3 Size estimation of the 2 Mb gap of C57BL/6J mouse Chromosome 8

The genomic sequence of Chromosome 8:18,508,450-24,203,501 was downloaded from Vega (v. 23, NCBIM36), and N's in the sequence, which represent gap regions, were replaced by a space. The sequence was then digested *in silico* at *Swa*I-predicted sites using RestrictionMapper (v. 3) (115). The ordered digest fragments were aligned with *Swa*I-digested DNA fragments from the same region of the Optical Map for murine C57BL/6J Chromosome 8, which was kindly provided by Steve Goldstein (Genome Center of Wisconsin, University of Wisconsin-Madison). The alignment and agreement of fragment sizes between the predicted and *in vitro* digests was determined manually and the results displayed in Excel. The presence of the inverted tandem repeat (Contig AC152164.14) was determined by S. Goldstein (*personal communication*).

2.2.4 Genome browser/ database gene set comparison

The Vega mouse α-defensin genes annotated here were used to query the gene sets contained within the Ensembl (v. 50), MGI (v. 4.11) and NCBI (Build 37.1) databases. Both Vega gene symbols and database identifiers (OTTIDs) were used because, depending on the database and gene, different results were sometimes returned with either search. As well, not all databases recognize the Vega accession number if a gene name has not been assigned. A gene linked to Vega indicates that there is a reference to the Vega Gene/ OTTID as being the same gene (i.e. mapped to the same position on the chromosome), but has been given a database-specific name. If the database had referenced another database to obtain the OTTID but there is no acknowledgement to the mapping in Vega, the gene was not considered linked and put into the Additional Gene column. Genes that were identified in the searches but were different to the Vega gene were put into the Additional Gene column. NCBI was searched with OTTIDs and gene name because some Vega accession numbers return bacterial artificial chromosome (BAC) clones in the results. NCBI does not cross-reference Vega, rather they obtain their information from MGI, so none of the genes are directly linked to Vega; genes linked to the Vega-linked Ensembl genes are therefore listed.

2.2.5 Genetic analysis of α- and CRS-defensins

Nucleotide and protein sequences of α- and CRS-defensins were aligned using ClustalW2 multiple sequence alignment software, with default parameters (116). Nucleotide ClustalW2 alignment files (.aln) were visualized using the GeneDoc software (117); protein ClustalW2 alignment files were visualized in .aln format.

Basic local alignment search tool (BLAST) searches were performed using NCBI BLAST 2.2.18+ and WU-BLAST2, with default and variable parameters, with both CRS-defensin genomic and peptides sequences. BLASTP (2.2.18+) searches only returned annotated defensin peptides therefore unannotated nucleotide databases (whole genome shotgun, expressed sequence tag, and draft genomes) were queried using TBLASTN. CRS-defensin peptide sequences were also used to query Pfam (v. 22.0) using the batch search option, with default parameters.

2.2.6 Transcriptional expression profiling of α- and CRS-defensins in C57BL/6J mice

2.2.6.1 Universal defensin primer design

Primers designed to amplify all annotated α- and CRS-defensin cDNAs in a single PCR reaction can be found in Table 2.1; these primers are denoted as universal defensin primers. The 26 mouse α- and CRS-defensin proteins were aligned using ClustalW2 (v. 2.0.8) with default parameters (116), and then PAL2NAL (v. 11) used to align the corresponding coding sequences according to their relative codon alignments (118). The output PAL2NAL alignment was visualized with the GeneDoc software (117). Additionally, the α- and CRS-defensin 5'-untranslated region (UTR) sequences were downloaded from Ensembl (v. 50) with BIOMART using Ensembl Gene IDs, corresponding to their Vega Gene IDs, and then aligned using ClustalW2 (v. 2.0.8).

Table 2.1. Primer sequences for universal defensin expression profiling by capillary sequencing.

Primer Name	Sequence (5'-3')
UDefF1	ATGAAGACAYTWGTCCTCCTCTCTG
UDefR	GCAGCACAGATACGACTCACG
UDefR-oligo-dT	GCAGCACAGATACGACTCACGTTTTTTTTTTTTTTTTTVN

The consensus sequence for the start of the coding sequence was identified as the longest stretch of 100% identity between all genes. OligoCalc was used to calculate GC content, melting temperature and secondary structure formation of potential primers (119). The UDefF1 primer comprised the nucleotide sequence for the first 25 bases following the start of each defensin coding sequence, which are all, except four positions, conserved across all defensin sequences. At these positions, the base included in the primer sequence was obtained either from the consensus sequence or else left as an ambiguous base. The differentiation was dependent on the relative number of genes with a substitution in that particular position. The first nine and last ten bases of the UDefF1 primer have 100% identity across all defensin genes, which should ensure accurate and specific amplification regardless of the intervening ambiguities.

Reverse transcription was performed using the UDefR-oligo-dT hybrid primer. The oligo-dT sequence, based on literature searches (120, 121), and commercially available kits, contains 18 thymidines, as well as [AGC][N] to anchor the primer at the end of the transcript. The non-dT or anchor part of the hybrid primer, denoted UDefR, was designed to be equivalent to the UDefF1 forward primer in length, GC content and melting temperature so that following reverse transcription, amplification was straightforward using the UDefF1 and UDefR primers. The UDefR sequence was derived from the 3' RACE Adaptor sequence, 5'-GCGAGCACAGAATTAATACGACTCACTATAGGT$_{12}$VN-3', from the FirstChoice® RLM-RACE Kit (Ambion, Inc.). This sequence was modified according to BLAST results to obtain a sequence with no significant homology to the mouse genome or transcriptome, whilst still having a similar melting temperature to that of the forward UDefF1 primer (119).

2.2.6.2 Animals

Naïve C57BL/6J mice, three female and three male denoted as M1, M2, M3 and M4, M5, M6, respectively, were used at 4-8 weeks of age. The mice were killed by carbon dioxide asphyxiation, and the small intestine removed and divided into seven equal segments. Faeces were removed from each segment, which was then cut open along its length and flattened. A tissue culture cell scraper was used to scrape along the entire segment to remove epithelial and Paneth cells. The scrapings were immersed in 500 µl RNAlater (Qiagen) in a 1.5 ml Eppendorf tube, and stored at 4°C for two days, upon which time RNA was extracted.

2.2.6.3 RNA extraction

RNA was extracted using the RNeasy Mini kit with QIAshredder columns (Qiagen), as per the manufacturer's instructions. Briefly, the samples in RNAlater were microcentrifuged at

maximum speed for 10 minutes to pellet the tissue, and the RNAlater supernatant removed. RLT lysis buffer (600 µl) containing 143 mM β-mercaptoethanol (Sigma Aldrich) was added to the sample and vortexed well. Tissues difficult to lyse were also pipetted to aid lysis. The entire lysate was transferred to a QIAshredder column, microcentrifuged at maximum speed for 2 minutes, added to a new tube containing 600 µl 70% ethanol (EtOH), and mixed by pipetting. The sample (600 µl) was transferred to an RNeasy spin column and microcentrifuged at maximum speed for 30 seconds; this step was repeated for each sample using the same column. The rest of the RNeasy Mini Handbook was followed, with on-column DNase digestion for 30 minutes at room temperature using 15 µl RNase-Free DNaseI and 115 µl RDD buffer (Qiagen) per sample. Each column was eluted twice with 50 µl nuclease-free water (NF-H_2O) (Ambion, Inc.) giving a final volume of approximately 100 µl per sample. The RNA quality was assessed using the Agilent 2100 Bioanalyzer 2100 Expert Version B.02.05.SI360 (Agilent Technologies), and the concentration determined using the Thermo Scientific NanoDrop™ ND-1000 Spectrophotometer (Thermo Fisher Scientific Inc.), both as per the manufacturers' instructions.

2.2.6.4 Reverse transcription

First strand synthesis reverse transcription (RT) was performed using the Qiagen QuantiTect Reverse Transcription kit, as per the manufacturer's instructions. Briefly 1000 ng total small intestinal RNA per mouse (samples M1-M6) was pooled from the individual extractions in a final volume of 12 µl. A further DNase digestion was carried out with 2 µl gWipeout buffer at 42°C for 2 minutes. The RT reaction contained 5 µl RT buffer, 4 µl 10 mM UDefR-oligo-dT primer (0.5 µg) and 1 µl RT enzyme, and was carried out at 42°C for 60 minutes, followed by 3 minutes at 95°C; the cDNA was stored at -20°C. Four RT reactions were performed for each sample, as well as quarter-sized no RT enzyme reactions, making up the volume with NF-H_2O, as the control for DNA contamination within the RNA extractions. Aliquots of cDNA were then pooled for each sample.

2.2.6.5 PCR

The mouse small intestinal cDNA from samples M1-M6 were amplified using Platinum Taq DNA Polymerase High Fidelity system (Invitrogen). Each reaction contained 1X High Fidelity PCR buffer, 200 µM each dNTP, 2 mM $MgSO_4$, 200 nM each primer (UDefF1, UDefR), 1 U Platinum Taq High Fidelity polymerase and 2.5 µl pooled cDNA in a volume of 50 µl. Cycling conditions were as follows: 94°C for 2 minutes, 35 cycles of 94°C for 30 seconds,

68°C for 30 seconds, 68°C for 60 seconds, a final extension of 68°C for 10 minutes, then 4°C. PCR products (5 µl) were visualized by agarose gel electrophoresis (1% agarose, 80 V, 70 min).

2.2.6.6 TOPO cloning

The PCR products from each sample (M1-M6) were cloned into the pCR®4Blunt-TOPO® plasmid, according to the Zero Blunt® TOPO® PCR Cloning Kit for Sequencing instructions (Invitrogen). Ligation reactions performed at room temperature for 30 minutes contained 4 µl PCR product, 1 µl pCR4BluntII-TOPO vector and 1 µl dilute (1/4) salt solution. A pCR4BluntII-TOPO vector-only control was also included, making up the volume with 4 µl NF-H_2O. Each ligation reaction (2 µl) was then added to 50 µl One Shot® TOP10 Electrocomp™ *E. coli* cells (Invitrogen). The entire volume was added to a 0.1 mm cuvette (Cell Projects Limited, UK) and electroporated using the Bio-Rad Gene Pulser Xcell Electroporation System (Bio-Rad Laboratories, Inc.) with the pre-set *E. coli* bacterial electroporation program (25 µF, 200 Ω, 1800 V, exponential decay). For the recovery, 250 µl super optimal broth with catabolite repression (SOC, Invitrogen) was added and the entire volume transferred to a 15 ml Falcon tube and incubated at 37°C with shaking at 200 rpm, for 90 minutes. A positive control (1 µl pUC19 plasmid, Invitrogen) was also electroporated under the same conditions to ensure TOP10 cells were transformation competent. The *E. coli* cells (20, 50 or 80 µl replicates) were spread onto low salt Luria Bertani (LB-Luria) agar (1.5%) plates containing 50 µg/ml ampicillin (Roche Applied Science) and 50 µg/ml kanamycin (Gibco) for TOPO ligations, and 100 µg/ml ampicillin for pUC19, and incubated at 37°C overnight. For each sample, 96 individual colonies were picked and grown overnight at 37°C and 200 rpm in 500 µl LB-Luria broth containing 50 µg/ml ampicillin and 50 µg/ml kanamycin in a 96-well deep well plate (BD Biosciences).

2.2.6.7 Capillary sequencing of TOPO clone inserts

Plasmid extraction from the overnight cultures (96-well format), and capillary sequencing of the inserts using standard T3 and T7 primers was performed by the WTSI core sequencing facility. Capillary sequencing (Applied Biosystems 3730XL capillary sequencer) was performed using BigDye Terminator BDTv3.1 sequencing chemistry. The sequencing facility (Team 56) processed the sequence trace files using the sequencing production software Asp and then deposited those that passed quality control into a central repository. I performed further analyses of the sequences.

2.2.6.8 Capillary sequencing analysis

The T3 and T7 sequences were converted from experimental file format to FASTA file format using the WTSI program exp-piece and following command, where n = 1-6 for each M1, M2, M3, M4, M5, M6 sample, the prefix for each experimental file is T15_lr5_090929UDefMn and the prefix for each resulting FASTA file is 100429_UDEF_Mn_exp-piece_FASTA. The exp-piece option (left right) clips the sequences, providing the first and last useful bases.

```
for read in `indir | grep "T15_lr5_090929UDefMn"`; do exp-piece left right $read
>> ~lr5/1793864-1793875/100429_UDEF_Mn_exp-piece_FASTA
```

The α- and CRS-defensin transcript FASTA sequences were downloaded from Vega (v. 37) using Ensembl Biomart. A database was created from the text file of the sequences using the following command, where –i is the input file, –p is the type of file (nucleotide, therefore false), and –o allows creation of the indices in FASTA format.

```
formatdb –i 100428_defensin_transcript_OTTIDs_FASTA_Vega37.txt –p F –o T
```

BLAST was then used to batch query each Mn_exp-piece_FASTA file (n=1-6), against the defensin transcript database using the following commands, where blastall is the program, –p indicates program name for type of blast (nucleotide), –d is the database name, –i is the input query file name, –o is the name of the output file, –b is the number of matches to retrieve (5), –m is the alignment output format (0 is pairwise, 8 is tabular, 9 is tabular with headings), –e is the expected value for significant alignments, and –F indicates any sequence filtering (F(alse) turns off filtering to prevent masking of repetitive sequences).

```
blastall –p blastn –d 100428_defensin_transcript_OTTIDs_FASTA_Vega37.txt –i
100430_UDEF_Mn_exp-piece_clipped_FASTA –o 100430_UDEF_M1_exp-
piece_clipped_v_transcript_5_alignments –b 5 –m 0 –e 1e-05 –F F
```

```
blastall –p blastn –d 100428_defensin_transcript_OTTIDs_FASTA_Vega37.txt –i
100430_UDEF_Mn_exp-piece_clipped_FASTA –o 100430_UDEF_Mn_exp-
piece_clipped_v_transcript_5_tabular –b 5 -m 0 –e 1e-05 –F F
```

All sequences with significant BLAST results were inspected manually using the trace viewing software Trev (v. 1.9), which is part of the Staden Package and displays both the base calls and confidence values. If a base-calling error was observed, the sequence position was noted in the tabular output file and the sequence corrected in the BLAST alignment file. As the

BLAST analysis was performed against the database of full-length defensin transcript sequences, any mismatches in the UDefF1 primer region, either due to primer synthesis defects or ambiguous base incorporation, were ignored. The percentage identity of the best BLAST hit for each (edited) sequence read was noted and the results tabulated.

2.2.6.9 Universal defensin 454 primer design and multiplex identifier tag incorporation

The primers used for defensin cDNA amplification prior to 454 sequencing were the same as those used prior to capillary sequencing, except that sequences for 454-specific adaptors and GS multiplex identifier (GSMID) tags were incorporated (Table 2.2). The Adaptor A and B sequences were added 5' to the UDefF1 and UDefR primer sequences, respectively, and the sequences for the GSMID tags were added between the adaptor and UDefF1 sequences. This allowed 454 sequencing of the defensin PCR products in the forward direction, with respect to transcript orientation, beginning at the start codon. Sequencing of the reverse strand is not possible because of the homopolymeric thymidine region, due to the transcript polyA tail. The GSMID sequences were added to each forward primer for multiplex sequencing of pooled PCR products. Standard Adaptor and GSMID sequences were obtained from Roche.

Table 2.2. Primer sequences for universal defensin expression profiling by 454 amplicon sequencing. Adaptor sequences are in italics, GSMID sequences are underlined, followed by the UDefF1 or UDefR sequences for the UDefF1-454 and UDefR-454 primers, respectively. The Roche standard adaptor and MID nomenclature are indicated.

Primer Name	Sequence (5'-3')	Adaptor	GSMID
UDefF1-454-1	*CCATCTCATCCCTGCGTGTCTCCGACTCAG* ACGAGTGCGTATGAAGACAYTWGTCCTCCTCTCTG	A	1
UDefF1-454-2	*CCATCTCATCCCTGCGTGTCTCCGACTCAG* ACGCTCGACAATGAAGACAYTWGTCCTCCTCTCTG	A	2
UDefF1-454-3	*CCATCTCATCCCTGCGTGTCTCCGACTCAG* AGACGCACTCATGAAGACAYTWGTCCTCCTCTCTG	A	3
UDefF1-454-4	*CCATCTCATCCCTGCGTGTCTCCGACTCAG* AGCACTGTAGATGAAGACAYTWGTCCTCCTCTCTG	A	4
UDefF1-454-5	*CCATCTCATCCCTGCGTGTCTCCGACTCAG* ATCAGACACGATGAAGACAYTWGTCCTCCTCTCTG	A	5
UDefF1-454-6	*CCATCTCATCCCTGCGTGTCTCCGACTCAG* ATATCGCGAGATGAAGACAYTWGTCCTCCTCTCTG	A	6
UDefR-454	*CCTATCCCCTGTGTGCCTTGGCAGTCTCAG* GCAGCACAGATACGACTCACG	B	n/a

2.2.6.10 454 sequencing amplicon PCR

The conditions for the 454 sequencing amplicon PCR were the same as previously described, except the UDefF1 primer was replaced with the UDefF1-454 primers (1-6), and each sample (M1-M6) amplified with the corresponding numbered UDefF1-454 primer. The same UDefR primer was used in all reactions. Four PCR reactions were set-up for each sample, and the products (5 μl) visualized by agarose gel electrophoresis (1% agarose, 80 V, 35 min). For each sample 10 μl of each PCR reaction was pooled, and then 2 μl of each sample pool was

pooled. The final M1-M6 PCR pool was quantified using the Quanti-iT™ PicoGreen dsReagent Kit (Qiagen), as per the manufacturer's instructions.

2.2.6.11 454 amplicon sequencing

The final M1-M6 PCR pool (18 ng/µl, 11 µl) was submitted to the WTSI core sequencing facility (Team 130) for GS FLX Titanium Series Amplicon 454 Sequencing (Roche). All subsequent sequencing protocols, data generation and quality control were performed by staff of Team 130. Briefly, the M1-M6 PCR pool was denatured to give single-stranded fragments, which were then immobilized onto DNA Capture Beads (Roche). The beads were subjected to emulsion PCR (emPCR) and then loaded onto a PicoTiter Plate (Roche). Sequencing was performed with the Genome Sequencer FLX Instrument (Roche). The data were then split into six files based on their GSMID tags (i.e. M1-M6), and deposited into a central repository as Standard Flowgram Format (SFF) files. I performed all subsequent analyses.

2.2.6.12 454 amplicon sequence analysis

The SFF files contain the flowgram, base-called sequences and quality scores for each sequencing read (Roche). FASTA files were extracted from each MIDn.sff file, where n=1-6, and saved in a new file using the following command.

```
sffinfo –s Region8.MID1.sff > 100512_Region8.MID1.sff.fa
```

All sequences in each FASTA file were counted with the following command, where n=1-6, and the results tabulated.

```
grep ">" 100512_Region8.MIDn.sff.fa | wc –l
```

The FASTA files were filtered for those over 300 nucleotides using a perl script kindly provided by Gregory Baillie (WTSI), and the following command, where n=1-6. The options for the get_seqs_within_length_range.pl script are minimum and maximum length, to which 300 and arbitrarily 10000 were assigned.

```
~gb7/bioinformatics/perl_scripts/get_seqs_within_length_range.pl
100512_Region8.MIDn.sff.fa 100512_Region8.MIDn_over300.sff.fa 300 10000
```

All sequences in each filtered FASTA file were counted with the following command, where n=1-6, and the results tabulated.

```
grep ">" 100512_Region8.MIDn_over300.sff.fa | wc -l
```

A database of α- and CRS-defensin sequences was created using the following command, and the same FASTA sequences as for the capillary sequencing analysis, except that any 5'-UTR and the UDefF1 primer sequences were removed (trimmed transcripts).

```
formatdb –i 100512_ defensin_transcript_trimmed_OTTIDs_FASTA_Vega37.txt –p F –o T
```

Nucleotide BLAST analysis was then performed with each 300 nucleotide filtered MID FASTA file against the trimmed transcript database, for the top hit in tabular form using the following command, where MIDn=MID1-6.

```
blastall –p blastn –d 100512_defensin_transcript_trimmed_OTTIDs_FASTA_Vega37.txt –i 100512_Region8.MIDn_over300.sff.fa –o 100512_Region8.MIDn_over300.sff.fa_v_trimmed_transcript –b 1 –m 8 –e 1e-05 –F F
```

For each MID, new files of all sequencing reads with 100% identity to any of the reference sequences, over an alignment length of greater than 300 nucleotides, were created and sorted, and then counted using the following commands, where n=1-6.

```
awk '{if ($3 == 100 && $4 > 300) {print $2 "\t" $3}}' 100512_Region8.MIDn_over300.sff.fa_v_trimmed_transcript | sort –k 1,1 > 100512_Region8.MIDn_over300.sff.fa_v_trimmed_transcript_100hits

wc –l 100512_Region8.MIDn_over300.sff.fa_v_trimmed_transcript_100hits
```

The number of sequencing reads matching each reference sequence with 100% identity was quantitated for all MID files with the following command. The MID1-6 variable was first defined and then the filtered BLAST files were then sorted according to reference sequence and the number of unique entries counted.

```
for mid in 1 2 3 4 5 6; do echo "Mid $mid"; sort 100512_Region8.MID${mid}_over300.sff.fa_v_trimmed_transcript_100hits | uniq –c; done
```

The BLAST analysis was repeated with the previous blastall command with alignments as the output (-m 0), however this also contains alignments under the 300 nucleotide cut-off used for the previous analyses.

2.2.6.13 Identification of unannotated C57BL/6J defensin transcripts

The 454 amplicon sequencing reads that did not match any of the defensin reference transcript sequences with 100% identity were further analyzer for putative novel variants. Peptide analysis of the C57BL/6J murine α- and CRS-defensins has revealed an α-defensin, termed Crp27, whose corresponding cDNA sequence, termed *Defcr27*, has not been annotated to the reference genome (*personal communication*, Andre Ouellette and Michael Shanahan). BLAST analysis was therefore performed to search the 454 sequences for matches to the Crp27 peptide.

A database of peptide sequences, including Crp27, and the six other α-defensin peptides found to be present in the C57BL/6J mouse as controls (*personal communication*, A. Ouellette and M. Shanahan), was created using the following command.

```
formatdb –i 100527_Crp_protein_7_FASTA –p T –o T
```

Each extracted MID FASTA file, without 300 nucleotide filtering, was then blasted against the peptide database using the following command, where n=1-6.

```
blastall –p blastx –d 100527_Crp_protein_7_FASTA –i
100512_Region8.MIDn.sff.fa –o 100527_Region8.MIDn.sff.fa_v_100527_
Crp_protein_7_FASTA –m 8 –e 1e-05 –F F
```

The output files were filtered and counted for those that translate with 100% identity to the full length of Crp27 (35 amino acids) using the following command.

```
awk ‘{if ($2 ~ /Crp27/ && $3 == 100 && $4 >= 35) {print}}’
100527_Region8.MIDn.sff.fa_v_100527_ Crp_protein_7_FASTA | wc –l
```

BLAST analysis was repeated for each MID FASTA with 300 nucleotide filtering, and the results tabulated for 100% identity for each of the seven peptides using the following command, where $4 = amino acid length of each peptide (Crp3=35, Crp5=36, Crp20=42, Crp21=36, Crp23=35, Crp24=35, Crp27=35).

```
for mid in 1 2 3 4 5 6; do echo "Mid $mid"; awk ‘{if ($2 ~ /Crp3/ && $3 == 100
&& $4 == 35) {print}}’ 100601_Region8.MID${mid}_
over300.sff.fa_v_100601_ Crp_protein_7_FASTA | wc –l; done
```

2.2.7 In vitro activity of the human defensin DEFA1

2.2.7.1 Reagents

DEFA1 (ACYCRIPACIAGERRYGTCIYQGRLWAFCC) was synthesized by *N*-(9-fluorenyl)methoxy carbonyl (F-moc) chemistry, and folded, as previously described (122), at the Biomedical Research Centre at the University of British Columbia (UBC). LL-37 (LLGDFFRKSKEKIGKEFKRIVQRIKDFFRNLVPRTES) was synthesized by F-moc chemistry at the Nucleic Acid Protein Service Unit at UBC. The lyophilized peptides were resuspended in endotoxin-free water, and their concentrations determined by amino acid analysis. LPS was purified from *Pseudomonas aeruginosa* strain H103 as per the Darveau-Hancock method (123).

2.2.7.2 Cell culture

Human venous blood (100 ml) was collected from healthy volunteers in Vacutainer collection tubes containing sodium heparin (BD Biosciences) in accordance with UBC ethical approval and guidelines. The blood was mixed with an equal volume of complete RPMI media, and 20 ml layered over 10 ml Ficoll-Paque® Plus (Amersham Biosciences) or Lymphoprep (Cedar Lane Labs) in 50 ml Falcon Tubes. PBMCs in the buffy coat were isolated by gradient centrifugation. The tubes were centrifuged at 1450 rpm for 20 minutes in a Beckman Coulter Allegra 6 Centrifuge, and the buffy coat removed. The PBMCs were washed twice with Dulbecco's PBS (D-PBS), spinning in between at 1600 rpm for 8 minutes. The cells were resuspended in complete RPMI and 1 X 10^6 cells in 1 ml seeded into 24-well tissue culture plates (BD Falcon), and rested at 37°C with 5% CO_2 for 1-2 hours prior to stimulation. PBMCs were then treated for 24 hours with DEFA1, LL-37 or LPS at the indicated concentrations. Endotoxin-free water served as the vehicle control.

2.2.7.3 Detection of cytokines and chemokines

Following incubation with the various stimuli, the culture supernatants were removed, centrifuged at 1000 g for 10 minutes, and then stored at -20°C. The concentrations of IL8, IL10 and TNF in the supernatants were measured using enzyme linked immunosorbent assay (ELISA), as per the manufacturer's instructions (IL8: BioSource International, Camarillo, CA; TNF and IL10: eBioscience, San Diego).

2.3 Results

2.3.1 Annotation of the defensin clusters of the mouse C57BL/6 reference genome

2.3.1.1 Genomic overview of the annotated region on mouse Chromosome 8

Manual annotation was performed for the genomic region on mouse Chromosome 8:18.9-23.2 Mb, which consists of 18 finished BAC clones that span approximately 2.4 Mb of the NCBIM36 assembly (and subsequently NCBIM37) reference sequence (Figure 2.1A). To date this is the only known region associated with murine α-defensins and was therefore chosen as the initial starting point for defensin annotation in the mouse genome. This region also contains two β-defensin gene clusters, which flank the α-defensin gene cluster. Known MGI nomenclature was used to name genes only if there was a 100% cDNA match. Otherwise genes were referred to by an interim Vega database identifier such as OTTMUSG00000018259. The Vega database identifier is stable, versioned and will remain unchanged after further naming or assembly updates.

In total, annotation of this genomic region revealed the existence of 54 and 44 loci in the α-defensin and β-defensin gene clusters, respectively, which includes both predicted genes and pseudogenes (Tables 2.3 and 2.4). The entire defensin gene cluster is flanked by *Xkr5* (X Kell blood group precursor-related family, member 5) and *Ccdc70* (coiled-coil domain containing 70). This region also contains three gaps of various sizes ranging from 50 kb to 2 Mb where additional defensin genes could be located. Following alignment of the *in silico Swa*I-digested genomic sequence with the fragment sizes for the corresponding region of the optical map of Chromosome 8, the estimation of the intervening gap between the defensin clusters is approximately 0.76 Mb (Figure 2.2 and Appendix A.1), which is considerably lower the current 2 Mb size. The 'finished' C57BL/6J genome was published in 2009, which included the same gap alignment analysis for NCBIM36 with a size estimation of 0.86 Mb (124).

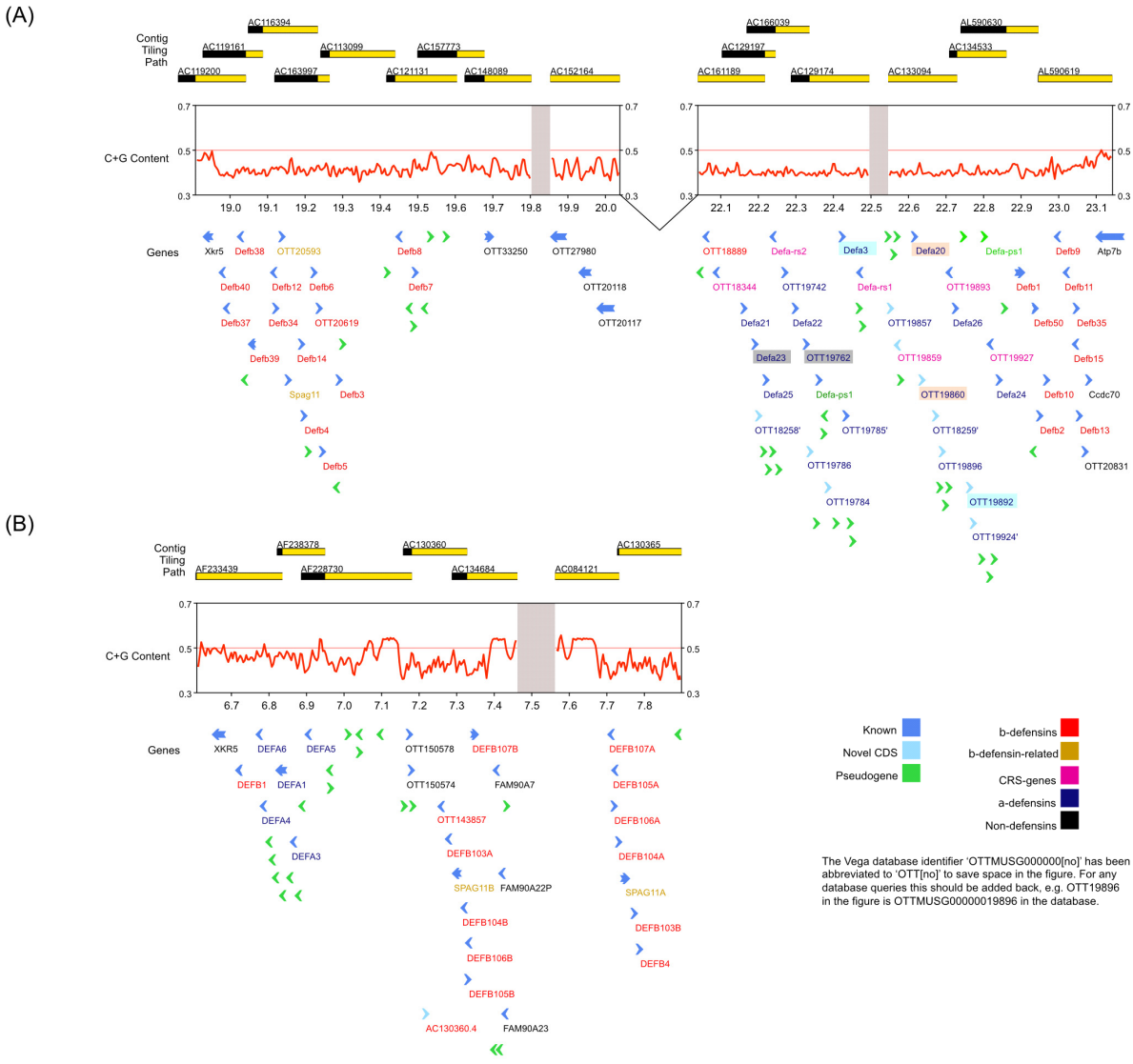


Figure 2.1. Overview of the Chromosome 8 defensin gene cluster region in mouse and human reference genomes. A clone tiling path is shown for the corresponding regions in mouse (**A**) and human (**B**). Clones are displayed in yellow but regions overlapping with adjacent clones are shown in black. Genes are indicated by blue arrows and pseudogenes by green arrows. The font colour of the gene identifier classifies the genes according to their defensin subfamily or non-defensin status (see colour legend for more details). Genes in shadowed boxes are duplicated and the colour indicates the pairs. A -'- highlights all potential *Defcr5* genes. The mouse assembly is based on NCBIM37, in which three gaps currently exist; two gaps are represented by grey bars and the biggest gap between the two clusters is joined by a 'V'. During writing of this thesis, MGNC implemented our suggestion of changing the α- and CRS-defensin gene symbol root from Defcr to Defa. This figure represents the most current version of the annotation, however all other analyses, and subsequent figures and tables generated, were performed using the Defcr root.

Table 2.3. Genes annotated within the α-defensin and β-defensin clusters on C57BL/6 mouse Chromosome 8. The Vega (v. 38) genomic coordinates of the region between *Xkr5* and *Ccdc70* are 8:18,932,729-21,221,369 bp. The order of genes in the table corresponds to their 5'-3' order within the genome. Proposed gene symbols are included, some of which have been implemented by the Mouse Genomic Nomenclature Committee (MGNC), as well as analyses concerning polymorphisms and orthology, where known or predicted. Approved gene symbols are shown in bold. "AC" gene symbols are provisional and based on their position on the clone. Nine α-defensins in rat do not have mouse orthologues; n/a, corresponding rat gene not found. Chr, Chromosome.

Organism/ Species/ Chr	Gene Symbol	Gene Name	Proposed Gene Symbol	Exon Count	Vega-Gene ID	TATA-Box	TATA-Box Genomic Coordinates	Polymorphic/ CNV/ Duplication	Human Orthologue	Rat Orthologue (potential)
Mouse C57BL/6J Chr8	**Defb40**	defensin beta 40		2	OTTMUSG00000020772	-				Defb40
Mouse C57BL/6J Chr8	**Defb37**	defensin beta 37		2	OTTMUSG00000020771	-				Defb37
Mouse C57BL/6J Chr8	**Defb38**	defensin beta 38		2	OTTMUSG00000020768	-				Defb38
Mouse C57BL/6J Chr8	**Defb39**	defensin beta 39		2	OTTMUSG00000020770	-				Defb39
Mouse C57BL/6J Chr8	**Defb12**	defensin beta 12		3	OTTMUSG00000020595	TATAAATG	19114860 - 19114865 (strand -1)		DEFB105 (MGI & ref)	Defb12
Mouse C57BL/6J Chr8	**Defb34**	defensin beta 34		2	OTTMUSG00000020594	-				n/a
Mouse C57BL/6J Chr8	AC116394.3	sperm-associated antigen 11c/h (Spag11c/h)		2 and 3	OTTMUSG00000020593					NP_0010329 41.1 (20 aa shorter-deletion)
Mouse C57BL/6J Chr8	**Spag11**	sperm associated antigen 11		2	OTTMUSG00000020599	TATAAATG	19157855 - 19157862		SPAG11B	Spag11
Mouse C57BL/6J Chr8	**Defb14**	defensin beta 14		2	OTTMUSG00000020600	-			DEFB103 (MGI & ref)	Defb14
Mouse C57BL/6J Chr8	**Defb4**	defensin beta 4		2	OTTMUSG00000020596	-			DEFB4 (ref)	Defb4
Mouse C57BL/6J Chr8	**Defb6**	defensin beta 6		2	OTTMUSG00000020598					n/a
Mouse C57BL/6J Chr8	AC163997.1	beta-defensin 53 (Defb53)	Defb#	2	OTTMUSG00000020619	-				n/a
Mouse C57BL/6J Chr8	**Defb5**	defensin beta 5		2	OTTMUSG00000020669	-				n/a
Mouse C57BL/6J Chr8	**Defb3**	defensin beta 3		2	OTTMUSG00000020668	-				Defb3
Mouse C57BL/6J Chr8	**Defb8**	defensin beta 8		2	OTTMUSG00000020716	-				n/a
Mouse C57BL/6J Chr8	**Defb7**	defensin beta 7		2	OTTMUSG00000020722	-				n/a
Mouse C57BL/6J Chr8	AC140205.1	CRS4C-6 (cryptdin-related sequence peptide)	Defa-rs#	2	OTTMUSG00000018344	TATAAATG	22084456 - 22084463 (strand -1)			n/a
Mouse C57BL/6J Chr8	AC140205.2	Beta-defensin 51 (Defb51)	Defb#	3	OTTMUSG00000018889	-				Defb51

Organism/ Species/ Chr	Gene Symbol	Gene Name	Proposed Gene Symbol	Exon Count	Vega-Gene ID	TATA-Box	TATA-Box Genomic Coordinates	Polymorphic/ CNV/ Duplication	Human Orthologue	Rat Orthologue (potential)
Mouse C57BL/6J Chr8	AC140205.3	beta-defensin 52 (Defb52)	Defb#	4	OTTMUSG00000018888	-				Defb52
Mouse C57BL/6J Chr8	AC140205.4	beta-defensin 33 (Defb33)	Defb#	3	OTTMUSG00000018925	-				Defb33
Mouse C57BL/6J Chr8	**Defcr21**	defensin related cryptdin 21	**Defa21**	2	OTTMUSG00000019489	TATAAATA	22165193 - 22165200			n/a
Mouse C57BL/6J Chr8	**Defcr23**	defensin related cryptdin 23	**Defa23**	2	OTTMUSG00000019488	TATAAATG	22194686 - 22194693			n/a
Mouse C57BL/6J Chr8	AC129197.1	novel protein similar to defensin related cryptdin 5	Defa5[suffix]	2	OTTMUSG00000018258	TATAAATG	22204671 - 22204678			n/a
Mouse C57BL/6J Chr8	**Defcr25**	defensin related cryptdin 25	**Defa25** (Defcr2 name should be removed)	2	OTTMUSG00000019700					n/a
Mouse C57BL/6J Chr8	AC129197.2	novel defensin related sequence cryptdin peptide CRS1C	Defa-rs2	2	OTTMUSG00000018260	TATAAATG	22235796 22235803 (strand -1)			n/a
Mouse C57BL/6J Chr8	AC166039.1	novel defensin related cryptdin	Defa#	2	OTTMUSG00000019742	TATAAAGG	22274268 - 22274275			n/a
Mouse C57BL/6J Chr8	**Defcr22**	defensin related cryptdin 22	**Defa22**	2	OTTMUSG00000019763	TATAAATA	22301925 - 22301932			n/a
Mouse C57BL/6J Chr8	AC166039.3	novel defensin related cryptdin **identical to Defcr23**	Defa23[suffix]	2	OTTMUSG00000019762	TATAAATG	22331219 - 22331226	Yes		n/a
Mouse C57BL/6J Chr8	AC129174.1	novel defensin related cryptdin	Defa#	2	OTTMUSG00000019786	TATAAATG	22341209 - 22341216			n/a
Mouse C57BL/6J Chr8	AC129174.4	novel defensin related cryptdin	Defa#	2	OTTMUSG00000019784	TATACATA	22389367- 22389374			n/a
Mouse C57BL/6J Chr8	**Defcr3**	defensin related cryptdin 3	**Defa3**	2	OTTMUSG00000019782	TATAAATG	22427048 - 22427055			n/a
Mouse C57BL/6J Chr8	AC129174.7	novel protein similar to defensin related cryptdin 5	**Defa5**[suffix]	2	OTTMUSG00000019785	TATAAATG	22437044 - 22437051			n/a
Mouse C57BL/6J Chr8	**Defcr-rs1** (alias CRS1C-2)	Defensin related sequence cryptdin peptide	**Defa-rs1**	2	OTTMUSG00000019792	TATAAATG	22466721 - 22466728 (strand -1)			n/a
Mouse C57BL/6J Chr8	AC133094.3	novel defensin related cryptdin	Defa#	2	OTTMUSG00000019857	-				n/a
Mouse C57BL/6J Chr8	AC133094.5	novel defensin related sequence cryptdin peptide CRS1C	Defa-rs4	2	OTTMUSG00000019859	TATAAATG	22567056 - 22567063 (strand -1)	Yes, identical to AC133094.13		n/a
Mouse C57BL/6J Chr8	**Defcr20**	defensin related cryptdin 20	**Defa20**	2	OTTMUSG00000019856	TATAAATG	22619699 - 22619706			n/a
Mouse C57BL/6J Chr8	AC133094.9	novel defensin related cryptdin **identical to Defcr20**	Defa20[suffix]	2	OTTMUSG00000019860	TATAAATG	22639473 - 22639480	Yes		n/a
Mouse C57BL/6J Chr8	AC133094.1	novel protein similar to defensin related cryptdin 5		2	OTTMUSG00000018259	TATAAATG	22675812 - 22675819	Yes		n/a
Mouse C57BL/6J Chr8	AC133094.11	novel defensin related cryptdin		2	OTTMUSG00000019896	-				n/a
Mouse C57BL/6J Chr8	AC133094.13	novel defensin related sequence cryptdin peptide CRS1C	Defa-rs5[suffix]	2	OTTMUSG00000019893	TATAAATG	22705004 - 22705011 (strand -1)	Yes, identical to AC133094.5		n/a

Organism/ Species/ Chr	Gene Symbol	Gene Name	Proposed Gene Symbol	Exon Count	Vega-Gene ID	TATA-Box	TATA-Box Genomic Coordinates	Polymorphic/ CNV/ Duplication	Human Orthologue	Rat Orthologue (potential)
Mouse C57BL/6J Chr8	**Defcr26**	defensin related cryptdin 26	**Defa26**	2	OTTMUSG00000019889	TATACATA	22728581-22728588			n/a
Mouse C57BL/6J Chr8	AC134533.2	novel defensin related cryptdin **identical to Defcr3**	Defa3[suffix]	2	OTTMUSG00000019892	TATAAATG	22766199 - 22766206	Yes		n/a
Mouse C57BL/6J Chr8	AC134533.3	novel protein similar to defensin related cryptdin 5	Defa5[suffix]	2	OTTMUSG00000019924	TATAAATG	22776195 - 22776202	Yes		n/a
Mouse C57BL/6J Chr8	AC134533.6	novel defensin related cryptdin (CRS1C-3)	Defa-rs3	2	OTTMUSG00000019927	TATAAATG	22814141 - 22814148 (strand -1)			n/a
Mouse C57BL/6J Chr8	**Defcr24**	defensin related cryptdin 24	**Defa24**	2	OTTMUSG00000019980	TATAAATG	22844935 - 22844942			Defcr24
Mouse C57BL/6J Chr8	**Defb1**	defensin beta 1		2	OTTMUSG00000019983	TATAAAAA	22887029 - 22887036		DEFB1 (MGI & ref)	Defb1
Mouse C57BL/6J Chr8	**Defb50**	defensin beta 50		2	OTTMUSG00000019981	TATAAATC	22933984 - 22933991			Defb50
Mouse C57BL/6J Chr8	**Defb2**	defensin beta 2		2	OTTMUSG00000020785	-				Defb2
Mouse C57BL/6J Chr8	**Defb10**	defensin beta 10		2	OTTMUSG00000020783	-				Defb9 53% Defb10 53%
Mouse C57BL/6J Chr8	**Defb9**	defensin beta 9		2	OTTMUSG00000020782	-				Defb11 53% questionable
Mouse C57BL/6J Chr8	**Defb11**	defensin beta 11		2	OTTMUSG00000020784	-				Defb9 50% Defb10 50% Both questionable
Mouse C57BL/6J Chr8	**Defb15**	defensin beta 15		2	OTTMUSG00000020830	TATAAAGG	23057269 - 23057262 (strand -1)		DEFB106 (MGI & ref)	Defb15 (syntenic) NP_0010326 09.1 (12 aa longer; not syntenic) Both questionable
Mouse C57BL/6J Chr8	**Defb35**	defensin beta 35		2	OTTMUSG00000020829	-				n/a
Mouse C57BL/6J Chr8	**Defb13**	defensin beta 13		2	OTTMUSG00000020827				DEFB107 (ref)	Defb13 75% questionable

Table 2.4. Pseudogenes annotated within the α-defensin and β-defensin clusters on C57BL/6 mouse Chromosome 8. The Vega (v. 38) genomic coordinates of the region between *Xkr5* and *Ccdc70* are 8:18,932,729-21,221,369 bp. The order of pseudogenes in the table corresponds to their 5'-3 order within the genome.

Organism/ Species/ Chr	Clone	Vega-Gene ID	MGI Name	Pseudogene Name	Pseudogene Type
Mouse C57BL/6J Chr8	AC140205.9	AC140205.9		defensin pseudogene similar to Defa7	Unprocessed
Mouse C57BL/6J Chr8	AC129197	OTTMUSG00000019818	Defa-ps3	defensin related cryptdin pseudogene	Unprocessed
Mouse C57BL/6J Chr8	AC166039	OTTMUSG00000019815	Defa-ps4	pseudogene similar to part of defensin related cryptdin	-
Mouse C57BL/6J Chr8	AC129174	OTTMUSG00000019783	Defa-ps5	novel defensin related cryptdin pseudogene	Unprocessed
Mouse C57BL/6J Chr8		OTTMUSG00000019780	Defa-ps6	defensin, alpha, pseudogene 1	Transcribed_unprocessed
Mouse C57BL/6J Chr8		OTTMUSG00000019817	Defa-ps7	pseudogene similar to part of defensin related cryptdin	-
Mouse C57BL/6J Chr8		OTTMUSG00000019781		novel defensin related cryptdin pseudogene	Unprocessed
Mouse C57BL/6J Chr8		OTTMUSG00000019794		novel defensin related cryptdin pseudogene	Unprocessed
Mouse C57BL/6J Chr8		OTTMUSG00000019793		novel defensin related cryptdin pseudogene	Unprocessed
Mouse C57BL/6J Chr8		OTTMUSG00000019795		pseudogene similar to part of defensin related cryptdin	Pseudogene
Mouse C57BL/6J Chr8	AC133094	OTTMUSG00000019855		defensin related cryptdin pseudogene	Unprocessed
Mouse C57BL/6J Chr8		OTTMUSG00000019858		pseudogene similar to part of defensin related cryptdin	Pseudogene
Mouse C57BL/6J Chr8		OTTMUSG00000019890		defensin alpha pseudogene	Unprocessed
Mouse C57BL/6J Chr8	AC134533	OTTMUSG00000019894		defensin alpha pseudogene	Unprocessed
Mouse C57BL/6J Chr8		OTTMUSG00000019925	Defa-ps9	defensin alpha pseudogene	Unprocessed
Mouse C57BL/6J Chr8		OTTMUSG00000019923	Defa-ps1	defensin alpha pseudogene **identical to Defa-ps1**	Transcribed_unprocessed
Mouse C57BL/6J Chr8		OTTMUSG00000019929	Defa-ps10	pseudogene similar to part of defensin related cryptdin	Pseudogene
Mouse C57BL/6J Chr8		OTTMUSG00000019982	Defa-ps11	defensin alpha pseudogene	Unprocessed
Mouse C57BL/6J Chr8	AC116394	OTTMUSG00000020597	Defa-ps12	defensin beta 46 pseudogene	Unprocessed
Mouse C57BL/6J Chr8	AC113099	OTTMUSG00000020671	Defa-ps13	defensin beta 54 pseudogene (Defb54-ps)	Unprocessed
Mouse C57BL/6J Chr8		OTTMUSG00000020719	Defa-ps14	novel defensin alpha pseudogene	Unprocessed
Mouse C57BL/6J Chr8	AC121131	OTTMUSG00000020718	Defa-ps15	novel defensin pseudogene	Unprocessed

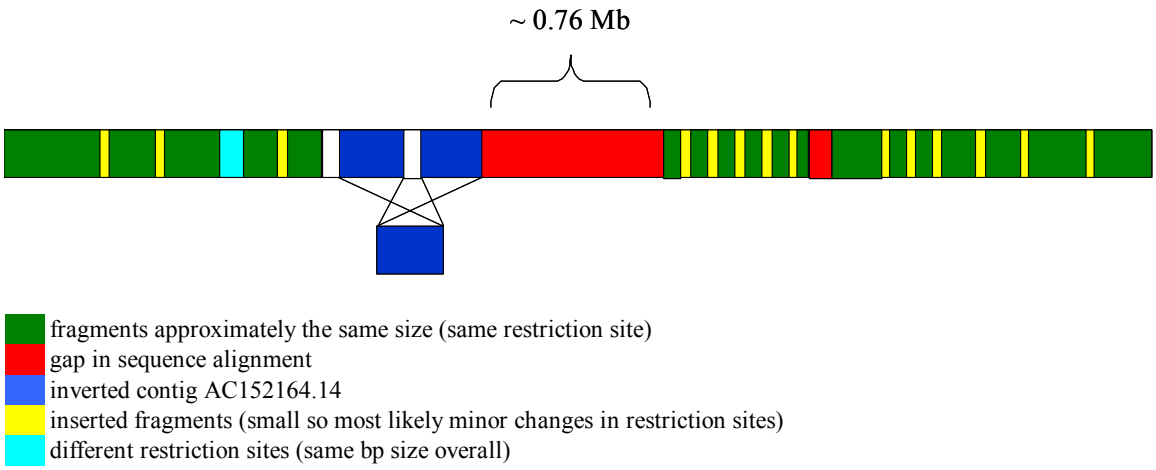


Figure 2.2. Schematic for size estimation of the gap in murine C57BL/6J Chromosome 8 located between genomic position 20 and 22 Mb. The genomic sequence for Chr8:18,508,450-24,203,501 bp was downloaded from Vega (v. 23), and then digested *in silico* at *Swa*I-predicted sites. The ordered digest fragments were aligned with *Swa*I-digested DNA fragments from the same region of the Optical Map for Chromosome 8. See Appendix A.1 for the original spreadsheet alignment.

2.3.1.2 α-Defensin gene cluster

Nineteen apparently intact α-defensin genes, eleven of which are novel, and 22 defensin-related pseudogenes were observed within the mouse α-defensin cluster (Figure 2.1A and 2.3). The α-defensin peptide alignment (Figure 2.3A) shows increased sequence variability of the mature peptide compared to the most recent publication (Figure 2.3B) (13). Additionally a conserved thymidine (T) residue has been identified at the position preceding the fourth cysteine, and a conserved glutamine (G) at the position 2-3 residues following the forth cysteine is not conserved within all murine α-defensins.

A.

```
OTTMUSG00000019784 (noveldef)      MKTLVLLSALFLLAFQVQADPIQKTDEETNTEVQPEEEEQAMSVSFGNPE 50
Defcr26 (OTTMUSG00000019889)       MKTLVLLSALFLLAFQVQADPIQNTDEETNTEVQPQEEDQAVSVSFGNPE 50
Defcr25 (OTTMUSG00000019700)       MKTLVLLSALALLAFQVQADPIQNRDEESKIDEQPGKEDQAVSVSFGDPE 50
OTTMUSG00000019896 (noveldef)      MKTLVLLSALALLAFQVQADPIQNRDEESKIDEQPGKEDQAVSVSFGDPE 50
Defcr24 (OTTMUSG00000019980)       MKTLILLSALVLLAFQVQADPIQNTDEETKTEEQPGEEDQAVSVSFGDPE 50
OTTMUSG00000019742 (noveldef)      MKTLILLSALVLLAFQVQADPIQNTDEETKTEEQPGEEDQAVSVSFGDPE 50
Defcr3 (OTTMUSG00000019782)        MKTLVLLSALVLLAFQVQADPIQNTDEETKTEEQPGEDDQAVSVSFGDPE 50
OTTMUSG00000019892 (Defcr3dupl)    MKTLVLLSALVLLAFQVQADPIQNTDEETKTEEQPGEDDQAVSVSFGDPE 50
Defcr23 (OTTMUSG00000019488)       MKTLVLLSALILLAFQVQADPIQNTDEETKTEEQPGKEDQAVSVSFGDPE 50
OTTMUSG00000019762 (Defcr23dupl)   MKTLVLLSALILLAFQVQADPIQNTDEETKTEEQPGKEDQAVSVSFGDPE 50
OTTMUSG00000019785 (novelDefcr5)   MKTFVLLSALVLLAFQAQADPIHKTDEETNTEEQPGEEDQAVSISFGGQE 50
OTTMUSG00000018258 (novelDefcr5)   MKTFVLLSALVLLAFQVQADPIHKTDEETNTEEQPGEEDQAVSISFGGQE 50
OTTMUSG00000018259 (novelDefcr5)   MKTFVLLSALVLLAYQVQADPIHKTDEETNTEEQPGEEDQAVSISFGGQE 50
OTTMUSG00000019924 (novelDefcr5)   MKTIVLLSALVLLAFQVQADPIQKTDEETNTEEQPGEEDQAVSISFGGQE 50
OTTMUSG00000019786 (noveldef)      MKTFVLLSALVLLAFQAQADPIHKTDEETNTEEQPGEEDQAVSISSGGQE 50
Defcr21 (OTTMUSG00000019489)       MKTLVLLSALILLAYQVQTDPIQNTDEETNTEEQPGEDDQAVSVSFGGQE 50
Defcr22 (OTTMUSG00000019763)       MKTLVLLSALILLAYQVQTDPIQNTDEETNTEEQPGEEDQAVSVSFGGQE 50
Defcr20 (OTTMUSG00000019856)       MKTLVLLSALVLLAFQVQADPIQNTDEETNTEEQPGEEDQAVSVSFGDPE 50
OTTMUSG00000019860 (Defcr20dupl)   MKTLVLLSALVLLAFQVQADPIQNTDEETNTEEQPGEEDQAVSVSFGDPE 50
                                   ***::***** ***:*.*:***:: ***:: : ** :::**:*:* *. *

OTTMUSG00000019784 (noveldef)      GSDLQEES---LRDLGCYCRKRGCTRRERINGTCRKGHLMYTLCCL----- 93
Defcr26 (OTTMUSG00000019889)       GSDLQEES---LRDLGCYCRKRGCTRRERINGTCRKGHLMYTLCCL----- 93
Defcr25 (OTTMUSG00000019700)       GSSLQEEC----EDLICYCRTRGCKRRERLNGTCRKGHLMYMLWCC----- 92
OTTMUSG00000019896 (noveldef)      GSSLQEESSSALRDRICYCRTS-CKKRERLNGTCRKGHLMYKLCCR----- 95
Defcr24 (OTTMUSG00000019980)       GASLQEES---LRDLVCYCRARGCKGRERMNGTCSKGHLLYMLCCR----- 93
OTTMUSG00000019742 (noveldef)      GSSLQEES---LRDLVCYCRARGCKGRERMNGTCSKGHLMYMLCCR----- 93
Defcr3 (OTTMUSG00000019782)        GSSLQEES---LRDLVCYCRKRGCKRRERMNGTCRKGHLMYTLCCR----- 93
OTTMUSG00000019892 (Defcr3dupl)    GSSLQEES---LRDLVCYCRKRGCKRRERMNGTCRKGHLMYTLCCR----- 93
Defcr23 (OTTMUSG00000019488)       GSSLQEES---LRDLVCYCRTRGCKRRERMNGTCRKGHLIYTLCCR----- 93
OTTMUSG00000019762 (Defcr23dupl)   GSSLQEES---LRDLVCYCRTRGCKRRERMNGTCRKGHLIYTLCCR----- 93
OTTMUSG00000019785 (novelDefcr5)   GSALHEEL---SKKLICYCRIRGCKRRERVFGTCRNLFLTFVFCCS----- 93
OTTMUSG00000018258 (novelDefcr5)   GSALHDEL---SKKLICYCRIRGCKRRERVFGTCRNLFLTFVFCCS----- 93
OTTMUSG00000018259 (novelDefcr5)   GSALHEEL---SKKLICYCRIRGCKRRERVFGTCRNLFLTFVFCCS----- 93
OTTMUSG00000019924 (novelDefcr5)   GSALHEEL---SKKLICYCRIRGCKRRERVFGTCRNLFLTFVFCCS----- 93
OTTMUSG00000019786 (noveldef)      GSALHEEL---SKKLICYCRIRGCKRRECVFGTCRNLFLTFVFCCS----- 93
Defcr21 (OTTMUSG00000019489)       GSALHEKL---SRDLICLCRNRRCNRGELFYGTCAGPFL---RCCRRRR-- 93
Defcr22 (OTTMUSG00000019763)       GSALHEKL---SRDLICLCRKRRCNRGELFYGTCAGPFL---RCCRRRR-- 93
Defcr20 (OTTMUSG00000019856)       GSALHEKS---SRDLICYCRKGGCNRGEQVYGTCSGRLL---FCCRRRHRH 95
OTTMUSG00000019860 (Defcr20dupl)   GSALHEKS---SRDLICYCRKGGCNRGEQVYGTCSGRLL---FCCRRRHRH 95
                                   *: *:::     ..  * **   *.  * . ***    *     *
```

B.

(i) $X_{5\text{-}10}\ C\ X\ C\ R\ X_{2\text{-}3}\ C\ X_3\ E\ X_3\ G\ T\ C\ X_{1\text{-}3}\ (G)\ X_{4\text{-}7}\ C\ C\ X_{0\text{-}6}$

(ii) $X_{1\text{-}2}\ C\ X\ C\ R\ X_{2\text{-}3}\ C\ X_3\ E\ X_3\ G\ X\ C\ X_3\ G\ X_5\ C\ C\ X_{1\text{-}4}$

Figure 2.3. Murine α-defensin peptides. A multiple protein alignment of the murine α-defensin prepropeptides (**A**). Protein sequences were deduced from the nucleotide sequences. The six canonical cysteine residues are highlighted in yellow. Mature peptides that have been isolated from C57BL/6J mice are in blue (*personal communication*, A. Ouellette and M. Shanahan). Genes identified for the first time in this study are tagged as noveldef and duplicated genes are tagged as dupl; for clarity, duplicated gene peptide products are not shown in blue. The peptide alignment shows increased sequence variability (**B**) of the murine α-defensin mature peptide (i) compared to the most recent publication (ii) (13). N-terminal variability is based on the sequences of the mature peptides isolated by Ouellette and Shanahan.

Furthermore six MYM-Type zinc finger protein pseudogenes as well as three ribosomal protein pseudogenes are also located in this region. Within the α-defensin gene cluster there is a region containing several genes very similar to *Defcr5* but no identical match to the Swiss-Prot entry P28312.2 for Defcr5, which is derived from the genomic sequence of the 129 mouse strain (Figure 2.4). Two of these loci, OTTMUSG00000019785 and OTTMUSG00000018259 show only one amino acid difference in their signal peptides compared to the Defcr5 Swiss-Prot entry P28312.2. Locus OTTMUSG00000018258 shows one amino acid difference in its pro-segment to P28312.2 and locus OTTMUSG00000019924 differs in one amino acid in the signal peptide and one in the pro-segment compared to P28312.2. These genes all have identical mature peptides compared to the P28312.2 Defcr5 sequence and have therefore been denoted as novel protein similar to defensin related cryptdin 5. Questions arise as to whether a common sequence for the mature peptide qualifies these genes to be named the same as a published sequence, whether they have the same functionality and how differences in the signal- and/or pro-segment might affect their expression. These *Defcr5* loci might be the result of chromosomal duplications or involved in copy number variation similar to a number of defensin genes where we observed 100% identity throughout the entire coding sequence (see below).

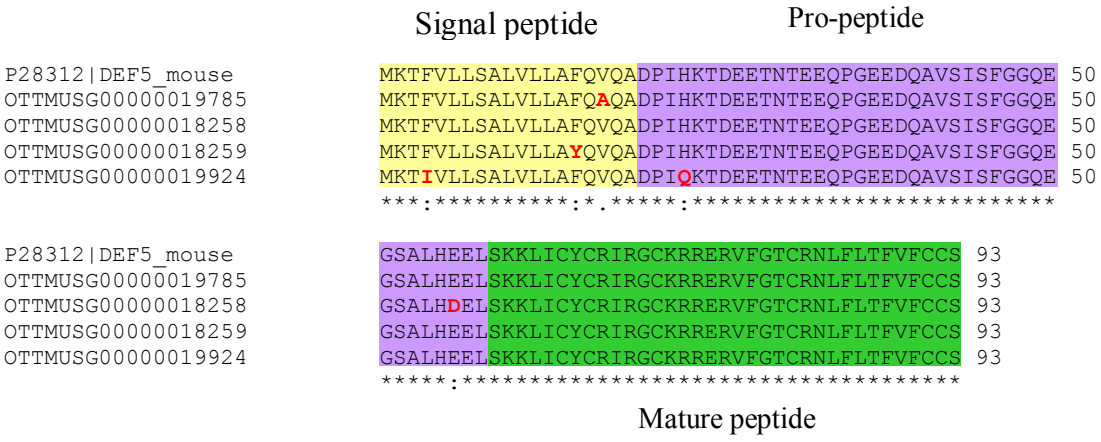


Figure 2.4. The polymorphic Defcr5 peptides. A protein alignment of the deduced prepropeptides of all potential Defcr5 copies and the Swiss-Prot sequence of Defcr5 P28312.2. Variations in amino acids are highlighted in red. The mature peptide cleavage site is based on the Swiss-Prot annotation however the preceding leucine residue was found in the mature Crp5 isolated by Ouellette and Shanahan.

Locus OTTMUSG00000019786 also has a best match to Defcr5 but there are three amino acid differences, one in the signal peptide, one in the pro-segment and another one in the mature peptide compared to P28312.2. Therefore, this locus has been annotated as a novel

defensin related cryptdin without commenting on any similarity to Defcr5, since there are clear precedents for applying different names to defensins with small sequence changes. Three genes were observed to have coding sequences with 100% identity to known genes. One example of this is represented by two copies for *Defcr23*, which are identical in their coding and 3'-UTR regions (Figure 2.5). To clarify this situation, one copy has been tagged as *Defcr23* and the other as 'novel defensin related cryptdin identical to *Defcr23*' or *Defcr23*dupl.

```
OTTMUSG00000019488      TCCTGCTCACCAATCCTCCAGGTGACTCCCAGCCATGAAGACACTAGTCCTCCTCTCTGC 60
OTTMUSG00000019762      TCCTGCTCACCAATCCTCCAGGTGACTCCCAGCCATGAAGACACTAGTCCTCCTCTCTGC 26
                        ************************************************************

OTTMUSG00000019488      CCTCATCCTGCTGGCCTTCCAGGTCCAGGCTGATCCTATCCAAAACACAGATGAAGAGAC 120
OTTMUSG00000019762      CCTCATCCTGCTGGCCTTCCAGGTCCAGGCTGATCCTATCCAAAACACAGATGAAGAGAC 86
                        ************************************************************

OTTMUSG00000019488      TAAAACTGAGGAGCAGCCAGGGAAAGAGGACCAGGCTGTGTCTGTCTCTTTTGGAGACCC 180
OTTMUSG00000019762      TAAAACTGAGGAGCAGCCAGGGAAAGAGGACCAGGCTGTGTCTGTCTCTTTTGGAGACCC 146
                        ************************************************************

OTTMUSG00000019488      AGAAGGCTCTTCTCTTCAAGAGGAATCGTTGAGAGATCTGGTATGCTATTGTAGAACAAG 240
OTTMUSG00000019762      AGAAGGCTCTTCTCTTCAAGAGGAATCGTTGAGAGATCTGGTATGCTATTGTAGAACAAG 206
                        ************************************************************

OTTMUSG00000019488      AGGCTGCAAAAGAAGAGAACGCATGAATGGGACCTGCAGAAAGGGTCATTTAATATACAC 300
OTTMUSG00000019762      AGGCTGCAAAAGAAGAGAACGCATGAATGGGACCTGCAGAAAGGGTCATTTAATATACAC 266
                        ************************************************************

OTTMUSG00000019488      GCTCTGCTGTCGCTGAACATGGAGACCACAGAGGACAAGACGAGCATGAGTACTGAGGCC 360
OTTMUSG00000019762      GCTCTGCTGTCGCTGAACATGGAGACCACAGAGGACAAGACGAGCATGAGTACTGAGGCC 326
                        ************************************************************

OTTMUSG00000019488      ACTGATGCTGGTGCCTGATGACCACTTCTCAATAAATTGTTCGCAATATGC 411
OTTMUSG00000019762      ACTGATGCTGGTGCCTGATGACCACTTCTCAATAAATTGTTCGCAATATGC 360
                        ************************************************
```

Figure 2.5. Duplicated *Defcr23* transcript sequences. These genes have been annotated as duplicated because their coding sequences, indicated in blue, are identical. However the transcripts can not be differentiated in either the capillary or 454 sequencing assays as their 3'-UTR sequences are also identical. The sequence highlighted in yellow has not been annotated as part of the 5'-UTR, however it is in the 5' upstream genomic sequence, and is included for confirmation of the duplicated nature of these two genes. *Defcr23* is OTTMUSG00000019488 and *Defcr23dupl* is OTTMUSG00000019762. There are three bases different within the entire sequence of these two genes, which are all within Intron1-2.

Two other α-defensin genes, *Defcr3* and *Defcr20*, also have apparent duplications in the mouse genome, however each of these pairs has one base pair difference in the 3'-UTR region (Figure 2.6). The duplicated genes have been denoted *Defcr3*dupl and *Defcr20*dupl because their predicted peptides would be indistinguishable from Defcr3 and Defcr20, respectively. Genes with duplicated copies are ideal candidates for copy number variation.

A.

```
OTTMUSG00000019782      TCCTGCTCACCAATCCTCCAGGTGACTCCCAGCCATGAAGACACTAGTCCTCCTCTCTGC 60
OTTMUSG00000019892      TCCTGCTCACCAATCCTCCAGGTGACTCCCAGCCATGAAGACACTAGTCCTCCTCTCTGC 47
                        ************************************************************

OTTMUSG00000019782      CCTCGTCCTGCTGGCCTTCCAGGTCCAGGCTGATCCTATCCAAAACACAGATGAAGAGAC 120
OTTMUSG00000019892      CCTCGTCCTGCTGGCCTTCCAGGTCCAGGCTGATCCTATCCAAAACACAGATGAAGAGAC 107
                        ************************************************************

OTTMUSG00000019782      TAAAACTGAGGAGCAGCCAGGGGAAGACGACCAGGCTGTGTCTGTCTCTTTTGGAGACCC 180
OTTMUSG00000019892      TAAAACTGAGGAGCAGCCAGGGGAAGACGACCAGGCTGTGTCTGTCTCTTTTGGAGACCC 167
                        ************************************************************

OTTMUSG00000019782      AGAAGGCTCTTCTCTTCAAGAGGAATCGTTGAGAGATCTGGTATGCTATTGTAGAAAAAG 240
OTTMUSG00000019892      AGAAGGCTCTTCTCTTCAAGAGGAATCGTTGAGAGATCTGGTATGCTATTGTAGAAAAAG 227
                        ************************************************************

OTTMUSG00000019782      AGGCTGCAAAAGAAGAGAACGCATGAATGGGACCTGCAGAAAGGGTCATTTAATGTACAC 300
OTTMUSG00000019892      AGGCTGCAAAAGAAGAGAACGCATGAATGGGACCTGCAGAAAGGGTCATTTAATGTACAC 287
                        ************************************************************

OTTMUSG00000019782      ACTCTGCTGTCGCTGAACATGGAGACCACAGAGGACAAGACGAACATGAGTACTGAGGCC 360
OTTMUSG00000019892      ACTCTGCTGTCGCTGAACATGGAGACCACAGAGGACAAGACGAACATGAGTACTGAGGCC 347
                        ************************************************************

OTTMUSG00000019782      ACTGATGCTGGTGCCTGATGACTACCTCGCAATAAATTGTTCGCAATATG 410
OTTMUSG00000019892      ACTGATGCTGGTGCCTGATGACCACCTCGCAATAAATTGTTCGCAATATG 397
                        ********************* **************************
```

B.

```
OTTMUSG00000019860      ACATTGGGCTCCTGCTCACCAATTCTCCAGGTGACTCACAGCCATGAAGACACTTGTCCT 60
OTTMUSG00000019856      ACATTGGGCTCCTGCTCACCAATTCTCCAGGTGACTCACAGCCATGAAGACACTTGTCCT 60
                        ************************************************************

OTTMUSG00000019860      CCTCTCTGCCCTCGTCCTGCTGGCCTTCCAGGTCCAGGCTGATCCTATCCAAAACACAGA 120
OTTMUSG00000019856      CCTCTCTGCCCTCGTCCTGCTGGCCTTCCAGGTCCAGGCTGATCCTATCCAAAACACAGA 120
                        ************************************************************

OTTMUSG00000019860      TGAGGAGACTAATACTGAGGAGCAGCCAGGGGAGGAGGACCAGGCTGTGTCTGTCTCCTT 180
OTTMUSG00000019856      TGAGGAGACTAATACTGAGGAGCAGCCAGGGGAGGAGGACCAGGCTGTGTCTGTCTCCTT 180
                        ************************************************************

OTTMUSG00000019860      TGGAGACCCAGAAGGATCTGCTCTTCATGAAAAATCGTCGAGAGATCTGATATGCTATTG 240
OTTMUSG00000019856      TGGAGACCCAGAAGGATCTGCTCTTCATGAAAAATCGTCGAGAGATCTGATATGCTATTG 240
                        ************************************************************

OTTMUSG00000019860      TAGAAAAGGAGGCTGCAATAGAGGAGAACAAGTTTATGGGACCTGCTCAGGACGACTTTT 300
OTTMUSG00000019856      TAGAAAAGGAGGCTGCAATAGAGGAGAACAAGTTTATGGGACCTGCTCAGGACGACTTTT 300
                        ************************************************************

OTTMUSG00000019860      GTTCTGCTGCCGCCGCCGCCACCGCCACTGAACATGCAGATGGCAGAGATATGACAACCA 360
OTTMUSG00000019856      GTTCTGCTGCCGCCGCCGCCACCGCCACTGAACATGCAGATGACAGAGATATGACAACCA 360
                        ****************************************** ****************

OTTMUSG00000019860      TCAGCTCTGAGGCTACTGATGCTGGGGCCTGATAACCACTTCTCAATAAATTGTTTGCAA 420
OTTMUSG00000019856      TCAGCTCTGAGGCTACTGATGCTGGGGCCTGATAACCACTTCTCAATAAATTGTTTGCAA 420
                        ************************************************************

OTTMUSG00000019860      TATGC 425
OTTMUSG00000019856      TATGC 425
                        *****
```

Figure 2.6. Duplicated *Defcr3* and *Defcr20* transcript sequences. *Defcr3*/*Defcr3dupl* (**A**) and *Defcr20*/*Defcr20dupl* (**B**) have been annotated as duplicated because their coding sequences, indicated in blue, are identical. However the transcripts can be differentiated in both the capillary and 454 sequencing assays due to one base difference, highlighted in red, within the 3'-UTR. The sequence highlighted in yellow has not been annotated as part of the 5'-UTR, however it is in the 5' upstream genomic sequence. In (**A**), *Defcr3* is OTTMUSG00000019782 and *Defcr23dupl* is OTTMUSG00000019892, and in (**B**), *Defcr20* is OTTMUSG00000019856 and *Defcr20dupl* is OTTMUSG00000019860. For each pair there are additionally two bases different within the entire gene sequence, which are all within Intron1-2.

2.3.1.3 β-Defensin clusters

The mouse β-defensin cluster contains 28 β-defensin genes, including sperm-associated antigen 11 (Spag11) and OTTMUSG00000020593 (Spag11c/h). Spag11 genes encode β-defensin-like peptides and have tissue- and species-specific alternative splicing in primate species (125). Furthermore, three β-defensin pseudogenes and nine other pseudogenes, as well as a gene coding for a novel protein with tubby-like domains have been annotated. This manual annotation confirms the current mouse β-defensin repertoire (13, 126). In the human genome, most β-defensin genes have been recently duplicated but in the mouse genome the annotation did not reveal any 100% identical β-defensin genes. This analysis is limited by the current mouse genome assembly as the most recent duplications may not be apparent. Finishing of this region may reveal duplicated β-defensin genes similar to those in the α-defensin gene set.

2.3.1.4 Cryptdin-related sequences

Within the α-defensin gene cluster genes that show similarity to the prosegment of α-defensins but have a different number and spacing pattern of cysteines within the mature peptide region were identified (21, 22, 127). Such genes are referred to as cryptdin-related sequences (*CRS*) or Defcr-related sequence (*Defcr-rs*). Six genes, three of which are novel, belonging to two groups of cryptdin-related sequences, CRS1C and CRS4C, have been annotated in the C57BL/6J mouse (Figure 2.7) (25, 127).

```
Defcr-rs1(OTTMUSG00000019792)     MKTLVLLSALVLLAFQVQADPIQNTDEETKTEEQPEEEDQAVSVSFGGTE 50
OTTMUSG00000018260(novelCRS1C)    MKTLVLLSALALLAFQVQADPIQNTDEETKTEEQPEEEDQAVSVSFGGTE 50
OTTMUSG00000019927(CRS1C-3)       MKTLVLLSALALLAFQVQADPIKNTDEETKTGEQPEEEDQAVSVSFGGTE 50
OTTMUSG00000019893(novelCRS1C)    MKTLVLLSALALLALQVQADPIQNTDEETKTQEQPGEEDQAVSVSFGGTE 50
OTTMUSG00000019859(novelCRS1C)    MKTLVLLSALALLALQVQADPIQNTDEETKTQEQPGEEDQAVSVSFGGTE 50
OTTMUSG00000018344(CRS4C-6)       MKTLVLLSALVLLAFYVQADSTQETDEETKTDDQPGEEDQGVSVSFEDPE 50
                                  **********.***: ****. ::******* :** ****.***** ..*

Defcr-rs1(OTTMUSG00000019792)     GSALQDVAQRRFPWCRKCRVCQKCQVCQKCPVCPTCPQCPKQPLCEERQN 100
OTTMUSG00000018260(novelCRS1C)    GSALQDVAQRRFLWCRKCPVCQKCQVCQKCPVCPTCPQCPKQPLCEERQN 100
OTTMUSG00000019927(CRS1C-3)       GSALQYVAQRRFPWCRKCPVCQKCQVCQKCPVCPTCPQCPKLPLCKERQN 100
OTTMUSG00000019893(novelCRS1C)    GSALQDVAQRRFPWCRKCRVCQKCEVCQKCPVCPTCPQCPKQPLCKERQN 100
OTTMUSG00000019859(novelCRS1C)    GSALQDVAQRRFPWCRKCRVCQKCEVCQKCPVCPTCPQCPKQPLCKERQN 100
OTTMUSG00000018344(CRS4C-6)       RYVLQVSGLGKPPQCPKCPVCSKCPQCPQCPQCPGCPRCN----CMTK-- 94
                                     .**  . :   * ** **.**  * :** ** **:*     *  :

Defcr-rs1(OTTMUSG00000019792)     KTAITTQAPNTQHKGC 116
OTTMUSG00000018260(novelCRS1C)    KTAITTQAPNTQHKGC 116
OTTMUSG00000019927(CRS1C-3)       KSAITTQAPNTQHKGC 116
OTTMUSG00000019893(novelCRS1C)    KTAITTQAPNTHHKGC 116
OTTMUSG00000019859(novelCRS1C)    KTAITTQAPNTHHKGC 116
OTTMUSG00000018344(CRS4C-6)       ----------------
```

Figure 2.7. CRS-defensin peptides. A multiple protein alignment of the deduced CRS-defensin prepropeptides. Cysteine residues are highlighted in yellow. Members of the CRS1C family contain eleven conserved cysteines; CRS4C-6 belongs to the CRS4C family but contains ten instead of the usual nine cysteines for this group.

Nine cysteine residues characterize CRS peptides of the CRS4C class purified from C3H/HeN murine gastrointestinal mucosa (101); CRS4C-6, annotated here OTTMUSG00000018344 belongs to the subfamily CRS4C; however, CRS4C-6, which harbors ten predicted cysteine residues has not been included in prior studies (21, 22, 127). The remaining five cryptdin-related sequences annotated here (OTTMUSG00000019792, OTTMUSG00000018260, OTTMUSG00000019927, OTTMUSG00000019893, OTTMUSG00000019859) contain eleven predicted cysteines residues. OTTMUSG00000018260 shows three amino acid differences, one in the signal sequence and two in the mature peptide, compared to Defcr-rs1 (OTTMUSG00000019792). Two identical genes have been assigned as coding for novel CRS1C peptides, OTTMUSG00000019859 and OTTMUSG00000019893, which show 100% identity to each other throughout their transcript sequences (Figure 2.8).

```
OTTMUSG00000019893      ACATTGAGCTCCTGCTCACTAATCTTCCAGGTGACTCCCAGTCATGAAGACACTTGTCCT 60
OTTMUSG00000019859      ACATTGAGCTCCTGCTCACTAATCTTCCAGGTGACTCCCAGTCATGAAGACACTTGTCCT 60
                        ************************************************************

OTTMUSG00000019893      CCTCTCTGCCCTTGCCCTGCTGGCCTTACAGGTCCAGGCTGATCCTATCCAAAACACAGA 120
OTTMUSG00000019859      CCTCTCTGCCCTTGCCCTGCTGGCCTTACAGGTCCAGGCTGATCCTATCCAAAACACAGA 120
                        ************************************************************

OTTMUSG00000019893      TGAAGAGACTAAAACTCAGGAGCAGCCAGGGGAAGAGGACCAGGCTGTTTCTGTCTCCTT 180
OTTMUSG00000019859      TGAAGAGACTAAAACTCAGGAGCAGCCAGGGGAAGAGGACCAGGCTGTTTCTGTCTCCTT 180
                        ************************************************************

OTTMUSG00000019893      TGGAGGCACAGAAGGCTCTGCTCTTCAAGATGTAGCCCAAAGAAGATTTCCGTGGTGCCG 240
OTTMUSG00000019859      TGGAGGCACAGAAGGCTCTGCTCTTCAAGATGTAGCCCAAAGAAGATTTCCGTGGTGCCG 240
                        ************************************************************

OTTMUSG00000019893      GAAGTGCCGAGTGTGCCAGAAGTGCGAAGTGTGCCAGAAGTGCCCTGTGTGCCCGACATG 300
OTTMUSG00000019859      GAAGTGCCGAGTGTGCCAGAAGTGCGAAGTGTGCCAGAAGTGCCCTGTGTGCCCGACATG 300
                        ************************************************************

OTTMUSG00000019893      CCCCCAGTGCCCAAAGCAGCCATTGTGCAAAGAAAGGCAAAATAAAACTGCTATCACCAC 360
OTTMUSG00000019859      CCCCCAGTGCCCAAAGCAGCCATTGTGCAAAGAAAGGCAAAATAAAACTGCTATCACCAC 360
                        ************************************************************

OTTMUSG00000019893      CCAAGCTCCAAATACACATCATAAAGGCTGTTGAGCTGAATGTGGAATCTGGGTTGAGAT 420
OTTMUSG00000019859      CCAAGCTCCAAATACACATCATAAAGGCTGTTGAGCTGAATGTGGAATCTGGGTTGAGAT 420
                        ************************************************************

OTTMUSG00000019893      GACCATTTGCCTTTGGTCTCCACGATCTCTTTGTGCTTAGCCTCAATTGCAATTCCTTCT 480
OTTMUSG00000019859      GACCATTTGCCTTTGGTCTCCACGATCTCTTTGTGCTTAGCCTCAATTGCAATTCCTTCT 480
                        ************************************************************

OTTMUSG00000019893      CTCATAAACTCCTTGCTGAAAAATCA 506
OTTMUSG00000019859      CTCATAAACTCCTTGCTGAAAAATCA 488
                        *************************
```

Figure 2.8. Duplicated novel *CRS1C* transcript sequences. These genes have been annotated as duplicated because their coding sequences, indicated in blue, are identical. However the transcripts can not be differentiated in either the capillary or 454 sequencing assays as their 3'-UTR sequences are also identical. The sequence highlighted in yellow has not been annotated as part of the 3'-UTR, however it is in the 3' downstream genomic sequence, and is included for confirmation of the duplicated nature of these two genes. There are three bases different within the entire sequence of these two genes, which are all within Intron1-2.

All CRS1C peptides known-to-date encode 116 amino acid proteins in contrast to α-defensins that encode proteins with 92-95 amino acids. Purification of the predicted CRS peptides annotated here from C57BL/6J mice will reveal whether processing occurs as expected when compared to other strains of mice (101). Additionally CRS4C peptides purified from C3H/HeN mice form covalent homo- and heterodimers, *in vitro* and *in vivo*, and kill commensal and pathogenic bacteria, *in vitro* (101). The function of the novel putative C57BL/6J CRS peptides, and indeed the novel putative C57BL/6J α-defensin peptides, remains to be determined.

As three novel CRS genes were annotated, and the expression of this family of defensin-like peptides has only been reported in mice, it was of interest to determine whether other species, rat in particular, have homologous peptides. The CRS family cluster separately from other mammalian α-defensins and it has been argued that the mouse CRS family and rat α-defensins have evolved separately with a common gene ancestor prior to speciation (12). Databases were searched for annotated or experimental evidence to identify novel rat gene or peptide sequences homologous to that of the CRS family, however for all of the CRS peptides, the only significant hit (based on e-value) is the defensin_propep domain, which is general and not species specific. Individual searches returned specific defensins but none aligned at the C-terminal end of the CRS peptides. Extensive BLAST searches using both genomic and peptide CRS sequences suggested rat α-defensins as potential homologues. However results of all BLAST searches, including those against unannotated databases, showed significant hits to the 5' and N-terminal region of the genomic and peptide sequences, respectively, but poor matches to 3' and C-terminal regions (Tables 2.5 and 2.6). An additional protein alignment of mouse CRS and rat α-defensin peptides confirmed this lack of homology (Figure 2.9). At this time, it appears the CRS family is unique to mouse.

Table 2.5. Pfam query results for the defensin-related CRS peptide sequences. Pfam (v. 22.0) was queried using the batch search option with default parameters, and the results returned were tabulated. Defensin_propep, defensin propeptide; VirB, VirB/Tra/Trw family (Type IV secretion system); GASA, Gibberellin regulated protein; Conotoxin, small neurotoxic peptides with disulfide bridges; MucB_RseB, regulators of the anti-sigma E protein RseD

Sequence ID	Sequence		Pfam	HMM		Alignment	Bit Score	E-value	Pfam ID
	Start	End	Accession	Start	End	Mode			
OTTMUSG00000018260	1	52	PF00879.9	1	54	ls	97.6	3.90E-26	Defensin_propep
	1	14	PF08139.3	1	15	fs	5.2	0.59	VirB
	1	81	PF02704.5	1	114	ls	-52.6	0.64	GASA
OTTMUSG00000018344	1	52	PF00879.9	1	54	ls	76.6	8.00E-20	Defensin_propep
	1	14	PF08139.3	1	15	fs	5.6	0.45	VirB
	1	81	PF02950.8	1	84	ls	-5.6	0.63	Conotoxin
OTTMUSG00000019792	1	52	PF00879.9	1	54	ls	102.8	1.10E-27	Defensin_propep
	1	81	PF02704.5	1	114	ls	-51.2	0.45	GASA
	1	14	PF08139.3	1	15	fs	5.6	0.45	VirB
	1	21	PF03888.5	1	26	fs	3.8	0.54	MucB_RseB
OTTMUSG00000019859	1	52	PF00879.9	1	54	ls	104.5	3.30E-28	Defensin_propep
	1	81	PF02704.5	1	114	ls	-48.1	0.2	GASA
	1	14	PF08139.3	1	15	fs	5.2	0.59	VirB
OTTMUSG00000019893	1	52	PF00879.9	1	54	ls	104.5	3.30E-28	Defensin_propep
	1	81	PF02704.5	1	114	ls	-48.1	0.2	GASA
	1	14	PF08139.3	1	15	fs	5.2	0.59	VirB
OTTMUSG00000019927	1	52	PF00879.9	1	54	ls	91.5	2.70E-24	Defensin_propep
	1	14	PF08139.3	1	15	fs	5.2	0.59	VirB
	1	81	PF02704.5	1	114	ls	-53	0.71	GASA

Table 2.6. Summary of the best non-mouse BLAST hits for CRS peptide sequences. All BLAST searches gave similar information regarding potential homologues in rat, which, appeared to be none. From these tests it was decided that WU-BLAST2 Nucleotide and Protein BLASTS (default parameters) using BLASTN and BLASTP algorithms, respectively, would give the most useful information with regards to full coverage of the gene and peptide sequences in the output. The best matches for both the genomic and peptide sequence for each gene are indicated.

Otter ID	Gene	Peptide	Score	% Identity	% Positive	E-Value
OTTMUSG00000018260	Defa9	n/a	1167	81	81	2.3e-84
	n/a	Defa9	90	37	45	3.5e-06
OTTMUSG00000019792	Defa9	n/a	1299	80	80	2.5e-104
	n/a	Defa9	98	39	46	5.3e-07
OTTMUSG00000019927	Defa9	n/a	1410	81	81	2.0e-104
	n/a	Defa9	109	40	50	6.3e-08
OTTMUSG00000018344	Defa6	n/a	1533	69	69	4.3e-90
	n/a	Defa9	98	41	45	5.4e-04
OTTMUSG00000019893	Defa9	n/a	1395	81	81	1.2e-106
	n/a	Defa6	151	48	54	5.8e-12
OTTMUSG00000019859	Defa9	n/a	1395	81	81	3.3e-96
	n/a	Defa6	151	48	54	5.8e-12

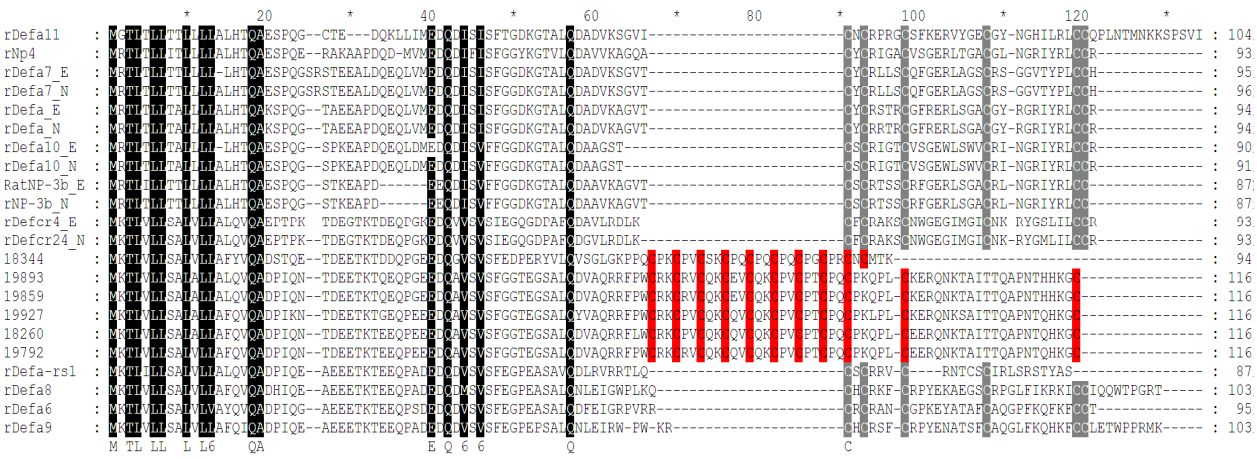


Figure 2.9. Alignment of mouse CRS-defensin and rat α-defensin peptides. The alignment shows that despite a shared homology in their signal and pro-peptide regions, mouse CRS-defensin and rat α-defensin peptides were significantly different peptides, and none seemed to be orthologous even though they most likely shared a common ancestor. Mouse sequences are denoted by the last 5 numbers of their OTTID (e.g. 18344 – OTTMUSG00000018344) and rat sequences are denoted by their gene name preceded by an r (e.g. rDefa11); E and N denote sequences in Ensembl and NCBI, respectively, when the peptides shared the same name in both databases but had different peptide sequences. Residues that showed 100% conservation across all sequences are highlighted in black. Classical cysteine residues of rat α-defensins are highlighted in grey and cysteine residues of the mouse CRS peptides are highlighted in red.

2.3.1.5 Identification of new splice variants within the mouse defensin genes

To complete the defensin gene set in the C57BL/6J mouse reference genome, all other defensin loci on Chromosomes 1, 2 and 14 were also annotated. These loci contain only β-defensin genes and as such were not subjected to in-depth analysis. Interestingly however, novel splice variants were annotated for *Defb30* and *Defb42* on Chromosome 14, which was in contrast to the family members on Chromosome 8. *Defb30* had four different splice variants, one of which was previously known; three variants were tagged as "putative coding" as they have a different first exon compared to the known variant. Two pairs of variants shared the same 5' exon but differed in their 3' exons. In each pair, one variant consisted of three exons and the other one of two (Figure 2.10). For *Defb42* two coding and two non-coding variants were identified and annotated. One of the transcripts that seemed to lack coding properties was tagged as a transcript likely to be subject to nonsense-mediated mRNA decay (NMD). All four *Defb42* variants had differentially spliced 5' first exons and only one had previously been identified in other gene sets. Tissue-specific and species-specific alternative splicing was previously implicated for primate SPAG11 (125). The β-defensin *Defb42* has been discovered and characterized in mice and its expression has been shown to be epididymis-specific (128). Looking at the origin of the manually annotated spice variants for *Defb42*, it was noticeable that all cDNA clones

representing the main coding variant were derived from the adult male reproductive tract, specifically the epididymis. However, there was one coding cDNA with an alternative 5' UTR exon compared to the main variant that was derived from the spleen of a four week old male mouse. The potential NMD splice variant is a two cells egg cDNA and another overlapping non-coding transcript is based on an 11 days embryo whole body cDNA. This observation suggests that alternative splicing for *Defb42* is likely to also be development stage specific. An unusual feature was observed for *Defb17* and *Defb41* on Chromosome 1. These genes shared the same start and first exon but differed in their second exon, which was crucial since it encoded the mature peptide. According to Vega general annotation guidelines, these two genes would normally be merged and the two transcripts would represent splice variants of the same gene since they share the first coding exon. Differential splicing seems to be a rare event for defensin genes; however, the observed examples here indicate potential functional differences for the affected transcripts.

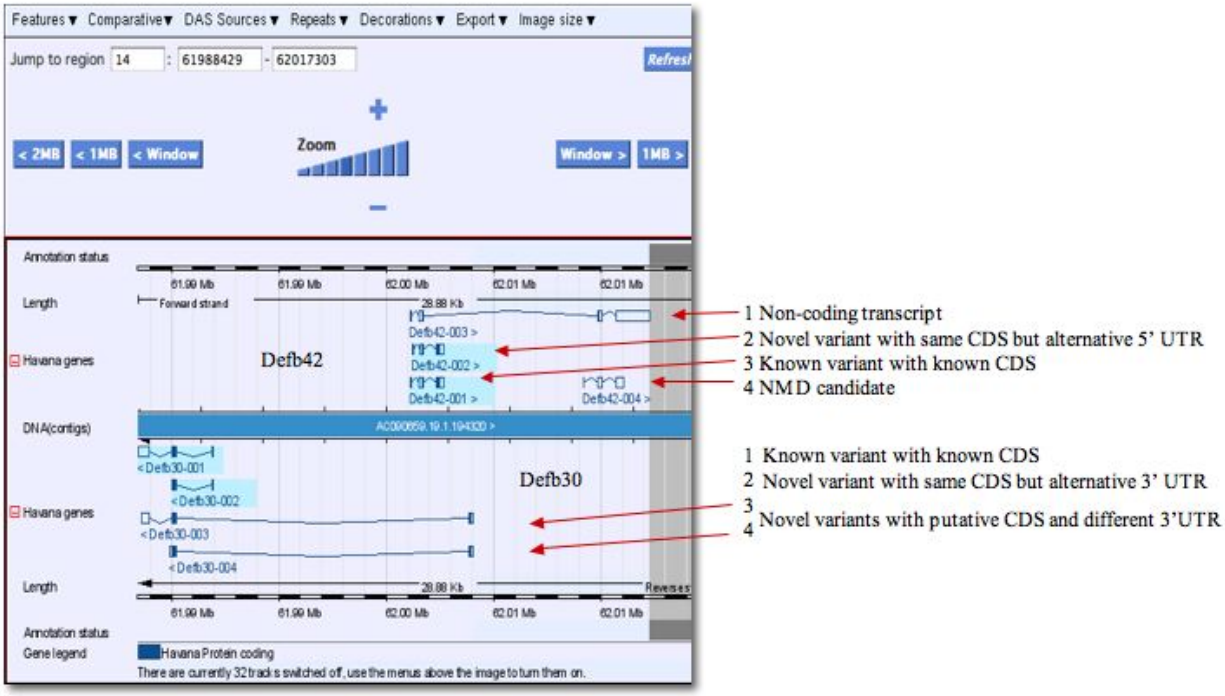


Figure 2.10. Novel coding and non-coding murine β-defensin splice variants. Vega screenshot of *Defb30* and *Defb42*, where three new variants per locus were annotated. *Defb30*: Variant 1 was a known variant with known CDS, variant 2 was a novel variant with the same CDS as variant 1 but had an alternative 3' UTR, variant 3 and 4 were novel variants with putative CDS and different 3'UTR. *Defb42*: Variant 1 represented a non-coding transcript, variant 2 was a novel variant with the same CDS as the known transcript (3) but with an alternative 5' UTR, variant 3 was a known variant with known CDS and variant 4 was a NMD candidate.

2.3.2 Genomic structures of annotated defensins on Chromosome 8

2.3.2.1 TATA boxes

Annotation of TATA boxes was based on motifs verified experimentally and published previously for five defensin genes in mouse (26, 27), and two defensin genes in human (129). The position of TATA box motifs for several more defensin genes, 28 in total including four β-defensin, Spag11, 17 α-defensin and all six CRS genes, can be found in Table 2.3. Somewhat unsurprisingly, duplicated defensin genes contained the same TATA box sequence, which suggested recent duplication events. A strong TATA box motif was identified for the majority of the α- and CRS-defensins. However for a novel α-defensin gene, OTTMUSG00000019784, and for *Defcr26* a TATA box with a weaker consensus was identified (Figure 2.11), which would be predicted to reduce the expression of these genes. TATA box-containing genes are significantly more likely to change in expression and are biased towards spontaneous mutations (130).

A.

```
OTTMUSG00000019784              CTTTTCTCTGCCATATACATATGGGCTGACTAATCACACTCCACACATTGGGCTCCTGTT 60
Defcr26(OTTMUSG00000019889)     CTTTTCTCTGCCATATACATATGTGCTGACTAATCACACTCCACACATTGGGCTCCTGCT 60
                                ********************** ********************************** *

OTTMUSG00000019784              CCCCAATCCCCCAGGTGACTCCCAGCCATG 90
Defcr26(OTTMUSG00000019889)     CCCCAATCCCCCAGGTGACTCCCAGCCATG 90
                                ******************************
```

B.

```
Defcr3(OTTMUSG00000019782)      CCTTTCTCTGTCCTATAAATGCAGGCTGGATATTCACTCTCCACACATTGGGCTCCTGCT 60
OTTMUSG00000019785              CCTTTCTCTGCCCTATAAATGCAAGTTGGCTACTCACTCTCCACACATTGGGCTCCTGCT 60
                                ********** ************ * *** ** ************************

Defcr3(OTTMUSG00000019782)      CACCAATCCTCCAGGTGACTCCCAGCCATG 90
OTTMUSG00000019785              CAACAATTCTCCAGGTGACCCCCAGCCATG 90
                                ** **** *********** **********
```

Figure 2.11. TATA-boxes annotated within the potential promoter region of murine defensin genes. A weak TATA box motif could be identified for two genes OTTMUSG00000019784 and *Defcr26* (**A**). A strong TATA box motif was found for 27 defensin genes; an example is shown for *Defcr3* and OTTMUSG00000019785, a novel defensin gene (**B**). TATA box motifs are shown in red/blue and start codons are underlined.

Several TATA box-less genes (*Defcr25*, OTTMUSG00000019857 and OTTMUSG00000019896) and also two pseudogenes (OTTMUSG00000019793 and OTTMUSG00000019923) had an identical 5' UTR/promoter region to that of the previously reported 'Crypi' (26, 27), which is presumed to be non-functional because of a pre-mature stop codon. The question that arose was whether these loci represented new pseudogenes or whether their expression was regulated by an alternative promoter. The gene OTTMUSG00000019857

had a divergent C-terminus/mature peptide compared to all other defensin genes (Figure 2.12). It had a coding potential for 112 amino acids but three of the consensus cysteine residues were missing. These data suggested that this gene was possibly pseudogenic in the reference genome.

```
OTTMUSG00000019857(noveldef)      MKTLVLLSALALLAFQVQADPIQNRDEESKIDEQPGKEDQAVSVSFGDPE 50
                                  GSSLQEECEDRICYCRTSCKKKRTPDWDLQKGSFNVQALLPLNMETTEDK 98
                                  TAMSTEATDAGA 112
```

Figure 2.12. Novel murine prepropeptide related to α- and CRS-defensins. A novel sequence (OTTMUSG00000019857) was annotated within the defensin gene cluster region, which shared homology with the signal and pro-region of α- and CRS-defensin peptides. OTTMUSG00000019857 contained four cysteine residues, highlighted in yellow, but lacked all the canonical cysteines in any known number and spacing pattern.

2.3.2.2 Pseudogenes

A total of 22 defensin pseudogenes were annotated in the major mouse defensin gene cluster region on mouse Chromosome 8, whereas ten pseudogenes were annotated in the corresponding human defensin cluster. Approximately 10% of the pseudogenes annotated in the Encyclopedia of DNA Elements (ENCODE) project were from genes involved in the immune response (131), thus the high frequency of defensins pseudogenes (>25%) was notable.

The Vega annotation guidelines divide pseudogenes into two categories, processed and unprocessed, each with two subcategories, transcribed or untranscribed. A locus is annotated as pseudogene when clear homology is shown to proteins but the coding sequence is disrupted, resulting in frameshifts or in-frame stop codons. A locus can also be tagged as a pseudogene when one or more parent genes that show spliced gene structure can be found elsewhere in the genome whereas the pseudogene locus is a single exon encoding for the corresponding protein.

Pseudogenes of the defensin gene cluster are unprocessed as they have likely evolved through duplication of functional genes and have accumulated mutations over time and become non-functional. Of the 22 pseudogenes annotated in this region five were only partial and one, *Defa-ps1* was tagged as a transcribed_unprocessed pseudogene. Here, protein homologies suggested this locus is a pseudogene, but overlapping locus-specific transcription evidence (cDNAs) indicated expression. There is recent evidence that regulatory interdependency exists between transcribed pseudogenes and their parent gene. For example a targeted knockdown of the transcribed ABC transported pseudogene *ABCC6P1* results in a significant reduction of the parent gene *ABCC6* expression levels (132).

Annotation of TATA box-like motifs was also performed for the defensin related pseudogenes. Two defensin pseudogenes contained a strong TATA box motif (TATAAA TG),

and four defensin pseudogenes contained a weak TATA box motif (TATACA TA/G), but for the majority nothing similar could be identified. Looking at the transcribed pseudogene, *Defa-ps1*, the homology broke down 48 bp upstream of the start codon and no TATA box-like motif could be identified. Generally, the TATA box-lacking defensin pseudogenes had 5' upstream sequences similar to the potential promoter regions of TATA box-lacking active defensin genes.

2.3.3 Comparative analysis of murine defensin gene sets

To illustrate the difficulties created in naming the defensins, by comparison to the assembled data from databases and the literature, four major gene sets were cross-referenced, and gene symbols and gene IDs, for the mouse α-defensin genes annotated herein, assembled (Table 2.7). These included the manually curated data set from Vega (v.30), automatic gene annotation from Ensembl (v.49), cDNA evidence from NCBI RefSeq, and a merged gene set encompassing UCSC, RefSeq and Ensembl from MGI (4.01) (133). The gene *Defcr25* had a cross-reference to gene *Defcr2* in NCBI's Entrez Gene indicating that this gene was also known as *Defcr2*. However, the protein sequences for Defcr25 (MGI:3630385; Swiss-Prot:Q5G864.1) and Defcr2 (MGI:94882; Swiss-Prot:P28309.2) were different and were derived from different mouse strains. Therefore, this locus was annotated as *Defcr25*, since the sequence in the reference mouse genome was identical to this gene. Another example of ambiguities between databases are *Defcr16* and *Defcr17*, which have been associated by MGI with OTTMUSG00000019742 and OTTMUSG00000019892 respectively; in case of *Defcr16* evidence for its expression was derived from C3H/HeJ strain only while the association of *Defcr17* with the Vega model is incorrect as OTTMUSG00000019892 was a duplicate of *Defcr3*. Additional examples are listed in Table 2.7.

Table 2.7. Genome browser comparison of mouse α-defensin genes. See Sections 2.2.4 and 2.3.3 in the text for the complete description of the comparison.

		Ensembl Annotation				MGI Annotation				NCBI Annotation			
		Gene Linked to Vega		Additional Gene(s)		Gene Linked to Vega		Additional Gene(s)		Gene Linked to Ensembl		Additional Gene(s)	
Vega Gene Symbol	Vega OTTER Gene ID	Gene Symbol	Gene ID	Gene Symbol	Gene ID	Gene Symbol	Gene ID	Gene Symbol	Gene ID	Gene Symbol	Gene ID	Gene Symbol	Gene ID
AC140205.1	OTTMUSG00000018344	AY761185	ENSMUSG00000079120	-	-	AY761185	MGI:3630303	-	-	AY761185	503556	-	-
Defcr21	OTTMUSG00000019489	Defcr21	ENSMUSG00000074447	-	-	-	-	Defcr21	MGI:1913548	Defcr21	66298	-	-
Defcr23	OTTMUSG00000019488	Defcr23	ENSMUSG00000074446	Defcr23 Defcr-rs7	ENSMUSG00000074442 OTTMUSG00000019762	Defcr23	MGI:3630381	-	-	Defcr23	497114	-	-
AC129197.1	OTTMUSG00000018258	Defcr5	ENSMUSG00000061845	-	-	-	-	OTTMUSG00000018258	MGI:3711900	OTTMUSG00000018258	100041688	Defcr5	13239
Defcr25	OTTMUSG00000019700	Defcr25	ENSMUSG00000071164	Defa-ps1 AC129174.9 AC134533.5	OTTMUSG00000019780 OTTMUSG00000019793 OTTMUSG00000019923	-	-	Defcr25	MGI:3630385	Defcr25 (aka Defcr2)	13236	-	-
AC129197.2	OTTMUSG00000018260	Defcr-rs1	ENSMUSG00000061958	-	-	-	-	OTTMUSG00000018260	MGI:3709605	OTTMUSG00000018260	634825	Defcr-rs1	13218
AC166039.1	OTTMUSG00000019742	Defcr16	ENSMUSG00000074444	Defcr24	ENSMUSG00000064213	Defcr16	MGI:99585	-	-	Defcr16 LOC100038927	13235 100038927	-	-
Defcr22	OTTMUSG00000019763	Defcr22	ENSMUSG00000074443	-	-	Defcr22	MGI:3639039	-	-	Defcr22	382059	-	-
AC166039.3	OTTMUSG00000019762	Defcr-rs7	OTTMUSG00000019762	Defcr23	ENSMUSG00000074442	Defcr-rs7	MGI:102509	-	-	Defcr-rs7 (aka CRS4C-2, CRS4C2b, CRS4C2c, CRS4C2d, CRS4C3a, CRS4C3b, CRS4C3c, CRS4C3d, CRS4C3e, CRS4C3a2)	13226	-	-
AC129174.1	OTTMUSG00000019786	novel	ENSMUSG00000079116	-	-	OTTMUSG00000019786	MGI:3705230	-	-	OTTMUSG00000019786	100041759	-	-
AC129174.4	OTTMUSG00000019784	novel	ENSMUSG00000074441	-	-	OTTMUSG00000019784	MGI:3708769	-	-	OTTMUSG00000019784	100041787	-	-
Defcr3	OTTMUSG00000019782	Defcr3	ENSMUSG00000074440	Defcr3	ENSMUSG00000060208	Defcr3	MGI:94883	-	-	Defcr3	13237	-	-
AC129174.7	OTTMUSG00000019785	NP_031877.2	ENSMUSG00000074439	AC134533.3	OTTMUSG00000019924	Defcr5	MGI:99583	-	-	Defcr5	13239	-	-
AC129174.10	OTTMUSG00000019792	Defcr-rs1	ENSMUSG00000074437	-	-	Defcr-rs1	MGI:94881	-	-	Defcr-rs1	13218	-	-
AC133094.3	OTTMUSG00000019857	novel	ENSMUSG00000074436	-	-	OTTMUSG00000019857	MGI:3642785	-	-	-	-	-	-
AC133094.5	OTTMUSG00000019859	novel	ENSMUSG00000079114	novel	ENSMUSG00000079113	EG665927	MGI:3645033	-	-	EG665927	665927	-	-
Defcr20	OTTMUSG00000019856	Defcr20	ENSMUSG00000065958	Defcr20	ENSMUSG00000065957	Defcr20	MGI:1915259	-	-	Defcr20 (aka cryptdin 4)	68009	-	-
AC133094.9	OTTMUSG00000019860	Defcr20	ENSMUSG00000065957	-	-	OTTMUSG00000019860	MGI:3709042	-	-	OTTMUSG00000019860	100041890	Defcr20 (aka cryptdin 4)	68009
AC133094.1	OTTMUSG00000018259	Defcr5	ENSMUSG00000065956	Defcr5	ENSMUSG00000061845	OTTMUSG00000018259	MGI:3705236	-	-	OTTMUSG00000018259	100041895	Defcr5	13239
AC133094.11	OTTMUSG00000019896	novel	ENSMUSG00000074434	-	-	EG626682	MGI:3646688	-	-	EG626682	626682	-	-
AC133094.13	OTTMUSG00000019893	novel	ENSMUSG00000079113	-	-	EG665956	MGI:3648003	-	-	EG665956	665956	-	-
Defcr26	OTTMUSG00000019889	Defcr26	ENSMUSG00000060070	-	-	Defcr26	MGI:3630390	-	-	Defcr26	626708	-	-
AC134533.2	OTTMUSG00000019892	Defcr3	ENSMUSG00000060208	Defcr17	OTTMUSG00000019892	Defcr17 ENMUSG00000060208	MGI:1345152	-	-	Defcr17	23855	Defcr3	13237
AC134533.3	OTTMUSG00000019924	novel	ENSMUSG00000063206	NP_031877.2	ENSMUSG00000074439	OTTMUSG00000019924	MGI:3709048	-	-	OTTMUSG00000019924	100041952	-	-
AC134533.6	OTTMUSG00000019927	AY761184	ENSMUSG00000058618	-	-	AY761184	MGI:3611585	-	-	AY761184	382000	-	-
Defcr24	OTTMUSG00000019980	Defcr24	ENSMUSG00000064213	-	-	Defcr24	MGI:3630383	-	-	Defcr24	503491	-	-

This analysis has highlighted the necessity in accounting for strain differences when deriving gene annotation based on cDNAs aligned to the C57BL/6J reference genome. The differences become very obvious when looking at the α-defensin region in MGI Genome Browser. The number of α-defensin genes there is higher than the number annotated here, but looking at the origin of many of the genes revealed that they are derived from strains distinct from the reference genome. An example is the Crp4 peptide, which was first isolated from Outbred Swiss mice (134) and corresponds to Defcr4 in MGI; this gene has been annotated in the 129X1/SvJ strain but has not been annotated in the reference strain (Table 2.8). The reference strain contained two presumed Crp4 peptide variants termed Crp4-B6a and Crp4-B6b because they were all missing three codons between the forth and fifth cysteine residues (135). MGNC has named these variants Defcr20 and Defcr21, respectively, however the relationship of these peptides between the two mouse strains is not obvious.

Table 2.8. Defensin genes currently 'missing' from the mouse reference genome. α- and CRS-defensin genes that did not map to the C57BL/6J reference genome, mostly likely because of mouse strain differences or as a result of the incomplete genome sequence within the defensin cluster of C57BL/6J. The strain origin of genomic, transcript and peptide sequences are indicated, where applicable. n/a, not applicable; n/s, not specified.

Gene Symbol	Genomic Strain	Transcript Strain	Peptide Strain	Cross-reference Validation	Entrez Gene
Defa1	129X1/SvJ	Swiss albino	n/a	missing in Vega and no 100% BLAST match	no annotation info
Defcr2	129	C3H/HeJ	n/a	missing in Vega and no 100% BLAST match	no annotation info
Defcr4	129X1/SvJ	129/SvJ	n/a	missing in Vega and no 100% BLAST match	not annotated
Defcr6	129X1/SvJ	129	n/a	missing in Vega and no 100% BLAST match	not annotated
Defcr7	n/a	C3H/HeJ	n/a	missing in Vega and no 100% BLAST match	not annotated
Defcr8	n/a	C3H/HeJ	n/a	missing in Vega and no 100% BLAST match	not annotated
Defcr9	n/a	C3H/HeJ	n/a	missing in Vega and no 100% BLAST match	not annotated
Defcr10	n/a	C3H/HeJ	n/a	missing in Vega and no 100% BLAST match	not annotated
Defcr11	n/a	C3H/HeJ	n/a	missing in Vega and no 100% BLAST match	not annotated
Defcr12	n/a	C3H/HeJ	n/a	missing in Vega and no 100% BLAST match	not annotated
Defcr13	n/a	C3H/HeJ	n/a	missing in Vega and no 100% BLAST match	not annotated
Defcr14	n/a	C3H/HeJ	n/a	missing in Vega and no 100% BLAST match	not annotated
Defcr15	n/a	C3H/HeJ	n/a	missing in Vega and no 100% BLAST match	not annotated
Defcr-rs2	n/a	Swiss albino	n/a	missing in Vega and no 100% BLAST match	no annotation info
Defcr-rs4	129X1/SvJ	C3H/HeN	C3H/HeN	missing in Vega and no 100% BLAST match	not annotated
Defcr-rs5	n/a	129X1/SvJ	n/a	missing in Vega and no sequence entry	not annotated
Defcr-rs6	n/a	129X1/SvJ	n/a	missing in Vega and no sequence entry	not annotated
Defcr-rs8	n/a	129X1/SvJ	n/a	missing in Vega and no sequence entry	not annotated
Defcr-rs9	n/a	129X1/SvJ	n/a	missing in Vega and no sequence entry	not annotated
Defcr-rs10	129	n/s	n/a	missing in Vega and no 100% BLAST match	not annotated
Defcr-rs11	n/a	129X1/SvJ	n/a	missing in Vega and no sequence entry	not annotated
Defcr-rs12	129	129	n/a	missing in Vega and no 100% BLAST match	no annotation info

Data displayed in the MGI Genome Browser is a combination of downloaded information from the UCSC's Genome browser (136) and that generated at MGI. MGI and UCSC do not filter any strain-specific data, and mapping of defensin genes has not been carried out stringently with the aim of displaying a comprehensive set of all existing genes. This is a valuable resource;

however, as a result several genes which shared the same nucleotide sequence coding for the signal peptide and the pro-segment have been mapped together. However, as the region encoding the mature peptide showed some differences, these genes can not be considered the same.

To determine the ease of cross-species comparison, genomic alignments and putative orthologues were searched for in both human and rat genomes compared to mouse α-defensins. There are only six defensin genes with defined orthologues between human and mouse found on the Mouse Chromosome 8 Linkage Map (137). It was appreciated that for the mouse α-defensin family, orthology was especially hard to predict because of the high intraspecies similarity for these genes. The same human and rat genes aligned with most of the mouse genes and/or were predicted to be orthologues for the mouse peptides. In particular, DEFA7P was predicted to be the human orthologue for the majority of the mouse cryptdins by Ensembl but this gene lacked a start codon (12) and was therefore designated a pseudogene by manual annotation. These alignments and orthologues were predicted by Ensembl and may be an artefact of Ensembl's naming scheme or of the high similarity between members of this gene family.

2.3.4 Conserved synteny within defensin clusters

Interspecies comparison is important not only from a biological perspective, but also to ascertain whether the human and/or rat genome assemblies can be used to facilitate closing of the estimated 0.76 Mb gap of mouse Chromosome 8, intervening the two α-defensin loci. Ensembl (v. 49) was used to display regions of conserved synteny between the mouse, human and rat α-defensin regions. The conserved syntenic region of the human genome is Chromosome 8p23.1. Figure 2.2B shows the arrangement of the human defensin genes within this region (8: 6.6 – 7.9 Mb), which includes one gap of approximately 100 kb in size. This region is also flanked at the 5' end by *XKR5*, the homologue of the mouse *Xkr5*. However, whereas in the mouse *Ccdc70*, *Atp7b* and *Alg11* were found 3' to the defensin gene cluster on Chromosome 8, *CCDC70* did not flank the 3' end in human, but rather is located on human Chromosome 13q14.3 where the two telomeric gene homologues, *ATP7B* and *ALG11*, have also been mapped. There are no defensin genes in this region of human Chromosome 13, which indicates the breakpoint occurred telomeric to the defensin cluster, and a survey of this region shows a complete assembly.

The quality of the human genome assembly near the defensin regions appears to be better than that of mouse. This may be due to a larger number of defensin genes within the mouse genome as compared to the human genome, as well as the high similarity between the mouse α-

defensin genes in particular. To investigate this further, we used Ensembl to analyze the conserved syntenic regions of the rat genome compared to the mouse (Chr16: 73.7 – 75.8 Mb). Nine rat α-defensin genes and one defensin-related gene were located on Chromosome 16q12.5. There were a few gaps in the region adjacent the defensin genes; one in particular of about 120 kb in the middle of the α- and β-defensin cluster, and another 5' to the defensin region about 200 kb in size. The assembly of the rat reference sequence appears to be more similar to that of the human sequence, compared to the mouse sequence, with respect to the level of completion of defensin-rich regions. However it is important to note that the rat genome sequence is a draft sequence which differs with respect to the finished sequence of both human and mouse genomes (138). The approach for the sequencing of the rat genome was a combination whole genome shotgun (WGS) and BAC sequencing, and the authors argue that this approach has generated sequence with quality near to that of finished sequence (138). Another caveat is that the human and mouse genomes aided the assembly of the rat genome in difficult regions (138), therefore any errors in either the human and mouse assemblies within the defensin regions could translate into errors in the rat genome. A coordinated effort has been undertaken for the generation of a new rat genome build and the anticipated release is imminent (139). Re-examination of the defensin clusters will determine whether our analyses and observations hold true for these regions, in particular.

Similar to the annotation of mouse α-defensins only on Chromosome 8, human and rat α-defensins have only been identified on Chromosome 8 and 16, respectively. This is in contrast to the presence of human, rat and mouse β-defensins on multiple chromosomes (human – Chromosomes 6p21, 8p23.1, 20q11.1 and 20p13, rat – 3q41, 9q13, 15p12 and 16q12.5 and mouse – 1A4, 2H1, 8A3 and 14D1). The assembly of human Chromosomes 6p21 and 20p13 are complete with no gaps, but there is a gap 5' to the β-defensin cluster on 20q11.1; however this gap is also near to the centromere, which was not targeted by the genome projects due to the difficulty in sequencing highly repetitive α-satellite DNA in heterchromatic regions (140, 141). Additional rat β-defensins are located on Chromosomes 3q41, 9q13, 15p12 and 16q12.5; these regions appear complete, with the exception of 3q41, which has an 11 kb gap within the β-defensin cluster and 16q12.5, which contains both α- and β-defensins genes, as previously described.

The assemblies of the mouse, rat and human genomes are more complete near regions of β-defensins compared to α-defensins since β-defensins are not as genetically similar as α-

defensins. β-defensins have had more time for movement associated with chromosomal rearrangements and multiple duplication events as compared to α-defensins; however mouse α-defensins have undergone a rapid expansion that has not occurred to the same extent in human and rat. Rapidly changing regions are interesting in evolutionary terms but are difficult to assemble into finished sequence (138), and additional defensin genes may be present within gaps in the assembly. These factors reinforce the biological importance and need for further characterization.

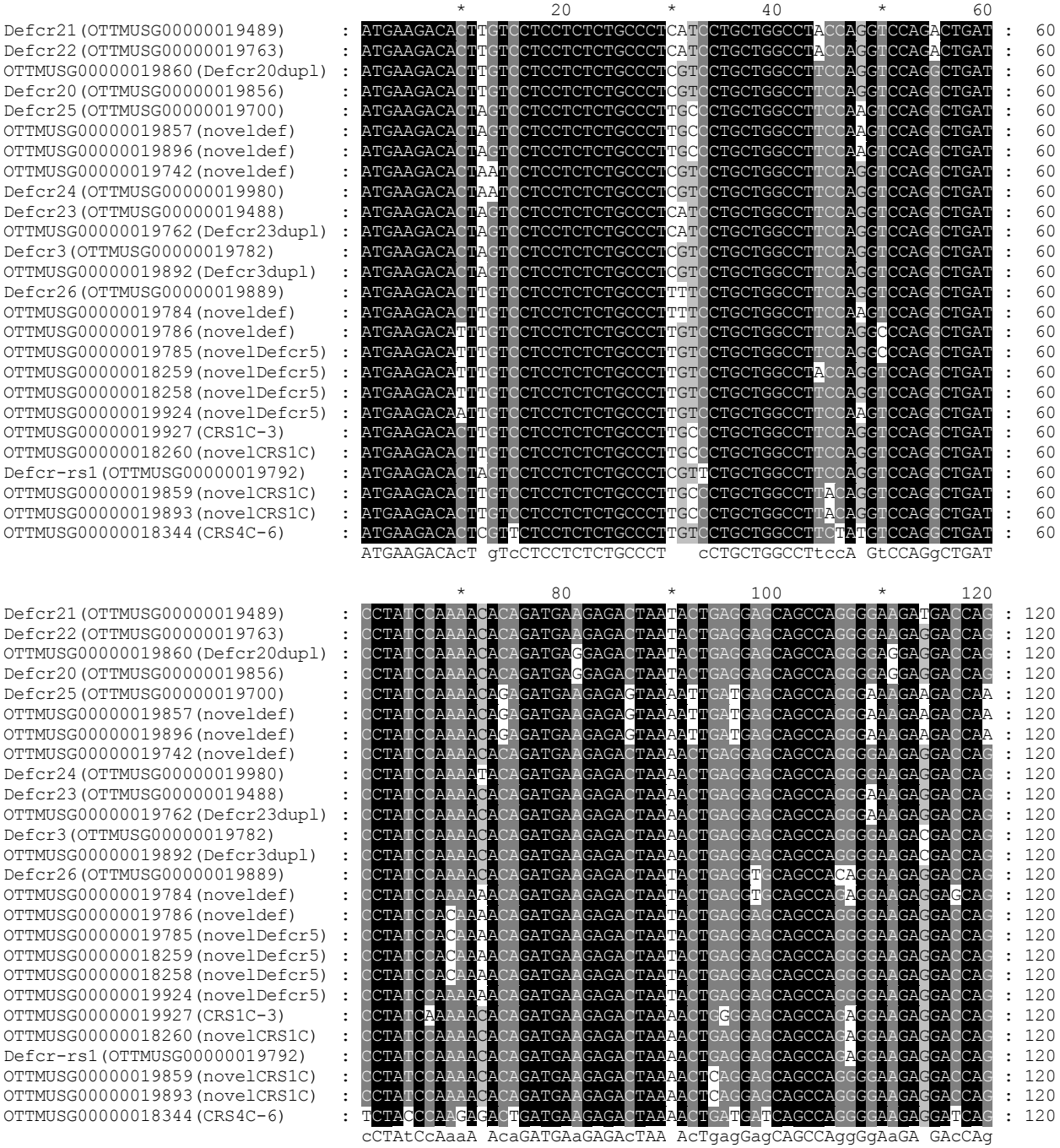
2.3.5 Defensin transcriptional expression profiling

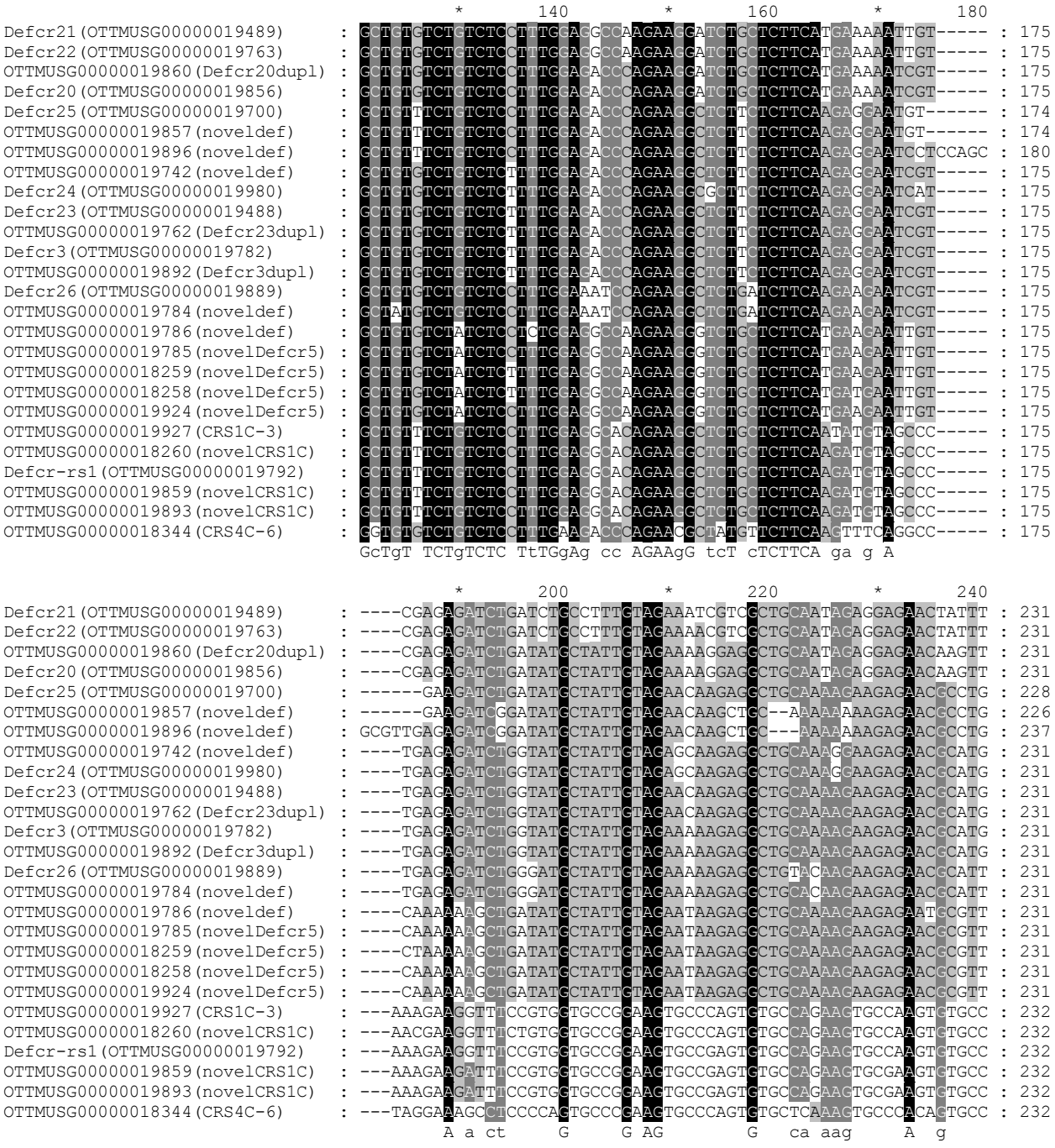
There are advantages and disadvantages of both capillary and 454 sequencing. Longer sequence reads are generated by capillary sequencing however 454 sequencing is free from clonal biases and generates greater coverage depth. Therefore both technologies were used in parallel to determine α-defensin expression in C57BL/6J mice.

2.3.5.1 Genetic analysis of α- and CRS-defensins necessitates universal primer design

The nucleotide alignment of all α- and CRS-defensin coding sequences (Figure 2.13) was used to design primers with which to investigate gene expression following the genomic annotation. The alignment showed high conservation at the 5' end of the coding sequences between all defensins, which ends approximately 160 nucleotides following the start codon. The 5'-UTR region of the 26 defensins was also inspected for primer design in order to encompass full transcript length in the amplification, but the conservation was lower than that of the start of the coding sequence. Additionally, there was variability in the length and composition of the 5'-UTR sequence when present (ten genes did not have the 5'-UTR annotated) so this region was not considered for design of the forward primer. There was also nucleotide conservation at the 3' end, however it was found only for defensin subgroups and did not occur throughout all defensins. The 3' ends of the genes were also quite repetitive, with respect to nucleotide arrangement (e.g. large stretches of Ts), so this would also hinder primer design. The variability of the 3' region would be lost when designing optimal primers by moving further in the 5' direction and this region was needed for differentiation between expressed defensin genes. Ideally the primers needed to be designed outside of this region to facilitate sequencing of the mature peptide coding region.

Interestingly the nucleotide alignment of the α- and CRS-defensin intronic regions also showed a high degree of similarity (Appendices A.2 and A.3), which supported their recent duplication.




* 140 * 160 * 180
Defcr21(OTTMUSG00000019489) : GCTGTGTCTCTCTCCTTTGGAGGCCAAGAAGGATCTGCTCTTCATGAAAATTGT----- : 175
Defcr22(OTTMUSG00000019763) : GCTGTGTCTCTCTCCTTTGGAGGCCAAGAAGGATCTGCTCTTCATGAAAATTGT----- : 175
OTTMUSG00000019860(Defcr20dupl) : GCTGTGTCTCTCTCCTTTGGAGACCCAGAAGGATCTGCTCTTCATGAAAATCGT----- : 175
Defcr20(OTTMUSG00000019856) : GCTGTGTCTCTCTCCTTTGGAGACCCAGAAGGATCTGCTCTTCATGAAAATCGT----- : 175
Defcr25(OTTMUSG00000019700) : GCTGTTTCTCTCTCCTTTGGAGACCCAGAAGGCTCTTCTCTTCAAGAGGAATGT------ : 174
OTTMUSG00000019857(noveldef) : GCTGTTTCTCTCTCCTTTGGAGACCCAGAAGGCTCTTCTCTTCAAGAGGAATGT------ : 174
OTTMUSG00000019896(noveldef) : GCTGTTTCTCTCTCCTTTGGAGACCCAGAAGGCTCTTCTCTTCAAGAGGAATCCTCCAGC : 180
OTTMUSG00000019742(noveldef) : GCTGTGTCTCTCTCTTTTGGAGACCCAGAAGGCTCTTCTCTTCAAGAGGAATCGT----- : 175
Defcr24(OTTMUSG00000019980) : GCTGTGTCTCTCTCTTTTGGAGACCCAGAAGGCGCTTCTCTTCAAGAGGAATCAT----- : 175
Defcr23(OTTMUSG00000019488) : GCTGTGTCTCTCTCTTTTGGAGACCCAGAAGGCTCTTCTCTTCAAGAGGAATCGT----- : 175
OTTMUSG00000019762(Defcr23dupl) : GCTGTGTCTCTCTCTTTTGGAGACCCAGAAGGCTCTTCTCTTCAAGAGGAATCGT----- : 175
Defcr3(OTTMUSG00000019782) : GCTGTGTCTCTCTCTTTTGGAGACCCAGAAGGCTCTTCTCTTCAAGAGGAATCGT----- : 175
OTTMUSG00000019892(Defcr3dupl) : GCTGTGTCTCTCTCTTTTGGAGACCCAGAAGGCTCTTCTCTTCAAGAGGAATCGT----- : 175
Defcr26(OTTMUSG00000019889) : GCTGTGTCTCTCTCCTTTGGAAATCCAGAAGGCTCTGATCTTCAAGAAGAATCGT----- : 175
OTTMUSG00000019784(noveldef) : GCTATGTCTCTCTCCTTTGGAAATCCAGAAGGCTCTGATCTTCAAGAAGAATCGT----- : 175
OTTMUSG00000019786(noveldef) : GCTGTGTCTATCTCCTCTGGAGGCCAAGAAGGGTCTGCTCTTCATGAAGAATTGT----- : 175
OTTMUSG00000019785(novelDefcr5) : GCTGTGTCTATCTCCTTTGGAGGCCAAGAAGGGTCTGCTCTTCATGAAGAATTGT----- : 175
OTTMUSG00000018259(novelDefcr5) : GCTGTGTCTATCTCTTTTGGAGGCCAAGAAGGGTCTGCTCTTCATGAAGAATTGT----- : 175
OTTMUSG00000018258(novelDefcr5) : GCTGTGTCTATCTCTTTTGGAGGCCAAGAAGGGTCTGCTCTTCATGATGAATTGT----- : 175
OTTMUSG00000019924(novelDefcr5) : GCTGTGTCTATCTCCTTTGGAGGCCAAGAAGGGTCTGCTCTTCATGAAGAATTGT----- : 175
OTTMUSG00000019927(CRS1C-3) : GCTGTTTCTCTCTCCTTTGGAGGCACAGAAGGCTCTGCTCTTCAATATGTAGCCC----- : 175
OTTMUSG00000018260(novelCRS1C) : GCTGTTTCTCTCTCCTTTGGAGGCACAGAAGGCTCTGCTCTTCAAGATGTAGCCC----- : 175
Defcr-rs1(OTTMUSG00000019792) : GCTGTTTCTCTCTCCTTTGGAGGCACAGAAGGCTCTGCTCTTCAAGATGTAGCCC----- : 175
OTTMUSG00000019859(novelCRS1C) : GCTGTTTCTCTCTCCTTTGGAGGCACAGAAGGCTCTGCTCTTCAAGATGTAGCCC----- : 175
OTTMUSG00000019893(novelCRS1C) : GCTGTTTCTCTCTCCTTTGGAGGCACAGAAGGCTCTGCTCTTCAAGATGTAGCCC----- : 175
OTTMUSG00000018344(CRS4C-6) : GGTGTGTCTCTCTCCTTTGAAGACCCAGAACGCTATGTTCTTCAACTTTCAGGCC----- : 175
GcTgT TCTgTCTC TtTGgAg cc AGAAgG tcT cTCTTCA ga g A
* 200 * 220 * 240
Defcr21(OTTMUSG00000019489) : ----CGAGAGATCTGATCTGCCTTTGTAGAAATCGTCGCTGCAATACAGGAGAACTATTT : 231
Defcr22(OTTMUSG00000019763) : ----CGAGAGATCTGATCTGCCTTTGTAGAAAACGTCGCTGCAATACAGGAGAACTATTT : 231
OTTMUSG00000019860(Defcr20dupl) : ----CGAGAGATCTGATATGCTATTGTAGAAAAGGAGGCTGCAATACAGGAGAACAAGTT : 231
Defcr20(OTTMUSG00000019856) : ----CGAGAGATCTGATATGCTATTGTAGAAAAGGAGGCTGCAATACAGGAGAACAAGTT : 231
Defcr25(OTTMUSG00000019700) : ------GAAGATCTGATATGCTATTGTAGAACAAGAGGCTGCAAAACAAGAGAACCCCTG : 228
OTTMUSG00000019857(noveldef) : ------GAAGATCGGATATGCTATTGTAGAACAAGCTGC--AAAAAAAAGAGAACCCCTG : 226
OTTMUSG00000019896(noveldef) : GCGTTGAGAGATCGGATATGCTATTGTAGAACAAGCTGC---AAAAAAAGAGAACCCCTG : 237
OTTMUSG00000019742(noveldef) : ----TGAGAGATCTGGTATGCTATTGTAGAGCAAGAGGCTGCAAAGCAAGAGAACCCATG : 231
Defcr24(OTTMUSG00000019980) : ----TGAGAGATCTGGTATGCTATTGTAGAGCAAGAGGCTGCAAAGCAAGAGAACCCATG : 231
Defcr23(OTTMUSG00000019488) : ----TGAGAGATCTGGTATGCTATTGTAGAACAAGAGGCTGCAAAACAAGAGAACCCATG : 231
OTTMUSG00000019762(Defcr23dupl) : ----TGAGAGATCTGGTATGCTATTGTAGAACAAGAGGCTGCAAAACAAGAGAACCCATG : 231
Defcr3(OTTMUSG00000019782) : ----TGAGAGATCTGGTATGCTATTGTAGAAAAAGAGGCTGCAAAACAAGAGAACCCATG : 231
OTTMUSG00000019892(Defcr3dupl) : ----TGAGAGATCTGGTATGCTATTGTAGAAAAAGAGGCTGCAAAACAAGAGAACCCATG : 231
Defcr26(OTTMUSG00000019889) : ----TGAGAGATCTGGGATGCTATTGTAGAAAAAGAGGCTGTACAACAAGAGAACCCATT : 231
OTTMUSG00000019784(noveldef) : ----TGAGAGATCTGGGATGCTATTGTAGAAAAAGAGGCTGCACAACAAGAGAACCCATT : 231
OTTMUSG00000019786(noveldef) : ----CAAAAAAGCTGATATGCTATTGTAGAATAAGAGGCTGCAAAACAAGAGAATCCGTT : 231
OTTMUSG00000019785(novelDefcr5) : ----CAAAAAAGCTGATATGCTATTGTAGAATAAGAGGCTGCAAAACAAGAGAACCCGTT : 231
OTTMUSG00000018259(novelDefcr5) : ----CTAAAAAGCTGATATGCTATTGTAGAATAAGAGGCTGCAAAACAAGAGAACCCGTT : 231
OTTMUSG00000018258(novelDefcr5) : ----CAAAAAAGCTGATATGCTATTGTAGAATAAGAGGCTGCAAAACAAGAGAACCCGTT : 231
OTTMUSG00000019924(novelDefcr5) : ----CAAAAAAGCTGATATGCTATTGTAGAATAAGAGGCTGCAAAACAAGAGAACCCGTT : 231
OTTMUSG00000019927(CRS1C-3) : ---AAAGAAGGTTTCCGTGGTGCCGGAAGTGCCCAGTGTGCCAGAAGTGCCAAGTCTGCC : 232
OTTMUSG00000018260(novelCRS1C) : ---AACGAAGGTTTCTGTGGTGCCGGAAGTGCCCAGTGTGCCAGAAGTGCCAAGTGTGCC : 232
Defcr-rs1(OTTMUSG00000019792) : ---AAAGAAGGTTTCCGTGGTGCCGGAAGTGCCGAGTGTGCCAGAAGTGCCAAGTGTGCC : 232
OTTMUSG00000019859(novelCRS1C) : ---AAAGAAGATTTCCGTGGTGCCGGAAGTGCCGAGTGTGCCAGAAGTGCGAAGTGTGCC : 232
OTTMUSG00000019893(novelCRS1C) : ---AAAGAAGATTTCCGTGGTGCCGGAAGTGCCGAGTGTGCCAGAAGTGCGAAGTGTGCC : 232
OTTMUSG00000018344(CRS4C-6) : ---TAGGAAAGCCTCCCCAGTGCCCGAAGTGCCCAGTGTGCTCAAAGTGCCCACAGTGCC : 232
A a ct G G AG G ca aag A g

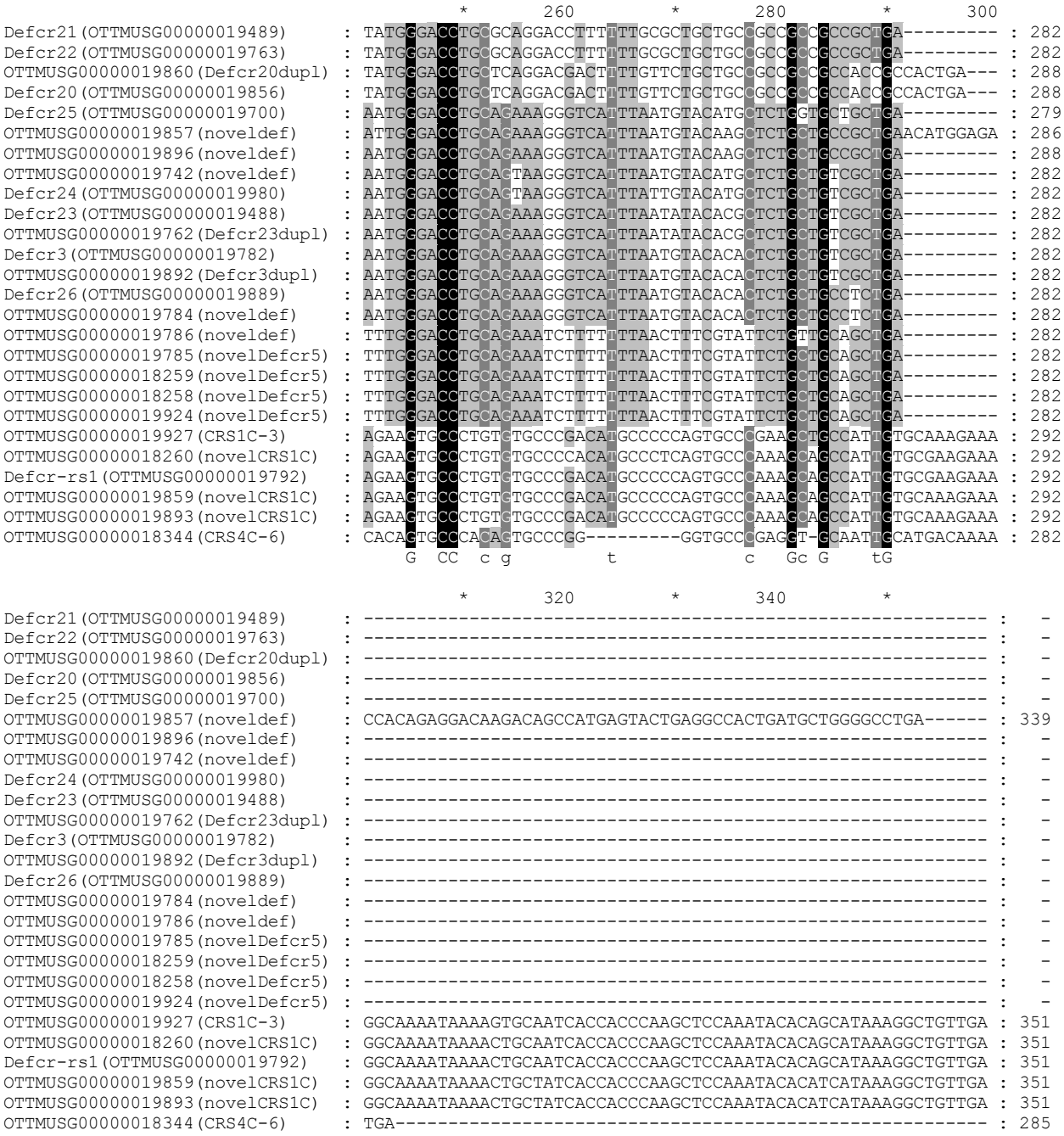


Figure 2.13. Nucleotide alignment of all murine α- and CRS-defensin coding sequences. Defensin gene sequences downloaded from Vega v.37 (Biomart) and aligned using ClustalW2. Alignment file visualized with GeneDoc (Multiple Sequence Alignment Editor and Shading Utility Version 2.7.000) software to view percentage conservation by shading (100% black, 80% dark grey, 60% light grey).

Designing defensin-specific primers for interrogation of transcriptional expression was not possible, either for each transcript individually or as a population. Potentially, primers for a few of the subgroups of defensins could have been designed, but this would have omitted the remainder, and also prevented the opportunity for novel defensin discovery. An alternative

method was to use the hybrid UDefR-oligo-dT primer in the reverse transcription. Due to the high conservation of all α- and CRS-defensins at the start of the coding sequence, the forward universal primer UDefF1 was designed at the start of the defensin coding sequences (Figure 2.14).

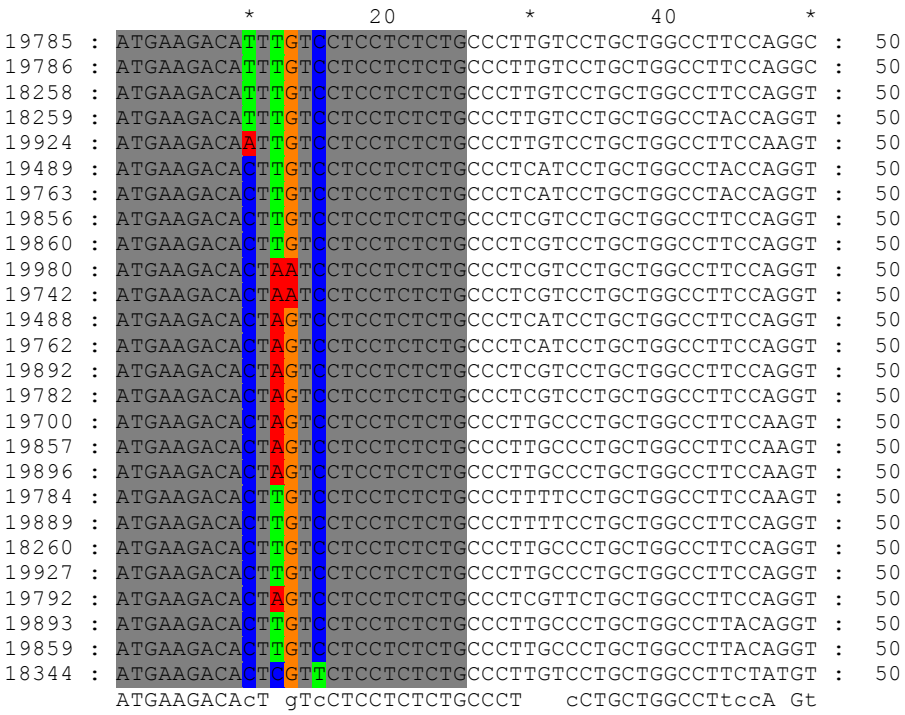
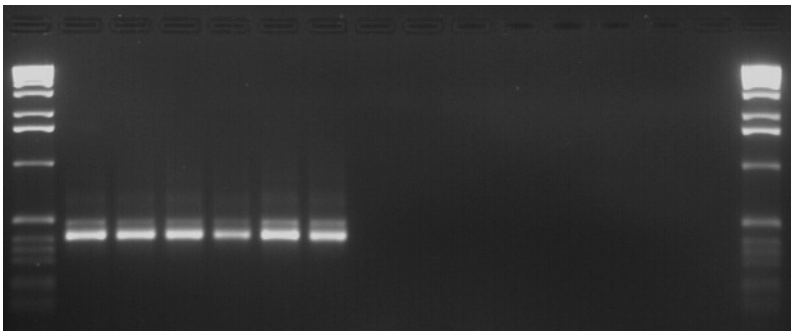


Figure 2.14. Murine α- and CRS-defensin coding sequence start with UDefF1 primer sequence. The GeneDoc display of the PAL2NAL alignment of the start of the defensin coding sequences with the consensus sequence shown below the alignment. Upper case letters indicate 100% identity between all sequences, lower case letters indicate the consensus of all sequences and spaces indicate no consensus between sequences. The universal UDefF1 primer is highlighted in grey, and the non-consensus positions coloured according to base identity. The defensin sequences are named by their OTTIDs, and should be preceded by OTTMUSG000000 for any database queries.

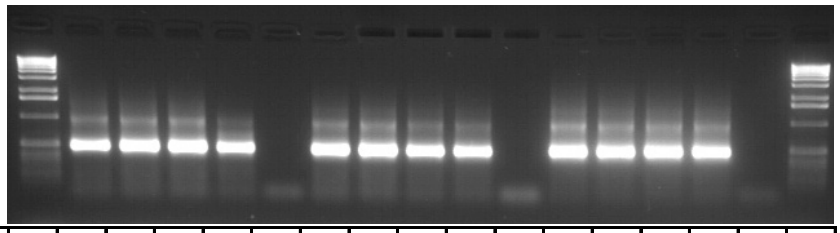
Following reverse transcription, amplification with the UDefF1 and UDefR primers resulted in distinct α- and CRS-defensin PCR product pools, with approximate sizes of 400 and 500 bp, respectively. (Figure 2.15A). Amplification with the UDefF1-454 and UDefR-454 primers also resulted in distinct α- and CRS-defensin PCR product pools, however the products were larger in comparison due to the additional length of the 454 primers (Figure 2.15B).

A.

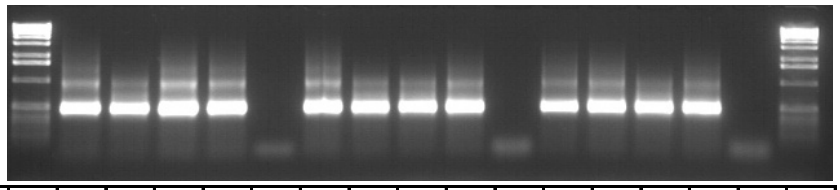


UDefF1 primer	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
UDefR primer	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
defensin cDNA template	-	M1	M2	M3	M4	M5	M6	-	M1	M2	M3	M4	M5	M6	-	-
RT enzyme	-	+	+	+	+	+	+	n/a	-	-	-	-	-	-	n/a	-
1 Kb DNA ladder (0.5 μg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

B.



UDefF1-454-1 primer	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
UDefF1-454-2 primer	-	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-	-
UDefF1-454-3 primer	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	-
UDefR primer	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
defensin cDNA template	-	M1	M1	M1	M1	-	M2	M2	M2	M2	-	M3	M3	M3	M3	-	-
1 Kb DNA ladder (0.5 μg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+



UDefF1-454-4 primer	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
UDefF1-454-5 primer	-	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-	-
UDefF1-454-6 primer	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	-
UDefR primer	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
defensin cDNA template	-	M4	M4	M4	M4	-	M5	M5	M5	M5	-	M6	M6	M6	M6	-	-
1 Kb DNA ladder (0.5 μg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

Figure 2.15. Generation of C57BL/6J murine universal α- and CRS-defensin pools for capillary and 454 amplicon sequencing. RNA was extracted from the small intestines of 3 female and 3 male mice and reverse transcribed using the UDefR-oligo-dT primer. PCR amplification was performed using either the UDefF1 and UDefR primers or UDefF1-454 and UDefR primers for capillary sequencing (**A**) and 454 sequencing (**B**), respectively, and products visualized by agarose gel electrophoresis. The no RT enzyme control was performed using the UDefF1 and UDefR primers (**A**).

2.3.5.2 Capillary sequencing of universal defensin cDNA

Unlike subsequent efforts, there were a high proportion of failed capillary sequences per sample. Additionally, some clones had only one good quality read, as it was common for reads initiated at the polyA tail to have messy sequences with higher background levels. Only those reads greater than 350 nucleotides were included in the analysis as these were considered full-length. Reads with short alignment length were excluded from the analysis because gaps in the PCR product were detected, which suggests bacterial-mediated deletion. Table 2.9 summarizes the capillary sequencing read quality and number of clones analyzed for each sample.

Table 2.9. Universal defensin capillary sequencing summary. The number of sequencing reads obtained from sequencing 96 TOPO clones per sample (M1-M6) with T3 and T7 primers. The number of reads that passed Asp (WTSI quality control), corresponding to the number of individual clones is indicated. Manual inspection of all the sequencing reads that passed Asp further reduced the number of high quality reads for analysis due to messy sequence, high background or less than 350 nucleotide alignment length.

Sample	Reads Passed Asp (/192)	Reads with Significant BLAST Hits (/192)	Clones with Significant BLAST Hits (/96)	Clones Analyzed (/96)	Clones Analyzed T3 and T7 Reads
M1	178	35	20	15	11
M2	183	65	36	20	15
M3	187	56	28	20	15
M4	189	53	27	17	11
M5	185	72	37	22	9
M6	190	65	34	31	12

The significant BLAST matches for the capillary sequences shared 98-100% identity with the 26 known α- and CRS-defensin transcripts (Table 2.10), however only those with 100% identity were further considered. None of the transcripts were detected in all samples M1-M6, however all but eleven transcripts, (OTTMUSG000000)18258, 18344, 19700, 19784, 19786, 19792, 19857, 19860, 19889, 19896 and 19924, were detected in at least one sample. Seven of these, 18344, 19700, 19786, 19792, 19857, 19896 and 19924, were also not detected by 454 sequencing (see below).

Table 2.10. Universal defensin transcript expression determined by capillary sequencing of TOPO clones. Relative murine α- and CRS-defensin transcript expression in six naïve C57BL/6J mice (M1-M6), determined by PCR amplification of small intestinal cDNA using universal defensin primers, cloning of PCR products into TOPO4 and capillary sequencing. TOPO clone sequencing reads are expressed as percentage identity with respect to each transcript reference sequence. > indicates identity was greater than the percentage listed but less than the next highest category (e.g. > 99.5% ≡ 100% < x ≥ 99.5%). Duplicated genes that could not be differentiated are listed together. * indicates an ability to differentiate between duplicated genes, and ** indicates an inability to differentiate between very similar genes.

	100% Identity		>99.5% Identity		>99.0% Identity		>98.5% Identity		>98% Identity	
Sample	OTTID	No. Reads (%)	OTTID	No. Reads (%)	OTTID	No. Reads (%)	OTTID	No. Reads (%)	OTTID	No. Reads (%)
M1	19892 * 19489 19763 19782 * 19856 * 19859/ 19893	3 (20.0) 1 (6.7) 1 (6.7) 1 (6.7) 1 (6.7) 1 (6.7)	19892 *	2 (13.3)	19980 19892 * 19488/ 19762	1 (6.7) 1 (6.7) 1 (6.7)	19763 19892 *	1 (6.7) 1 (6.7)	n/a	n/a
M2	19489 19763 19856 * 19892 * 18259 19782 * 19859/ 19893 19488/ 19762	2 (10.0) 2 (10.0) 2 (10.0) 2 (10.0) 1 (5.0) 1 (5.0) 2 (10.0) 1 (5.0)	18259 19489 19856 *	1 (5.0) 1 (5.0) 1 (5.0)	19763	1 (5.0)	19892 *	3 (15.0)	n/a	n/a
M3	19489 19856 * 18259 18260 19488/ 19762	2 (10.0) 2 (10.0) 1 (5.0) 1 (5.0) 1 (5.0)	19856 * 19742 19489	5 (25) 2 (10.0) 1 (5.0)	19742 19892 *	1 (5.0) 1 (5.0)	19860 * 19762 19488 **	1 (5.0) 2 (10.0)	n/a	n/a
M4	19489 19763 18259 19856 * 19892 *	2 (11.8) 2 (11.8) 1 (5.9) 1 (5.9) 1 (5.9)	18259 19860 * 19892 * 19927 19489/ 19763 **	1 (5.9) 1 (5.9) 1 (5.9) 1 (5.9) 2 (11.8)	19856 * 19489 19892 *	1 (5.9) 1 (5.9) 1 (5.9)	19489	1 (5.9)	n/a	n/a
M5	18259 18260 19742 19763	1 (4.5) 1 (4.5) 1 (4.5) 1 (4.5)	19856 * 18259 19763 19859/ 19893	3 (13.6) 2 (9.1) 1 (4.5) 1 (4.5)	19489 18259 19856 * 19489/ 19763 **	1 (4.5) 1 (4.5) 1 (4.5) 1 (4.5)	19860 * 19856 * 19980	2 (9.1) 1 (4.5) 1 (4.5)	19782 * 18260 19742/ 19980 **	1 (4.5) 1 (4.5) 1 (4.5)
M6	18259 19742 19763 19785 19856 * 19892 * 19927 18260 19489 19980 19488/ 19762 19859/ 19893	2 (6.5) 2 (6.5) 2 (6.5) 2 (6.5) 2 (6.5) 2 (6.5) 2 (6.5) 1 (3.2) 1 (3.2) 1 (3.2) 1 (3.2) 1 (3.2)	19856 * 19892 * 19488/ 19762	1 (3.2) 1 (3.2) 3 (9.7)	18259 19856 * 19892 *	1 (3.2) 1 (3.2) 1 (3.2)	19782 * 19489 19856 *	2 (6.5) 1 (3.2) 1 (3.2)	n/a	n/a

2.3.5.3 Deep sequencing by 454 of universal defensin PCR products

The 454 sequencing of the universal defensin amplicons generated a total of 101,525 sequences for the M1-M6 samples, based on their GSMID tag sequence identifier. The minimum and maximum number of reads for any individual sample was 13,671 and 20,145, respectively. The number of reads for each sample decreased by about half once they were filtered for those over 300 nucleotides, which was considered full-length, taking into account the lengths of the trimmed transcript (minus UDefF1 primer sequence and any 5'-UTR) reference sequences. One transcript (OTTMUSG00000019700) was shorter than this cut-off following trimming (254 nt), so the analysis was repeated filtered for those reads over 250 nucleotides, however there were still no 100% matches to this transcript. All other results were reported as per the 300 nucleotide filtering. Overall, approximately 3% of the filtered reads matched at least one of the trimmed transcripts with an alignment length of at least 300 nucleotides and 100% identity. Sequencing reads that did not match any defensin reference sequence were candidates for novel, unannotated genes. Pyrosequencing is prone to base calling errors in repetitive and homopolymeric regions of sequence, and these errors would also be contained within the non-100% identity category. The 454 sequencing results summary, including the numbers of reads generated for each sample and the number of identical matches to the trimmed defensin transcripts, can be found in Table 2.11.

Table 2.11. Universal defensin 454 amplicon deep sequencing summary. The total numbers of 454 sequencing reads for each C57BL/6J mouse (M1-M6). Unfiltered sequences that passed quality control processing steps (WTSI sequencing facility) were then filtered for those at least 300 nucleotides in length. The filtered sequences were further subdivided based on their identity to the defensin reference transcript sequences, trimmed to exclude any 5'-UTR sequence as well as the UDefF1 primer sequence. The sequencing reads with 100% identity (ID) with an alignment length of at least 300 nucleotides were quantitated for relative defensin transcript numbers. The percentage of unfiltered reads was compared to the total run number, the percentage of filtered reads were compared to the total unfiltered reads, and both the percentage of 100% and non-100% ID reads were compared to the total filtered reads.

	Number of Sequence Reads			
Sample	**Unfiltered**	**Filtered**	**100% ID**	**Non-100% ID**
M1	13732	7436	345	7091
M2	13671	7358	456	6902
M3	19753	10268	181	10087
M4	20145	10123	152	9971
M5	16867	8792	141	8651
M6	17357	8467	148	8319
Total	101525	52444	1423	51021
Percentage	99.6	51.7	2.7	97.3
TOTAL RUN	101928			

Throughout the six samples, variations within the numbers of each of the 26 transcripts were observed (Table 2.12 and Figure 2.16), however there were overall trends in the degree of transcript expression. The α-defensin genes found to have the highest expression included *Defcr3*dupl (OTTMUSG00000019892), *Defcr*21 (OTTMUSG00000019489), *Defcr23*/*Defcr23*dupl (OTTMUSG00000019488/OTTMUSG00000019762), and *Defcr20* (OTTMUSG00000019856). Moderate α-defensin gene expression was observed for *Defcr22* (OTTMUSG00000019763), *Defcr20*dupl (OTTMUSG00000019856), *Defcr24* (OTTMUSG00000019980), and *Defcr3* (OTTMUSG00000019782). α-Defensin genes with low expression included OTTMUSG00000018259 (novel*Defcr5*), *Defcr26* (OTTMUSG00000019889), OTTMUSG00000019785 (novel*Defcr5*) and OTTMUSG00000019784 (novelDef). α-Defensin genes for which no expression was detected were OTTMUSG00000018258 (novel*Defcr5*), OTTMUSG00000019924 (novel*Defcr5*), *Defcr25* (OTTMUSG00000019700), OTTMUSG00000019786 (novel*Def*), and OTTMUSG00000019896 (novel*Def*). The CRS-defensin genes expressed at moderate to low levels were OTTMUSG00000018260 (novel*CRS1C*), OTTMUSG00000019859/OTTMUSG00000019893

(novel*CRS1C*/novel*CRS1C*dupl), and *CRS1C*-3 (OTTMUSG00000019927). The expression of *CRS4C-6* (OTTMUSG00000018344) and *Defcr-rs1* (OTTMUSG00000019792) was not detected. The expression of OTTMUSG00000019857 (novel*Def*), which has a predicted mature peptide with only three cysteines, was also not detected.

Table 2.12. Universal defensin transcript expression determined by 454 amplicon deep sequencing. Relative murine α- and CRS-defensin transcript expression in six naïve C57BL/6J mice (M1-M6), determined by PCR amplification of small intestinal cDNA using 454-adapted universal defensin primers and 454 sequencing, expressed as a percentage of the total number of sequencing reads with 100% identity to each transcript. Defensin reference transcripts are denoted by the last five digits of their gene OTTIDs, which are preceded by OTTMUSG000000 in the Vega database genome browser. * and ** indicates genes annotated as duplications, which can be distinguished based on their 3'-UTR sequence. Duplicated genes that could not be differentiated are listed together.

	M1		M2		M3		M4		M5		M6		TOTAL	
OTTID	**Reads**	**%**	**Reads**	**%**	**Reads**	**%**	**Reads**	**%**	**Reads**	**%**	**Reads**	**%**	**Reads**	**%**
19892 * (Defcr3dupl)	76	22.0	124	27.2	71	39.2	43	28.3	33	23.4	18	12.2	365	25.7
19489 (Defcr21)	13	3.8	20	4.4	40	22.1	60	39.5	69	48.9	79	53.4	281	19.7
19488/ 19762 (Defcr23/Defcr23dupl)	126	36.5	118	25.9	3	1.7	3	2.0	4	2.8	4	2.7	258	18.1
19856 ** (Defcr20)	63	18.3	78	17.1	6	3.3	6	3.9	3	2.1	12	8.1	168	11.8
19763 (Defcr22)	26	7.5	41	9.0	2	1.1	3	2.0	2	1.4	2	1.4	76	5.3
19860 ** (Defcr20dupl)	22	6.4	39	8.6	2	1.1	1	0.7	2	1.4	1	0.7	67	4.7
19980 (Defcr24)	0	0.0	0	0.0	32	17.7	18	11.8	11	7.8	4	2.7	65	4.6
19782 * (Defcr3)	4	1.2	21	4.6	9	5.0	4	2.6	5	3.5	0	0.0	43	3.0
18260 (novelCRS1C)	3	0.9	3	0.7	4	2.2	7	4.6	5	3.5	16	10.8	38	2.7
18259 (novel Defcr5)	4	1.2	1	0.2	3	1.7	2	1.3	2	1.4	2	1.4	14	1.0
19859/ 19893 (novelCRS1C)	1	0.3	0	0.0	3	1.7	2	1.3	4	2.8	3	2.0	13	0.9
19742 (novelDefcr)	2	0.6	8	1.8	2	1.1	0	0.0	0	0.0	0	0.0	12	0.8
19889 (Defcr26)	3	0.9	1	0.2	3	1.7	2	1.3	0	0.0	2	1.4	11	0.8
19927 (CRS1C-3)	0	0.0	0	0.0	1	0.6	0	0.0	1	0.7	4	2.7	6	0.4
19784 (novelDefcr)	2	0.6	1	0.2	0	0.0	0	0.0	0	0.0	0	0.0	3	0.2
19785 (novel Defcr5)	0	0.0	1	0.2	0	0.0	1	0.7	0	0.0	1	0.7	3	0.2
19792 (Defcr-rs1)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
18344 (CRS4C-6)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
19857 (novelDefcr)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
19896 (novelDefcr)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
19786 (novelDefcr)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
19700 (Defcr25)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
19924 (novel Defcr5)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
18258 (novel Defcr5)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
TOTAL	345	100	456	100	181	100	152	100	141	100	148	100	1423	100

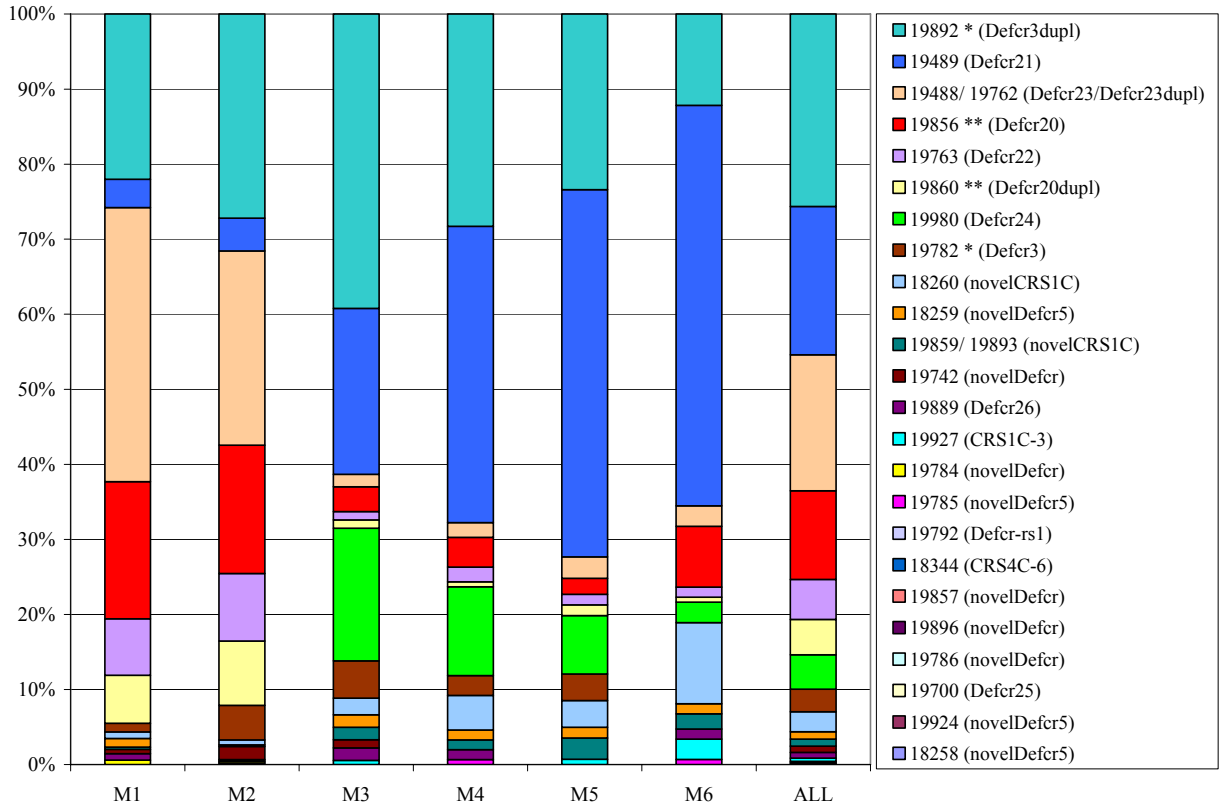


Figure 2.16. Universal defensin transcript expression determined by 454 amplicon deep sequencing. Relative murine α- and CRS-defensin transcript expression in six naïve C57BL/6J mice (M1-M6), determined by PCR amplification of small intestinal cDNA using 454-adapted universal defensin primers and 454 sequencing, expressed as a percentage of the total number of sequencing reads with 100% identity to each transcript. Defensin reference transcripts are denoted by the last five digits of their gene OTTIDs, which are preceded by OTTMUSG000000 in the Vega database genome browser. * and ** indicates genes annotated as duplications, which could be distinguished based on their 3'-UTR sequence. Duplicated genes that could not be differentiated are listed together. Table 2.12 lists the exact numbers and percentages of 454 sequencing reads for each transcript reference sequence. Each set of duplicated genes had identical mature peptide regions, which therefore could not be differentiated in upon translation.

The 454 sequencing confirmed the gene expression of six out of the seven α-defensin peptides recently purified from the small intestine of C57BL/6J mice (*personal communication*, A. Ouellette and M. Shanahan). These included Crp3, Crp5, Crp20, Crp21, Crp23, and Crp24, which are coded for by genes *Defcr3*, *Defcr5*, *Defcr20*, *Defcr21*, *Defcr23* and *Defcr24*, respectively. These genes comprise the most highly and moderately expressed transcripts identified through 454 sequencing, including the duplicated genes. One exception was the highly expressed transcript from *Defcr22*, for which the corresponding predicted peptide Crp22 was not identified by Ouellette and Shanahan. However the mature peptide regions of Crp21 and Crp22 differ by only one amino acid at position 11 which corresponds to position 68 in the prepropeptide. At this position Crp21 contains an asparagine (polar, neutral), whereas Crp22

contains a lysine (polar, positive), which both have similar hydropathy indices (142). The peptides were separated by reverse-phase high performance liquid chromatography (RP-HPLC) with an increasing acetonitrile gradient, thus Crp21 and Crp22 may have eluted together.

The defensin transcript data generated in this work confirms the genomic annotation for those genes detected, and their relative expression suggests that the polymorphisms observed between strains are dependent not only on their genomic content but also on the regulation of their transcription.

None of the translated-454 sequences for any of the six mouse samples matched with 100% identity to the putative novel Crp27 peptide purified from C57BL/6J mice (Table 2.13). In contrast, the 454 analysis identified two other moderately expressed transcripts, OTTMUSG00000019742 (novel*Def*) and *Defcr26* (OTTMUSG00000019889), which were not identified as peptides (*personal communication*, A. Ouellette and M. Shanahan).

Table 2.13. Blastx analysis of 454 transcripts against the mature Crp peptides. The amino acid sequences for the mature peptides purified from C57BL/6J small intestine were obtained from A. Ouellette and M. Shanahan. A database of the peptide sequences was created, against which the translated filtered 454 sequences (over 300 nt) for each sample (M1-M6) were blasted. The translated unfiltered sequences were also blasted against the Crp27 peptide sequence, but no matches were detected (not shown). The lengths of the mature peptides are as follows. Crp3, 35 a.a; Crp5, 36 a.a.; Crp20, 42 a.a.; Crp21, 36 a.a.; Crp23, 35 a.a.; Crp24, 35 a.a.; Crp27 35 a.a.

	Number of 100% Identity Reads					
Peptide	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
Crp3	702	772	580	415	298	168
Crp5	40	48	62	75	41	60
Crp20	1858	1937	430	460	444	465
Crp21	107	104	174	322	292	288
Crp23	640	517	243	144	122	80
Crp24	21	19	179	76	52	15
Crp27	0	0	0	0	0	0
TOTAL	3368	3397	1668	1492	1249	1076

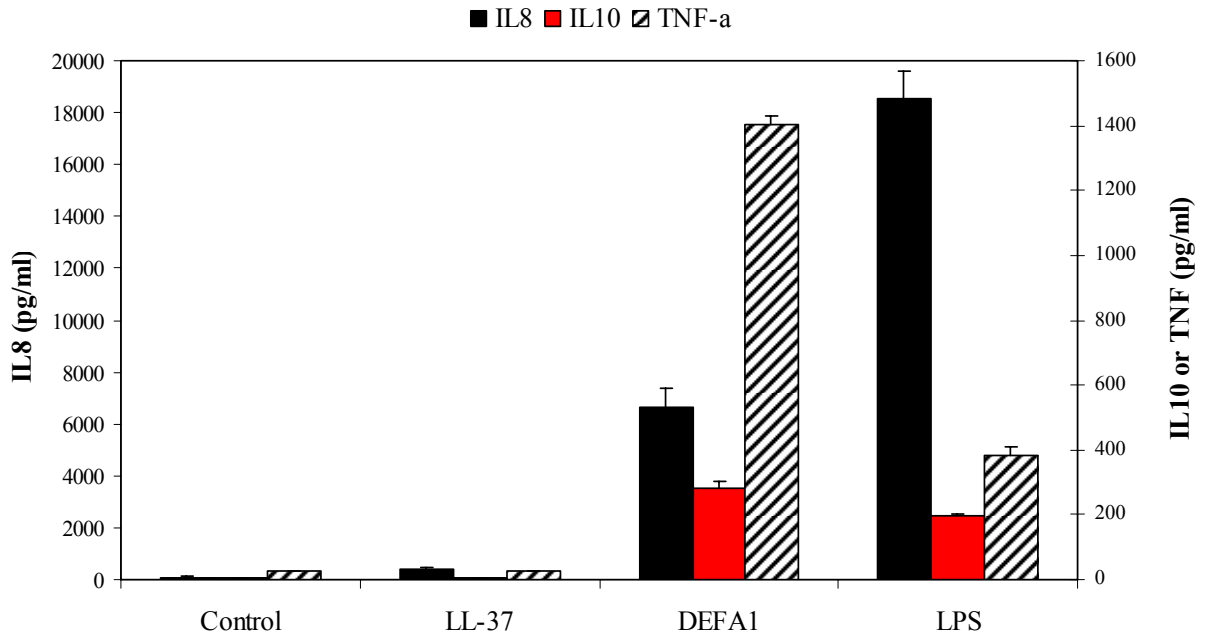
There are several possible explanations for these apparent discrepancies. It seems possible that the *Defcr27* cDNA sequence is incorrect since the GenBank accession is from Expressed Sequence Tag evidence. Conversely, the predicted and actual mass spectrometry atomic mass units (A.M.U.) were not as close for Crp27 as for the other six peptides detected (*personal communication*, A. Ouellette and M. Shanahan). It is also possible that the *Defcr27* transcript might have been produced at levels below the limits of detection of the 454 sequencing or

conversely the transcripts identified by 454 sequencing were produced as peptides but at quantities below the limits of detection by RP-HPLC. If the cDNA sequence is incorrect, then it is possible that the 454 sequencing results could aid in the identification of the correct sequence. The predicted molecular weights of the deduced mature peptides of OTTMUSG00000019742 (novelDef) and Defcr26 (OTTMUSG00000019889) are 4085.0 and 4150.1 A.M.U., respectively (143). The molecular weight of OTTMUSG00000019742 was similar to the predicted (4086.7 A.M.U.) and actual (4077.0 A.M.U.) values obtained for the putative novelCrp27 peptide (*personal communication*, A. Ouellette and M. Shanahan). Thus, OTTMUSG00000019742 could potentially be a novel α-defensin identified at the genomic, transcript and peptide levels.

2.3.6 The human α-defensin DEFA1 induces IL8 and IL10 release in vitro

DEFA1 induced the release of IL8 and TNF, as well as IL10 from human PBMCs (Figure 2.17A). This was dose-dependent for IL10 and TNF (IL8 not tested) (Figure 2.17B). The specificity of peptide action was verified by proteinase K digestion of DEFA1, which significantly reduced the TNF production by human PBMCs (Appendix A.4). Proteinase K digestion of LPS under the same conditions did not significantly reduce the TNF production, except at the highest concentration of proteinase K tested, which is likely due to excessive residual enzymatic activity in the tissue culture supernatant or to a contaminant in the proteinase K. Cytotoxicity of the peptide was not observed by WST-1 cellular proliferation assay, and no endotoxin contamination was detected by LAL assay (Hycult Biotechnology). Preliminary experiments suggested the involvement of NF-κB and mitogen-activated protein kinase (MAPK) 14 signalling, as DEFA1-induced IL8, IL10 and TNF release from human PBMCs was inhibited by one hour pretreatment with 10 μM of the Iκ-Bα phosphorylation inhibitor, Bay 11-7085, or 10 μM of the MAPK14 inhibitor, SB208530 (Biomol International) (Appendix A.5).

A.



B.

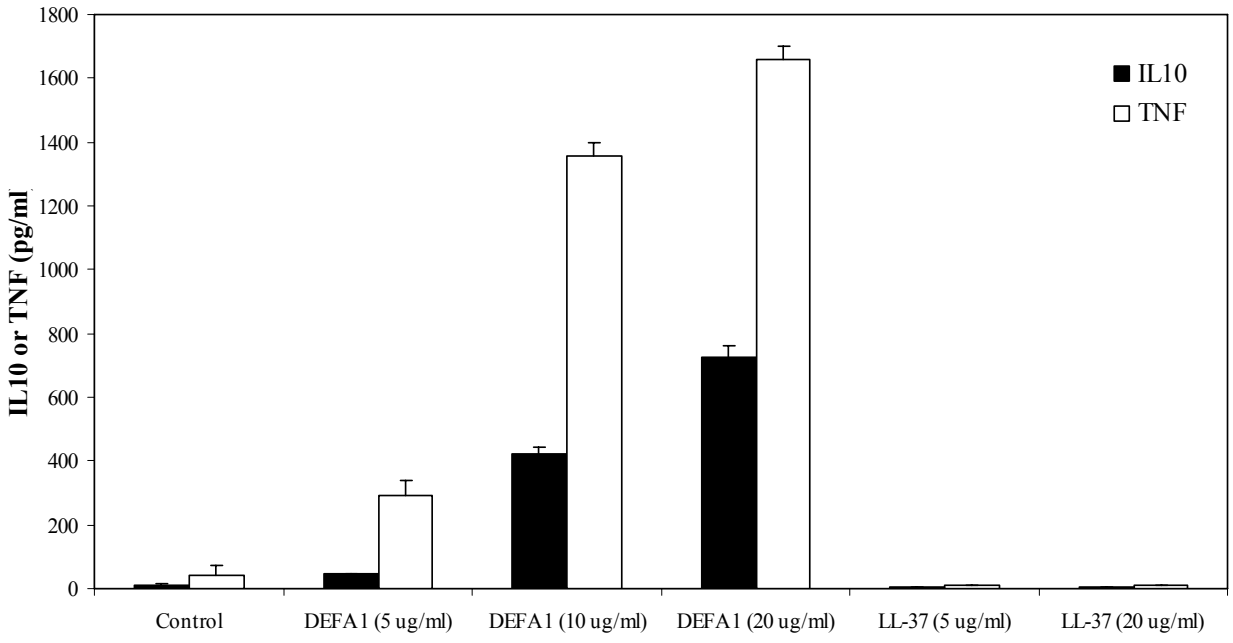


Figure 2.17. The human α-defensin DEFA1 induced cytokine and chemokine release *in vitro*. Human PBMCs treated with DEFA1 (20 μg/ml) for 24 hours release IL8, IL10 and TNF into the cell culture supernatant, which was measured by ELISA (**A**). The quantity of IL8 produced is reported on the left-hand axis and the quantities of IL10 and TNF are reported on the right-hand axis. The production of IL10 and TNF, measured by ELISA, by PBMCs treated for 24 hours with 5-20 μg/ml DEFA1 was dose-dependent (**B**). The release of IL8 by PBMCs in a dose-dependent manner was not tested. LPS and LL-37 were included as controls. Results are representative of two independent experiments for A and four for B, each with duplicate wells per treatment.

2.4 Discussion

Historically the organization of mouse α-defensin genes has been poorly defined, which is due in part to the difficulties in assembling the reference genome in these regions. Additionally consistency and standardization in naming genes and peptides has been an issue between research groups, journals and genome browsers, and is most likely due to a combination of factors. Defensins were first discovered as peptides or from cDNA. With the completion of large-scale genome sequencing projects, it has become possible to mine these genomes for defensin genes by scanning translated genomic sequence for the six conserved cysteine residues. Some defensins have only been identified at the genomic level, without subsequent peptide or RNA expression data. Searching the literature gives the impression that the naming of murine α-defensins is orderly and systematic although searching major databases for the corresponding information is difficult. The problem arises because most of the experimental data comes from mouse strains that are different from the C57BL/6J reference genome. I have found evidence to confirm that there are strain and possibly CNV differences within the defensin gene family and therefore annotating/mapping genes and peptides discovered in non-reference strains remain a challenge. Since defensin genes are involved in copy number variation as well as revealing polymorphisms between strains, a one to one orthology between species is hard to predict.

Prior to this analysis, there was no reliable reference gene set available for the mouse strain C57BL/6J defensin genes, demonstrating that manual intervention is still critical for the annotation of complex gene families and heavily duplicated regions. This manual annotation of the defensin gene cluster enabled the identification of novel α- and CRS-defensin genes, and pseudogenes. Comparison of the mouse α-defensins in the three main mouse reference gene sets Ensembl, MGI, and NCBI RefSeq revealed significant inconsistencies in annotation and nomenclature. Accurate gene annotation is facilitated by the annotation of pseudogenes and regulatory elements. The manually curated gene models described here will be incorporated into the Ensembl and Consensus Coding Sequence (CCDS) reference sets.

This work has highlighted the need to establish a standardized defensin nomenclature system, applicable to all organisms, which can then be implemented by the major genome centres, genome browsers and journals. For mouse, human and many other species, the abbreviation DEFB/Defb and DEFA/Defa has been used to tag β- and α-defensins, respectively. Both the Hugo (Human Genome Organisation) Gene Nomenclature Committee (HGNC) and Mouse Genomic Nomenclature Committee (MGNC) groups have approved the DEFB/Defb and

DEFA/Defa naming schemes for human and mouse (144). This system is ideal compared to taxonomy-based naming (Table 2.14), which only complicates matters further.

Table 2.14. Species-specific defensin nomenclature. Examples of the diverse names and symbols historically used to describe defensins within several species including mouse, most of which can be found in current literature. n/a, no symbol associated with name.

Species	Name	Symbol
human	human neutrophil peptide	HNP-
	human defensin	HD-
	alpha-defensin	DEFA-
	human beta-defensin	HBD-
	beta-defensin	DEFB-
	corticostatin	CS
rabbit	neutrophil peptide	NP-
	microbicial cationic protein	MCP-
	corticostatin	CS
	"rabbit kidney"	RK-
mouse	crypi	crypi
	cryptidin	Cryp-
	cryptdin	Cryp-
	defensin-related cryptdin	Defcr-
	cryptdin-related sequence	CRS-
	defensin-related cryptdin-related sequence	Defcr-rs-
	alpha defensin	Defa-
	beta defensin	Defb-
rat	corticostatin	CS
	rat neutrophil peptide	RatNP-
	"rat peptide from bone marrow"	R-
	alpha-defensin	Defa-
	beta-defensin	Defb-
guinea pig	guinea pig neutrophil peptide	GPNP-
	guinea pig cationic peptide	GNCP-
	corticostatin	CS
cow	tracheal antimicrobial peptide	TAP
	neutrophil beta-defensin	BNBD-/BNDB-
	beta-defensin	DEFB-/DB-/BD-
	lingual antimicrobial peptide	LAP
	bovine beta-defensin	BBD-
chicken	chicken heterophil peptide	CHP1/AMP1_CHICK
	gallinacin	GAL-/GLL-
	beta-defensin	n/a
	avian beta-defensin	AvBD-
ostrich	ostricacins	OSP-
	ostrich gallinacin	n/a
	ostrich avian beta-defensin	ostrich AvBD-
turkey	turkey heterophil peptide	THP-
	turkey beta-defensin	n/a
	turkey avian beta-defensin	turkey AvBD-

If possible, any naming scheme should encompass copy number variants and define relationships between peptides in different strains of mice (e.g. Crp4 and Defcr20/Defcr21). Suggestions for a clearer and unified naming scheme have been submitted by us to MGNC, and are being discussed by both HGNC and the Rat Genome Nomenclature Committee (RGNC). MGNC has started to assign symbols to most defensin pseudogenes identified here and has implemented some of the proposed nomenclature changes (Tables 2.3 and 2.4), e.g. the Defcr root has been changed to Defa for mouse α-defensins and defensin-related cryptdin sequences. Elucidation of the genomic structure of this complex gene cluster on the mouse reference sequence, and adoption of a clear and unambiguous naming scheme, will provide a valuable tool to support studies on the evolution, regulatory mechanisms and biological functions of defensins *in vivo*.

The human α- and β-defensin genes are subject to copy number variation. Evolution, duplication and allelic variation of defensin genes are currently under investigation (145-147). Since many diseases/disorders appear to be associated with copy number variation, the review and standardization of CNV nomenclature is critical to future studies. Several analyses have shown that large-scale copy number polymorphisms are a major source of genetic variation (148-150). One of these polymorphisms involves the human β-defensin cluster on 8p23.1. Whereas carriers of a euchromatic variant that is cytogenetically visible have nine to twelve copies of the region (151, 152), most people have two to seven copies (145). Correlation of β-defensin copy number with expression levels suggests that variable expression levels could cause different predisposition and susceptibility to infectious diseases. A recent genetic mapping approach confirms two distinct β-defensin CNV loci, approximately 5 Mb apart on human Chromosome band 8p23.1 (146). The authors state that this contradicts the current genome assembly. As a follow up, we analyzed the region surrounding the genomic coordinates indicated in the aforementioned study and found five known β-defensin genes (126) and one β-defensin pseudogene (two copies of *Defb130*, *Defb134*, *Defb136* and *Defb137*, data not shown). The relationship between this cluster and the duplicated CNV region is unclear. In a study analyzing the expression levels of human α-defensins *DEFA1* and *DEFA3*, the relative proportions of *DEFA1*:*DEFA3* mRNA were associated with the respective gene copy numbers (153). However, combined levels of *DEFA1* and *DEFA3* were not correlated with gene copy number, and distinct transcription factor regulation was suggested to explain the differential expression of these genes (153). Conversely total *DEFA1* and *DEFA3* copy numbers were correlated with the amount of DEFA1-3 peptides purified from human neutrophils (147), which might explain the apparent

constitutive nature of human neutrophil α-defensin expression. Total *DEFA1* and *DEFA3* copy numbers have been reported to vary between four and 11 in 111 individuals (153) and five to 14 in 27 individuals (147), whilst the *DEFA3* allele was absent in 10% (153) or 26% (147) of the individuals tested. The proportion of individuals lacking *DEFA3* appears to be population-dependent and has been shown to vary between 10-37% (154). It has been speculated that populations with distinct ecological histories carry different defensin gene copy numbers derived from the selective pressure presented on their historical geographic regions (147). However a direct correlation between the copy number variation in the defensin region and geographic origin has not been established.

Detection of copy number variation and structural variation in mouse, similar to what has been observed in human, has proven to be difficult because of the limited sequence information on various mouse strains and as well as the large gap in Chromosome 8. There has also been speculation that the current assembly may be collapsed due to its highly repetitive nature and the putative tandem repeat of Contig AC152164.14 is such an example. Current mouse tile-path arrays do not have the capability to resolve individual gene copy numbers and their design is limited by the current mouse genome assembly. Genes that have been annotated as 100% identical are obvious candidates for copy number variants, but sequence information from additional mouse strains is needed for verification. Between the existing α-defensins, we have observed apparent polymorphisms in different mouse strains. A project currently being undertaken at the Wellcome Trust Sanger Institute aims to sequence the genomes of 17 common mouse strains. However, it is reasonable to assume that there will be difficulties in assembling Chromosome 8 proximal to α-defensin gene regions. Nevertheless, comparisons between sequence available for these strains will start to define the copy number and polymorphic variation of mouse α-defensins.

The difficulties in characterizing mouse α-defensins have hindered exploration of their *in vivo* function. Designing unique vectors for individual knockouts of α-defensin genes is impossible. However, in order to circumvent the potential redundancy in defensin function *in vivo*, I explored the rationale for generating an α-defensin deficient mouse. The Mutagenic Insertion and Chromosome Engineering Resource (MICER) utilizes hypoxanthine/aminopterin/thymidine (HAT) resistance mediated by hypoxanthine guanine phosphoribosyl transferase (*Hprt*) for murine embryonic stem cell gene targeting (155). Chromosomal deletions can also be engineered however two rounds of microinjection with the

MICER targeting vectors are required; one vector contains Exons1-2 of *Hprt*, a loxP site, and a neomycin resistance cassette, and the second vector contains Exon3-9 of *Hprt*, a loxP site (in the same orientation as the first), and a puromycin resistance cassette. Cre-mediated recombination between the loxP sites allows for deletion of the intervening DNA. Serendipitously there are MICER clones in the correct location and orientation to delete the entire α-defensin cluster (~0.7 Mb). However the combination of multiple targeting events and, if successful, low frequency of the deletion event, negated this approach. Additionally, until the assembly issues have been addressed, manipulation of this region of the genome is likely to come under a lot of scrutiny, as the exact complement of the α-defensin genes within the mouse reference genome is still unknown. Therefore it was decided that any attempts to generate an α-defensin deficient mouse would not be prudent at this time.

At present only a small fraction of defensin genes have peptide products that have been purified from any strain of laboratory mouse. The small size, redundancy and variable expression levels of these peptides may be the reasons for the difficulty in isolating the peptides. Transcriptional profiling enables differentiation between those genes expressed and those for which peptides have been detected. Traditional cloning and capillary sequencing methods have provided solid evidence of murine α- and CRS-defensin expression in strains other than C57BL/6J (20-22, 25-27, 88, 127, 156), although the antimicrobial nature of defensins is likely to bias this method. Single nucleotide polymorphisms were detected in the capillary sequences obtained here that have 98-99.9% identity to reference sequences. Additionally capillary sequencing only allows for presence/absence determination of gene expression or at most a very rough estimate of transcript numbers.

In-depth analyses of α- or CRS-defensin transcription have not been previously performed in any mouse strain. This is therefore the first study to provide quantitative gene expression data. The 454 sequencing analysis was performed in order to validate the genomic annotation, especially for novel genes, as well as to quantify the relative numbers of transcripts produced. The results confirmed what is known about the expression levels of defensin peptides in strains of mice other than the C57BL/6J, including Swiss Outbred, 129, C3H (28, 29, 88, 156), and also gives a rational explanation for this apparent discrepancy between genomic content and peptide repertoire. The control and degree of defensin gene expression appears crucial to their detection as peptides. The annotation of TATA boxes provides the starting point for promoter analysis, however this does not appear to be a main regulatory factor as the transcripts expressed

in high quantities had both strong or weak TATA boxes. There is a trend, however, that transcripts expressed in low amounts or those not detected did not have a TATA box. This is likely due to the rapid expansion and gene duplication events that have occurred in mouse defensins, leading to loss or gain of regulatory elements. A more in-depth analysis the defensin promoter regions is required, as well as experimental validation of those transcription factors involved in defensin gene expression. Copy number variation also appears to be a defining factor in the detection of the defensin genes as peptides. However, 454 sequencing has also shown that CNV does not necessarily correlate with peptide expression levels. CNV has been implicated in many human diseases, and defensin polymorphisms could indeed play a significant role in immunity and disease susceptibility. However a better understanding of the regulation of defensins, especially with respect to CNV, is needed in order to address their role in disease or disease susceptibility.

The Blastx analysis showed a higher proportion of the filtered reads that matched known peptides compared to the number of filtered reads that matched the known transcripts (Table 2.13). This is not surprising when considering that the peptide alignment was less than a third of the length of the full-length transcript alignment, as only the sequence for the mature peptide region, and not the full-length pro-peptide, has been reported (*personal communication*, A. Ouellette and M. Shanahan). This indicates that the high proportion of less-than-100% matches in the filtered sequences could be the result of base-calling errors, which are inherent in 454 sequencing of those sequences with homopolymeric regions. The identification of novel variants amongst the non-100% matches remains to be determined. Regardless, 454 sequencing appeared to be the most appropriate of the next generation sequencing technologies for α- and CRS-defensins. The reads obtained from Illumina sequencing are not long enough to differentiate the defensin transcripts, due to their high similarity while the very large number of reads obtained by 454 sequencing allowed for relative quantification of the α- and CRS- defensins in the C57BL/6J mouse.

In vitro tissue culture experiments revealed that DEFA1 induces IL8, IL10 and TNF release from human PBMCs. DEFA1 may mediate cytokine production through the activation of NF-κB signalling pathways. Similarly, in response to DEFA1 human bronchial epithelial cells produce *IL1B* and *IL8*, stimulate IL8 release, and promote NF-κB-DNA binding (157). Both pro- and anti-inflammatory functions have been proposed for DEFA1 due to its ability to selectively induce cytokine production, chemoattract monocytes, naïve T cells, and immature dendritic cells,

and act as an adjuvant, but also to inhibit the release of IL1B from LPS-treated human monocytes, and inhibit complement activation (42). These experiments provided evidence for the function of DEFA1. However the *in vivo* role of DEFA1, and indeed other defensins, needs further elucidation. Due to the potential of host defence peptides, including defensins, as novel therapeutics against infection (158), *in vivo* models need to be developed in order to test their function amidst the sea of peptide and cytokine redundancy. The development of knock-in mice expressing IL8 or Il10 will allow for the testing of peptide mechanism in the treatment of infection under conditions of inflammation or immunosuppression.

3 TARGETED KNOCK-IN OF INTERLEUKIN 8 AND INTERLEUKIN 10 INTO MURINE EMBRYONIC STEM CELLS TO GENERATE MICE WITH CONDITIONAL INTESTINAL-SPECIFIC GENE EXPRESSION

3.1 Introduction

Many different stimuli induce IL8 production *in vitro* from human primary cells or cell lines, and as such IL8 is a common output for inflammatory assays. Additionally, IL8 has been implicated in a number of human enteric diseases, in particular Shigellosis, Salmonellosis and chronic inflammation (e.g. Crohn's disease) (78, 159, 160), as well as in certain cancers due to its angiogenic and tumorigenic properties (46). However the lack of Il8 in the common laboratory model organisms, mouse and rat, has made it difficult to study its role *in vivo*. In addition, the lack of Il8 is hypothesized to be one reason why models of intestinal disease do not necessarily mimic the pathology seen in the corresponding human disease, particularly with regards the lack of neutrophil infiltration (78).

Alternative models have been developed to address this issue, but their use has not been widespread. Guinea pigs, which express an IL8 orthologue, were the first rodents used to study infectious diseases, and for some diseases provide the most comparable model to the human disease (161). However mice are most often used for experimentation due to their relatively low cost of maintenance, as well as the large number of laboratory reagents (e.g. antibodies), genetic tractability and bioinformatics tools (e.g. genome sequence), available for their study. The completion of the mouse reference genome sequence and its annotation has made it possible to generate specific gene knock-outs in murine embryonic stem cells, which can then be used to generate gene deficient mice in a targeted manner (162).

There has been considerable interest in administering recombinant IL8 to mice in order to induce neutrophil infiltration and the pathology associated with intestinal bacterial disease (79), as well as generating transgenic mice expressing human IL8 (163, 164). The human apolipoprotein (APO) E promoter/ hepatocyte control region (HCR) enhancer, or the rat fatty acid binding protein (Fabp) promoter were used to create mice expressing IL8 in the liver or intestine, respectively (163). However for both constructs, considerable variability of gene expression was determined in different F1 heterozygous lines (163). Random transgene integration occurred for both constructs, and likely the number of insertions varied between the F1 lines since insertion of a similar APOE/ HCR construct resulted in 1-90 copies of the transgene (165). Additionally, there was no external control over the expression of IL8 in either of these constructs and while constitutive expression resulted in IL8 in the serum and

subsequently higher numbers of neutrophils, further neutrophil movement was inhibited (163), potentially due to receptor desensitization. Another IL8 transgenic mouse has been generated, by random transgene insertion, whereby IL8 is expressed in the distal ileum, caecum, and proximal colon of the intestine under the control of the rat Fabp promoter and the reverse tetracyline transactivator (164). IL8 expression can therefore be induced using doxycycline, although expression does not occur throughout the length of the intestine. This model is not appropriate for bacterial infection studies, as doxycycline is likely to affect the normal flora of the mouse as well as the course of infection (166). Additionally, all of these models were created by random insertion of an engineered transgene, therefore genetic elements in proximity to the site(s) of insertion could affect expression of the transgene, and conversely insertion of the transgene could disrupt expression of endogenous genetic elements by insertional mutagenesis (166).

Mice deficient in either the Cxcr1 or Cxcr2 genes have been generated. There have been numerous studies using the Cxcr2 deficient mouse, however there have been only limited studies on the Cxcr1 deficient mouse and suggest that homozygous mutants do not show any abnormal phenotype (167). In general, Cxcr2 deficient mice have decreased neutrophil recruitment, which is partially tissue- and time-dependent, but cell numbers vary according to type of inflammation, whether it be induced by infection or chemical (52). In some cases neutrophil chemotaxis is almost completely abrogated, which would suggest Cxcr2 as solely responsible for neutrophil chemotaxis under certain conditions (52). Murine neutrophils are capable of *in vivo* chemotaxis following administration of recombinant IL8 (79), indicating sufficient binding of Cxcr1 or Cxcr2 to IL8, or conversely indicating that *in vivo* cross-talk occurs between Cxcr1 and Cxcr2 upon IL8 engagement. These experiments provide a way of determining the *in vivo* function of Cxcr1 and Cxcr2, which may be extrapolated in part to define that of IL8. However the fact that both Cxcr1 and Cxcr2 bind other ligands, in addition to IL8, makes this an indirect method for determining specific ligand function as a whole.

The importance of IL10 in the regulation of the immune system becomes apparent upon loss of the gene product in various knockout mouse models. Il10-deficient mice, created either by gene knockout or administration of recombinant anti-Il10 antibodies, are generally more resistant to bacterial infection, especially intracellular bacteria (63, 168). However these mice are also prone to excessive immune stimulation and can develop spontaneous colitis, which is mediated largely by CD4+ T cells (169). Interestingly the Il10 produced by some $CD4^+$ T cell subsets, Th1 and Treg, inhibits activity of these cells as well as macrophages and DCs (64, 170). The role of the normal flora of the mice also plays an important role, as spontaneous colitis does

not develop in germ-free Il10 deficient mice (171). IL10 is also critical for tolerance to self, which if broken, can lead to other autoimmune disorders (172). The recognition of and differentiation between self, commensal organism and pathogen is crucial to mount the proper immune response and limit damage to the host.

A transgenic Il10-Foxp3 reporter mouse with CD90.1 and green fluorescent protein (GFP) knocked-in at the endogenous *Il10* and *Foxp3* loci, respectively, has been developed to monitor Treg differentiation *in vivo* (173). The lung and liver have Tregs expressing Foxp3 but not Il10, whereas the small intestine has Treg expression of Il10 but not Foxp3 and the large intestine has Tregs expressing both Il10 and Foxp3 (173). Tregs can develop in the absence of Il10 *in vivo*, which contrasts to many *in vitro* studies (173), therefore it is important to develop models with which to study Il10 function. Although Il10 production was disrupted in the above described reporter mouse, another Il10 reporter mouse has been generated wherein GFP was knocked-in downstream of *Il10* (174). This model was used to elegantly show Il10 production by intestinal lymphocytes in the small and large intestine following T cell receptor (TCR) stimulation (174). Il10 production therefore occurs normally *in vivo*, which is advantageous for the study of its induction. However study of its function following such induction is not as straightforward. A transgenic model in which Il10 is expressed only in the intestinal tract was generated by introducing the rat Fabp intestinal (Fabpi) promoter upstream of endogenous *Il10* and then crossing with an Il10 deficient mouse (175). Constitutive Il10 expression was expected in enterocytes from the small intestine and the proximal region of the large intestine, although expression could only be demonstrated in the small intestine (175). Regardless, there was an increase in $CD3^+$ T cells and $CD4^+CD25^+$ Treg cells, and a decrease in tumor necrosis factor (Tnf) and interferon-γ (Ifng) production, but increased transforming growth factor-β1 (Tgfb1) production (175). Importantly the authors noted that lymphocyte populations in the spleen were not affected indicating that Il10 acted locally in the small intestine (175). A caecal ligation and puncture model of sepsis demonstrated improved survival in the case of these Il10 transgenic mice. However, most physiological parameters tested were equivalent for both transgenic and wild type mice, except that the transgenic mice had lower Ccl2 and Il6 production from splenocytes stimulated with anti-CD3/CD28 but not LPS, and lower numbers of peripheral neutrophils (176). Critically, this Il10 transgenic model was limited by the constitutive expression of Il10 in the small intestine, which could affect cellular differentiation and normal flora composition throughout development and in subsequent experimentation.

The existing knock-in mutant IL8 or Il10 mice have been very informative in investigations regarding the function of these important cytokines and particularly with IL8, instrumental in showing proof of principle that the expression of IL8 results in neutrophil influx *in vivo*. The proposed critical involvement of chemokines and Il10 in the *in vivo* action of immunomodulatory host defence peptides, including defensins, would make these models a useful adjunct to attempts to define mechanism *in vivo*. However, these models have significant limitations. For this reason, it was decided here to build on these initial studies to develop new transgenic mice expressing either IL8 or Il10 in a controllable manner throughout the intestinal tract. Figure 3.1 depicts the targeting strategy employed in this work.

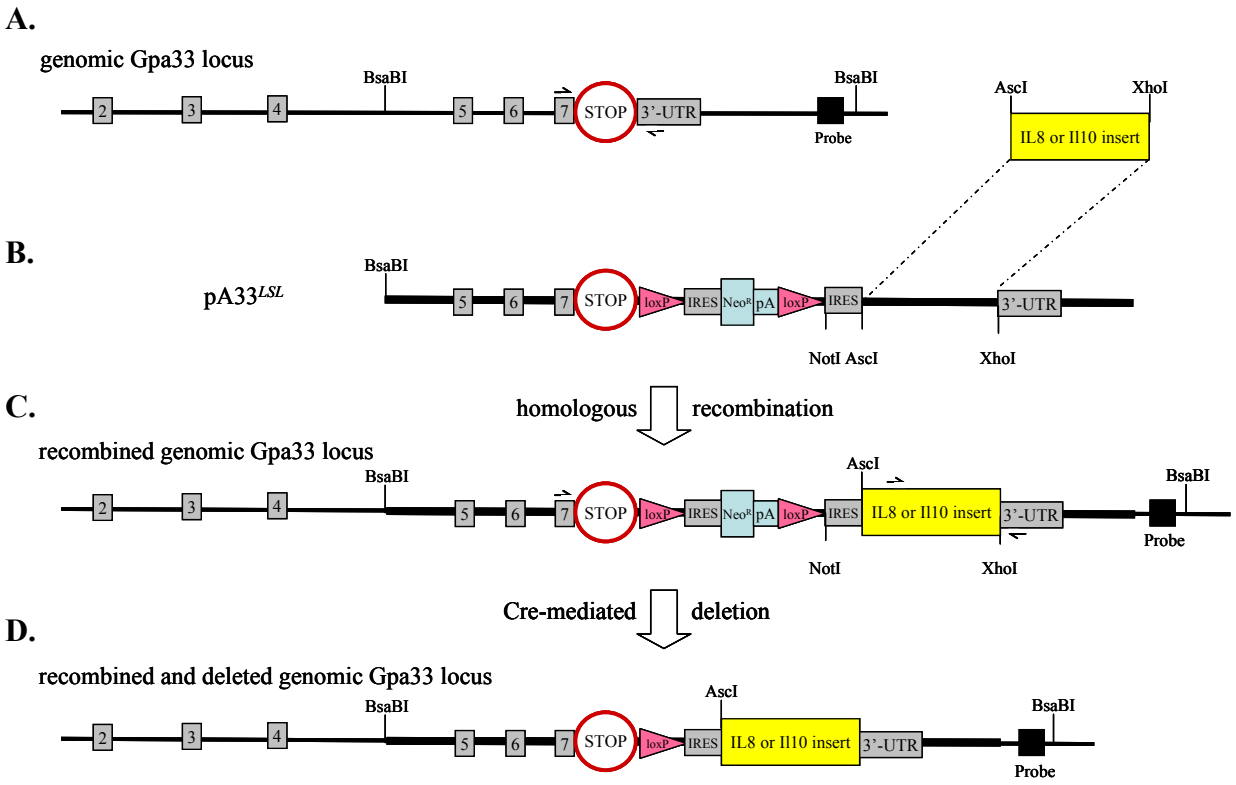


Figure 3.1. pA33LSL targeting vector and genomic locus of recombination. The targeting strategy employed for the generation of mice with conditional, intestinal-specific IL8 or Il10 expression [reproduced and modified with permission from (177)]. The murine *Gpa33* genomic locus (A) is shown from exon 2 to the 3'-UTR. The *BsaB*I-linearized pA33LSL targeting vector (B) contains the 5' and 3' arms of homology with intervening loxP flanked neomycin resistance cassette (Neo^R) and internal ribosome entry site (IRES) to ensure distinct translation from the *Gpa33* mRNA. An additional IRES was cloned into pA33LSL (David Adams, WTSI) to allow direct cloning of the desired insert. Cloning of the *IL8* and *Il10* inserts into pA33LSL produced the final targeting vectors used for transfection of JM8.N4 murine embryonic stem cells. Homologous recombination resulted in the creation of either *IL8* or *Il10* conditional alleles (C) and Cre-mediated deletion of the neomycin resistance cassette (D) allows for bicistronic *Gpa33-IL8* or *Gpa33-Il10* transcription and subsequent translation of the IL8 or Il10. Restriction enzyme sites for insert cloning, *Asc*I/ *Xho*I, and Southern blot confirmation of the targeted locus, *BsaB*I, are indicated. The Southern blot radiolabelled probe would bind to the region boxed in black; the probe position allowed for discrimination between wildtype and recombined alleles.

The pA33LSL vector targets the murine glycoprotein A33 (transmembrane) (*Gpa33*) gene (177) located on Chromosome 1: 168,060,548-168,096,640 (178). Gpa33 is a transmembrane glycoprotein expressed during development in the inner cell mass of the blastocyst and in the endoderm cell layer (179). Strong expression of Gpa33 occurs throughout both the small and large intestine, in the crypts and villi, and elsewhere a very small amount of expression has been detected from stomach and bladder epithelial cells; Gpa33 has been designated a marker for the basolateral intestinal epithelium (179-181). Other intestinal-specific genes, in particular intestinal and liver Fabp, are differentially expressed throughout the small and large intestine as well as in the crypts and villi (181), which is likely the reason for variable expression in the IL8 transgenic mice previously described. Expression of IL8 or Il10 under control of the *Gpa33* promoter is therefore advantageous in the intestinal-specific models developed here. The function of Gpa33 has not been well characterized but it is a member of the immunoglobulin superfamily (179), and as such may play a role in the immune system. Mice deficient in Gpa33 lose the ability to repair chemically-induced intestinal epithelial damage which is concurrent with decreased cellular infiltration of polymorphonuclear cells despite similar mononuclear cell migration (182). It is therefore important that the pA33LSL targeting vector does not to disrupt the *Gpa33* gene upon homologous recombination (177), as this would severely hamper any infection studies carried out with the resulting mice. It was necessary to use the same methodology in the generation of both mouse strains to permit direct comparison of experiments involving them. In addition to probing the role of IL8 and Il10 *in vivo*, these mice will provide models of hyper- and hypo-inflammation, respectively, for testing potential therapeutics.

3.2 Methods and Materials

3.2.1 Human interleukin-8 and mouse interleukin-10 cDNA source

The human IL8 TrueClone™ cDNA plasmid (OriGene, Rockville, MD, USA; catalog number TC119807) contains human *IL8* cDNA within the multiple cloning site (MCS) between the *EcoR*I and *Sal*I restriction sites, as well as an ampicillin resistance gene (β–lactamase) for selection. The plasmid, supplied as 1 mg lyophilized pCMV6-XL5 vector, was resuspended in 15 µl nuclease-free water (NF-H_2O) (Ambion/ Applied Biosystems) to give a concentration of 66 ng/µl, as per the manufacturer's instructions. See Appendix B.1 for the vector map of pCMV6-

XL5 (a), pCMV6-XL5 MCS for *IL8* insertion (b) and the *IL8* NCBI reference cDNA sequence NM_000584.2 with the coding sequence indicated (c).

The mouse cytokine plasmid pUMVC3-mIL10 (Aldevron, Fargo, ND, USA; catalog number 4032) contains mouse *Il10* cDNA, cloned into the MCS of pUMVC3 between *Sal*I and *BamH*I, as well as a kanamycin resistance gene (aminoglycoside 3'-phosphotransferase) for selection. The plasmid, supplied at 3.1 mg/ml in TE buffer, was diluted to a working stock of 10 ng/µl in NF-H_2O. See Appendix B.2 for the vector map of pUMVC3-mIL10 (a) and the *Il10* NCBI reference cDNA sequence NM_010548.1 with the coding sequence indicated (b).

3.2.2 Primer design to engineer IL8 and Il10 inserts for directional cloning and expression detection

Primers were designed to amplify the *IL8* and *Il10* coding sequence from their respective cDNA plasmids, incorporating a 5' FLAG-tag as well as two flanking restriction enzyme sites. *Asc*I (GGCGCGCC) and *Xba*I (TCTAGA) restriction sites were added at the 5' end, and *Xho*I (CTCGA) and *Asc*I restriction sites were added at the 3' end of the coding sequence. The 24 base pair (bp) FLAG-tag sequence, 5'-GACTACAAGGATGACGATGACAAG-3', corresponding to the N-DYKDDDDK-C protein sequence (183), was preceded by a start codon (ATG). The final engineered coding sequences are referred to as *IL8* and *Il10* inserts.

Primer sequences for IL8F-Flag and IL8R-Flag, used to PCR amplify the *IL8* insert, can be found in Table 3.1. Amplification of the *Il10* insert was optimized using the forward primer, IL10F-Flag, paired with five different reverse primers, IL10R-Flag2-6; all primer sequences for *Il10* insert amplification can be found in Table 3.1. The initial IL10R-Flag1 primer was redesigned, as the original sequence obtained from Vega was shorter than that in NCBI's RefSeq; the Il10 sequence was amended in the subsequent Vega (v.29) release. Primer design took into account GC content (50-60%), presence of 3' GC handles (G, C, GC or CG at 3' end) and nearest neighbour melting temperature (T_m), and eliminated repetitive sequences or bases, if possible. An additional base (A) was added to the 5' end of each primer to account for any defect in primer synthesis due to their unusual length. For each primer, the T_m and potential for secondary structure formation were calculated using OligoCalc (119), and specificity against the *Mus musculus* reference assembly genome was verified using NCBI Primer-BLAST (184), with default parameters. Lyophilized primers (Sigma-Genosys) were reconstituted at 100 µM and

diluted to a working stock of 10 μM in NF-H_2O. Annealing temperatures were calculated by subtracting 5°C from the T_m, and then optimizing accordingly, if necessary.

Table 3.1. Primer sequences for the generation and confirmation of IL8 and Il10 knock-in mice at the *Gpa33* genomic locus. * denotes the original reverse primer designed against the *Il10* transcript downloaded from Vega. A longer transcript, which is in agreement with Ensembl v.48, was subsequently downloaded from the Vega v.29 update, therefore IL10R-Flag2-6 were designed. ** denotes these primers and their sequences were provided by the WTSI core sequencing facility.

Primer Name	Sequence (5'-3')
IL8F-Flag	AGGCGCGCCTCTAGAATGGACTACAAGGATGACGATGACAAGATGACTTCCAAGCTGGCCG
IL8R-Flag	AGGCGCGCCCTCGAGTTATGAATTCTCAGCCCTCTTC
IL10F-Flag	AGGCGCGCCTCTAGAATGGACTACAAGGATGACGATGACAAGATGCCTGGCTCAGCACTGC
IL10R-Flag1 *	AGGCGCGCCCTCGAGGCCTGGGGCATCACTTCTACCAGG
IL10R-Flag2	AGGCGCGCCCTCGAGTTAGCTTTTCATTTTGATCATCATGTATGCTTC
IL10R-Flag3	AGGCGCGCCCTCGAGTTAGCTTTTCATTTTGATCATCATGTATGC
IL10R-Flag4	AGGCGCGCCCTCGAGTTAGCTTTTCATTTTGATCATCATG
IL10R-Flag5	AGGCGCGCCCTCGAGTTAGCTTTTCATTTTGATC
IL10R-Flag6	AGGCGCGCCCTCGAGTTAGCTTTTCATTTTG
M13F **	TGTAAAACGACGGCCAGT
M13R **	CAGGAAACAGCTATGACC
A33F	GGACGTGGTTTTCCTTTGAA
A33R	TTTCATTGGAAAGGCTGGTC
IL10-1	GGCCATGCTTCTCTGCCT
IL10-2	CCAGCTGGACAACATACTGCT
IL8-5	GCAGCTCTGTGTGAAGGT
IL8-6	CCTTGGGGTCCAGACAGA
IL8-7	CGCGCCTCTAGAATGGACTAC
IL8-8	GTTATGAATTCTCAGCCCTCTTC
IL8-9	GTGCAGTTTTGCCAAGGAGT
IL8-10	TAATTTCTGTGTTGGCGCAG
A33p-F2	CAGTGTGACCTTGACATGGG
A33p-R2	CTGCATCTTGAAAGGCAACA
NeoF	CTGAATGAACTGCAGGACGA
NeoR	AATATCACGGGTAGCCAACG
A33-E7-F	AGGATCGGTGGAGCTCGGGG
A33-3U-R	AGGCCCTGCCAGCCGTAAGA

3.2.3 PCR amplification of IL8 and Il10 cDNA plasmids to incorporate the FLAG-tag and restriction sites

The *IL8* and *Il10* inserts were amplified from pCMV6-XL5 and pUMVC3-mIL10, respectively, using the Expand High Fidelity PCR System (Roche Applied Science), as per the manufacturer's instructions. The final PCR reaction mix contained 1X Expand High Fidelity buffer containing 1.5 mM $MgCl_2$, 200 μM each dNTP (PCR Nucleotide Mix 10 mM each dNTP, Roche Applied Science), 1 μM each forward and reverse primer, 2.6 U Expand High Fidelity enzyme mix and 20 ng template plasmid in a final volume of 50 μl. Cycling conditions were as follows: an initial denaturation step at 95°C for 2 minutes, followed by 10 cycles of 94°C for 30

seconds, 65°C for 30 seconds, 72°C for 60 seconds, and then 20 cycles of 94°C for 30 seconds, 65°C for 30 seconds, 72°C for 60 seconds adding 5 seconds onto the extension for each successive cycle and a final elongation 72°C for 10 minutes. PCR amplification was performed using a DNA Engine Tetrad 2 Peltier Thermal Cycler (MJ Research).

The *IL8* PCR products (5 µl) were analyzed by agarose gel electrophoresis (0.5% agarose, 100 V, 60 min) to confirm successful amplification. Unless otherwise indicated agarose gels contained ethidum bromide (EtBr) dispensed from a 10 mg/ml stock solution (Sigma-Aldrich) by dipping approximately 0.5 cm of the end of a pipette tip into the solution and then swirling in the melted and slightly cooled agarose. Electrophoresis was performed using the MGU-502T gel tank apparatus (C.B.S. Scientific, Co., CA, USA) in 1X Tris-Acetate-EDTA (40 mM Tris acetate, 1 mM EDTA, TAE) buffer pH 8.0, and 1X TAE agarose gels visualized under ultraviolet (UV) light using the UVItec gel doc system and UVIproMV version 11.03 software (UVItec Limited, Cambridge, UK). The 1 Kb DNA ladder (0.5 µg in 1X BlueJuice® Gel Loading Buffer) (Invitrogen) was run in all agarose gels for size determination. The remaining PCR product (45 µl) was gel extracted using the TOPO® XL gel purification kit (Invitrogen), as per the manufacturer's instructions. Samples prepared in 1X final crystal violet loading buffer were subjected to agarose gel electrophoresis (0.4% agarose containing 90 µl crystal violet, 90 V for 60 min, 60 V for 40 min). Gel slices from the two PCR products were pooled and the PCR product extracted using the QIAquick Gel Extraction Kit (Qiagen), eluting in a final volume of 50 µl Buffer EB (10 mM Tris-Cl, pH 8.5); the purified product (5 µl) was visualized by agarose gel electrophoresis (0.5% agarose, 110 V, 45 min). Small-scale centrifugation steps were carried out in a Micro 240A microfuge (Denville Scientific, Inc., NJ, USA).

The *Il10* PCR products (5 µl) visualized by agarose gel electrophoresis (0.8% agarose, 1X TAE buffer, 100 V, 60 min) to confirm amplification. The PCR product obtained using the IL10F-Flag and IL10R-Flag6 primers gave the cleanest product therefore was purified using the QIAquick gel extraction kit eluting in a final volume of 50 µl Buffer EB; the purified product (2.5 µl) was visualized by agarose gel electrophoresis (1% agarose, 80 V, 90 min).

3.2.4 Cloning of IL8 and Il10 PCR products into pCR®-BluntII-TOPO® and confirmation by PCR, restriction digestion and capillary sequencing

3.2.4.1 IL8 insert TOPO® cloning and confirmation

The purified *IL8* PCR product was cloned into the pCR®-BluntII-TOPO® vector (Invitrogen) as per the manufacturer's instructions; the ligated vector will be referred to as IL8_pCR-BluntII-TOPO, with registered trademark symbols assumed. Ligation reactions performed at room temperature for 20 minutes consisted of 1 µl dilute (1:4) salt solution, 1 µl pCR-BluntII-TOPO vector and either 1, 2 or 4 µl purified PCR product in a final volume of 6 µl made up with NF-H_2O, if necessary; 2 µl of ligated product was then added to 50 µl One Shot® TOP10 Electrocomp™ *E. coli* cells (Invitrogen). The entire volume was added to a 0.1 mm cuvette (Cell Projects Limited, UK) and electroporated using the Bio-Rad Gene Pulser Xcell Electroporation System (Bio-Rad Laboratories, Inc.) with the pre-set *E. coli* bacterial electroporation program (25 µF, 200 Ω, 1800 V, exponential decay). For the recovery, 250 µl super optimal broth with catabolite repression (SOC, Invitrogen) was added and the entire volume transferred to a 15 ml Falcon tube and incubated at 37°C with shaking at 200 rpm, for 160 minutes. A positive control (1 µl pUC19 plasmid, Invitrogen) was also electroporated under the same conditions to ensure TOP10 cells were transformation competent. The *E. coli* cells were diluted 1:10 in SOC and 100 µl each of the neat and diluted cells were spread onto two low salt Luria Bertani (LB-Luria) agar (1.5%) plates containing 50 µg/ml kanamycin (Gibco) for TOPO ligations, and 100 µg/ml ampicillin (Roche Applied Science) for pUC19, and incubated at 37°C overnight. Thirty of resulting colonies were screened using a rapid boil DNA extraction method, colony PCR. Briefly, each colony was added to 10 µl NF-H_2O and heated for 10 minutes at 100°C. The final 28.5 µl reaction contained 5 µl of the boiled colony, 0.9X Platinum PCR Supermix (Invitrogen) and 175 nM each of IL8F-Flag and IL8R-Flag primers; amplification was performed as previously described, and the products (10 µl) visualized by agarose gel electrophoresis (0.5% agarose, 100 V, 60 min).

Ten positive IL8_pCR-BluntII-TOPO clones (1-10) were each inoculated into 25 ml sterile universal tubes (Sarstedt) containing 5 ml LB-Luria broth with 50 µg/ml kanamycin and incubated overnight at 37°C with shaking at 200 rpm. Glycerol stocks (25%) were made by adding 500 µl of the overnight culture to 500 µl of 50% glycerol; high quality plasmid was extracted from the remainder of the overnight culture using the QIAprep® Miniprep kit

(Qiagen), eluting in a final volume of 50 µl Buffer EB. The large-scale centrifugation step to pellet the overnight culture was carried out in a Sorvall Legend RT (Thermo Scientific Inc.) with rotor 6446. The plasmids were quantified using the Thermo Scientific NanoDrop™ ND-1000 Spectrophotometer (Thermo Fisher Scientific Inc.) as per the manufacturer's instructions; unless otherwise indicated all concentrations were determined using the Nanodrop. IL8_pCR-BluntII-TOPO plasmids were subjected to restriction digestion with either *Asc*I or *Eco*RI (New England Biolabs, NEB) for verification of correct insert size. For each final 25 µl enzyme mix, 87-269 ng of plasmid (1 µl) was digested with 10 U of enzyme in the appropriate 1X NEB buffer at 37°C for 60 minutes, followed by 65°C for 20 minutes to heat inactivate the enzymes. Agarose gel electrophoresis (0.5% agarose, 100 V, 70 min) of 15 µl of each digest and 1 µl uncut plasmid for each IL8_pCR-BluntII-TOPO clone was performed to confirm the correct insert size for further sequence confirmation. Negative mastermix controls without plasmid template were also run to ensure no cross-sample contamination.

3.2.4.2 IL8 insert capillary sequencing

Capillary sequencing of the *IL8* inserts was performed by the WTSI core sequencing facility using IL8F-Flag and IL8R-Flag primers, as well as standard M13F and M13R primers (Table 3.1) supplied by the facility. IL8_pCR-BluntII-TOPO-1-10 plasmids were submitted at 100 ng/µl, diluting where necessary with NF-H_2O; IL8F-Flag and IL8R-Flag primers were submitted at 5 µM concentration. Capillary sequencing (Applied Biosystems 3730XL capillary sequencer) was performed using BigDye Terminator BDTv3.1 sequencing chemistry. The sequencing facility processed the sequence trace files using the sequencing production software Asp and then deposited those that passed quality control into a central repository. I performed further analyses of the sequences.

For each IL8_pCR-BluntII-TOPO clone a file of filenames (fofn) was created to include all the names of the sequence traces from the four primers as well as the name of the text file containing the 355 bp sequence of the *IL8* insert. The sequence traces for each clone and the *IL8* insert sequence text file were assembled with the Genome Assembly Program, Gap4, using the normal shotgun assembly function with default parameters (185); a separate Gap4 database was created for each clone. In each database, the *IL8* insert sequence (5'-3') was set as the reference. The assemblies were checked manually for sequence quality and the identification of any discrepancies between the trace sequences and the reference sequence. Any discrepancies

between bases in the reference insert and those within the sequence traces were verified by two or more sequence reads or re-sequenced.

The positive IL8_pCR-BluntII-TOPO-10 plasmid was cultured overnight at 37°C with shaking at 200 rpm in a 250 ml Erlenmeyer flask (Corning, Inc) containing 100 ml LB-Luria broth and 50 μg/ml kanamycin. A larger-scale plasmid preparation was carried out using the Qiagen Midi Plasmid kit, following manufacturer's instructions. The final plasmid pellet was resuspended in 150 μl 10 mM Tris-Cl pH 8.0 and the concentration determined.

3.2.4.3 Il10 insert TOPO® cloning and confirmation

The purified *Il10* PCR product was cloned into the pCR®-BluntII-TOPO® vector as per the manufacturer's instructions; the ligated vector will be referred to as Il10_pCR-BluntII-TOPO, again with registered trademark symbols assumed. The pCR-BluntII-TOPO ligation reactions were set up as using 1 μl and 3 μl volumes of Il10 PCR product in a final volume of 6 μl at room temperature for 15 minutes. A pCR-BluntII-TOPO alone control and Il10 PCR product alone control (3 μl) were also set-up under the same conditions. One shot® TOP10 Electrocomp™ *E. coli* cells were electroporated in a 0.1 mm cuvette using the Bio-Rad pre-set E. coli bacterial program, as previously described. The cells were allowed to recover in 250 μl SOC for 90 minutes at 37°C with shaking at 200 rpm, and then diluted 1:10 in SOC The neat and diluted transformations (100 μl) were then spread onto LB-Luria agar plates containing 25 μg/ml kanamycin and 25 μg/ml zeocin (100 mg/ml stock, Invitrogen). A pUC19 control was also transformed, as previously described.

Fourteen isolated colonies were screened by colony PCR as previously described, except the colony was added straight to the reaction instead of pre-boiling. Final reaction conditions consisted of 1X Platinum PCR Supermix and 200 nM of each M13F and M13R primer in a volume of 25 μl with an annealing temperature of 55°C. The PCR products (5 μl) were visualized by agarose gel electrophoresis (1% agarose, 80 V, 90 min). The same colonies were also inoculated into 25 ml universal tubes containing 5 ml LB-Luria with 50 μg/ml kanamycin and 25 μg/ml zeocin and grown overnight at 37°C with shaking at 200 rpm; 25% glycerol stocks were prepared as previously described. Plasmids were extracted using the QIAprep Miniprep kit eluting in 50 μl Buffer EB and their concentrations determined. The Il10_pCR-BluntII-TOPO plasmids (5 μl 1:10 dilution; 56-169 ng) were digested with *Hind*III and *Xmn*I in a final digest reaction of 25 μl consisting of 1X NEB buffer 2, 1X BSA and 20 U each of *Hind*III and *Xmn*I

(NEB). Digestion was performed for 4 hours at 37°C and fragments visualized by agarose gel electrophoresis (0.8% agarose, 100 V, 40 min); the entire 25 µl digest volume and 1 µl (111.5-338.06 ng) of uncut plasmid was loaded for each Il10_pCR-BluntII-TOPO clone.

3.2.4.4 Il10 insert capillary sequencing

Capillary sequencing of the *Il10* inserts from PCR- and digest-positive plasmids was performed by the WTSI core sequencing facility, as previously described. Il10_pCR-BluntII-TOPO plasmids and the IL10F-Flag, IL10R-Flag6 primers were submitted at 100 ng/µl and 5 µM, respectively. M13F and M13R primers were supplied by the facility. Additional primers, IL10-1 and IL10-2 (Table 3.1), were designed in order to get full coverage of the *Il10* insert. Samples were processed by the WTSI core facility and I assembled the sequences with the 592 bp *Il10* insert sequence text file and performed subsequent analysis, as previously described.

The positive Il10_pCR-BluntII-TOPO-10 plasmid was cultured overnight at 37°C with shaking at 200 rpm in a 250 ml Erlenmeyer flask containing 100 ml LB-Luria broth and 50 µg/ml kanamycin and 25 µg/ml zeocin. A larger-scale plasmid preparation was carried out using the Qiagen Midi Plasmid kit, following manufacturer's instructions. The final plasmid pellet was resuspended in 300 µl Buffer EB and the concentration determined.

3.2.5 Digestion of the pA33LSL embryonic stem cell targeting vector for IL8 and Il10 insert ligation

The pA33LSL targeting vector (Figure 3.1) (177, 186), already transformed into *E. coli*, was kindly provided by David Adams (WTSI). The culture was streaked onto LB-Luria agar containing 100 µg/ml ampicillin and grown overnight at 37°C; a 25% glycerol stock was prepared. An isolated colony was inoculated into a 25 ml universal tube containing 5 ml LB-Luria broth and 100 µg/ml ampicillin and grown overnight at 37°C with shaking at 200 rpm. The pA33LSL plasmid was isolated using the QIAprep Miniprep kit eluting with 50 µl Buffer EB, and the concentration determined. Separate pA33LSL plasmid preparations were used for the *IL8* and *Il10* cloning reactions; inoculation into LB-Luria broth was always with a single colony from a streak plate of the glycerol stock.

For the *Il8* and *Il10* ligations, 0.9-5 µg pA33LSL was digested for 4 hours at 37°C in a final volume of 50 µl containing 1X NEB buffer 4, 25 U of *Asc*I (NEB) and 50 U of *Xho*I (NEB), then dephosphorylated for one hour at 37°C by adding 10 U calf intestinal alkaline phosphatase (NEB) directly to the reaction mix. Subsequently the digested pA33LSL was gel

purified after agarose gel electrophoresis of the digest reaction (*Il8* ligation:1% agarose gel, 110 V, 40 min; *Il10* ligation: 1% agarose, 90 V, 90 min) using the QIAquick gel extraction kit [as per the manufacturer's instructions with the following modification: 750 µl QG buffer was used as the standard volume regardless of gel slice weight]. The final purified plasmids were eluted with 50 µl Buffer EB, and quantified.

3.2.6 Digestion of IL8_pCR-BluntII-TOPO and Il10_pCR-BluntII-TOPO for IL8 and Il10 insert ligation with the pA33LSL targeting vector

3.2.6.1 IL8_pCR-BluntII-TOPO digest

IL8_pCR-BluntII-TOPO-10 (50 µg) was digested for 4 hours at 37°C in a final volume of 75 µl containing 1X NEB buffer 4, 200 U *Asc*I (NEB) and 400 U *Xho*I (NEB). The entire digest was subjected to agarose gel electrophoresis (2% agarose, 100 V, 75 min), and the 351 bp IL8_*Asc*I_*Xho*I insert gel extracted, as previously described, eluting with 50 µl Buffer EB. The purified product was quantified, and 5 µl analyzed by agarose gel electrophoresis (1% agarose, 80 V, 90 min).

3.2.6.2 Il10_pCR-BluntII-TOPO digest

Il10_pCR-BluntII-TOPO-10 (5 µg) was digested for 4 hours at 37°C in a final volume of 50 µl containing 1X NEB buffer 4, 25 U *Asc*I and 50 U *Xho*I. The entire digest was subjected to agarose gel electrophoresis (1% agarose, 90 V, 90 min) and the 588 bp Il10_*Asc*I_*Xho*I insert gel extracted, as previously described, eluting in a final volume of 50 µl EB buffer. The purified product was quantified, and 5 µl visualized by agarose gel electrophoresis (1% agarose, 90 V, 120 min).

3.2.7 Ligation of the AscI- and XhoI-digested pA33LSL vector with each of the AscI- and XhoI-digested IL8 and Il10 inserts

3.2.7.1 IL8_AscI_XhoI insert ligation

The ligation reaction for *IL8* with pA33LSL consisted of 18 ng of digested pA33LSL, 46 ng digested *IL8* insert, 1X T4 DNA ligase reaction buffer and 400 CEU (cohesive end unit) T4 DNA ligase (NEB) in a final volume of 10 µl, and was carried out at 4°C for 4 hours. Vector- and insert-only control ligations were also set-up, using NF-H_2O to make up the volume.

Transformations were performed using 1 µl of each ligation and 20 µl ElectroMAX™ DH10B™ competent *E. coli* (Invitrogen) in a 0.1 mm cuvette using the Bio-Rad pre-set *E. coli*

bacterial program, as previously described. One ml SOC was added to each transformation and 100 µl of undiluted culture plated out onto LB-Luria agar containing 100 µg/ml ampicillin and incubated overnight at 37°C. The rest of the transformation was incubated at room temperature overnight and the following day 200 µl was plated onto three LB-Luria agar plates containing 100 µg/ml ampicillin. Six individual colonies were analyzed for the correct ligation event; each colony was inoculated into a 25 ml universal tube containing 5 ml LB-Luria broth and 100 µg/ml ampicillin and grown overnight at 37°C with shaking at 200 rpm. Plasmids were isolated using the QIAprep Miniprep kit eluting with 50 µl Buffer EB, and their concentrations determined. The presence and insert size was confirmed by PCR amplification of approximately 20 ng plasmid with 1X Platinum PCR Supermix and 200 nM each of A33F and A33R primers (Table 3.1), which amplify the insert from within the pA33LSL vector, in a final volume of 25 µl,. The products were visualized by agarose gel electrophoresis (1% agarose, 100 V, 55 min).

3.2.7.2 Il10_AscI_XhoI insert ligation

The ligation reaction for *Il10* into pA33LSL was carried out at 4°C for 4 hours using 50 ng of digested pA33LSL, 50 ng digested *Il10* insert, 1X T4 DNA ligase reaction buffer and 400 CEU (cohesive end unit) T4 DNA ligase in a final volume of 10 µl. Vector- and insert-only negative control ligations were also performed, making up the extra volume with NF-H_2O.

Each ligation was transformed into DH10B *E. coli* cells made competent in-house. Briefly, the 25% glycerol stock (made from ElectroMAX™ DH10B™ competent cells, Invitrogen) was streaked onto an LB-Luria agar plate containing 50 µg/ml streptomycin (Sigma-Aldrich), as this strain is streptomycin resistant due to a chromosomal mutation in *rpsL150* (187), and incubated overnight at 37°C. An isolated colony was inoculated into two 25 ml universal tubes each containing 5 ml LB-Luria broth and 50 µg/ml streptomycin and incubated overnight at 37°C with shaking at 200 rpm. The next day the 5 ml culture was added to 200 ml LB-Luria broth in a one litre Erlenmeyer flask (Corning, Inc.) and grown at 37°C with shaking at 200 rpm for approximately 2 hours until the *E. coli* reached mid-log phase (optical density approximately 0.5). The culture was split into four 50 ml Falcon tubes and put on ice. All remaining steps were carried out at 4°C and all reagents and the centrifuge were pre-chilled to 4°C. The tubes were centrifuged at 4000 rpm for 10 minutes. The supernatant was poured off and each pellet resuspended in 1 ml MilliQ H_2O (Synthesis A10, Millipore) and then topped up to 50 ml with MilliQ H_2O, and centrifuged at 4000 rpm for 10 minutes. The pellets were resuspended in 1 ml MilliQ H_2O and transferred to 2 ml Eppendorf tubes. The tubes were centrifuged at

13000 rpm for 2 minutes and the cells washed five times with MilliQ H_2O; the first wash was resuspended in 1 ml, the second in 0.5 ml (two tubes pooled), the third in 1 ml, the forth in 0.5 ml (two tubes pooled) and the fifth in 1 ml, giving one tube containing 1 ml cells. After a final spin at 13000 rpm for 2 minutes, the cells were resuspended in 100 µl MilliQ H_2O.

Transformations were performed immediately using 50 µl competent cells and 1 µl ligation reaction in 0.1 mm cuvettes and the Bio-Rad pre-set *E. coli* bacterial program, as previously described. One ml SOC was added and the cells recovered at 37°C with shaking at 200 rpm for 2 hours. The cells were plated onto LB-Luria plates containing 100 µg/ml ampicillin, two with 100 µl neat culture and one with 200 µl neat culture, and incubated overnight at 37°C. One colony was tested using the rapid boil DNA extraction method with PCR amplification using IL10F-Flag and IL10R-Flag6 primers, as previously described. Plasmids were then extracted from a further six colonies, using the QIAprep Miniprep kit, and inserts confirmed by PCR amplification using 1X Platinum PCR Supermix and A33F/ A33R and IL10F-Flag/ IL10F-Flag6 primer pairs, as previously described; products were visualized by agarose gel electrophoresis (0.8% agarose, 90 V, 45 min).

3.2.8 Capillary sequencing of the IL8_pA33LSL and Il10_pA33LSL targeting vectors

PCR-positive IL8_pA33LSL and Il10_pA33LSL plasmids were capillary sequenced by the WTSI core facility, as previously described, for confirmation of the insert sequences of the final murine embryonic stem cell (ESC) targeting vectors. Plasmids were submitted at 100 ng/µl and primers at 5 µM. IL8_pA33LSL plasmids were sequencing using IL8F-Flag, IL8R-Flag, as well as additional IL8 insert-specific IL8-5, IL8-6, IL8-7, IL8-8, IL8-9 and IL8-10 primers (Table 3.1), and Il10_pA33LSL plasmids were sequencing using IL10F-Flag, IL10R-Flag6, IL10-1 and IL10-2 primers. The additional primers circumvented difficulties encountered obtaining full sequence coverage of the inserts. I performed all subsequent sequence analyses using Gap4 assembly and trace viewing software, as previously described.

3.2.9 Linearization and preparation of the IL8 and Il10 final targeting vectors for embryonic stem cell transfection

Cultures of the positive IL8_pA33LSL-1 and Il10_pA33LSL-1 DH10B clones were grown overnight from single colonies at 37°C with 200 rpm shaking in one litre Erlenmeyer flasks containing 200 ml of LB-Luria broth and 100 µg/ml ampicillin. Plasmids were isolated

using the Qiagen Plasmid Midi kit, following the manufacturer's instructions, and plasmid pellets resuspended in 100 µl NF-H_2O. The concentration of a 1:10 dilution of each plasmid was determined, from which the concentration of the undiluted plasmid was extrapolated.

Plasmid DNA was prepared prior to ESC electroporation by linearization with *Sal*I restriction enzyme (NEB). In a final volume of 100 µl, 100 µg of vector was digested with 200 U of *Sal*I in 1X NEB buffer 3 and 1X BSA at 37°C overnight, and then inactivated at 65°C for 20 minutes. The linear vectors were purified by adding 2.5 times the volume of 100% EtOH (250 µl), vortexing and incubating on ice for 5 minutes. The precipitate was centrifuged at 14000 rpm for two minutes at room temperature and the supernatant discarded. The resulting pellet was washed twice with 70% EtOH (750 µl each), centrifuged at 14000 rpm for two minutes between washings and the supernatant removed. The pellets were resuspended in 100 µl of sterile tissue-culture grade PBS (Invitrogen) at 4°C overnight, giving a final vector concentration of one µg/µl. Digests and uncut plasmids (1 µg) were visualized by agarose gel electrophoresis (0.6% agarose, 90V, 60 min).

3.2.10 JM8.N4 murine embryonic stem cell culture reagents

A 100X β-mercaptoethanol (BME) solution was prepared by adding 72 µl β-mercaptoethanol stock (Sigma-Aldrich) to 100 ml Dulbecco's phosphate buffered saline (D-PBS) without magnesium chloride or calcium chloride (Invitrogen) and filter sterilizing (0.2µm); BME was stored at 4°C and used within two weeks of preparation.

A 1X trypsin solution was prepared by adding 0.1 g EDTA (Sigma-Aldrich) to 500 ml D-PBS and filter sterilizing (0.2 µm). To this 5 ml chicken serum (Invitrogen) and 10 ml 2.5% trypsin (Invitrogen) were added. 2X trypsin was prepared by adding 4 ml 2.5% trypsin to 200 ml of 1X solution. Aliquots of both 1X and 2X trypsin were stored at -20ºC, and then at 4ºC once thawed. Trypsin solutions and aliquots were prepared by Theodore Whipp (WTSI).

A glutamine, penicillin and streptomycin (GPS) mix was prepared by adding 58.4 g glutamine (Amresco), 10 g streptomycin, 6 g penicillin (Sigma-Aldrich) to two litres of MilliQ water, and filter sterilized (0.2µm). Single-use aliquots (6 ml) were stored -20ºC and thawed just prior to media preparation. GPS solution and aliquots prepared by Theodore Whipp (WTSI).

0.1% Gelatin was prepared by adding 25 ml of 2% gelatin (Sigma-Aldrich) to 500 ml D-PBS and stored at 4°C.

M15 media contains 500 ml Knockout DMEM (Invitrogen), 90 ml fetal bovine serum (Invitrogen), 6 ml GPS, 6 ml 100X β-mercaptoethanol and 120 µl leukocyte inhibitory factor (LIF); LIF prepared in-house by Theodore Whipp (WTSI) using a modified protocol obtained from Janet Rossant (Hospital for Sick Children, Toronto, Canada). M15 was stored at 4°C and used within 2-3 weeks of preparation.

Freezing medium was prepared with 90% M15 media and 10% DMSO (Sigma-Aldrich) and filter sterilized (0.2µm); freezing media was stored at 4°C and used within 48 hours.

3.2.11 JM8.N4 murine embryonic cell line culture

The JM8.N4 mouse ESC line was kindly provided by David Adams (WTSI). Throughout the entire procedure, feeder-free JM8.N4 ESC were cultured on 0.1% gelatin at 37°C with 5% CO_2 in M15 media. For selection steps, JM8.N4 cells were cultured in M15 media containing 175 µg/ml Geneticin or 3 µg/ml puromycin (Sigma-Aldrich), as indicated. All centrifugations for ESC work performed in an Eppendorf 5702 microfuge as indicated.

3.2.12 JM8.N4 ESC transfection and selection

Prior to electroporation, the JM8.N4 cells were cultured in a T150 flask (Corning, Inc.) until 90% confluent, washed once with D-PBS and trypsinized with 5 ml of 1X trypsin for 5 minutes at 37°C and 5% CO_2, after which time 15 ml of M15 media was added to neutralize the trypsin. A volume of 6 ml of JM8.N4 cells, corresponding to $1.75x10^7$ cells, was added to each of two 25 ml universal tubes, which were centrifuged at 1200 rpm for 3 minutes. The supernatant was aspirated and each pellet resuspended in 700 µl D-PBS. For each vector, 20 µl (i.e. 20 µg) of the linear plasmid was added to a new 1.5 ml Eppendorf tube to which the 700 µl cell suspension was added. The cells/ DNA were mixed twice with a P1000 pipettor and the entire contents added to a 0.4 mm cuvette (Bio-Rad). The mixture was electroporated at 800 V, 25 µF with infinite resistance and then incubated at room temperature for 10 minutes.

The cells were washed out of the cuvette with 2 ml of selection media (M15 with Geneticin) and the entire volume added to a 50 ml Falcon tube containing 47 ml selection media. The cuvette was rinsed with another 1 ml of selection media, which was added to the Falcon tube bringing the final volume to approximately 50 ml. Following thorough mixing by inversion, 10 ml was added to each of five 10 cm tissue culture (TC) dishes (Corning, Inc.); the dishes were swirled to ensure even distribution of cells upon settling and adherence. The final number of

cells per dish was approximately $3.5x10^6$ million. Media changes were made daily for the entire duration of the Geneticin selection period.

3.2.13 Colony picking and clone expansion

Following 10 days of Geneticin selection, colonies of JM8.N4 ESC, presumably undergone homologous recombination and clonal expansion were distinctly visible to the naked eye. The media was aspirated from the dishes and replaced with 1X D-PBS. Well-spaced colonies were first identified by eye and then their quality checked under an Olympus TH4-20 microscope to ensure even, uniform edges. Each colony was removed from the dish in one motion by sliding a pipette tip underneath it and then gently sucking up with 54 µl (to account for the volume of the colony) D-PBS without breaking up the colony. The entire volume including the colony was ejected into one well of a pre-gelatinized 48-well plate (BD Falcon). This picking procedure was repeated to fill two 48-well plates, giving 96 colonies for each targeting construct.

To each colony 50 µl 2X trypsin was added and the plates incubated at 37°C with 5% CO_2 for 7 minutes. M15 media, without Geneticin selection, was added to fill each well approximately three-quarters full and mixed thoroughly (6-8 times) with a P1000 pipettor, changing tips between each well; the plate was then returned to the incubator. The trypsinizing procedure was repeated with the remaining three 48-well plates.

The media was changed the following day (Day 11) and everyday thereafter until Day 14, upon which time the majority of clones were 60-90% confluent. Media was aspirated and the cells washed once with D-PBS. The cells were trypsinized using 100 µl of 1X trypsin and incubated at 37°C with 5% CO_2 for 4 minutes in the incubator. M15 media was added to fill the wells three-quarters full, and each well mixed by pipetting up and down 5-6 times with a P1000 pipettor. The volume of cells/ media was approximately split between two 24-well plates (BD Falcon), one for freezing (F1-8) and the other for DNA extraction (DNA1-8). Where necessary, additional M15 media was added so all wells were approximately three-quarters full. The media was changed every day until most of the wells were confluent enough for freezing (80-90%) or for extracting DNA (~100%). On Day 17, plates F1-8 were frozen since the majority were ~90% confluent; some clones were less confluent but if the plate was left any longer the others would have become overgrown and risk the potential of losing their pleuripotency. The media was aspirated and the cells washed once with D-PBS. A volume of 100 µl of 1X trypsin was added to

each well and incubated at 37°C with 5% CO_2 for 4 minutes in the incubator. Freezing media (500 µl) was added to each well and the plates sealed and stored at -80°C until positive clones were identified by Southern blotting and PCR of the targeted allele.

The DNA1-8 plates were cultured until Day 19 when the maximum confluency for any one clone was ~100%, upon which time the DNA was extracted.

3.2.14 Embryonic stem cell DNA extraction and quantification

The media was aspirated from plates DNA1-8 and the cells washed once with D-PBS, which was then flicked out of the plates. A volume of 500 µl of tail extraction buffer, consisting of 50 mM Tris pH 8, 50 mM EDTA pH 8, 0.5% sodium dodecyl sulphate (Ultrapure 10% SDS Solution, Gibco) containing 1 mg/ml fungal proteinase K (Invitrogen) was added to each well and the plates incubated at room temperature with shaking at 40 rpm for 5 minutes to detach the cells. Cell lysates were added to 1.5 ml Eppendorf tubes and incubated for 22 hours at 55°C and then stored at 4°C for 24 hours. The entire lysate was added to 900 µl of 100% EtOH and mixed by inversion. The DNA was swirled onto a glass rod (sealed capillary tube made of soda glass, Samco), washed by dunking in 500 µl 70% EtOH and then in 500 µl 100% EtOH, and inverted to dry for one hour at room temperature. The DNA was resuspended in 150 µl 10mM Tris-HCl pH 8.0 for three days at 4°C, and then stored at -20°C.

The samples were thawed at room temperature; in some tubes the DNA had not resuspended so all samples were incubated 55°C for an additional two hours. The DNA concentrations were determined, and then the samples were stored at 4°C to avoid freeze-thaw cycles until the restriction digest was performed.

Wild type (WT) JM8.N4 DNA was extracted as described above for probe generation and Southern blot WT control.

3.2.15 Restriction digestion of ESC DNA for Southern blotting

The genomic sequence surrounding the *Gpa33* targeting locus (Exon 4 to 3'-UTR plus 6260 bp; Chromosome 1:168087644-168102900 bp) was downloaded through Ensembl v.51 (NCBIM37 Assembly). Appendix B.3 is a screenshot of this region taken from Ensembl v.51. The genomic sequence (Appendix B.4) was subjected to *in silico* digestion using NEBcutter (188) to identify a restriction enzyme(s) that could be used to differentiate between WT and targeted knock-in (KI) allele by Southern blotting. ESC DNA (10 µg) was digested with 20 U

*BsaB*I in 1X NEB buffer 4 in a final volume of 50 µl for 16.5 hours at 60°C, followed by a 20 minute heat inactivation at 80°C. JM8.N4 WT DNA was also digested under the same conditions to serve as the positive control (+/+ genotype) and the *BsaB*I master mix alone served as the negative control. The human IL8 KI and mouse Il10 KI alleles are denoted as hIL8:c and mIL10:c, respectively, indicating that the KI allele is conditional; human (h) and mouse (m) designations are made for the alleles so that the species origin of the gene is clear on the WTSI Mouse Database.

3.2.16 Gel electrophoresis and transfer for Southern blot analysis

*BsaB*I-digested DNA (10 µg) was subjected to gel electrophoresis (0.6% agarose gel, 60 V, 6 hours, 0.25 mg/ml EtBr). The gel was washed twice, with gentle agitation, in one litre depurination buffer (0.242 M HCl) for 10 minutes each, and once in one litre neutralization buffer (0.4M NaOH, 1M NaCl) for 15 minutes, and then transferred overnight (14.5 hours) to positively charged nylon membrane (Hybond-XL, Amersham/ GE Healthcare) by capillary action. The membrane was agitated in wash buffer (0.5M Tris-HCl, 1M NaCl) for 15 minutes and then baked for three hours at 80°C between Whatman filter paper.

3.2.17 Southern blot probe labelling and hybridization

The transferred and baked Hybond-XL nylon membrane was pre-hybridized in a large glass bottle (300 X 35 mm, Thermo Scientific) containing 20 ml Rapid-hyb™ buffer (Amersham/ GE Healthcare) at 65°C for 2 hours rotating in a Hybaid Shake and Stack hybridization chamber (Thermo Scientific). One aliquot of the A33 probe (described below) was labelled with α-^{32}P-dCTP (Redivue deoxycytidine 5'-[A32p] triphosphate, triethylammonium salt 3000Ci/mmol 10mCi/ml EasyTide, 250 µCi; Perkin Elmer) using the Prime-It® II Random Primer Labeling Kit (Agilent Technologies, Stratagene Products) according to the manufacturer's instructions. Briefly, 10 µl of Random 9-mer primers was added to the 25 µl probe aliquot (50 ng), boiled at 100ºC for 5 minutes and then placed on ice for 1 minute. On ice, 10 µL of 5X dCTP buffer (containing 0.1 mM of each of dATP, dGTP and dTTP) and 1 µL of Klenow polymerase were added. In the radiation room, 5µl (50 µCi) of αP^{32}–dCTP was added and incubated at 37ºC for 1 hour 22 minutes. The probe was then cleaned through an Illustra ProbeQuant™ G-50 Micro column (GE Healthcare) to remove unincorporated nucleotides, primers and the Klenow polymerase. The column was prepared by snapping off the bottom and

spinning in a collection tube for 2 minutes at 4000 rpm. The column was placed in a new 1.5 ml tube and add the entire probe reaction added to the gel resin in the column and spun again for 2 minutes at 4000 rpm. The cleaned probe (eluant) was denatured for 5 minutes at 100ºC and then placed on ice.

The entire labelled probe was then added to the pre-hybridized membrane and hybridized by rotating at 65°C for 5 hours and 15 minutes. Low stringency wash (LSW) and high stringency wash (HSW) buffers were pre-warmed to 65°C. Following hybridization the membrane was washed by rotation for 10 minutes with each of LSW (2X SSC, 0.1% SDS) and HSW (0.5X SSC, 0.1% SDS) buffers at 65°C. The membrane was wrapped in cling film and exposed to Hyperfilm MP (Amersham/ GE Healthcare) at -80°C for 15 hours 15 minutes. The film was developed using a Compact X4 Automatic X-ray Film Processor (Xograph Healthcare Ltd, UK).

3.2.18 Southern blot probe design and amplification

The genomic sequence downloaded for the *Gpa33* targeting locus encompassed 6260 bp past the 3' UTR, and subsequently past the end of the 3' arm of homology of pA33LSL, to ensure that the Southern blot probe hybridized to the genomic sequence outside the region of homologous recombination. This was to detect both and distinguish between WT and KI alleles using the same probe. Within the genomic sequence, exons 4-7, 3'-UTR, 5' and 3' arms of homology, repeat elements and *BsaB*I restriction sites were identified (Appendix B.4). The genomic sequence following the 3'-UTR contains four repeat elements (Type 1 Transposons) ranging from 55-157 bp in length. The sequence after this was free of repeats for a long enough region to design primers to amplify a probe optimal for Southern blot hybridization (500-1000 bp). The primer sequences for A33pF2 and A33pR2 can be found in Table 3.1 and the probe amplicon is highlighted in Appendix B.4. The T_m and potential for secondary structure formation for each primer were calculated using OligoCalc (119) and specificity against the *Mus musculus* genomic reference assembly determined using NCBI Primer-BLAST (184).

The A33 probe (A33p) was amplified from 50 ng WT JM8.N4 DNA in a final volume of 50 µl under the following final reaction conditions: 1X Pfx reaction buffer, 300 µM dNTPs, 300 nM each of A33pF2 and A33pR2 primers, 5 U Pfx50 DNA polymerase (Invitrogen). The cycling conditions were as follows: 94ºC for 2 minutes, 35 cycles of 94ºC for 15 seconds, 58ºC for 30 seconds, 68ºC for 30 seconds, with a final extension at 68ºC for 5 minutes. The PCR product (10 µl) was analyzed by agarose gel electrophoresis (0.8% agarose, 90 V, 55 min), and purified using

the QIAquick PCR purification kit (Qiagen), according to the manufacturer's instructions, and eluted using 50 µl Buffer EB. The pure product (1 µl) was then ligated into pCR-BluntII-TOPO for 20 minutes, as previously described, and then 2 µl of a 1/10 dilution transformed into 50 µl electrocompetent TOP10 *E. coli* cells, as previously described. A vector-only negative control was also ligated and transformed under the same conditions. SOC (250 µl) was added and the transformation allowed to recover at 37°C with shaking at 200 rpm for one hour, after which time 20, 50 and 100 µl were plated onto LB-Luria agar with 50 µg/ml kanamycin and incubated overnight at 37°C. Twelve individual colonies were each inoculated into 25 ml universal tubes containing 5 ml LB-Luria and 50 µg/ml kanamycin and grown overnight at 37°C, with shaking at 200 rpm. The plasmids were extracted using the QIAprep Miniprep kit, according to the manufacturer's instructions, eluted in 50 µl Buffer EB, and quantified.

The A33p_pCR-BluntII-TOPO plasmid insert size was confirmed by restriction digestion using *BamH*I and *EcoR*V, which both cut once in the pCR-BluntII-TOPO MCS to release the insert, as well as *EcoR*I, which cuts twice in the pCR-BluntII-TOPO MCS to release the insert and also cuts once within the A33p amplicon. Both restriction digests were carried out using 1 µg of each A33p_pCR-BluntII-TOPO plasmid in a final volume of 50 µl for 16.5 hours at 37°C, followed by 65°C for 20 minutes to inactivate the enzymes. The *BamH*I/ *EcoR*V double digest reaction consisted of 1X NEB buffer 3, 0.1 mg/ml BSA and 10 U each of *BamH*I and *EcoR*V (NEB); the *EcoR*I single digest consisted of 1X NEB *EcoR*I buffer and 20 U of *EcoR*I (NEB). The digests (10 µl) and 1 µl of their respective parent A33p_pCR-BluntII-TOPO plasmids (54-305 ng) were subjected to agarose gel electrophoresis (0.8% agarose, 90 V, 60 minutes).

The A33p_pCR-BluntII-TOPO plasmid insert and therefore probe identity was confirmed by PCR amplification using 1X Platinum PCR Supermix, 300 nM each of A33pF2 and A33pR2 primers and 5-30 ng plasmid in a final volume of 50 µl. WT JM8.N4 (25 ng) was included as the positive control. Amplification was carried out, as above for A33p TOPO cloning using Pfx50 DNA polymerase, except the annealing temperature was decreased to 55°C. Products (10 µl) were visualized by agarose gel electrophoresis (0.8% agarose, 80 V, 120 min).

As a final confirmation of primer specificity, the A33p_pCR-BluntII-TOPO plasmids confirmed to be restriction digest- and PCR-positive were capillary sequenced using A33pF2 and A33pR2 primers as well as standard M13F and M13R primers by the WTSI core sequencing facility. The plasmids and primers were diluted to 100 ng/µl and 5 µM, respectively, in NF-H_2O,

prior to sample submission. I performed all subsequent sequence analysis using Gap4 assembly and trace viewing software, as previously described.

A33p PCR products amplified from seven A33p_pCR-BluntII-TOPO positive plasmids using the A33pF2 and A33pR2 primers and Pfx50 DNA polymerase, as previously described for TOPO cloning, except 100-600 pg of plasmid was used as template. The resulting PCR products were column purified using the QIAquick PCR purification kit, as per the manufacturer's instructions, with a final elution of 50 µl Buffer EB, to remove unincorporated nucleotides, primers and Pfx50 DNA polymerase. Each purified product (5 µl) was visualized by agarose gel electrophoresis (0.8% agarose, 80 V, 90 or 120 min). A volume of 5 µl was pooled from each reaction and the pool concentration determined. The A33p PCR pool was diluted to 2 ng/µl with NF-H_2O and aliquots of 25 µl made for a final amount of 50 ng/ aliquot, the amount required for labelling in the Southern blot analysis.

3.2.19 PCR confirmation of Southern blot-positive clones

Clones determined to be positive by Southern blot analysis were further confirmed by PCR amplification using several primer combinations for maximum confidence in the targeting process. For each gene construct, insert-specific primers were used to amplify the *IL8* insert (IL8-7/ IL8-8) or the *Il10* insert (IL10F-Flag/ IL10R-Flag6), as well as the neomycin resistance cassette (aminoglycoside phosphotransferase) using NeoF/ NeoR primers (Table 3.1). The NeoF and insert-specific forward primers were each paired with the reverse A33p primer, A33pR2 (Table 3.1), which is downstream of the 3' arm of homology. Due to the nature of the targeting vector, design of a probe upstream of the 5' arm of homology is difficult to the necessity of a very long-range PCR. The 5' arm of homology is 4979 bp and ends within intron 4-5 of *Gpa33*. To ensure specificity the primer would have to be positioned within exon 4, the start of which is 7767 bp away from the conditional allele insertion site. The final 50 µl reactions consisted of 1X High Fidelity PCR buffer, 200 µM dNTPs, 2 mM $MgSO_4$, 200 nM each of forward and reverse primers, 1 U Platinum Taq High Fidelity DNA polymerase (Invitrogen) and 10 or 25 ng genomic DNA. The amplification conditions were as follows: 94°C for 2 minutes, 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds, 68°C for 6.5 minutes, a final extension of 68°C for 10 minutes, followed by 4°C. Products (5 µl) were analyzed by agarose gel electrophoresis (0.6% agarose, 90V, 1-2 hours).

3.2.20 Expansion of correctly targeted ESC clones

For each gene construct, ESC clones identified as positive by Southern blot and PCR were cultured as potential candidates for microinjection, as previously described. Frozen stocks were thawed and cells cultured in 6-well plates (BD Falcon) for 10 days prior to the microinjection target date. M15 media was changed daily and upon reaching ~80% confluency, every second day, cells were passaged at a 1/5 dilution. Cells were also passaged the day before microinjection at two different dilutions, regardless of the schedule, to ensure maximum cell health on the day of microinjection.

3.2.21 ESC clone microinjection and chimera F0 mouse generation

On the morning of microinjection the best of the positive clones for each gene construct (*IL8* or *Il10*) was chosen based on visual inspection by microscopy. This was judged by percentage confluency (70-80%), clone morphology (evenly-sized and spaced), and the least amount of dead or differentiating cells.

Media was aspirated from the chosen clone and the cells washed once with D-PBS. Trypsin (1X) was added to just cover the bottom of the well (<500 µl) and the cells were incubated for 4 minutes at 37°C with 5% CO_2. M15 media was added and the entire volume transferred to a 25 ml universal tube, topped up to 10 ml with M15 and pipetted well to mix. The tube was centrifuged for 4 minutes at 1500 rpm and the supernatant aspirated. The ESC pellet was resuspended in 1ml M15: HEPES solution (sterile-filtered cell culture grade, Sigma-Aldrich) (50:1) and the cells mixed well. An aliquot of 500 µl of the cell suspension was added to a cryovial and placed immediately on ice. The remainder of the cells were added to one well of a 6-well plate and checked under an Olympus CKX41 microscope.

The cryovial was submitted to the WTSI Research Support Facility (RSF), and Team 121 (Generating Mouse Mutants) performed the microinjection and surgical implantation; I was allowed to observe the procedures but it was not on a day that any of my constructs were microinjected. See Appendix B.5 for my observations. Briefly, ESC were microinjected into 30-40 blastocysts (5-15 ESC/ blastocyst); the microinjected blastocysts were then surgically implanted into pseudo-pregnant recipient females (3-4/ MI depending on total number of embryos microinjected).

3.2.22 Cre plasmid transfection for functional analysis of the targeted ESC loxP sites

A Cre-expressing plasmid containing puromycin selection was kindly provided by Kosuke Yusa and Wei X. Wang (WTSI). The plasmid was transformed into *E. coli* TOP10 electrocompetent cells, plated onto LB-Luria agar and incubated overnight at 37°C, as previously described. An isolated colony was inoculated into a 250 ml Erlenmeyer flask containing 50 ml LB-Luria broth and cultured overnight at 37°C, with shaking at 200 rpm. The Cre-puromycin plasmid was purified using the Qiagen Plasmid Midi kit, as per the manufacturer's instructions. The plasmid was resuspended in 150 µl 10 mM Tris-HCl pH 8.0, quantified and diluted to one µg/µl in NF-H_2O.

The first set of microinjected ESC clones were validated for functional Cre-mediated neomycin resistance cassette deletion. (i.e. IL8_Gpa33_JM8N4-12 and Il10_Gpa33_JM8N4-24). ESC clones were thawed and cultured in T25 flasks (Corning, Inc.). To prepare the ESC for transfection, old M15 media was aspirated from the T25 flasks and the cells washed once with D-PBS. One ml of 1X trypsin was added to each flask and incubated for 4 minutes at 37°C with 5% CO_2. Fresh M15 (9 ml) was added to each flask and the entire contents transferred to a 25 ml universal tube and mixed to create a single cell suspension. For each gene construct, two transfections were performed with $3X10^6$ ESC resuspended in 700 µl D-PBS and 20 µg Cre-puromycin plasmid. ESC were added to 20 µl pre-aliquoted plasmid, mixed twice and transferred to a 0.4 cm cuvette and electroporated with 230 V, 500 µF and infinite resistance. After 20 minutes recovery, ESC from each transfection were plated out onto a pre-gelatinized 10 cm TC dish in selection-free M15 media. The following day puromycin selection was started and continued for nine days, changing the media daily. On day 10 after transfection, 48 colonies were picked for each gene construct into 48-well plates, as previously described. These clones were expanded in selection-free media until 80-90% confluent and then split in half into 24-well plates. These plates were grown again until approximately 90% confluent and then one plate frozen and DNA extracted from the other for PCR confirmation of the Cre-mediated neomycin resistance cassette deletion. Positive clones were thawed, expanded for one passage and then frozen in one ml freezing media in liquid nitrogen for archiving. Clone DNA (25 ng) from each insert were tested by amplification using Platinum PCR Supermix, as previously described, with gene-specific or neomycin resistance cassette primers; the parent ESC clone (25 ng), WT JM8.N4 (25 ng) were included as controls. The PCR products (5 µl) were analyzed by agarose gel electrophoresis (1% agarose, 90 V, 45 min).

3.2.23 Chimera percentage determination and F0 matings

The genetic background of the embryos (blastocyst) used for microinjection is C57BL/6J-Tyr<c-Brd> (100%) (prefix CALB) and that of the recipient pseudo-pregnant mice is CBA (50%) and C57BL/6J (50%) (F1) (prefix BXCB). Both CALB and BXCB strains are albino whereas JM8.N4 ESC are derived from a black C57BL/6N strain; this allows resulting F0 chimeras to be genotyped by coat colour expressed as a percentage, which was performed by RSF staff. Chimera F0 mice were then mated with C57BL/6N Taconic USA (prefix CBLN) wild type mice and subsequent F1 pups genotyped for germline transmission of respective knock-in allele by PCR.

All mouse husbandry and matings performed by staff of the WTSI RSF.

3.2.24 Genotyping F1 offspring to detect germline transmission of hIL8:c or mIL10:c alleles

Ear clips from F1 pups were taken by WTSI RSF staff and stored at -20°C until DNA extraction. I performed subsequent processes.

DNA was extracted from ear clips using the DNeasy Blood & Tissue Kit (Qiagen), as per the manufacturer's instructions, with the following specifications: Proteinase K treatment was carried out for at least 5 hours and up to approximately 16 hours (overnight) at 55°C, DNA was eluted twice into the same tube with 100 µl Buffer AE, and the concentration determined.

DNA from the first F1 litters (LRHE5.1, 6.1, 7.1, 8.1 and LRIT6.1) was amplified using the neomycin resistance cassette primers (NeoF/ NeoR) as well as *IL8* or *Il10* gene-specific primers (IL8-7/ IL8-8 or IL10F-Flag/ IL10R-Flag6). A final reaction volume of 25 µl consisting of 1X Platinum PCR Supermix (Invitrogen), 200 mM each forward and reverse primer and 12.5 ng DNA was subjected to the following amplification conditions: 94°C for 2 minutes, 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds, a final extension of 72°C for 5 minutes. Amplicons (5 µl) were visualized by agarose gel electrophoresis (1% agarose, 90 V, 45 min).

F1 litters LRHE5.2 and LRHE8.2 were genotyped using a multiplex PCR to amplify both the WT and hIL8:c alleles in one reaction, which was confirmed using the neomycin resistance cassette and *IL8*-specific primers, as described above. The multiplex PCR contains a forward primer within exon 7 of *Gpa33* (A33-E7) and a reverse primer within the 3'-UTR of Gpa33

(A33-3U) (Table 3.1), as well either a IL8- or Il10-specific forward primer. A final reaction of 25 µl consisting of 1X Platinum PCR Supermix (Invitrogen), 100 mM each of A33-E7 and gene-specific forward primers, 200 mM A33-3U reverse primer and 2 µl undiluted DNA (28-60 ng) using the same amplification conditions as above. The PCR products (5 µl) were visualized by agarose gel electrophoresis (1% agarose, 90 V, 40 min). WT JM8.N4 and Southern blot-confirmed targeted IL8_Gpa33_JM8N4 ESC DNA were included as heterozygous and wild type controls.

All subsequent LRHE, LRIT and LRMT litters were genotyped using the multiplex PCR to discriminate between wild type, heterozygous and homozygous mice. The PCR conditions were further standardized and optimized. The DNA was diluted to 10 ng/µl for a final amount of 20 ng in each PCR reaction, the annealing temperature was increased to 57°C and the number of cycles decreased to 30.

3.2.25 F1 pairings for colony expansion and homozygous mouse generation

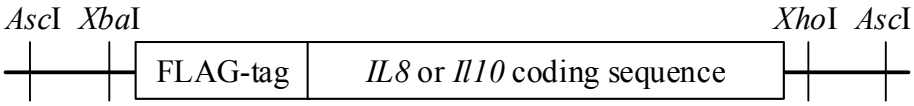
F1 wild type mice that did not have transmission of the hIL8:c or mIL10:c allele were culled by staff of the RSF. F1 hIL8:c/+ heterozygous mice were paired, and the resulting F2 offspring genotyped, as previously described, for the generation of homozygous mice, hIL8:c/-. The homozygous genotype is denoted as hIL8:c/-, indicating that the mice carry two copies of the conditional allele and no WT allele. Homozygous pairings were set up. Once these matings have produced 2-3 litters they will be paired with Cre-expressing mice.

3.3 Results

3.3.1 PCR amplification of engineered IL8 and Il10 inserts from cDNA plasmids

The *IL8* and *Il10* coding sequences were successfully amplified from their respective cDNA plasmids using the primers designed to incorporate two restriction sites at both the 5' and 3' ends, as well as a 5' FLAG-tag. Figure 3.2 depicts the final gene inserts with relative positions of the modifications indicated (A), the *IL8* coding sequence (B), the engineered *IL8* insert (C), the *Il10* coding sequence (D) and engineered *Il10* insert (E). Optimization of PCR amplification was performed using different high fidelity DNA polymerases; the Expand High Fidelity PCR system gave the optimal amplification for both the *IL8* insert (Figure 3.3A), and *Il10* insert (Figure 3.4A), and the gel extracted *IL8* (Figure 3.3B) and *Il10* (Figure 3.4B) insert products were used for subsequent TOPO cloning.

A.



B.

```
ATGACTTCCAAGCTGGCCGTGGCTCTCTTGGCAGCCTTCCTGATTTCTGCAGCTCTGTGTGAAGGTGCAGTTTTGCCAAGGAGTGC
TAAAGAACTTAGATGTCAGTGCATAAAGACATACTCCAAACCTTTCCACCCCAAATTTATCAAAGAACTGAGAGTGATTGAGAGTG
GACCACACTGCGCCAACACAGAAATTATTGTAAAGCTTTCTGATGGAAGAGAGCTCTGTCTGGACCCCAAGGAAAACTGGGTGCAG
AGGGTTGTGGAGAAGTTTTTGAAGAGGGCTGAGAATTCATAA
```

C.

```
GG▼CGCGCCT▼CTAGAATGGACTACAAGGATGACGATGACAAGATGACTTCCAAGCTGGCCGTGGCTCTCTTGGCAGCCTTCCTGAT
TTCTGCAGCTCTGTGTGAAGGTGCAGTTTTGCCAAGGAGTGCTAAAGAACTTAGATGTCAGTGCATAAAGACATACTCCAAACCTT
TCCACCCCAAATTTATCAAAGAACTGAGAGTGATTGAGAGTGGACCACACTGCGCCAACACAGAAATTATTGTAAAGCTTTCTGAT
GGAAGAGAGCTCTGTCTGGACCCCAAGGAAAACTGGGTGCAGAGGGTTGTGGAGAAGTTTTTGAAGAGGGCTGAGAATTCATAAC▼T
CGAGGG▼CGCGCC
```

D.

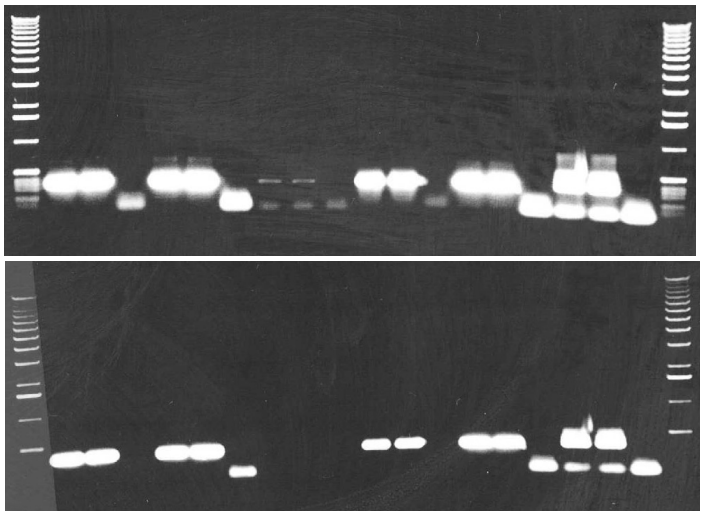
```
ATGCCTGGCTCAGCACTGCTATGCTGCCTGCTCTTACTGACTGGCATGAGGATCAGCAGGGGCCAGTACAGCCGGGAAGACAATAA
CTGCACCCACTTCCCAGTCGGCCAGAGCCACATGCTCCTAGAGCTGCGGACTGCCTTCAGCCAGGTGAAGACTTTCTTTCAAACAA
AGGACCAGCTGGACAACATACTGCTAACCGACTCCTTAATGCAGGACTTTAAGGGTTACTTGGGTTGCCAAGCCTTATCGGAAATG
ATCCAGTTTTACCTGGTAGAAGTGATGCCCCAGGCAGAGAAGCATGGCCCAGAAATCAAGGAGCATTTGAATTCCCTGGGTGAGAA
GCTGAAGACCCTCAGGATGCGGCTGAGGCGCTGTCATCGATTTCTCCCCTGTGAAAATAAGAGCAAGGCAGTGGAGCAGGTGAAGA
GTGATTTTAATAAGCTCCAAGACCAAGGTGTCTACAAGGCCATGAATGAATTTGACATCTTCATCAACTGCATAGAAGCATACATG
ATGATCAAAATGAAAAGCTAA
```

E.

```
GG▼CGCGCCT▼CTAGAATGGACTACAAGGATGACGATGACAAGATGCCTGGCTCAGCACTGCTATGCTGCCTGCTCTTACTGACTGG
CATGAGGATCAGCAGGGGCCAGTACAGCCGGGAAGACAATAACTGCACCCACTTCCCAGTCGGCCAGAGCCACATGCTCCTAGAGC
TGCGGACTGCCTTCAGCCAGGTGAAGACTTTCTTTCAAACAAAGGACCAGCTGGACAACATACTGCTAACCGACTCCTTAATGCAG
GACTTTAAGGGTTACTTGGGTTGCCAAGCCTTATCGGAAATGATCCAGTTTTACCTGGTAGAAGTGATGCCCCAGGCAGAGAAGCA
TGGCCCAGAAATCAAGGAGCATTTGAATTCCCTGGGTGAGAAGCTGAAGACCCTCAGGATGCGGCTGAGGCGCTGTCATCGATTTC
TCCCCTGTGAAAATAAGAGCAAGGCAGTGGAGCAGGTGAAGAGTGATTTTAATAAGCTCCAAGACCAAGGTGTCTACAAGGCCATG
AATGAATTTGACATCTTCATCAACTGCATAGAAGCATACATGATGATCAAAATGAAAAGCTAAC▼TCGAGGG▼CGCGCC
```

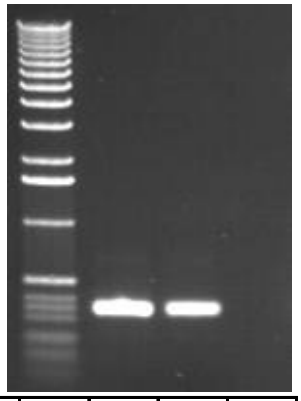
Figure 3.2. Schematic and nucleotide sequence of the *IL8* and *Il10* inserts. The *IL8* and *Il10* coding sequences were engineered by PCR to contain two sets of flanking restriction enzyme recognition sites as well as a 5'-FLAG tag; schematic in (A). The restriction enzyme sites facilitate directional cloning into two murine embryonic stem cell targeting vectors, pA33LSL (*Asc*I/ *Xho*I) and pBigT/pROSA26PA (*Xba*I/ *Xho*I). The final *IL8* and *Il10* inserts cloned into the pA33LSL eukaryotic expression vector therefore do not contain sequence for the 3'-*Asc*I recognition site. The 300 bp *IL8* coding sequence, NCBI accession number NM_000584.2, (B) was amplified from the pCMV6-XL5 cDNA plasmid to give the final 355 bp *IL8* insert (C) with the forward IL8F-Flag primer (underlined) incorporating 5' sequence for *Asc*I and *Xba*I restriction enzyme recognition sites and the FLAG-tag, with start codon, and the reverse IL8R-Flag primer (underlined) incorporating 3' sequence for *Xho*I and *Asc*I restriction enzyme sites. The 537 bp *Il10* coding sequence, NCBI accession number NM_010548.1, (D) was amplified from the pUMVC3-mIL10 cDNA plasmid to give the final 592 bp *Il10* insert (E) with the forward IL10F-Flag primer (underlined) incorporating 5' sequence for *Asc*I and *Xba*I restriction enzyme recognition sites and the FLAG-tag, with start codon, and the reverse IL10R-Flag6 primer (underlined) incorporating 3' sequence for *Xho*I and *Asc*I restriction enzyme sites (E). In (C and E), the *Asc*I recognition site is highlighted in teal, the *Xba*I recognition site is highlighted in pink, the FLAG-tag start codon is highlighted in yellow, the FLAG-tag sequence is highlighted in red and the *Xho*I recognition site is highlighted in green. Each restriction enzyme cut site is denoted by ▼.

A.



IL8 insert amplification	-	+	+	-	-	-	-	+	+	-	-	-	-	+	+	-	-	-	-	-
Il10 insert amplification	-	-	-	-	+	+	-	-	-	-	+	+	-	-	-	-	+	+	-	-
Mastermix alone negative	-	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-
DNA polymerase	-	HiFi			HiFi			HiFi Plus			HiFi Plus			Pwo			Pwo			-
1 Kb DNA ladder (0.5 μg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

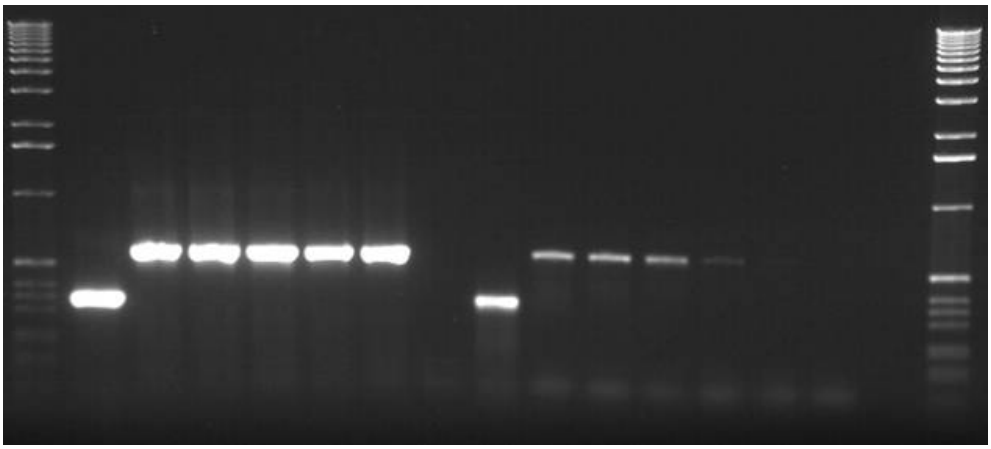
B.



IL8 Flag HiFi PCR product (column purified)	-	+	-	-
IL8 Flag HiFi PCR product (gel extracted)	-	-	+	-
blank	-	-	-	+
1 kb ladder (0.5 μg)	+	-	-	-

Figure 3.3. Generation of the *IL8* insert by PCR amplification. The final 355 bp *IL8* insert was amplified from the pCMV6-XL5 cDNA plasmid with the IL8F-Flag and IL8F-Flag primers. Optimization of *IL8* insert amplification with HiFi, HiFi Plus and Pwo DNA polymerase systems was performed and each product visualized by agarose gel electrophoresis (**A**); gel contains both *IL8* and *Il10* products but *Il10* insert amplification was repeated following primer re-design. The large primer-dimer products in some reactions are most likely due to the unusual length of the primers. The *IL8* HiFi PCR products were gel extracted and the purified *IL8* inserts analyzed by agarose gel electrophoresis (**B**).

A.



HiFi DNA polymerase	-	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-
HiFi Plus DNA polymerase	-	-	-	-	-	-	-	-	+	+	+	+	+	+	-	-	-
IL10F-Flag primer	-	+	+	+	+	+	+	-	+	+	+	+	+	+	-	-	-
IL10R-Flag(n) primer	-	1	2	3	4	5	6	-	1	2	3	4	5	6	-	-	-
PCR mastermix	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-
5X orange G loading buffer	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
1 kb ladder (0.5 µg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

B.

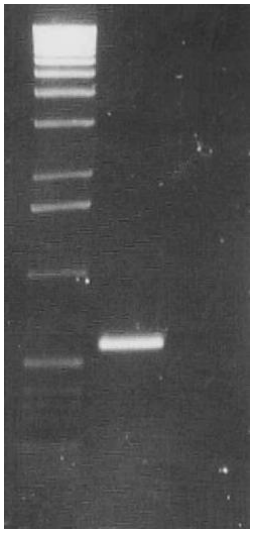


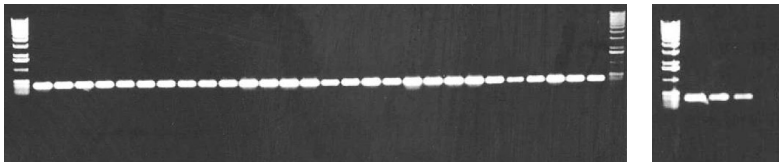
Figure 3.4. Generation of the *Il10* insert by PCR amplification. The final 592 bp *Il10* insert was amplified from the pUMVC3-mIL10 cDNA plasmid. Optimization of *Il10* insert amplification with HiFi and HiFi Plus DNA polymerase systems was performed with five primer sets; the forward primer IL10F-Flag was paired with the reverse primers IL10R-Flag2-6 as well as the original IL10R-Flag1 for confirmation of correct amplicon size; n denotes primer number. Each product was visualized by agarose gel electrophoresis (**A**). The HiFi *Il10* insert PCR product with IL10F-Flag and IL10R-Flag6 primers was gel extracted and the purified *Il10* insert analyzed by agarose gel electrophoresis (**B**).

3.3.2 Confirmation of IL8 and Il10 TOPO cloning

Successful TOPO cloning of the *IL8* insert was verified in a set of 30 colonies screened by PCR amplification (Figure 3.5A) and *Asc*I or *Eco*RI restriction digestion of the plasmids isolated from the first ten of these yielded eight putatively positive clones (Figure 3.5B). *Asc*I would be predicted to cut on either side of the *IL8* insert producing a 351 bp fragment, and

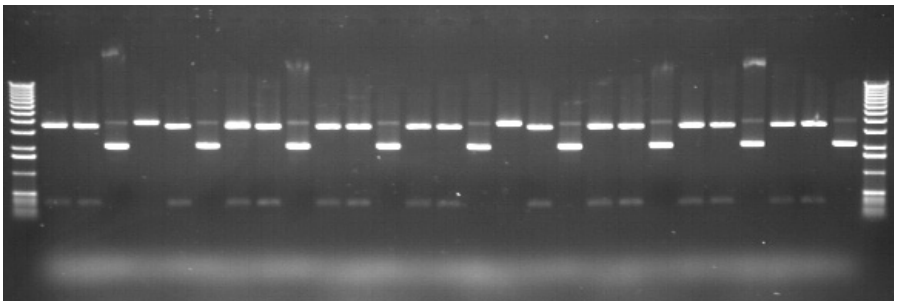
*EcoR*I should cut once on either side of the pCRBluntII-TOPO cloning site to give a 373 bp fragment. Based on these criteria, IL8_pCRBluntII-TOPO-1, 3, 4, 5, 7, 8, 9, 10 were both PCR- and digest-positive clones.

A.

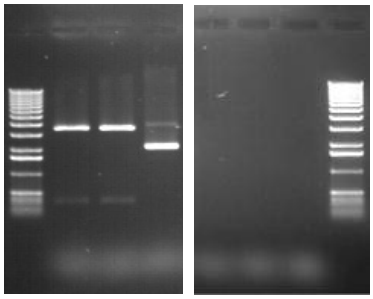


IL8_pCR-BluntII-TOPO clone	-	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	-		-	29	30	-	-
pCMV6-XL5 cDNA plasmid	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	+	-
PCR mastermix alone	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	+
1 kb ladder (0.5 µg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+		+	-	-	-	-

B.



IL8_pCR-BluntII-TOPO clone	-	1			2			3			4			5			6			7			8			9			-
AscI enzyme	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	-
EcoRI enzyme	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-
1 kb ladder (0.5 µg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

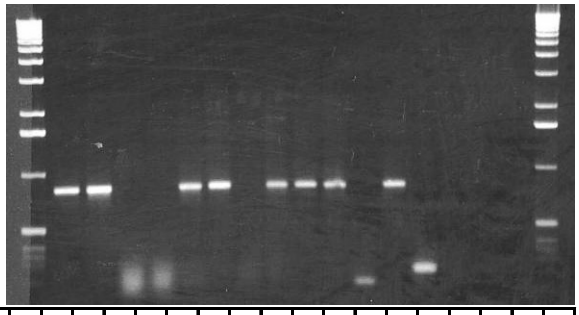


IL8_pCR-BluntII-TOPO clone	-	10			-			-
AscI enzyme	-	+	-	-	+	-	-	-
EcoRI enzyme	-	-	+	-	-	+	-	-
5X orange G loading buffer	-	-	-	-	-	-	+	-
1 kb ladder (0.5 µg)	+	-	-	-	-	-	-	+

Figure 3.5. Confirmation of *IL8* insert cloning into pCR-BluntII-TOPO. Thirty IL8_pCR-BluntII-TOPO clones were screened by colony PCR using IL8F-Flag/ IL8R-Flag primers (**A**); the pCMV6-XL5 cDNA plasmid was included as the positive control and the PCR mastermix alone was the negative control. Each reaction was subjected to agarose gel electrophoresis. All clones were PCR-positive therefore the first ten clones were grown overnight for plasmid isolation (miniprep). Restriction digestion of IL8_pCR-BluntII-TOPO-1-10 clones with each of *Asc*I and *EcoR*I was performed for confirmation of insert size (**B**). The digests were analyzed by agarose gel electrophoresis; uncut plasmids were also analyzed. *Asc*I releases a 351 bp fragment and *EcoR*I releases a 373 bp fragment; positive clones for both digests are IL8_pCR-BluntII-TOPO-1, 3, 4, 5, 7, 8, 9, 10.

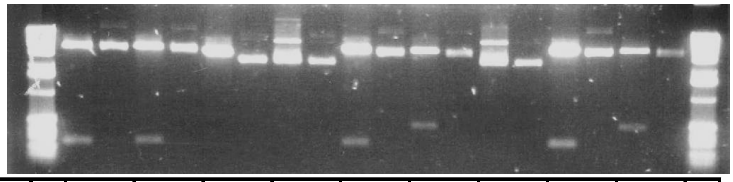
Fourteen colonies from the *Il10* insert TOPO cloning screened by PCR amplification (Figure 3.6A), and restriction digestion of the isolated plasmids with *Hind*III and *Xmn*I yielded eight putatively positive clones (Figure 3.6B). *Hind*III was predicted to cut once in the pCR-BluntII-TOPO vector only and *Xmn*I should cut once in the *Il10* insert only, giving fragments of 259 and 3853 bp or 451 and 3661 bp, depending on whether the insert is in forward or reverse orientation, respectively, with respect to the reference *Il10* coding sequence. From these criteria Il10_pCR-BluntII-TOPO-1, 2, 5, 6, 8, 9, 10 and 12 were both PCR- and digest-positive clones.

A.

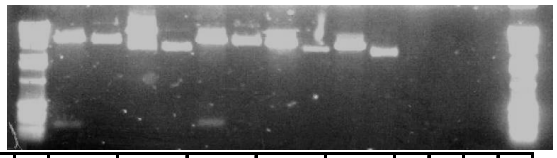


Il10_pCR-BluntII-TOPO clone	-	1	2	3	4	5	6	7	8	9	10	11	12	13	14	-	-	-
PCR mastermix alone	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
5X orange G loading buffer	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
1 kb ladder (0.5 μg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

B.



Il10_pCR-BluntII-TOPO clone	-	1		2		3		4		5		6		7		8		9		-
HindIII and XmnI enzymes	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	-
5X orange G loading buffer	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1 kb ladder (0.5 μg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+



Il10_pCR-BluntII-TOPO clone	-	10		11		12		13		14		-	-	-	-
HindIII and XmnI enzymes	-	+	-	+	-	+	-	+	-	+	-	+	+	-	-
5X orange G loading buffer	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
1 kb ladder (0.5 μg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+

Figure 3.6. Confirmation of *Il10* insert cloning into pCR-BluntII-TOPO. Fourteen Il10_pCR-BluntII-TOPO clones were screened by colony PCR using standard M13F and M13R primers (A) The pCR-BluntII-TOPO plasmid contains M13F/ M13R primer binding sites which are absent from the original pUMVC3-mIL10 cDNA plasmid. Each reaction was subjected to agarose gel electrophoresis. All colonies were grown overnight for plasmid isolation. Restriction digestion of the Il10_pCR-BluntII-TOPO-1-14 plasmids with *Hind*III and *Xmn*I (B) confirmed positive clones. The digests were analyzed by agarose gel electrophoresis; uncut plasmids were also analyzed. The sizes of digest fragments correspond to insert orientation within pCR-BluntII-TOPO; 259 bp indicates 5' orientation and 451 bp indicates 3' orientation. PCR- and digest-positive clones are Il10_pCR-BluntII-TOPO-1, 2, 5, 6, 8, 9, 10, 12.

3.3.3 Capillary sequencing of IL8 and Il10 TOPO inserts

Capillary sequencing, using gene-specific and M13 forward and reverse primers, of the putatively positive IL8_pCR-BluntII-TOPO and Il10_pCR-BluntII-TOPO clones identified those clones containing *IL8* or *Il10* inserts with the correct coding sequence and without any PCR-introduced mutations. The trace files within the assemblies were checked manually for discrepancies between the assemblies and reference sequence; some discrepancies were bona fide base pair changes (i.e. mutations) in the insert sequence, whereas others were errors in base calling by the Gap4 software. Merely taking the assembly consensus sequence as the final sequence would miss the discrimination between base pair changes and base calling errors by Gap4. Table 3.2 summarizes the eight IL8_pCR-BluntII-TOPO clones and Table 3.3 summarizes the eight Il10_pCR-BluntII-TOPO clones, indicating the number of discrepancies between the assembly consensus sequence and insert reference sequence, and from this the number of passes or fails as determined following manual checking. Two IL8_pCR-BluntII-TOPO clones (3, 10) were deemed to be acceptable to proceed with for subsequent cloning into the ESC targeting vector, pA33LSL. Six Il10_pCR-BluntII-TOPO clones (2, 5, 6, 9, 10, 12) were deemed suitable for pA33LSL cloning. Of the clones with correct sequences, IL8_pCR-BluntII-TOPO-10 and Il10_pCR-BluntII-TOPO-10 had high base calling confidence values throughout the reference sequence, and as such were used in subsequent cloning steps.

Table 3.2. IL8_pCR-BluntII-TOPO clone sequencing results summary. Sequence traces were assembled using the Gap 4 software and checked manually for any discrepancies between the *IL8* insert reference sequence. A discrepancy prevented further action when it was not resolved by at least two other sequencing reads. The overall suitability took into account the lack of fails as well as high sequencing quality throughout the insert length.

Sample	Contigs	Discrepancies	Number Fails	Number Passes	Overall Suitability	Subsequent Cloning
IL8_pCR-BluntII-TOPO-1	1	4	1	3	No	NO
IL8_pCR-BluntII-TOPO-3	1	1	0	1	Yes	NO
IL8_pCR-BluntII-TOPO-4	1	5	1	4	No	NO
IL8_pCR-BluntII-TOPO-5	1	4	3	1	No	NO
IL8_pCR-BluntII-TOPO-7	1	6	1	5	No	NO
IL8_pCR-BluntII-TOPO-8	1	4	2	2	No	NO
IL8_pCR-BluntII-TOPO-9	1	10	6	4	No	NO
IL8_pCR-BluntII-TOPO-10	1	6	0	6	Yes	YES

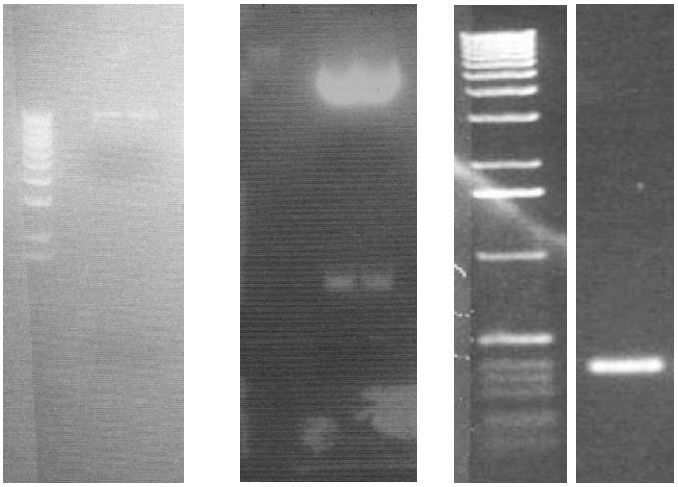
Table 3.3. Il10_pCR-BluntII-TOPO clone sequencing results summary. Sequence traces were assembled using the Gap 4 software and checked manually for any discrepancies between the *Il10* insert reference sequence. A discrepancy failed when it was not resolved by at least two other sequencing reads. The overall suitability took into account the lack of fails as well as high sequencing quality throughout the insert length.

Sample	Contigs	Discrepancies	Number Fails	Number Passes	Overall Suitability	Subsequent Cloning
Il10_pCR-BluntII-TOPO-1	1	16	2	14	No	NO
Il10_pCR-BluntII-TOPO-2	1	15	0	15	Yes	NO
Il10_pCR-BluntII-TOPO-5	1	13	0	13	Yes	NO
Il10_pCR-BluntII-TOPO-6	1	24	0	24	Yes	NO
Il10_pCR-BluntII-TOPO-8	1	17	2	15	No	NO
Il10_pCR-BluntII-TOPO-9	1	19	0	19	Yes	NO
Il10_pCR-BluntII-TOPO-10	1	15	0	15	Yes	YES
Il10_pCR-BluntII-TOPO-12	1	39	0	39	Yes	NO

3.3.4 IL8_pCR-BluntII-TOPO and Il10_pCR-BluntII-TOPO digestion and cloning into the pA33LSL targeting vector

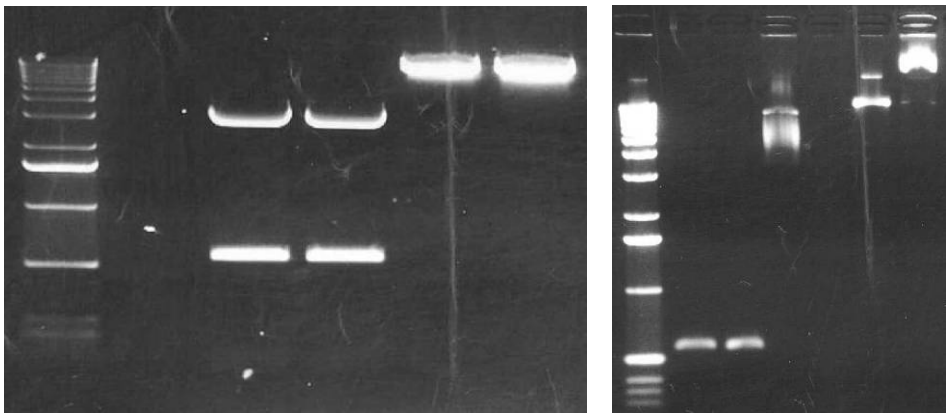
To permit cloning into the pA33LSL targeting vector, the modified *IL8* and *Il10* inserts were excised from their respective pCR-BluntII-TOPO vectors using *Asc*I and *Xho*I restriction enzymes. Figure 3.7 shows the agarose gels confirming the 340 bp *IL8* (A) and 577 bp *Il10* (B) digest-fragments and gel extracted products, as well as the *Asc*I- and *Xho*I-digested pA33LSL plasmids. The amplified fragments and pA33LSL vector were ligated together.

A.



pA33LSL	-	-	+	+		-	-	-	-		-	-
IL8_pCRII-BluntII-TOPO-10	-	-	-	-		-	-	+	+		-	-
AscI and XhoI enzymes	-	-	+	+		-	-	+	+		-	-
Purified IL8 insert	-	-	-	-		-	-	-	-		-	+
1 Kb DNA ladder (0.5 µg)	+	-	-	-		+	-	-	-		+	-

B.

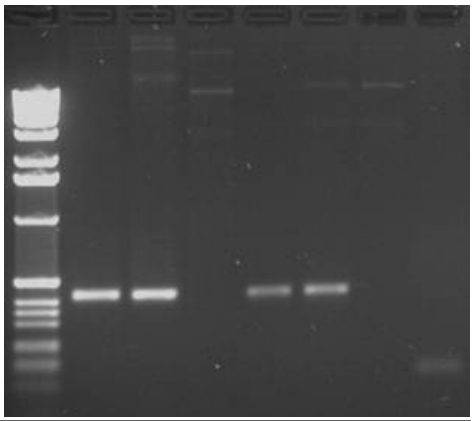


Il10_pCRII-BluntII-TOPO-10	-	-	+	+	-	-		-	-	-	+	-	-	-
pA33LSL	-	-	-	-	+	+		-	-	-	-	-	-	+
AscI and XhoI enzymes	-	-	+	+	+	+		-	-	-	-	-	-	-
Purified Il10 insert	-	-	-	-	-	-		-	+	+	-	-	-	-
Purified pA33LSL digest	-	-	-	-	-	-		-	-	-	-	-	+	-
1 Kb DNA ladder (0.5 μg)	+	-	-	-	-	-		+	-	-	-	-	-	-

Figure 3.7. Directional cloning of the *IL8* and *Il10* inserts into the pA33LSL murine embryonic stem cell targeting vector facilitated by *Asc*I and *Xho*I restriction digestion. The confirmed insert-positive IL8_pCR-BluntII-TOPO-10 and Il10_pCR-BluntII-TOPO-10 plasmids were digested with *Asc*I and *Xho*I restriction enzymes to release the final 340 bp *IL8* and 577 bp *Il10* inserts, respectively, for ligation with pA33LSL, also digested with *Asc*I and *Xho*I, to create the final targeting vectors. Agarose gel electrophoresis of the IL8_pCR-BluntII-TOPO-10 (**A**) and Il10_pCR-BluntII-TOPO-10 (**B**) digests confirmed insert size, which were then gel extracted prior to ligation. The linearized pA33LSL plasmid was also gel extracted in concert with each insert (**A**, **B**).

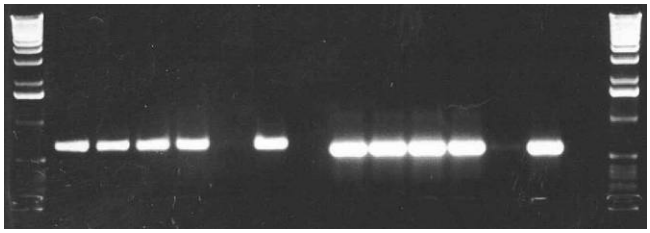
Successful amplification across the pA33LSL plasmid *Asc*I/ *Xho*I cloning site, as well as with insert-specific primers for the *Il10* clones, identified four putatively positive *IL8* clones, IL8_pA33LSL-1, 2, 4, 5 (Figure 3.8A), and five putatively positive *Il10* clones, Il10_pA33LSL-1, 2, 3, 4, 6 (Figure 3.8B).

A.



IL8_pA33LSL clone	-	1	2	3	4	5	6	-
PCR mastermix	-	+	+	+	+	+	+	+
1 Kb DNA ladder (0.5 μg)	+	-	-	-	-	-	-	-

B.



Il10_pA33LSL clone	-	1	2	3	4	5	6	-	1	2	3	4	5	6	-	-
A33F/ A33R primers	-	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-
IL10F-Flag/IL10R-Flag6 primers	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	-
PCR mastermix	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
1 Kb DNA ladder (0.5 μg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

Figure 3.8. Validation of *IL8* and *Il10* insert ligation with pA33LSL for the generation of final ESC targeting vector. Six of each of the IL8_pA33LSL (**A**) and Il10_pA33LSL (**B**) plasmids were amplified with Platinum PCR Supermix and primers targeting the pA33LSL cloning site (A33F/ A33R), as well as insert-specific primers for the *Il10* clones (IL10F-Flag/ IL10R-Flag6). Successful amplification identified four positive *IL8* clones, IL8_pA33LSL-1, 2, 4, 6, and five positive *Il10* clones, Il10_pA33LSL-1, 2, 3, 4, 6.

The capillary sequencing analysis for the PCR-positive IL8_pA33LSL and Il10_pA33LSL plasmid inserts is found in Tables 3.4 and 3.5, respectively. Two of the four IL8_pA33LSL and four of the five Il10_pA33LSL clones had the correct sequences and were suitable for use as the ESC targeting vector. The final vectors were chosen based the overall base calling confidence scores for the insert sequences. IL8_pA33LSL-1 and Il10_pA33LSL-1 were selected as the final ESC targeting vectors.

Table 3.4. IL8_pA33LSL clone sequencing results summary. Sequence traces were assembled using the Gap 4 software and checked manually for any discrepancies between the *IL8* insert, digested *in silico* with *Asc*I and *Xho*I, reference sequence. A discrepancy failed when it was not resolved by at least two other sequencing reads. The overall suitability took into account the lack of fails as well as high sequencing quality throughout the insert length. The IL8_pA33LSL-5 clone was not analyzed because the two sequences that passed Asp (WTSI quality control) were very short and did not assemble with the reference sequence.

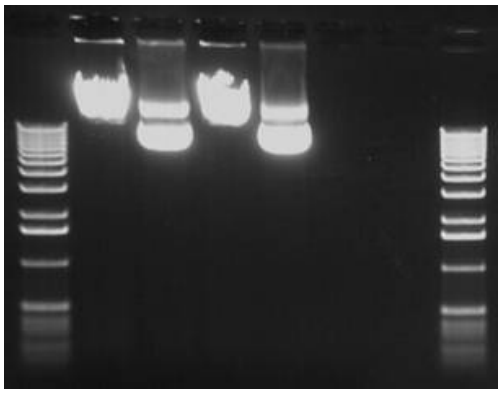
Clone	Contigs	Discrepancies	Number Fails	Number Passes	Overall suitability	ESC Targeting
IL8_pA33LSL-1	1	3	0	3	Yes	YES
IL8_pA33LSL-2	1	5	0	5	Yes	NO
IL8_pA33LSL-4	2	2	0	2	Maybe	NO
IL8_pA33LSL-5	3	n/a	n/a	n/a	No	NO

Table 3.5. Il10_pA33LSL clone sequencing results summary. Sequence traces were assembled using the Gap 4 software and checked manually for any discrepancies between the *Il10* insert, digested *in silico* with *Asc*I and *Xho*I, reference sequence. A discrepancy failed when it was not resolved by at least two other sequencing reads. The overall suitability took into account the lack of fails as well as high sequencing quality throughout the insert length; the sequences for IL10_pA33LSL-6 were of lower quality than those of the other clones and therefore not deemed suitable.

Clone	Contigs	Discrepancies	Number Fails	Number Passes	Overall suitability	ESC Targeting
Il10_pA33LSL-1	1	8	0	8	Yes	YES
Il10_pA33LSL-2	1	8	0	8	Yes	NO
Il10_pA33LSL-3	1	6	0	6	Yes	NO
Il10_pA33LSL-4	1	6	0	6	Yes	NO
Il10_pA33LSL-6	1	7	0	7	No	NO

3.3.5 Transfection of JM8.N4 murine embryonic stem cells with the IL8_pA33LSL and Il10_pA33LSL targeting vectors

Following transfection of JM8.N4 ESC with either the *Sal*I linearized IL8_pA33LSL or Il10_pA33LSL targeting vector (Figure 3.9), Geneticin section and ESC colony formation, microscopic inspection revealed that most colonies had round, well-defined edges. Lysed cells were visible at the surface of the cuvette, which had indicated successful electroporation. A few colonies were star-burst in appearance, indicative of differentiating cells. Generally speaking, colonies with differentiating cells were not picked for expansion to maintain ESC pluripotency.



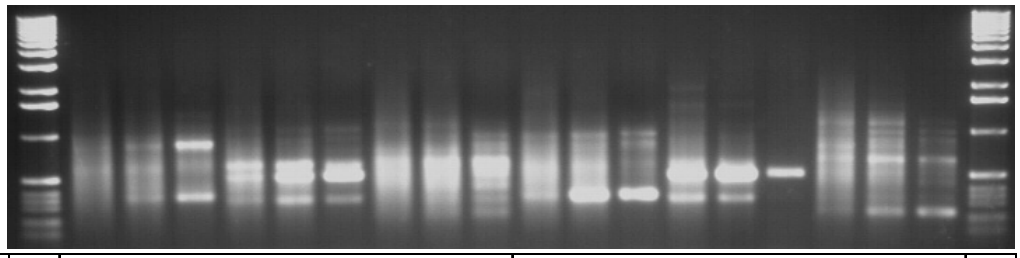
IL8_pA33LSL-1	-	+	+	-	-	-	-	-
Il10_pA33LSL-1	-	-	-	+	+	-	-	-
SalI enzyme	-	+	-	+	-	+	-	-
5X orange G loading buffer	-	-	-	-	-	-	+	-
1 Kb DNA ladder (0.5 µg)	+	-	-	-	-	-	-	+

Figure 3.9. Final *Sal*I-digested IL8_pA33LSL-1 and Il10_pA33LSL-1 targeting vectors. Each plasmid was linearized by *Sal*I restriction digestion for JM8.N4 ESC transfection and vector targeting at the murine *Gpa33* genomic locus. Agarose gel electrophoresis confirmed the single cut site for both IL8_pA33LSL-1 and Il10_pA33LSL-1 compared to uncut plasmids.

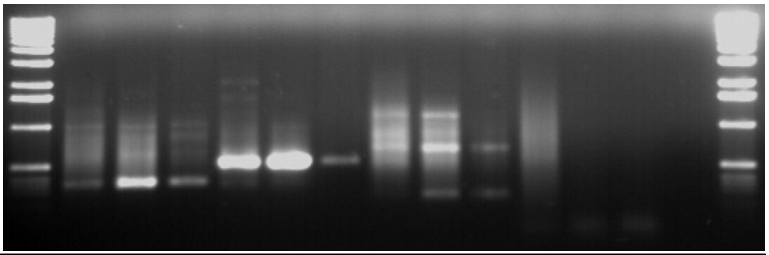
During colony expansion for recombination-positive clones, it was necessary to find a balance between maximum desired confluency (80-90%) and the maximum number of wells at this confluency, so that there would be the highest number of cells possible for freezing and DNA extraction without compromising ESC health. Of the 96 clones picked for each gene construct, 66, 52 and 0 of the IL8_Gpa33_JM8N4 ESC clones and 54, 29 and 2 of the Il10_Gpa33_JM8N4 ESC clones had at least 5, 10 and 20 μg of DNA, respectively, following extraction. Clones with concentrations sufficiently high enough to obtain 10 μg DNA in the required volume were further subjected to restriction digestion.

3.3.6 Southern blot probe PCR and sequencing confirmation

Optimization of the amplification conditions for the A33 probe (A33p) using Pfx50 DNA polymerase was performed at variable annealing temperatures (50, 55, 58°C), primer sets (A33pF1-3, A33pR1-3) and JM8.N4 DNA template amounts (25, 50, 100 ng) (Figure 3.10). An annealing temperature of 58°C, using A33pF2/ A33pR2 primers and 50 ng template gave the cleanest amplification.



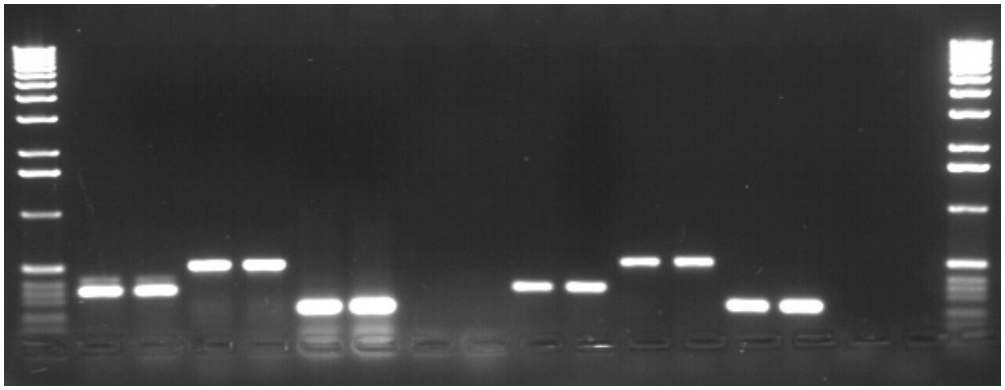
Annealing Temp (°C)	-	50									55									-
Primers	-	F1/R1	F1/R1	F1/R1	F2/R2	F2/R2	F2/R2	F3/R3	F3/R3	F3/R3	F1/R1	F1/R1	F1/R1	F2/R2	F2/R2	F2/R2	F3/R3	F3/R3	F3/R3	-
JM8.N4 ESC DNA (ng)	-	25	50	100	25	50	100	25	50	100	25	50	100	25	50	100	25	50	100	-
1 Kb ladder (0.5 μg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+



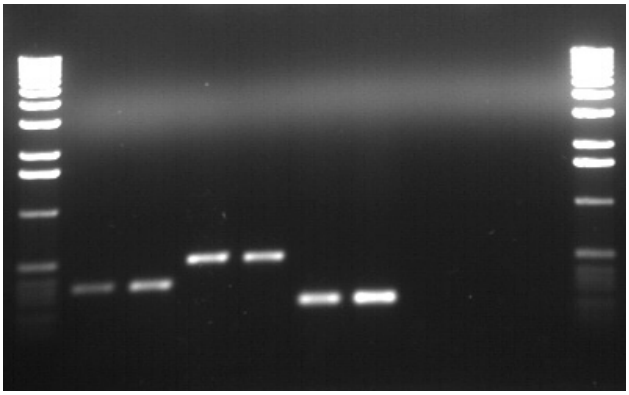
Annealing Temp (°C)	-	58									50	55	58	-	-
Primers	-	F1/R1	F1/R1	F1/R1	F2/R2	F2/R2	F2/R2	F3/R3	F3/R3	F3/R3	F1/R1	F2/R2	F3/R3	-	-
JM8.N4 ESC DNA (ng)	-	25	50	100	25	50	100	25	50	100	-	-	-	-	-
5X orange G loading buffer	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
1 Kb ladder (0.5 μg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+

Figure 3.10. PCR optimization of the Southern blot probe for ESC targeting confirmation. Three sets of forward and reverse primers were designed to amplify a region downstream of the *Gpa33* genomic locus and 3' arm of homology of the pA33LSL targeting vector to act as the Southern blot probe. Amplification was performed with Pfx50 DNA polymerase using variable primer pairs (F1/R1, F2/R2, F3/R3), JM8N4 ESC DNA template (25, 50, 100 ng), and annealing temperatures (50, 55, 58°C). The PCR products were visualized by agarose gel electrophoresis. Expected amplicon sizes from the F1/R1, F2/R2 and F3/R3 primer pairs were 356, 520 and 248 bp, respectively.

Amplification with Platinum PCR Supermix using the same annealing temperatures and primer combinations, but with 12.5 and 50 ng JM8.N4 DNA, confirmed the primer specificity (Figure 3.11). However the Pfx50 product was used for TOPO cloning due to the high fidelity of the Pfx50 DNA polymerase.



Annealing Temp (°C)	-	50								55								-
Primers	-	F1/R1	F1/R1	F2/R2	F2/R2	F3/R3	F3/R3	F1/R1	F2/R2	F1/R1	F1/R1	F2/R2	F2/R2	F3/R3	F3/R3	F1/R1	F3/R3	-
JM8.N4 ESC DNA (ng)	-	12.5	50	12.5	50	12.5	50	-	-	12.5	50	12.5	50	12.5	50	12.5	50	-
1 Kb ladder (0.5 µg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+



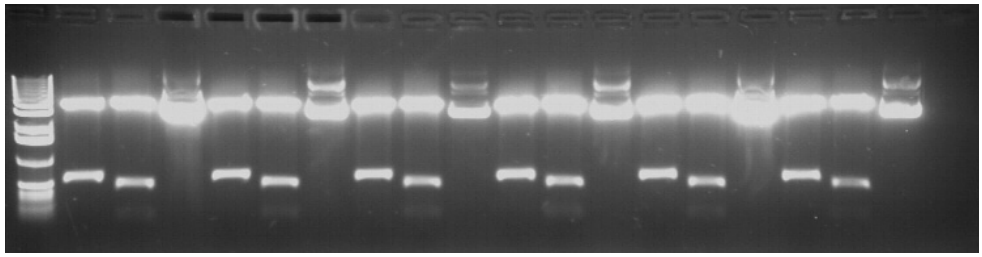
Annealing Temp (°C)	-	58								-	-
Primers	-	F1/R1	F1/R1	F2/R2	F2/R2	F3/R3	F3/R3	F2/R2	F3/R3	-	-
JM8.N4 ESC DNA (ng)	-	12.5	50	12.5	50	12.5	50	12.5	50	-	-
5X orange G loading buffer	-	-	-	-	-	-	-	-	-	+	-
1 Kb ladder (0.5 µg)	+	-	-	-	-	-	-	-	-	-	+

Figure 3.11. Confirmation of Southern blot probe primer specificity. The Southern blot probe primers were amplified using Platinum PCR Supermix to test their specificity as amplification with Pfx50 DNA polymerase was very smeary. This could also have been due to too much genomic DNA template so the same primer combinations (F1/R1, F2/R2, F3/R3) were tested (300 nM) at the three different annealing temperatures (50, 55, 58°C), however 12.5 and 50 ng JM8N4 ESC DNA was used as the template in a final volume of 50 µl. The PCR products (10 µl) were visualized by agarose gel electrophoresis (0.8% agarose, 80 V, 90 min). The F1/R1, F2/R2 and F3/R3 primers all amplified the expected amplicons of 356, 520 and 248 bp, respectively.

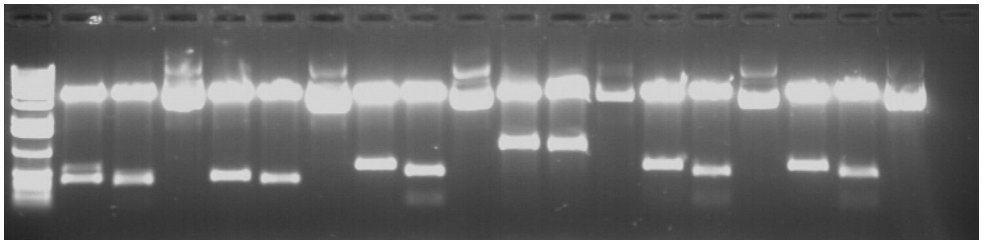
Successful cloning of the A33p PCR product into pCR-BluntII-TOPO was confirmed by restriction digestion with *BamH*I/ *EcoR*V (581 bp) and *EcoR*I (451 and 87 bp) and analysis of undigested clones, Figure 3.12A. Nine of the twelve clones analyzed gave fragments of expected

size and therefore contained the A33p insert. Undigested clones demonstrated ligation of the linear A33p_pCR-BluntII-TOPO vector upon cloning. A33p_pCRIIBlunt_TOPO clones 7, 8 and 10 did not give restriction fragments of expected size so these were not used in further analyses. The identity of the insert of the nine positive A33p_pCR-BluntII-TOPO plasmids was confirmed by PCR amplification with A33pF2 and A33pR2 primers (Figure 3.12B). A33p_pCRIIBlunt_TOPO clones 8 and 10 did not give amplicons of expected size and thus were not used in subsequent analyses. A further two A33p_pCR-BluntII-TOPO clones (3, 11) did not have sufficient DNA for further analysis and were consequently excluded. Putatively positive A33p_pCR-BluntII-TOPO clones were numbers 1, 2, 4, 5, 6, 9 and 12.

A.

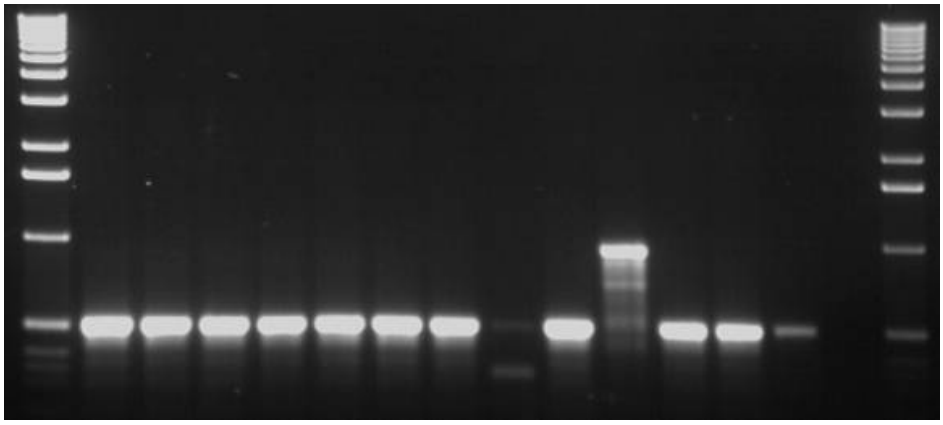


A33p_pCR_BluntII-TOPO clone	-	1			2			3			4			5			6			-
BamHI and EcoRV enzymes	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+
EcoRI enzyme	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-
1 Kb ladder (0.5 µg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-



A33p_pCR_BluntII-TOPO clone	-	7			8			9			10			11			12			-
BamHI and EcoRV enzymes	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	-
EcoRI enzyme	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	+
1 Kb ladder (0.5 µg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

B.



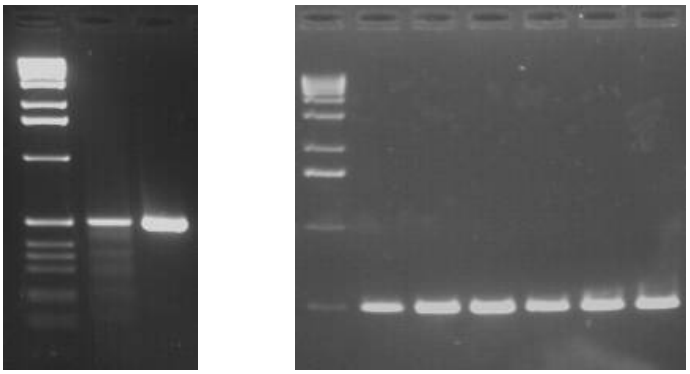
A33p_pCR_BluntII-TOPO clone	-	1	2	3	4	5	6	7	8	9	10	11	12	-	-	-
JM8.N4 WT DNA	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
PCR mastermix	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
1 Kb ladder (0.5 μg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

Figure 3.12. Confirmation of Southern blot probe cloning into pCR-BluntII-TOPO. Twelve isolated colonies were grown overnight for plasmid isolation (miniprep) and screened for correct insert size by restriction digestion (**A**) and PCR amplification (**B**). Restriction digestion was carried out with *BamH*I and *EcoR*V, which both cut once in the pCR-BluntII-TOPO multiple cloning site (MCS) to release the insert (expected size 581 bp), as well as *EcoR*I, which cuts twice in the pCR-BluntII-TOPO MCS to release the insert and once within the A33 amplicon, giving two digest fragments (expected sizes 87 and 451 bp). Digest-positive clones were A33p_pCR-BluntII-TOPO-1, 2, 3, 4, 5, 6, 9, 11, 12. PCR-positive clones were A33p_pCR-BluntII-TOPO-1, 2, 3, 4, 5, 6, 7, 9, 11, 12.

Capillary sequencing confirmed the identity of the seven A33p_pCR-BluntII-TOPO plasmids and a summary of sequence analysis is found in Table 3.6. Any discrepancies between sequence trace and A33p reference sequence were checked manually. In all but one plasmid, all questionable bases were verified to be those of the A33p sequence; A33p_pCRIIBlunt-5 contained an A to G mutation at position 520 with respect to the A33p reference sequence. This mutation is unlikely to affect hybridization of the probe and therefore all seven plasmids were deemed suitable for use in the generation of the Southern blot probe. Figure 3.13 shows the final purified A33p products amplified from the A33p_pCR-BluntII-TOPO clones, which were then pooled in order to maintain the composition of the Southern blot probe over several hybridizations, if necessary.

Table 3.6. A33p_pCR-BluntII-TOPO clone sequencing results summary. Sequence traces were assembled using the Gap 4 software and checked manually for any discrepancies between the *Gpa33* probe (A33p) reference sequence. A discrepancy failed when it was not resolved by at least two other sequencing reads. The overall suitability took into account the lack of fails as well as high sequencing quality throughout the insert length. A33p_pCRIIBlunt-5 was included in the probe despite having one discrepancy because this was an A to G mutation at the last position of the sequence that would not affect hybridization in the Southern blot assay.

Clone	Contigs	Discrepancies	Number Fails	Number Passes	Overall suitability	Southern Blot Probe
A33p_pCRIIBlunt-1	1	6	0	6	Yes	YES
A33p_pCRIIBlunt-2	1	5	0	5	Yes	YES
A33p_pCRIIBlunt-4	1	8	0	8	Yes	YES
A33p_pCRIIBlunt-5	1	5	1	4	Yes	YES
A33p_pCRIIBlunt-6	1	4	0	4	Yes	YES
A33p_pCRIIBlunt-9	1	7	0	7	Yes	YES
A33p_pCRIIBlunt-12	1	4	0	4	Yes	YES



A33p_pCR_BluntII-TOPO clone	-	1	1		-	2	4	5	6	9	12
Platimum PCR Supermix	-	+	-		-	-	-	-	-	-	-
Pfx50 DNA Polymerase	-	-	+		-	+	+	+	+	+	+
1 Kb ladder (0.5 μg)	+	-	-		+	-	-	-	-	-	-

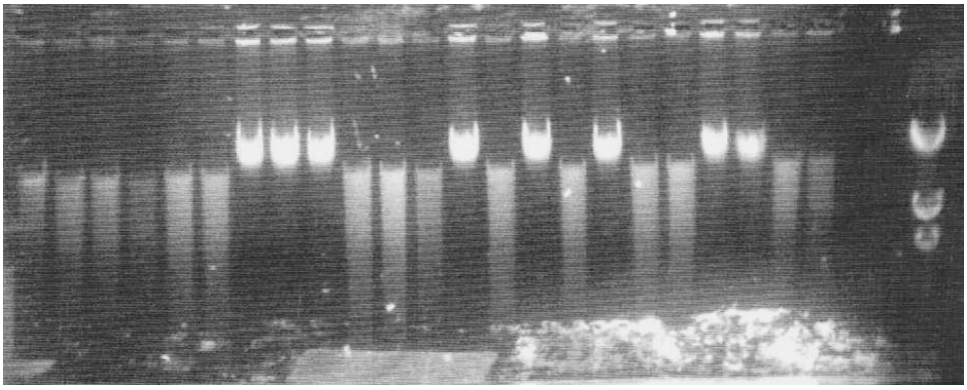
Figure 3.13. Purified A33 amplicons for use as Southern blot probe. The Southern blot probe was amplified from sequence-positive A33p_pCR-BluntII-TOPO plasmids with Pfx50 DNA polymerase and A33pF2/ A33pR2 primers, then column purified and pooled. The second lane in the left gel is the A33p_pCR-BluntII-TOPO-1 PCR product amplified with Platinum PCR Supermix and column purified as a test for what appeared to be degradation in the confirmation PCR (Figure 3.12B). No further use was made of the Platinum PCR Supermix products beyond insert confirmation.

3.3.7 Southern blot analysis of IL8 and Il10 ESC clone BsaBI-digested DNA

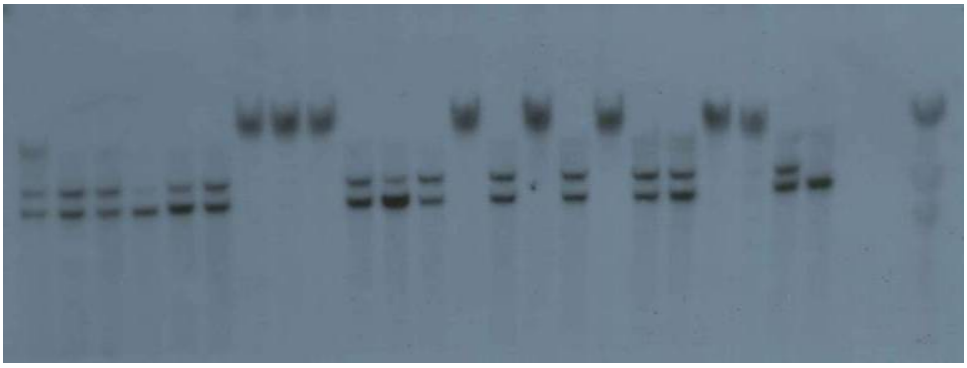
Figure 3.14 is the agarose gel (A) and film image (B) for the Southern blot analysis of the first eleven ESC clones of each gene construct. Of the *IL8* blotted clones, three had high molecular weight bands (approximately 23 kb in size) indicative of undigested DNA and seven were positive with hIL8:c allele of 13.1 kb and WT allele of 10.9 kb relatively equal in intensities. The identity of one clone was undetermined as it had both hIL8:c and WT alleles, but

the WT allele was stronger in intensity than the hIL8:c potentially indicative of a heterogeneous cell population and incomplete hIL8:c penetrance. Of the *Il10* blotted clones, five had undigested DNA and five were positive with mIL10:c allele of 13.3 kb and WT allele of 10.9 kb relatively equal in intensities. The identity of one clone was undetermined as it also had a much stronger intensity WT band compared to mIL10:c band.

A.



B.



Gene Insert	8	8	8	8	8	8	8	8	8	8	10	10	10	10	10	10	10	10	10	10	10	8	-	-	-	-
Clone ID	1	12	15	20	21	22	26	28	31	34	3	4	9	11	14	15	22	23	24	26	27	35	-	-	-	-
JM8.N4 WT DNA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
Genotype	H	H	H	ND	H	H	U	U	U	H	ND	H	U	H	U	H	U	H	H	U	U	H	WT	-	-	-
BsaBI MM	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-
λ DNA/ HindIII ladder	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

Figure 3.14. Southern blot analysis of *BsaB*I-digested IL8_Gpa33_JM8N4 and Il10_Gpa33_JM8N4 clone DNA. Each clone DNA was digested with the restriction enzyme *BsaB*I. JM8.N4 DNA was also digested under the same conditions for the wildtype (WT) genotype/ positive control and the *BsaB*I digest mastermix alone was the negative control. The λ DNA/ *Hind*III ladder (25 μg) was included to approximate the sizes of the hybridized digestion fragments; the sizes of the three visible λ DNA/ *Hind*III bands are 23130, 9416 and 6557 bp, from the largest at the top of the agarose gel (**A**) and Southern blot (**B**). The expected size of the WT allele (+/+) is 10.9 kb, the IL8 knock-in allele (hIL8:c) is 13.1 kb and the Il10 knock-in allele (mIL10:c) is 13.3 kb. The genotype of each clone is noted. Correctly targeted hIL8:c/+ or mIL10:c/+ clones are labelled H to denote their conditional heterozygous nature, clones labelled ND were not determined as they had a weaker knock-in allele compared to WT, clones labelled U were undigested (>23 kb single band) and therefore their genotype was also not determined. Both the *BsaB*I mastermix alone and blank lane negative controls were clear.

The Southern blot revealed *BsaB*I-digest-generated DNA fragments of expected size for both hIL8:c and mIL10:c alleles, and as such it can be assumed that homologous recombination occurred in the correct genomic loci since the Southern blot probe hybridizes to the genomic sequence outside of the targeting vector. A second method was needed for confirmation of this assumption and thus PCR amplifications in and around the locus of recombination were performed. The remaining *BsaB*I-digested clone DNA were not subjected to further analysis as the first Southern blot gave sufficient positive clones for PCR confirmation and subsequent ESC culture.

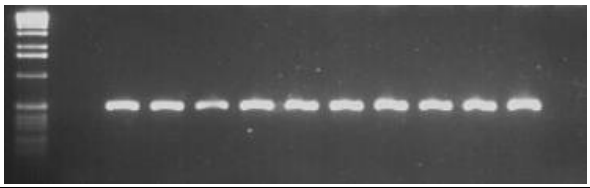
3.3.8 PCR confirmation of hIL8:c and mIL10:c alleles

Concomitant to Southern blot analysis, the locus of recombination of the first eleven clones was also analyzed by PCR. Primers used to amplify the knock-in allele, neomycin resistance cassette as well as a primer downstream of the pA33LSL vector arm of homology confirmed the clones identified as positive by Southern blot analysis. Seven IL8_Gpa33_JM8N4 and five Il10_Gpa33_JM8N4 ESC clones were thus identified as positive for correct homologous recombination at the genomic locus after the seventh exon of the Gpa33 gene. Figures 3.15 and 3.16 show the PCR results for the IL8_Gpa33_JM8N4 and Il10_Gpa33_JM8N4 ESC clones, respectively.

A.

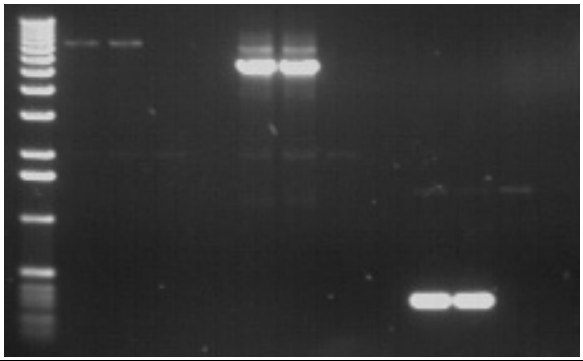
IL8_Gpa33_JM8N4 clone	1	12	15	20	21	22	26	28	31	34	35	-	-
WT JM8.N4 DNA	-	-	-	-	-	-	-	-	-	-	-	+	-
IL8-7/ IL8-8 primers	+	+	+	+	+	+	+	+	+	+	+	+	-
1 Kb DNA ladder (0.5 µg)	-	-	-	-	-	-	-	-	-	-	-	-	+

B.



IL8_Gpa33_JM8N4 clone	-	1	12	15	20	21	22	26	28	31	34	35	-
WT JM8.N4 DNA	-	-	-	-	-	-	-	-	-	-	-	-	+
NeoF/ NeoR primers	-	+	+	+	+	+	+	+	+	+	+	+	+
1 Kb DNA ladder (0.5 µg)	+	-	-	-	-	-	-	-	-	-	-	-	-

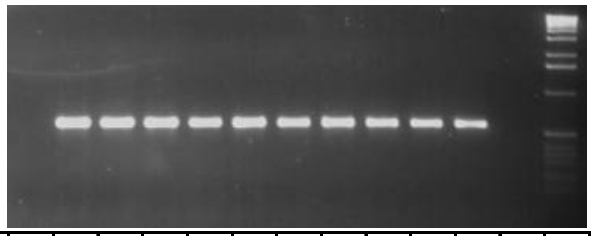
C.



IL8_Gpa33_JM8N4 clone	-	1	12	-	-	1	12	-	-	1	12	-	-
WT JM8.N4 DNA	-	-	-	+	-	-	-	+	-	-	-	+	-
PCR mastermix alone	-	-	-	-	+	-	-	-	+	-	-	-	+
NeoF/ A33pR2 primers	-	+	+	+	+	-	-	-	-	-	-	-	-
IL8-7/ A33pR2 primers	-	-	-	-	-	+	+	+	+	-	-	-	-
IL8-7/ IL8-8 primers	-	-	-	-	-	-	-	-	-	+	+	+	+
1 Kb DNA ladder (0.5 µg)	+	-	-	-	-	-	-	-	-	-	-	-	-

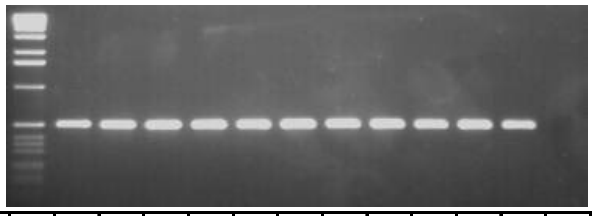
Figure 3.15. Confirmation of homologous recombination of the IL8_A33 targeting vector and JM8N4 ESC DNA at the expected genomic locus by PCR amplification. IL8_Gpa33_JM8N4 clone DNA was amplified with Platinum Taq DNA polymerase High Fidelity using several primer combinations to confirm the correct locus of recombination, as well as the identity of *IL8* itself. These include (**A, C**) *IL8*-specific primers (IL8-7/ IL8-8), (**B**) neomycin resistance cassette primers (NeoF/ NeoR), (**C**) neomycin resistance cassette or *IL8* insert to past the 3' recombination locus (NeoF/ A33pR2 and IL8-7/ A33pR2).

A.



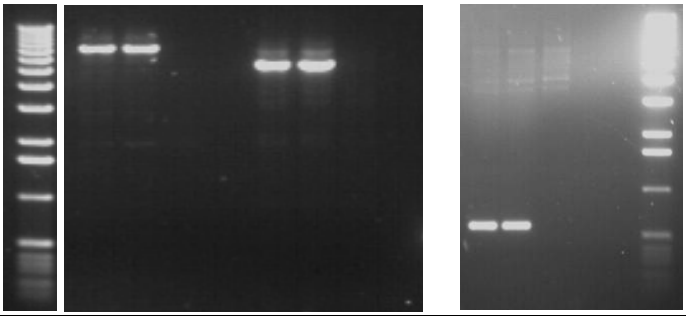
Il10_Gpa33_JM8N4 clone	3	4	9	11	14	15	22	23	24	26	27	-	-
WT JM8.N4 DNA	-	-	-	-	-	-	-	-	-	-	-	+	-
IL10F-Flag/ IL10R-Flag6 primers	+	+	+	+	+	+	+	+	+	+	+	+	-
1 Kb DNA ladder (0.5 µg)	-	-	-	-	-	-	-	-	-	-	-	-	+

B.



Il10_Gpa33_JM8N4 clone	-	3	4	9	11	14	15	22	23	24	26	27	-
WT JM8.N4 DNA	-	-	-	-	-	-	-	-	-	-	-	-	+
NeoF/ NeoR primers	-	+	+	+	+	+	+	+	+	+	+	+	+
1 Kb DNA ladder (0.5 µg)	+	-	-	-	-	-	-	-	-	-	-	-	-

C.



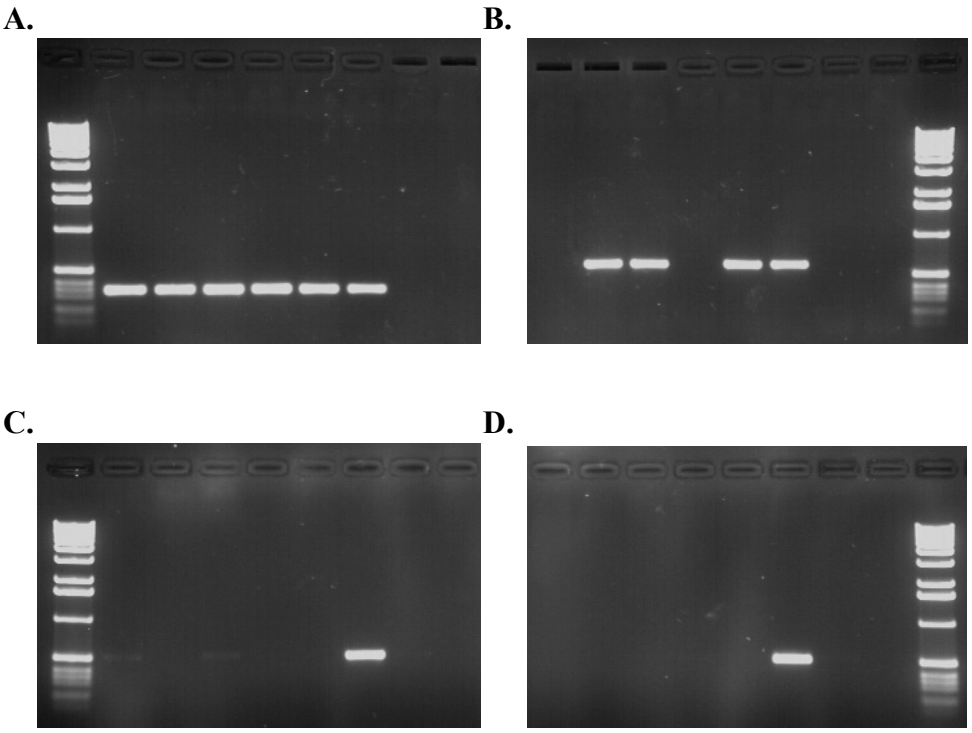
Il10_Gpa33_JM8N4 clone	-	3	4	-	-	3	4	-	-		3	4	-	-	-
WT JM8.N4 DNA	-	-	-	+	-	-	-	+	-		-	-	+	-	-
PCR mastermix alone	-	-	-	-	+	-	-	-	+		-	-	-	+	-
NeoF/ A33pR2 primers	-	+	+	+	+	-	-	-	-		-	-	-	-	-
ILl0F-Flag/ A33pR2 primers	-	-	-	-	-	+	+	+	+		-	-	-	-	-
IL10F-Flag/ IL10R-Flag6 primers	-	-	-	-	-	-	-	-	-		+	+	+	+	-
1 Kb DNA ladder (0.5 µg)	+	-	-	-	-	-	-	-	-		-	-	-	-	+

Figure 3.16. Confirmation of homologous recombination of the Il10_A33 targeting vector and JM8N4 ESC DNA at the expected genomic locus by PCR amplification. Il10_Gpa33_JM8N4 clone DNA was amplified with Platinum Taq DNA polymerase High Fidelity using several primer combinations to confirm the correct locus of recombination, as well as the identity of *Il10* itself. These include (**A**) *Il10*-specific primers (IL10F-Flag/ IL10R-Flag6), (**B**) neomycin resistance cassette primers (NeoF/ NeoR) and (**C**) neomycin resistance cassette or *Il10* insert to past the 3' recombination locus (NeoF/ A33pR2 and IL10F-Flag/ A33pR2).

3.3.9 Confirmation of functional loxP sites within the IL8_Gpa33_JM8N4 and Il10_Gpa33_JM8N4 clones in vitro

Following transformation of the Cre-puromycin plasmid into IL8_Gpa33_JM8N4 and Il10_Gpa33_JM8N4 ESC clones, 48 colonies were picked for selection. Compared to the IL8_pA33LSL or Il10_pA33LSL targeting vector transformations, there were fewer high quality colonies for picking; some colonies with differentiating cells around the edges were picked to make up the numbers and give the best chance of finding at least one with a deleted neomycin resistance gene. As a result, only 11 and 22 of the colonies picked from the IL8_Gpa33_JM8N4 and Il10_Gpa33_JM8N4 ESC transformations, respectively, grew to the desired confluency following clonal expansion; this is expected as differentiating cells lose their pleuripotent ability. Figure 3.17 shows the agarose gel electrophoresis of PCR products using gene-specific (A, B) and neomycin resistance cassette primers (C, D) for both IL8 and Il10 alleles, respectively. A clone was deemed positive if amplification occurred with the gene-specific but not neomycin resistance cassette primer pairs. All five IL8_Gpa33_JM8N4_Cre ESC clones had amplification of the IL8 insert and three of these had no amplification of the neomycin resistance cassette. Three out of five Il10_Gpa33_JM8N4_Cre ESC clones had amplification of the expected gene insert and all of these had no amplification of the neomycin resistance cassette. The two Il10_Gpa33_JM8N4_Cre ESC clones that demonstrated amplification of neither the gene insert nor the neomycin resistance cassette either did not have enough DNA for amplification or else more likely, something in the reaction inhibited the amplification. A WT JM8.N4 positive reaction would have been beneficial for this determination. Further optimization of these two samples was not performed because the other three clones demonstrated the functionality of the Cre/loxP system. The parent IL8_Gpa33_JM8N4 and Il10_Gpa33_JM8N4 ESC clones, as well as WT JM8N4 DNA were included as controls; the parent clones had both amplification of the gene insert as well as the neomycin resistance cassette and the WT DNA did not amplify with either set of primers.

The loxP sites are therefore intact in both the IL8_Gpa33_JM8N4 and Il10_Gpa33_JM8N4 ESC and Cre-mediated deletion of the neomycin resistance cassette functions as expected. It was therefore assumed that this process would occur in the same manner *in vivo* upon breeding of IL8 or Il10 knock-in mice with Cre-expressing and the subsequent expression of Cre-recombinase.



IL8_A33_JM8N4-12_Cre clone	-	1	13	20	27	30	-	-	-		-	-	-	-	-	-	-	-	-
IL8_A33_JM8N4-12	-	-	-	-	-	-	+	-	-		-	-	-	-	-	-	-	-	-
Il10_A33_JM8N4-24_Cre clone	-	-	-	-	-	-	-	-	-		6	14	15	16	18	-	-	-	-
Il10_A33_JM8N4-24	-	-	-	-	-	-	-	-	-		-	-	-	-	-	+	-	-	-
JM8.N4 WT DNA	-	-	-	-	-	-	-	+	-		-	-	-	-	-	-	+	-	-
PCR mastermix alone	-	-	-	-	-	-	-	-	+		-	-	-	-	-	-	-	+	-
1 Kb DNA ladder (0.5 µg)	+	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	+

Figure 3.17. ***In vitro*** **Cre-mediated deletion of the** ***IL8*** **and** ***Il10*** **targeted allele neomycin resistance cassette.** The initial IL8_Gpa33_JM8N4-12 and Il10_Gpa33_JM8N4-24 clones microinjected were transformed with a Cre-expressing plasmid, which also contains puromycin resistance for selection. Five colonies from each clone were tested by amplification using Platinum PCR Supermix with gene-specific or neomycin resistance cassette primers; the parent ESC clone, WT JM8.N4 and PCR mastermix alone were included as controls. All five of the IL8_Gpa33_JM8N4-12_Cre clones had amplification of the *IL8* insert with IL8-7 and IL8-8 primers (**A**) but only three had full deletion of the neomycin resistance cassette by the lack of product with NeoF and NeoR primers (**C**). Three of the Il10_Gpa33_JM8N4-24_Cre clones had amplification of the *Il10* insert with IL10F-Flag and IL10F-Flag6 primers (**B**) and all of the three had deletion of the neomycin resistance cassette by lack of product with NeoF and NeoR primers (**D**).

3.3.10 IL8_Gpa33_JM8N4 and Il10_Gpa33_JM8N4 ESC clone microinjection

The single-cell suspension submitted for microinjection contained very round and bright ESC. The WTSI technician who performed each microinjection additionally confirmed the quality of the ESC. IL8_Gpa33_JM8N4 ESC underwent two rounds of microinjection with two different positive clones to ensure the best chance of germline transmission (i.e. the microinjections were carried out on separate days using different clones with different passage numbers). Table 3.7 summarizes the two microinjections performed for the *IL8* gene construct including the ESC clone, number of ESC microinjected, the number of embryos implanted and

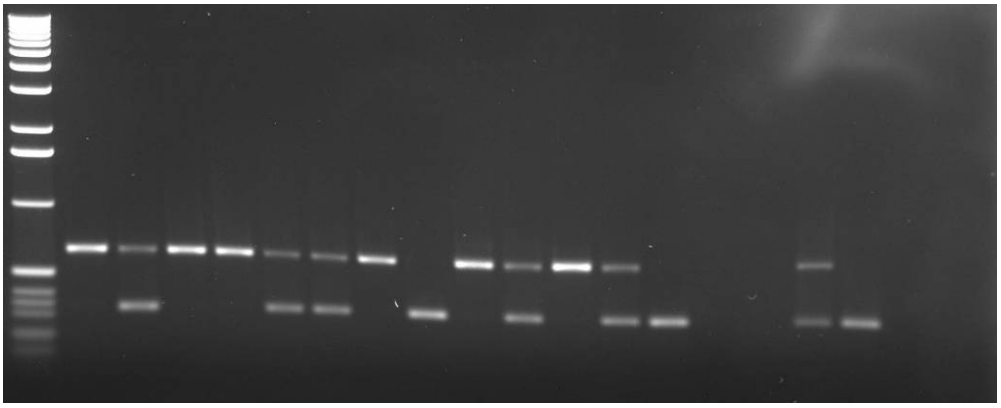
outcome for each recipient mouse. The first microinjection (MI2049) did not result in any of the three recipient mice getting pregnant. The repeat IL8_Gpa33_JM8N4 ESC microinjection (MI2098, Colony name LRHE: Linda Rehaume Human Interleukin-Eight) resulted in three male F0 offspring LRHE3.1a, LRHE4.1a, LRHE4.1b having 90%, 90% and 50% chimerism, respectively as evaluated by staff of the WTSI RSF. LRHE3.1a developed a growth/grey film over its left eye, but was otherwise healthy. LRHE4.1a developed the growth/grey film over both eyes and a bloated abdomen and was culled at 4.9 weeks of age due to the enlarged abdomen. A wild type, age-matched CBLN mouse was also culled as a control. For both mice, the eye, kidney, spleen and portions of the small and large intestine were removed, fixed in formaldehyde and embedded in wax. Upon dissection, the kidneys of the LRHE4.1a mouse were obviously swollen and full of liquid. Haematoxylin and eosin stained sections showed relatively normal kidney cellular integrity in areas surrounding the cavity that contained the liquid. The left eye of LRHE4.1a showed abnormal formation with loss of symmetry and extension of the sclera over the cornea. No further analysis of LRHE4.1a tissue was performed. It is known that sometimes injecting too many ESC can overwhelm the embryo resulting in abnormalities (*personal communication,* D. Adams), and this was therefore considered a sporadic finding.

Table 3.7. IL8_Gpa33_JM8N4 and Il10_Gpa33_JM8N4 microinjection and chimera generation summary. The embryo (3.5 dpc blastocyst) microinjection and implantation was performed by members of Team 121, WTSI. The embryo strain is C57BL/6J-Tyr<c-Brd> (100%), and the recipient strain is CBA (50%) and C57BL/6J (50%). dpc, days post coitus.

Colony Prefix	ID	MI Date	ESC Clone Injected	ESC Injected/ Embryo	Embryos Implanted/ Recipient	Recipient Status	Chimeras (M/F)	% Chimera
LRIE	2049	2009/05/19	IL8_Gpa33_JM8N4-21	5-15	11, 10, 10	none pregnant	n/a	n/a
LRHE	2097	2009/07/06	IL8_Gpa33_JM8N4-12	5-20	11, 11, 11, 11	all pregnant	3M	90, 90, 50
LRIT	2050	2009/05/20	Il10_Gpa33_JM8N4-24	5-15	11, 11, 11	all pregnant	2F	70, 50
LRMT	2098	2009/09/10	Il10_Gpa33_JM8N4-4	5-15	11, 10, 10	all pregnant	3M	90, 90, 30

The remaining two F0 IL8 chimera males (LRHE3.1a, 90%; LRHE4.1b, 50%) were each mated with two CBLN female mice, leading to each IL8 chimera male fathering two litters. LRHE3.1a fathered fifteen F1 pups in the LRHE5.1 and LRHE6.1 litters and LRHE4.1b fathered sixteen F1 pups in the LRHE7.1 and LRHE8.1 litters. None of the pups from LRHE4.1b were heterozygous for the IL8 gene. However there were six heterozygous pups from the first LRHE3.1a litters giving a germline transmission rate of 40%, which is close to the theoretical 50%, but slightly lower than the 62% reported for JM8.N4 ESC (189). Subsequently, LRHE3.1a

fathered eight pups in LRHE5.2 of which only one was heterozygous and unfortunately this pup was found dead at 2.5 weeks of age (w). LRHE3.1a also fathered litters LRHE6.2b and LRHE6.2c, which produced eight and three pups, respectively, of which five were heterozygous. LRHE4.1b fathered another LRHE8.2 litter producing ten more F1 pups, but none of these were heterozygous, therefore germline transmission was not successful with the 50% IL8 chimera. Pairings of F1 heterozygous hIL8:c/+ offspring were set up to generate mice homozygous for the IL8 insert, hIL8:c/-, and subsequently homozygous pairings have been set up. Figure 3.18 is an example of the multiplex genotyping for the litters LRHE15.1 and LRHE13.1, which is representative of all LRHE offspring. Table 3.8 summarizes the LRHE colony, which includes all mice regardless of fate, as of 2010/04/20.



LRHE15.1 litter	-	a	b	c	d	e	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LRHE13.1 litter	-	-	-	-	-	-	a	b	c	d	e	f	g	h	-	-	-	-	-	-
IL8_A33_JM8N4 ESC DNA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
JM8.N4 WT DNA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
Tweezer Control	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
PCR mastermix	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
NF-H_2O	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
1 Kb DNA ladder (0.5 μg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Genotype	n/a	Hom	Het	Hom	Hom	Het	Het	Hom	WT	Hom	Het	Hom	Het	WT	n/a	n/a	Het	WT	n/a	n/a

Figure 3.18. IL8 F1 (LRHE) offspring genotyping. Genomic DNA was amplified with Platinum PCR Supermix in a multiplex reaction consisting of an *IL8* insert forward primer (IL8-7), a *Gpa33* exon-7 forward primer (A33-E7) and a reverse *Gpa33* 3'-UTR primer (A33-3U). The PCR product obtained from the wildtype allele is 331 bp and that of the hIL8:c allele is 664 bp. Amplification of both products occurs for heterozygous mice, the 664 bp product for homozygous mice and the 331 bp product for wildtype mice. This genotyping is representative of all LRHE offspring.

Table 3.8. The IL8 LRHE mouse colony. Three male chimeras were produced out of the five F0 pups born to the four recipient mice (M00297577-9, M00297590). One of the 90% chimeras was culled due to sickness and the 50% chimera was culled due to unsuccessful germline transmission. The remaining 90% chimera (LRHE3.1a) had successful germline transmission, determined by PCR, upon mating with CBLN females resulting in an established breeding colony. The age of each mouse or age at time of death is given in weeks. The colony is current as of 2010/04/20. U, unsexed; F, female; M, male.

Name	Generation	Sex	Coat	Genotype	Status	Age	Mating	Full Strain Genetic Background
LRHE3.1a	F0	M	Albino/Black (90%)	hIL8:chimera 90%	Alive	38.7	LRHE6	C57BL/6N(100%)
LRHE5.1b	F1	M	Black	hIL8:c/+	Alive	28.4	LRHE9	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE5.1e	F1	F	Black	hIL8:c/+	Alive	28.4	LRHE9	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE5.1c	F1	M	Black	hIL8:c/+	Alive	28.4	LRHE10	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.1d	F1	F	Black	hIL8:c/+	Alive	25.9	LRHE10	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.1b	F1	M	Black	hIL8:c/+	Alive	25.9	LRHE11	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.1e	F1	F	Black	hIL8:c/+	Alive	25.9	LRHE11	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.2a	F1	M	Black	hIL8:c/+	Alive	18.7	LRHE12	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.2f	F1	F	Black	hIL8:c/+	Alive	18.7	LRHE12	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.2b	F1	M	Black	hIL8:c/+	Alive	18.7	LRHE13	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.2h	F1	F	Black	hIL8:c/+	Alive	18.7	LRHE13	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.1g	F2	F	Black	hIL8:c/+	Alive	15.7	LRHE14	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.1a	F2	M	Black	hIL8:c/-	Alive	15.7	LRHE14	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.1b	F2	M	Black	hIL8:c/-	Alive	15.7	LRHE15	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.1f	F2	F	Black	hIL8:c/+	Alive	15.7	LRHE15	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE9.2d	F2	F	Black	hIL8:c/-	Alive	11.1	LRHE16	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE9.2c	F2	M	Black	hIL8:c/-	Alive	11.1	LRHE17	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE9.2e	F2	F	Black	hIL8:c/-	Alive	11.1	LRHE17	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE13.1b	F2	M	Black	hIL8:c/-	Alive	6.6	LRHE18	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE13.1f	F2	F	Black	hIL8:c/-	Alive	6.6	LRHE18	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE13.1d	F2	M	Black	hIL8:c/-	Alive	6.6	LRHE19	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE15.1c	F3	F	Black	hIL8:c/-	Alive	6.9	LRHE19	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE15.1a	F3	M	Black	hIL8:c/-	Alive	6.9	LRHE20	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE15.1d	F3	F	Black	hIL8:c/-	Alive	6.9	LRHE20	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.3a	F2	M	Black	hIL8:c/+	Alive	3.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.3b	F2	M	Black	hIL8:c/+	Alive	3.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.3c	F2	M	Black	hIL8:c/+	Alive	3.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.3d	F2	F	Black	hIL8:c/+	Alive	3.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.3e	F2	F	Black	hIL8:c/+	Alive	3.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE15.1b	F3	M	Black	hIL8:c/+	Alive	6.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.4b	F1	M	Black	hIL8:c/+	Alive	5.6		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.4d	F1	F	Black	hIL8:c/+	Alive	5.6		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE15.2a	F3	U	Black	hIL8:untested	Alive	1.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE15.2b	F3	U	Black	hIL8:untested	Alive	1.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE15.2c	F3	U	Black	hIL8:untested	Alive	1.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE15.2d	F3	U	Black	hIL8:untested	Alive	1.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE15.2e	F3	U	Black	hIL8:untested	Alive	1.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE15.2f	F3	U	Black	hIL8:untested	Alive	1.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE15.2g	F3	U	Black	hIL8:untested	Alive	1.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.4a	F1	M	Black	+/+	To be culled	5.6		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.4c	F1	F	Black	+/+	To be culled	5.6		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.1e	F2	F	Black	+/+	Culled	3		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.2b	F2	M	Black	+/+	Culled	6.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.2h	F2	M	Black	+/+	Culled	6.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.2i	F2	F	Black	+/+	Culled	6.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE13.1c	F2	M	Black	+/+	Culled	5.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE13.1h	F2	F	Black	+/+	Culled	5.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE5.1a	F1	M	Black	+/+	Culled	10.6		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE5.1d	F1	M	Black	+/+	Culled	6.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE5.1f	F1	F	Black	+/+	Culled	10.6		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE5.1g	F1	F	Black	+/+	Culled	6.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE5.2a	F1	M	Black	+/+	Culled	3.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE5.2b	F1	M	Black	+/+	Culled	3.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)

Name	Generation	Sex	Coat	Genotype	Status	Age	Mating	Full Strain Genetic Background
LRHE5.2c	F1	F	Black	+/+	Culled	3.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE5.2d	F1	F	Black	+/+	Culled	3.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE5.2e	F1	F	Black	+/+	Culled	3.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE5.2f	F1	F	Black	+/+	Culled	3.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE5.2h	F1	F	Black	+/+	Culled	3.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.1a	F1	M	Black	+/+	Culled	4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.1c	F1	F	Black	+/+	Culled	4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.1f	F1	F	Black	+/+	Culled	4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.1g	F1	F	Black	+/+	Culled	4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.1h	F1	F	Black	+/+	Culled	4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.2c	F1	M	Black	+/+	Culled	5.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.2e	F1	F	Black	+/+	Culled	5.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.2g	F1	F	Black	+/+	Culled	5.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE7.1a	F1	M	Black	+/+	Culled	6.7		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE7.1b	F1	M	Black	+/+	Culled	6.7		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE7.1c	F1	M	Black	+/+	Culled	6.7		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE7.1d	F1	M	Black	+/+	Culled	6.7		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE7.1e	F1	M	Black	+/+	Culled	6.7		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE7.1f	F1	M	Black	+/+	Culled	6.7		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE7.1g	F1	F	Black	+/+	Culled	7.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE8.1a	F1	M	Black	+/+	Culled	6.7		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE8.1b	F1	M	Black	+/+	Culled	6.7		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE8.1c	F1	M	Black	+/+	Culled	6.7		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE8.1d	F1	M	Black	+/+	Culled	6.7		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE8.1e	F1	M	Black	+/+	Culled	6.7		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE8.1f	F1	M	Black	+/+	Culled	6.7		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE8.1g	F1	M	Black	+/+	Culled	6.7		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE8.1h	F1	F	Black	+/+	Culled	7.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE8.1i	F1	F	Black	+/+	Culled	7.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE8.2a	F1	M	Black	+/+	Culled	3.1		C57BL/6NTac/USA(50%);C57BL/6N(50%)
LRHE8.2b	F1	M	Black	+/+	Culled	3.1		C57BL/6NTac/USA(50%);C57BL/6N(50%)
LRHE8.2c	F1	M	Black	+/+	Culled	3.1		C57BL/6NTac/USA(50%);C57BL/6N(50%)
LRHE8.2d	F1	M	Black	+/+	Culled	3.1		C57BL/6NTac/USA(50%);C57BL/6N(50%)
LRHE8.2e	F1	M	Black	+/+	Culled	3.1		C57BL/6NTac/USA(50%);C57BL/6N(50%)
LRHE8.2f	F1	F	Black	+/+	Culled	3.1		C57BL/6NTac/USA(50%);C57BL/6N(50%)
LRHE8.2g	F1	F	Black	+/+	Culled	3.1		C57BL/6NTac/USA(50%);C57BL/6N(50%)
LRHE8.2h	F1	F	Black	+/+	Culled	3.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE8.2i	F1	F	Black	+/+	Culled	3.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE8.2j	F1	F	Black	+/+	Culled	3.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.1c	F2	M	Black	hIL8:c/+	Culled	12		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.1d	F2	M	Black	hIL8:c/+	Culled	12		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.2a	F2	M	Black	hIL8:c/+	Culled	6.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.2c	F2	M	Black	hIL8:c/+	Culled	6.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.2f	F2	M	Black	hIL8:c/+	Culled	6.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.2g	F2	M	Black	hIL8:c/+	Culled	6.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.2j	F2	F	Black	hIL8:c/+	Culled	6.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE13.1a	F2	M	Black	hIL8:c/+	Culled	5.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE13.1e	F2	F	Black	hIL8:c/+	Culled	5.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE13.1g	F2	F	Black	hIL8:c/+	Culled	5.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE15.1e	F3	F	Black	hIL8:c/+	Culled	6.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE5.2g	F1	F	Black	hIL8:c/+	Found Dead	2.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.2d	F1	M	Black	hIL8:c/+	Culled	15		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE9.2a	F2	M	Black	hIL8:c/+	Culled	7.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE9.2b	F2	M	Black	hIL8:c/+	Culled	7.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.2d	F2	M	Black	hIL8:c/-	Culled	6.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE10.2e	F2	M	Black	hIL8:c/-	Culled	6.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE3.1b	F0	U	Albino	hIL8:chimera	Culled	3.9		C57BL/6N(100%)
LRHE4.1c	F0	U	Albino	hIL8:chimera	Culled	2.6		C57BL/6N(100%)
LRHE4.1b	F0	M	Albino/Black (50%)	hIL8:chimera 50%	Culled Sick	32.1		C57BL/6N(100%)
LRHE4.1a	F0	M	Albino/Black (90%)	hIL8:chimera 90%	Culled Sick	4.9		C57BL/6N(100%)
LRHE11.1a	F2	U	Black	hIL8:untested	Missing	0.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE11.1b	F2	U	Black	hIL8:untested	Missing	0.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE11.1c	F2	U	Black	hIL8:untested	Missing	0.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE11.1d	F2	U	Black	hIL8:untested	Missing	0.4		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE11.2a	F2	U	Black	hIL8:untested	Missing	1.7		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE14.1a	F3	U	Black	hIL8:untested	Found Dead	0.3		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE14.1b	F3	U	Black	hIL8:untested	Found Dead	0.3		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE15.1f	F3	U	Black	hIL8:untested	Missing	1.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE15.1g	F3	U	Black	hIL8:untested	Missing	1.1		C57BL/6N(50%);C57BL/6NTac/USA(50%)

Name	Generation	Sex	Coat	Genotype	Status	Age	Mating	Full Strain Genetic Background
LRHE16.1a	F3	U	Black	hIL8:untested	Found Dead	0.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.3a	F1	U	Black	hIL8:untested	Missing	0.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.3b	F1	U	Black	hIL8:untested	Missing	0.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE6.3c	F1	U	Black	hIL8:untested	Missing	0.9		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE9.1a	F2	U	Black	hIL8:untested	Missing	0.6		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE9.1b	F2	U	Black	hIL8:untested	Missing	0.6		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE9.1c	F2	U	Black	hIL8:untested	Found Dead	0		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE9.1d	F2	U	Black	hIL8:untested	Found Dead	0		C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRHE9.1e	F2	U	Black	hIL8:untested	Found Dead	0		C57BL/6N(50%);C57BL/6NTac/USA(50%)
M00297577		F		+/+	Culled			CBA/Wtsi;C57BL/6JIco
M00297578		F		+/+	Culled			CBA/Wtsi;C57BL/6JIco
M00297579		F		+/+	Culled			CBA/Wtsi;C57BL/6JIco
M00297590		F		+/+	Culled			CBA/Wtsi;C57BL/6JIco
CBLN907.2e		F	Black	+/+	Alive	37.9	LRHE5	C57BL/6NTac/USA(100%)
CBLN907.2f		F	Black	+/+	Alive	37.9	LRHE6	C57BL/6NTac/USA(100%)
CBLN907.2h		F	Black	+/+	Culled	34.1		C57BL/6NTac/USA(100%)
CBLN907.2i		F	Black	+/+	Culled	34.1		C57BL/6NTac/USA(100%)

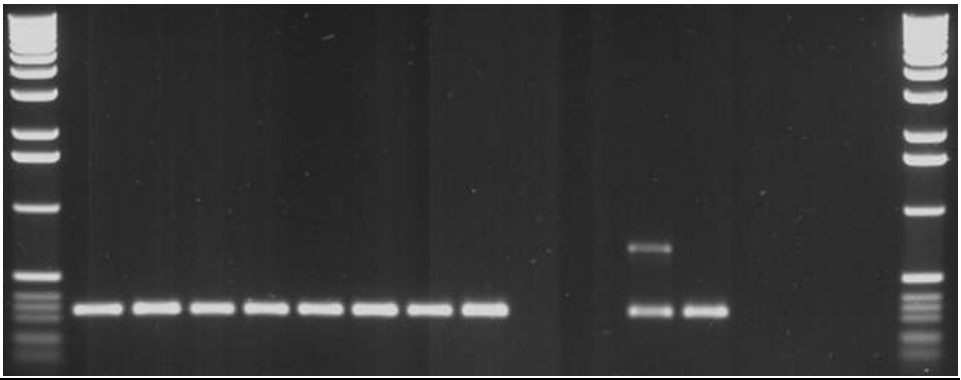
Il10_Gpa33_JM8N4 ESC also underwent two rounds of microinjection with two different positive clones. Table 3.7 summarizes the two microinjections performed for the *Il10* gene construct including the number of ESC microinjected, the number of embryos implanted and outcome for each recipient mouse. The first microinjection (ID 2050, Colony name LRIT: Linda Rehaume Interleukin-Ten) resulted in two female F0 offspring LRIT2.1a and LRIT2.1b having 70% and 50% chimerism, respectively. LRIT2.1a and LRIT2.1b were both mistakenly paired with C57BL6c-c- (prefix CALB) males, which are albino, and each had a litter of five and seven offspring, respectively, but in addition to them not being the desired genetic background, they were all albino and therefore culled. The chimera females were then both paired with male CBLN males. LRIT2.1a produced an F1 litter LRHE6.1 containing seven pups but none of these were heterozygous for the exogenous *Il10* gene; six of these pups were genotyped but one went missing so genotyping was not possible. LRIT2.1a produced two more litters LRHE6.2 and LRHE6.2 with four and two pups, respectively, but none of those were heterozygous for the mIL10:c allele. Unfortunately LRIT2.1a was found dead at 30.8w and thought to have died during littering. This colony was terminated due to several factors. The chimeric mice were seven months old, which is getting too old for breeding, there was no germline transmission of the mIL10:c allele from the 70% chimera (LRIT2.1a), which then died during what had already been intended as her final littering, and the 50% chimera (LRIT2.1b) did not produce any litters with the CBLN male. Table 3.9 summarizes the LRIT colony.

Table 3.9. The Il10 LRIT mouse colony. Two female chimeras were produced out of the eleven F0 pups born to the three recipient mice (M00267240-2). The chimeras were first mistakenly mated with CALB males producing only albino F1 offspring, which were not genotyped by PCR. Mating of the chimeras with CBLN males produced only wildtype (+/+) offspring, determined by PCR. The age of death for each mouse is given in weeks. U, unsexed; F, female; M, male.

Name	Generation	Sex	Coat	Genotype	Fate	Age	Full Strain Genetic Background
LRIT1.1a	F0	U	Albino	mIL10:chimera	Culled	1.4	C57BL/6N(100%)
LRIT1.1b	F0	U	Albino	mIL10:chimera	Culled	1.4	C57BL/6N(100%)
LRIT1.1c	F0	U	Albino	mIL10:chimera	Culled	1.4	C57BL/6N(100%)
LRIT1.1d	F0	U	Albino	mIL10:chimera	Culled	1.4	C57BL/6N(100%)
LRIT1.1e	F0	U	Albino	mIL10:chimera	Culled	1.4	C57BL/6N(100%)
LRIT1.1f	F0	U	Albino	mIL10:chimera	Culled	1.4	C57BL/6N(100%)
LRIT2.1a	F0	F	Albino/Black (70%)	mIL10:chimera 70%	Found Dead	30.7	C57BL/6N(100%)
LRIT2.1b	F0	F	Albino/Black (50%)	mIL10:chimera 50%	Culled	31.3	C57BL/6N(100%)
LRIT2.1c	F0	U	Albino	mIL10:chimera	Culled	3.6	C57BL/6N(100%)
LRIT2.1d	F0	U	Albino	mIL10:chimera	Culled	3.6	C57BL/6N(100%)
LRIT2.1e	F0	U	Albino	mIL10:chimera	Culled	3.6	C57BL/6N(100%)
LRIT4.1a	F1	U	Albino	mIL10:untested	Culled	1.4	C57BL/6J-Tyr<c-Brd>(50%);C57BL/6N(50%)
LRIT4.1b	F1	U	Albino	mIL10:untested	Culled	1.4	C57BL/6J-Tyr<c-Brd>(50%);C57BL/6N(50%)
LRIT4.1c	F1	U	Albino	mIL10:untested	Culled	1.4	C57BL/6J-Tyr<c-Brd>(50%);C57BL/6N(50%)
LRIT4.1d	F1	U	Albino	mIL10:untested	Culled	1.4	C57BL/6J-Tyr<c-Brd>(50%);C57BL/6N(50%)
LRIT4.1e	F1	U	Albino	mIL10:untested	Culled	1.4	C57BL/6J-Tyr<c-Brd>(50%);C57BL/6N(50%)
LRIT5.1a	F1	U	Albino	mIL10:untested	Culled	1.3	C57BL/6J-Tyr<c-Brd>(50%);C57BL/6N(50%)
LRIT5.1b	F1	U	Albino	mIL10:untested	Culled	1.3	C57BL/6J-Tyr<c-Brd>(50%);C57BL/6N(50%)
LRIT5.1c	F1	U	Albino	mIL10:untested	Culled	1.3	C57BL/6J-Tyr<c-Brd>(50%);C57BL/6N(50%)
LRIT5.1d	F1	U	Albino	mIL10:untested	Culled	1.3	C57BL/6J-Tyr<c-Brd>(50%);C57BL/6N(50%)
LRIT5.1e	F1	U	Albino	mIL10:untested	Culled	1.3	C57BL/6J-Tyr<c-Brd>(50%);C57BL/6N(50%)
LRIT5.1f	F1	U	Albino	mIL10:untested	Culled	1.3	C57BL/6J-Tyr<c-Brd>(50%);C57BL/6N(50%)
LRIT5.1g	F1	U	Albino	mIL10:untested	Culled	1.3	C57BL/6J-Tyr<c-Brd>(50%);C57BL/6N(50%)
LRIT6.1a	F1	M	Black	+/+	Culled	4.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRIT6.1b	F1	M	Black	+/+	Culled	4.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRIT6.1c	F1	M	Black	+/+	Culled	4.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRIT6.1d	F1	M	Black	+/+	Culled	4.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRIT6.1e	F1	F	Black	+/+	Culled	4.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRIT6.1f	F1	F	Black	+/+	Culled	4.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRIT6.1g	F1	F	Black	mIL10:untested	Missing	2.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRIT6.2a	F1	M	Black	+/+	Culled	6.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRIT6.2b	F1	M	Black	+/+	Culled	6.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRIT6.2c	F1	M	Black	+/+	Culled	6.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRIT6.2d	F1	M	Black	+/+	Culled	6.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRIT6.3a	F1	F	Black	+/+	Culled	2.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRIT6.3b	F1	F	Black	+/+	Culled	2.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
M00267240		F		+/+	Culled		CBA/Wtsi;C57BL/6JIco
M00267241		F		+/+	Culled		CBA/Wtsi;C57BL/6JIco
M00267242		F		+/+	Culled		CBA/Wtsi;C57BL/6JIco
CALB2596.1d	F22	M	Albino	+/+	Culled	11	C57BL/6J-Tyr<c-Brd>(100%)
CALB2596.1e	F22	M	Albino	+/+	Culled	11	C57BL/6J-Tyr<c-Brd>(100%)
CBLN779.2c		M	Black	+/+	Culled	22.3	C57BL/6NTac/USA(100%)
CBLN830.3a		M	Black	+/+	Culled	23.1	C57BL/6NTac/USA(100%)

The second Il10 Gpa33_JM8N4 ESC microinjection (ID 2098) was more successful in producing three males LRMT1.1a, LRMT1.1b, LRMT1.1c having 90%, 90% and 30% chimerism, respectively. LRMT1.1a and LRMT1.1b were put with two CBLN females each but the chances of LRMT1.1c transmitting the targeted allele to offspring is low, as observed for the 50% IL8 chimera, so this mouse was only paired with one CBLN female. Interestingly, only the CBLN paired with the 30% chimera (LRMT1.1c) produced any offspring from these initial matings; one pup was born but found dead at 0.8w. All three chimeras were put with younger

CBLN females, again two each for the 90% chimeras and one for the 30%, on two different occasions but again only the 30% chimera produced any offspring and thus the two 90% chimeras were presumed sterile. Figure 3.19 is an example of the multiple genotyping for the litter LRMT18.1, which is representative of both LRIT and LRMT offspring. Table 3.10 summarizes the LRMT colony, which has been terminated. Repeat microinjections were carried out on 2010/04/23 for two more Il10_Gpa33_JM8N4 ESC clones.



LRMT18.1 litter	-	a	b	c	d	e	f	g	h	-	-	-	-	-	-	-	-
Il10_A33_JM8N4 ESC DNA	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
JM8.N4 WT DNA	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
Tweezer Control	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
PCR mastermix	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-
NF-H_2O	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
5X orange G loading buffer	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
1 Kb DNA ladder (0.5 μg)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
Genotype	n/a	WT	WT	WT	WT	WT	WT	WT	WT	n/a	n/a	Het	WT	n/a	n/a	n/a	n/a

Figure 3.19. Il10 F0 chimera (LRMT) offspring genotyping. Genomic DNA was amplified with Platinum PCR Supermix in a multiplex reaction consisting of an *Il10* insert forward primer (IL10-2), a *Gpa33* exon-7 forward primer (A33-E7) and a reverse *Gpa33* 3'-UTR primer (A33-3U). The wildtype allele is 331 bp and the mIL10:c allele is 686 bp. All eight mice from this litter (LRMT18.1a-h) were wildtype. This genotyping is representative of all LRIT and LRMT offspring.

Table 3.10. The Il10 LRMT mouse colony. Three male chimeras were produced out of the eleven F0 pups born to the three recipient mice (M00320092, M00320120, M00320122). The two 90% chimeras were each trio'd with two CBLN females three separate times, but did not produce any offspring so presumed sterile. The 30% chimera was paired with one CBLN female three times but only produced wildtype (+/+) offspring, determined by PCR. The age of death for each mouse is given in weeks. U, unsexed; F, female; M, male.

Name	Generation	Sex	Coat	Genotype	Fate	Age	Full Strain
LRMT1.1a	F0	M	Albino/Black (90%)	mIL10:chimera 90%	To be culled	29	C57BL/6N(100%)
LRMT1.1b	F0	M	Albino/Black (90%)	mIL10:chimera 90%	To be culled	29	C57BL/6N(100%)
LRMT1.1c	F0	M	Albino/Black (30%)	mIL10:chimera 30%	To be culled	29	C57BL/6N(100%)
LRMT1.1d	F0	U	Albino	mIL10:chimera	Culled	1.4	C57BL/6N(100%)
LRMT1.1e	F0	U	Albino	mIL10:chimera	Culled	1.4	C57BL/6N(100%)
LRMT1.1f	F0	U	Albino	mIL10:chimera	Culled	1.4	C57BL/6N(100%)
LRMT1.1g	F0	U	Albino	mIL10:chimera	Culled	1.4	C57BL/6N(100%)
LRMT8.1a	F1	U	Black	mIL10:untested	Found Dead	0.7	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT18.1a	F1	M	Black	+/+	Culled	3.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT18.1b	F1	M	Black	+/+	Culled	3.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT18.1c	F1	M	Black	+/+	Culled	3.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT18.1d	F1	M	Black	+/+	Culled	3.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT18.1e	F1	M	Black	+/+	Culled	3.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT18.1f	F1	F	Black	+/+	Culled	3.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT18.1g	F1	F	Black	+/+	Culled	3.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT18.1h	F1	F	Black	+/+	Culled	3.3	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT19.1a	F1	M	Black	+/+	Culled	3.1	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT19.1b	F1	F	Black	+/+	Culled	3.1	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT19.1c	F1	F	Black	+/+	Culled	3.1	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT19.1d	F1	F	Black	+/+	Culled	3.1	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT19.1e	F1	F	Black	+/+	Culled	3.1	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT19.1f	F1	F	Black	+/+	Culled	3.1	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT19.1g	F1	F	Black	+/+	Culled	3.1	C57BL/6N(50%);C57BL/6NTac/USA(50%)
LRMT19.1h	F1	F	Black	+/+	Culled	3.1	C57BL/6N(50%);C57BL/6NTac/USA(50%)
M00320092	F1	F	Agouti	+/+	Culled	11.1	C57BL/6JIco(50%);CBA/Wtsi(50%)
M00320120	F1	F	Agouti	+/+	Culled	11.1	CBA/Wtsi(50%);C57BL/6JIco(50%)
M00320122	F1	F	Agouti	+/+	Culled	13.3	CBA/Wtsi(50%);C57BL/6JIco(50%)
CBLN1039.2h		F	Black	+/+	Culled	22.3	C57BL/6NTac/USA(100%)
CBLN1039.2i		F	Black	+/+	Culled	18	C57BL/6NTac/USA(100%)
CBLN1039.2k		F	Black	+/+	Culled	18	C57BL/6NTac/USA(100%)
CBLN1208.1k		F	Black	+/+	Culled	17.4	C57BL/6NTac/USA(100%)
CBLN1370.3h		F	Black	+/+	To be culled	14.6	C57BL/6NTac/USA(100%)
CBLN1421.1c		F	Black	+/+	Culled	12.7	C57BL/6NTac/USA(100%)
CBLN1421.1d		F	Black	+/+	Culled	12.7	C57BL/6NTac/USA(100%)
CBLN1421.1e		F	Black	+/+	Culled	12.7	C57BL/6NTac/USA(100%)
CBLN1421.1f		F	Black	+/+	Culled	12.7	C57BL/6NTac/USA(100%)
CBLN1659.1a		F	Black	+/+	To be culled	15.2	C57BL/6NTac/USA(100%)
CBLN1659.1b		F	Black	+/+	To be culled	15.2	C57BL/6NTac/USA(100%)
CBLN1659.1c		F	Black	+/+	To be culled	15.2	C57BL/6NTac/USA(100%)
CBLN1659.1d		F	Black	+/+	To be culled	15.2	C57BL/6NTac/USA(100%)
CBLN1659.1e		F	Black	+/+	To be culled	15.2	C57BL/6NTac/USA(100%)
CBLN918.4b		F	Black	+/+	Culled	13	C57BL/6NTac/USA(100%)
CBLN931.3c		F	Black	+/+	Culled	17.7	C57BL/6NTac/USA(100%)

All data from microinjection to colony maintenance is freely available on the WTSI Mouse Database. I have included all relevant information in the Figures, Tables, and the Appendix. Additionally the ESCs and mice are freely available to the research community as part of the WTSI data release policy.

3.4 Discussion

To generate mouse models for the study of the roles of IL8 and Il10 in intestinal infection, I set out to create four knock-in mouse strains: two containing IL8 and two containing Il10.

Human intestinal epithelial cells are a major source of IL8 during infection by some enteric pathogens (78, 83), therefore it was the intent of this project to generate mice expressing IL8 from these cells. Il10 is also produced by murine intestinal epithelial cells (190), and even though it is still not clear which cells are the most important source of Il10 (65), there is also strong evidence of Il10 expression in intestinal lymphocytes (174). Regardless of the main cellular sources of IL8 and Il10 during intestinal infection, their presence in the inflammatory milieu basolateral to the epithelial barrier has been confirmed (78, 173, 174).

Many events, including bacterial adhesion and translocation as well as host recognition and signalling, need to occur before intestinal epithelial cell cytokine production occurs. Generating a mouse in which the production of IL8 or Il10 occurs in 'all' cells simultaneously would not only fail to mimic events in man, but would also likely lead to overwhelming inflammation or immunosuppression, respectively, and overshadow the course of infection. As well, the production of either of these cytokines from birth would also likely skew cellular differentiation and development, again making interpretation of the results difficult. Therefore the source of IL8 or Il10 and the timing of their production were both desirable factors over which to have control.

The generation of mice expressing IL8 or Il10 in the intestinal tract was approached using two different methods. The first was to generate a knock-in behind a ubiquitously expressed promoter and then cross with an intestinal-specific Cre-expressing mouse and the second was to generate the knock-in behind an intestinal-specific promoter and then cross with a general Cre-expressing mouse. The advantage of the first method is that the original transgenic mouse is more versatile for the wider research community and can be used to generate mice expressing IL8 or Il10 in the tissue of interest, depending with which Cre-expressing mouse the transgenic is crossed. The advantage of the second method is that the intestinal-specific nature of IL8 or Il10 is created in the original mouse and there is no need to rely on the availability of an appropriate intestinal-specific Cre-expressing mouse. Both of these mouse models are applicable for the study of enteric diseases and mechanism of action of defensins and other host defence peptides. All of the above factors were considered for the design of the knock-in mice; even though it

might not be possible to directly mimic the production of either cytokine during infection, there is still the opportunity to probe their function through controlled experimentation.

Two targeting vector systems were chosen for each gene: the pBigT/ pROSA26PA two-vector system targets the ROSA26 locus, which allows ubiquitous expression of the inserted gene (191, 192); and the pA33LSL vector targets the *Gpa33* gene (177), which is located on Chromosome 1, and is constitutively expressed in the mouse intestine. Both *IL8* and *Il10* were successfully cloned into the pA33LSL vector, and the integrity of the genes was confirmed by restriction enzyme digestion, PCR, and sequencing. However, cloning of both *IL8* and *Il10* into the pBigT/ pROSA26PA plasmids was unsuccessful, and this approach will not be discussed further.

The *IL8*- and *Il10*-containing pA33LSL constructs were transfected into mouse embryonic stem cells from the JM8 line, and cells in which targeting events had occurred were selected by growth in Geneticin, resistance to which was conferred by the targeting vector. For both IL8 and Il10, numerous colonies were subsequently isolated and expanded, and successful targeting events were confirmed by Southern blotting and PCR.

A subset of the targeted clones was selected for microinjection into mouse blastocysts, for the generation of knock-in mice. Two rounds of microinjection were performed with both the *IL8*- and *Il10*-targeted ESC, and independent ESC clones were used for each round. The second round of microinjections, LRHE, produced three male chimeras, one of which transmitted the hIL8:c allele to its offspring. A successful breeding colony has been established producing both conditional heterozygous and conditional homozygous mice; colony expansion will continue until sufficient numbers have been generated to start breeding with vil-Cre-ER^{T2} mice. Germline transmission did not occur for the *Il10*-targeted ESC but these have been microinjected again, in a final attempt to obtain germline transmission.

The purpose of the Cre-recombinase-conditional system was to enable control over the context of expression of my genes of interest. The constitutive nature of Gpa33 expression throughout the intestine is a potential disadvantage of this construct if there is also concomitant IL8 or Il10 expression, which would likely overwhelm the intestinal mucosa and skew cellular differentiation. Any further immune system challenge (e.g. bacterial infection) may be difficult to assess because of the ensuing chronic hyper- or hypo-inflammation, respectively. However, the advantage of using the pA33LSL targeting vector is that the resulting IL8 and Il10 mice are conditional knock-in mutants and full expression of either cytokine occurs only with deletion of the neomycin resistance cassette. Read-through of the neomycin resistance cassette polyA tail

may occur resulting in transcription of *IL8* or *Il10* but this should occur only at a very low frequency. The neomycin resistance cassette was utilized during ESC transfection with the pA33LSL vector to ensure survival of those cells that take up the vector and undergo homologous recombination. Following the expansion of surviving ESC and the identification of correctly targeted clones, the neomycin resistance cassette is no longer necessary for general growth. It does however play a role in the conditional nature of the resulting mice; deletion of the neomycin resistance cassette *in vivo*, facilitated by the position and orientation of the loxP sites of pA33LSL, permits transcription of *Gpa33-IL8* or *Gpa33-Il10*.

I verified that the loxP sites of the targeted alleles were intact and *in vitro* the Cre-loxP system functioned to delete the neomycin resistance cassette. It is therefore assumed that these sites will be intact in the resulting mice and Cre-mediated neomycin resistance cassette deletion will occur *in vivo* following breeding with vil-Cre-ER^{T2} mice. The ESC colonies that formed following transfection with the Cre-puromycin plasmid had an increased number of differentiating cells compared to the pA33LSL transfection, which is an indication of lower ESC health. In addition to creating conditional knock-in mice, *in vivo* Cre-mediated neomycin resistance cassette deletion negates the need for a second electroporation thereby increasing the rate of germline transmission since the percentage of pluripotent ESC can decrease with each manipulation.

Expression of IL8 or Il10 *in vivo* requires breeding of the conditional IL8 and Il10 mice with those expressing Cre recombinase. In order to have full control over IL8 and Il10 expression, the expression of Cre recombinase has to be inducible and ideally only expressed in the intestine to prevent any *Gpa33-IL8* or *Gpa33-Il10* transcription in the stomach or bladder, regardless of whether this is negligible. A transgenic inducible C57BL/6N Cre-expressing mouse, vil-Cre-ER^{T2}, has been generated that expresses Cre recombinase under the control of a tamoxifen-dependent villin (vil) promoter (193). Administration of tamoxifen, which is a selective antagonist for the estrogen receptor (SERM), induces Cre expression throughout the intestinal tract with high expression in the small intestine and lower expression in the large intestine (193). However when these mice were crossed with Cre-dependent β-galactosidase reporter mice, X-gal (5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside) staining was visible throughout both the small and large intestine (193). Additionally Cre expression was observed in epithelial cells in both the crypt and villus (193). The use of tamoxifen to induce expression of Cre recombinase is desirable over more traditional tetracycline inducible systems since the antibiotic properties of tetracycline would interfere with bacterial infection challenges. Crossing

of the conditional IL8 and Il10 mice with vil-Cre-ER^{T2} mice will allow for controlled IL8 or Il10 expression.

Future experiments to study the role of IL8 in enteric *Shigella* and *Salmonella* infections will be discussed; those involving Il10 specifically will not be discussed since the mice have not been generated. However the Il10 conditional mice would have also been studied in parallel to the IL8 conditional mice in the described infections, so that the role of these cytokines could be compared in enteric infection as well as in administration of peptide therapeutics.

The ability to control IL8 or Il10 expression allows examination of cell activation kinetics, cellular recruitment and disease pathology following infectious challenge. Additionally the effects of IL8 or increased Il10 production on the efficacy of host defence peptides will be evaluated. There are two ways to control IL8 or Il10 expression in this system. First is the breeding of conditional mice with vil-Cre-ER^{T2} mice for tamoxifen-induced Cre-mediated neomycin resistance cassette deletion, as previously described. Second is the infection of conditional mice with Cre-expressing bacteria. *S. typhimurium* harbouring the plasmid pSB1881 express a SopE-Cre recombinase fusion protein (194). SopE, which, as part of the SPI-1 TTSS, is injected into epithelial cells upon infection (82, 194). The background of the *S. typhimurium* strain plays a significant role in the success of this system as the strain must be virulent enough to replicate following invasion but not kill the epithelial cells before IL8 or Il10 is expressed (194). The wild type strain SL1344 is hypothesized to be too virulent for the *in vivo* efficacy of this system (194), so pSB1881 will be transformed into *S. typhimurium* strain M525, which has moderate virulence compared to SL1344 or C5 strains, for a trial infection of the conditional IL8 mice. **(NB. Between submission and defence of this thesis, the experiments described above were performed. The results are shown in Appendices B.6 and B.7.)**

The SopE-Cre reporter system was developed to study the SPI-1 TTSS and the possibility of its use in vaccine development; however it is an ideal system for Cre-mediated neomycin resistance cassette deletion *in vivo*. The expression of either cytokine will only occur in infected cells, thus more closely resembling the process of natural infection. Administration of peptide, either prophylactically or therapeutically, and the characterization of intestinal cellular populations and induction of chemokines and cytokines will aid in determining the mechanism of action of host defence peptides in infection *in vivo*.

The *S. typhimurium* M525 strain expressing Cre recombinase will also be used to infect the IL8-vil-Cre-ER^{T2} mice. The comparison of IL8-vil-Cre-ER^{T2} mice infected with wild type *Salmonella* and IL8 conditional mice infected with Cre-expressing *Salmonella* could highlight

subtleties in host-pathogen interactions, in addition to controlling for potential immune modulating effects mediated by tamoxifen-dependent estrogen receptor engagement.

In summary, I have generated two embryonic stem cell lines harbouring insertions of the human *IL8* or mouse *Il10* gene behind the intestinal-specific *Gpa33* gene. Additionally this work has led to the generation of conditional heterozygous and homozygous mice carrying the hIL8:c allele(s), and efforts to generate conditional Il10 mice are ongoing. These cell lines and mouse strain will enable valuable studies into the roles of IL8 and Il10 in intestinal infection and peptide mechanism of action.

4 CHARACTERIZATION OF TRANSCRIPTIONAL PROFILES INVOLVED IN INTERLEUKIN 12- AND INTERFERON-γ-MEDIATED PRIMARY IMMUNODEFICIENCIES

4.1 Introduction

The ability of *Salmonella* species to penetrate the epithelial barrier (e.g. intestine, lung) is only the first step in their strategy for host invasion. As a facultative intracellular pathogen, the survival of *Salmonella* species also depends on their ability to invade and replicated within phagocytic cells (82). Conversely, the ability of the host to defend itself against invasion depends on cell-mediated immunity and type 1 cytokines. IFNG and TNF are key Th1 cytokines involved in the activation of macrophages for cell-mediated killing of *Salmonella* species and other intracellular pathogens such as mycobacterial species (82). The DEFA5 transgenic mouse has also highlighted the importance of defensins in immunity to *S. typhimurium* infection (43), which might be partially through regulation of the intestinal microbiota composition (96). However *S. typhimurium* reduces *Defa1* expression and α-defensin peptide release from FvB mice, in a SPI1-dependent manner (195). *S. typhimurium* has developed mechanisms for resistance to host defence peptides, including defensins. These include the PhoP-PhoQ two-component system (196), yejABEF operon (197), and sigma factor RpoE (198). This suggests that defensins have additional functions in protecting against infection. Consistent with their adjuvant antitumor activity (discussed previously), human neutrophil α-defensins augment IFNG and antigen-specific antibody production following immunization (199). Additionally human neutrophil α-defensins can promote macrophage IFNG and TNF production (200), and a chicken β-defensin can induce IL12B (IL12p40) production in a TLR4-NF-κB-dependent mechanism (201). Interestingly, recombinant IL12 and IFNG also induce human β-defensin expression (202, 203). Finally IL12B- and IFNG-deficient mice have dramatically reduced *Defb3* expression following infection with the intestinal pathogen, *Citrobacter rodentium* (204). These examples further strengthen the relationship between defensins and cytokines in shaping the immune response against intestinal pathogens.

Mutations in molecules within the IL12-dependent IFNG signalling pathways (Figure 4.1) increase susceptibility to the Bacillus Calmette-Guerin (BCG) tuberculosis vaccine, non-tuberculosis mycobacteria (NTM) and *Salmonella* infections (205, 206).

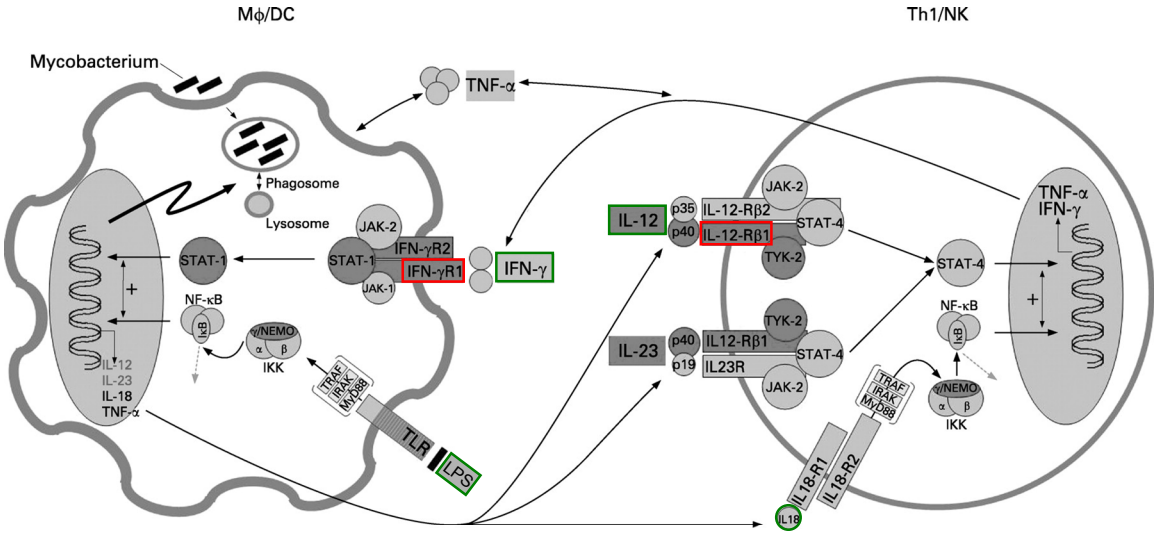


Figure 4.1. Key signalling molecules involved in IL12-dependent IFNG production. The primary immunodeficiencies, IL12RB1 and IFNGR1, of patients involved in this work are boxed in red. Stimuli used for *ex vivo* stimulation of PBMCs isolated from patients are boxed in green. Modified from (206) (permission obtained from BMJ Publishing Group Ltd.).

The clinical phenotype of these primary immunodeficiencies, which are normally autosomal recessive and present in early childhood, is largely dependent not only on the molecule involved but the genetic location of the mutation (205). The disease associated with complete deficiency is generally more severe than that of partial deficiency, and additionally people with partial deficiency can respond to high dose IFNG treatment (205). Genes in which mutations have been identified within the IL12-dependent IFNG pathway are *IFNGR1*, *IFNGR2*, *IL12RB1*, *IL12B*, signal transducer and activator of transcription 1 (*STAT1*), tyrosine kinase 2 (*TYK2*) and NF-κB essential modulator (*NEMO*). Even though *NEMO* mutations can predispose to NTM infection, there are other complicating factors that differentiate them from mutations in the other genes and thus this group will not be discussed further. The diagnostic assays involved in putatively identifying such mutations have been well established, which involve the treatment of peripheral blood mononuclear cells (PBMCs) or whole venous blood (206) with various stimuli, and measuring the cytokines produced (206). The output of the cytokine assay in conjunction with the clinical phenotype is used to predict the gene harbouring the mutation. Confirmation is required by gene sequencing. Figure 4.2 summarizes the treatments and expected outcomes for the aforementioned gene deficiencies.

Stimulus	Cytokine Responses from PBMCs or Blood of Individuals with Th1 Deficiencies				Readout
	Normal Controls	IL12RB1	IL12B	IFNGR1/2 STAT1	
Media					IFNG
IL12					
PHA					
PHA + IL12					
LPS					
LPS + IL12					
Media					IL12A/B (IL12p70)
IFNG					
LPS					
LPS + IFNG					
Media					TNF
IFNG					
LPS					
LPS + IFNG					
Media					IL12B
IFNG					
LPS					
LPS + IFNG					

Figure 4.2. Cytokine production as a diagnostic indicator of Th1 primary immunodeficiencies. The relative levels of each cytokine are indicated by shading. Black is the highest output set at 100%. Dark and light grey denote approximately 50% and 25% output, respectively, compared to the maximum. Modified from (206) (permission obtained from BMJ Publishing Group Ltd.).

The primary immunodeficiencies associated with mycobacterial and *Salmonella* infections and the genetic mutations within *IFNGR1*, *IFNGR2*, *IL12RB1*, *IL12B*, *STAT1* and *TYK2* have been well described (205-209). The signalling cascades involved in IFNG and TNF production are also well characterized, as well as the synergy required between IL12 and IL18 for efficient IFNG production (205-209). However the transcriptional responses which mediate these or other events, and the expression of other genes involved in the resulting immunodeficiencies are not well defined. Additionally synergies between LPS/IL12 or LPS/IFNG, if any, warrant further investigation. Therefore the aim of this project was to compare the gene expression of PBMCs, treated with various stimuli, from people with IFNGR1 and IL12RB1 deficiencies to those of healthy controls.

IFNGR1 is a 489 amino acid protein coded by 1470 nucleotides comprising seven exons. IFNGR1 forms a heterodimer with IFNGR2, which contains 337 amino acids coded by 1014 nucleotides also comprising seven exons. IFNGR1 binds IFNG, and both IFNGR1 and IFNGR2 are required to transduce the signal intracellularly through JAK1/2 and STAT1 phosphorylation

events (206). STAT1 homodimers translocated to the nucleus bind to gamma activation sequences and induce transcription (206). The two people included in this study with IFNGR1 deficiency, denoted as IFNGR1(1) and IFNGR1(2), both had heterozygous mutations in *IFNGR1*. IFNGR1(1) and IFNGR1(2) both had a heterozygous four nucleotide deletion (TTAA) at position 818 (coding sequence) within Exon 6 of *IFNGR1* (Figure 4.3A) (208). This deletion results in a frameshift mutation and premature stop codon just downstream of the receptor transmembrane domain (208). As a result, the truncated IFNGR1 is expressed at the surface of the cell. However due to the lack of the intracellular signalling and receptor recycling domains and normal binding to IFNG, IFNGR1 accumulates at the cell surface and exhibits dominant negative expression (208).

IL12RB1 is a 381 amino acid protein coded by 1146 nucleotides comprising 16 exons. IL12RB1 forms a heterodimer with IL12Rβ2, which contains 862 amino acids coded for by 2589 nucleotides comprising 15 exons. Both IL12RB1 and IL12Rβ2 are required for high affinity binding of active IL12 (IL12p70), which is composed of IL12A (IL12p35) and IL12B (IL12p40) subunits (206). IL12Rβ2 then translocates the signal via phosphorylation of JAK2, TYK2 and STAT4 (206). Homodimers of STAT4 translocated to the nucleus resulting in transcription. Transcription of STAT4 and NF-κB, from IL18-IL18R1/2 signalling, is necessary for full IFNG production (206). The two people included in this study with IL12RB1 deficiency, denoted as IL12RB1(1) and IL12RB1(2), both had homozygous mutations in *IL12RB1*. IL12RB1(1) contained a GC deletion and TT insertion at positions 1623-1624 (coding sequence) at the start of Exon 14 of *IL12RB1* (209). IL12RB1(2) contained a C to A substitution at position 962 (coding sequence) within Exon 9 of *IL12RB1* (209) (Figure 4.3B). Both mutations cause a pre-mature stop codon, from either a glutamine or serine amino acid, respectively, within the extracellular domain of IL12RB1 resulting in non-functional receptors (209) (Figure 4.3B).

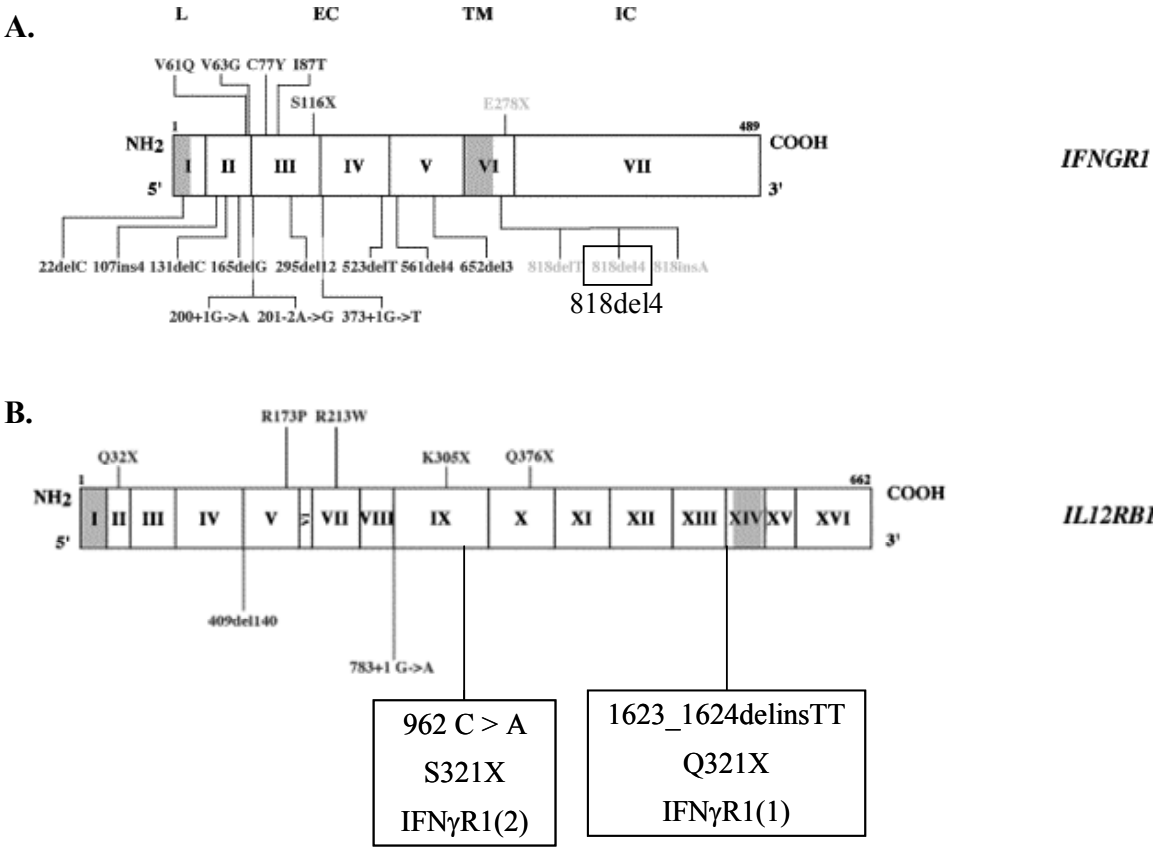


Figure 4.3. Location of known genetic mutations in *IFNGR1* and *IL12RB1*. The location of known mutations within the *IFNGR1* (**A**) and *IL12RB1* (**B**) coding sequences. The IFNGR1 deficient patients studied here both have 818del4 mutations. The IL12RB1 deficient patients studied here have either 1623_1624delinsTT or S321X mutations. Modified from (205) (permission obtained from Elsevier Publishing). The transmembrane domains are shaded in grey in exons 6 and 14 of *IFNG1* and *IL12RB1*, respectively. L, ligand binding domain; EC; extracellular domain; TM, transmembrane domain; IC, intracellular domain.

The study of people with mutations in IFNGR1 and IL12RB1 provide the opportunity to further elucidate key transcriptional processes involved in the control of intracellular bacteria in humans, and would lead to a better understanding of these two immunodeficiency diseases.

4.2 Methods and Materials

4.2.1 Patient clinical phenotypes

Both IFNγR1 patients are from Europe but are not related. IFNGR1(1) was born in Ireland in 1949. She has suffered from spinal infections (mycobacterial osteomyelitis) due to Mycobacterial avium-intracellulare complex (MAC) and undetermined mycobacterial species, at 15, 33 and 47 years of age. Each infection was successfully treated with antimycobacterial therapy, however the length of treatment required was longer for each successive infection. Additionally, splenic abscesses manifested during the last course of infection. IFNGR1(1) had an identical twin sister who developed disseminated MAC infection at 12 years of age, and died at

age 17 due to uncontrollable *M. avium* infection. Both sisters had been vaccinated with BCG without any adverse effects. IFNGR1(1)'s daughter also developed a MAC spinal infection at age 15, which was treated with antimycobacterial therapy. Currently both mother and daughter are well and treatment-free.

Detailed *in vitro* analysis of IFNGR1(1) cellular expression and activity has been performed. Total IFNγR1 mRNA was expressed at a similar level in the patient compared to controls. The normal and mutated alleles were expressed at similar levels in patient cells. There was increased IFNγR1 protein on cell surface of patient cells; expressed IFNγR1 had normal IFNγ affinity. Normal and smaller MW IFNγR1 protein was detected; smaller size is consistent with size of the predicted truncated IFNγR1. B cells respond to IFNγ, measured by STAT1 translocation, but much higher concentrations are needed for measurable translocation (e.g. 1 μg protein from control cells treated for 10 minutes with 10 UI/ml IFNγ compared to 10 μg protein from patient treated for 30 minutes with 100,000 UI/ml IFNg) (208).

IFNGR1(2) was born in England in 1991. He did not receive BCG vaccine but developed *Mycobactrium avium* pneumonia and osteomyelitis as a child (7 years old) due to M. avium-intracellulare complex (MAC). He responded to antimycobacterial therapy (clarithromycin); clarithromycin treatment is ongoing and the patient is well. Genetic analysis revealed neither parent as having the mutated allele; two siblings (brother, sister) were not available for genetic analysis. Since IFNGR1(2) has the same mutation as IFNGR1(1) it is assumed that *in vitro* results would be similar however this has not been performed except for the detection by flow cytometry of higher IFNγR1 surface expression.

IL12RB1(1) was born in Cyprus in 1969. He was vaccinated with BCG at 5 years old without any adverse effects. However he suffered from disseminated *Mycobacterium avium* infection at 17, 24 and 28 years, *M. triplex* at 32 years and *M. genevense* at 35 years, and *Salmonella enteritidis* infection at 11 and 20 years of age. Mycobacterial infection improved when IFNγ added to antibiotic therapy; currently the patient is well. IL12RB1(1) had a brother, who died of disseminated *M. avium* infection at 8 years of age. He also has two sisters, one of whom has had *Salmonella* infections; the other has had neither mycobacterial or *Salmonella* infections. His mother does not have the mutation; the father was not genetically tested.

IL12RB1(2) was born in Pakistan in 1983. He was vaccinated with BCG without any adverse effects. He had recurrent disseminated *Salmonella* (*S. enteritidis*) sepsis throughout childhood with full recovery upon antibiotic treatment, however he has not had any

mycobacterial infections. IFNγ treatment is ongoing and the patient is currently well. Neither parents nor siblings (two sisters, one brother) have the mutation.

In vitro studies showed that neither patient expressed IL12Rβ1 on the cell surface, and in response to BCG alone or in combination with IL12, their cells produced less IFNγ (209).

4.2.2 Patient and control sample collection

All experiments were approved by Addenbrooke's Hospital ethical approval and guidelines, including the movement of samples off-site to WTSI for processing and storage. Additionally all experiments were approved by the WTSI ethical committee for RNA extraction from human cell lysates, and storage of the acellular material.

Due to the rarity of these two primary immunodeficiencies, in each of the groups one person was sampled twice and the other person was sampled once, for three independent biological repeats.

4.2.3 Reagents

Dulbecco's PBS (D-PBS) and RPMI supplemented with 2 mM L-glutamine, denoted as RPMI media, were both purchased from Invitrogen Life Technologies. Recombinant IL12 was purchased from R&D Systems (Minneapolis, MN, USA), recombinant IL18 was purchased from MBL International (Woburn, MA, USA), and recombinant IFNG was purchased from Boehringer Ingelheim Limited (Bracknell, Berkshire, UK). Lipopolysaccharide (LPS) from *Salmonella minnesota* strain R595, highly purified, was purchased from List Biological Laboratories Incorporated (Campbell, CA, USA).

4.2.4 Cell isolation and culture

4.2.4.1 PBMC stimulation for microarray analysis

Human venous blood (~ 25 ml) was collected from patients and healthy volunteer control subjects in lithium-heparin blood collection tubes (BD Biosciences) in accordance with Addenbrooke's Hospital (Cambridge, UK) ethical approval and guidelines. For each experiment, patient and control samples were collected at the same time and processed together on the same day. All steps for cell isolation were performed at room temperature, and centrifugations carried out in a Varifuge® 3.0R (Heraeus Sepatech, Germany). A volume of 2 ml whole blood was removed from one tube for the whole blood stimulation assay (described below), then the tube centrifuged at 1700 rpm for 10 minutes to pellet the red blood cells. A volume of 1.5 ml plasma

was removed (to be used in place of 10% FBS because problems with batch to batch variability), then the tube mixed well to resuspend the cells (referred to as autologous plasma). The remainder of the blood was mixed with an equal volume of Dulbecco's PBS (D-PBS) (Gibco) in a sterile 50 ml Falcon tube. PBMCs were isolated by layering the mixture over Lymphoprep (Cedar Lane Labs), followed by gradient centrifugation at 1600 rpm for 30 minutes. The buffy coat layer (5-10 ml) was split into two 15 ml Falcon tubes and washed once with D-PBS (up to 14 ml), with a centrifugation of 1600 rpm for 10 minutes. The cell pellet was resuspended in RPMI + GlutaMax media and the viable cells counted by Trypan blue (0.4%) exclusion. The PBMCs were supplemented with 10% autologous plasma, to a final concentration of 2-3 million cells/ml/well in 24-well tissue culture plates (BD Falcon); patient and control cell numbers were standardized for each experiment. PBMCs were treated, at the final concentrations indicated, with media alone, IL12 (50 ng/ml), IL18 (100 ng/ml), IL12/IL18, LPS (2 μg/ml), LPS/IL12, IFNG (20000 U/ml), LPS/IFNG at 37°C with 5% CO_2 for 4 hours. Following the incubation, the 24-well plate was centrifuged at 500 g for 5 minutes at 4°C. Supernatants were removed and cells lysed with 1 ml Qiazol reagent (Qiagen). Cell lysates were frozen and stored at -70°C until RNA extraction.

4.2.4.2 Whole blood stimulation assay for analysis of secreted protein

In a final volume of 250 μl, whole blood (50 μl) was treated, at the final concentrations indicated making up the volume with RPMI media, with media alone, IL12 (20 ng/ml), IL18 (50 ng/ml), IL12/IL18, LPS (2 μg/ml), LPS/IL12, IFNG (10000 U/ml), LPS/IFNG at 37°C with 5% CO_2 for 24 hours. Following the incubation, the supernatants were removed and stored at -20°C. The cytokine assays, set up in parallel with the microarray assays, will serve to validate the transcriptional data.

4.2.5 RNA extraction and quality control

RNA was extracted according to the Qiazol Handbook instructions and RNA resuspended in 87.5 μl nuclease-free water (Ambion, Inc.). The RNA was treated in solution with RNase-free DNase (Qiagen) according to the instructions in the RNeasy MinElute kit (Qiagen). The RNA was then cleaned using the Qiagen MinElute kit, according to the manufacturer's instructions. RNA was eluted from the MinElute columns with 30 μl of nuclease-free water. RNA purity and integrity was assessed using the Agilent 2100 Bioanalyzer 2100 Expert Version B.02.05.SI360 (Agilent Technologies), and the concentration determined using the Thermo Scientific NanoDrop™ ND-1000 Spectrophotometer (Thermo Fisher Scientific Inc.),

both as per the manufacturers' instructions. The RNA was diluted to 30 ng/µl for microarray analysis.

4.2.6 Microarrays

Microarray analysis was performed by Peter Ellis (WTSI Microarray Facility, Team 77). Total RNA (500 ng) was amplified using the Illumina Total Prep RNA Amplification Kit (Ambion) according to the manufacturer's instructions. The biotinylated cRNA (1500 ng per sample) was applied to Illumina HumanWG-6 v3 Expression BeadChips and hybridized overnight at 58˚C. Chips were washed, detected and scanned using a BeadArray reader, according to the manufacturer's instructions. The scanner output imported into BeadStudio software (Illumina). Each HumanWG-6 v3 Expression BeadChip contains over 48,000 probes and provides coverage of the entire human genome.

4.2.7 Data analysis

4.2.7.1 Data pre-processing

Bead summary data was generated by Peter Ellis using the BeadStudio software (Illumina) without background correction (normalization) for the data analysis.

4.2.7.2 MetaGEX analysis pipeline

The microarray MetaGEX analysis software tool (Development Version 0.2), was written by Christopher Fjell (Hancock lab, UBC). The raw bead summary data was input by Christopher Fjell, and the analysis and results maintained on an internal server, which provides a web interface for data download. For each patient group, p-values and fold changes (adjusted or non-adjusted) are calculated from probe intensities across the three biological replicates. For each treatment, data is output as the comparison between patient and control, and will be referred to as the [treatment] comparison. For each treatment comparison, genes with negative fold change values will be referred to as downregulated and those with positive fold change values will be referred to as upregulated. Log (base 2) fold change values are output from the BeadStudio software, which were then converted to fold change. An addition transformation was performed so that values are reported as integers.

4.3 Results

4.3.1 Differentially expressed genes

For the comparisons between patient and control gene expression, the MetaGEX analysis output thresholds for significance were any fold change with a p-value (non-adjusted) of less than 0.05 for the difference in means. This permitted capture of as many significant genes as possible for analyses of over- or under-represented factors, e.g. transcription factor targets, which may not be apparent in the gene changes themselves but reflect possible signalling events leading to the changes observed in gene expression. This output is referred to as enriched with differentially expressed genes, and the distinction is made for putatively differentially expressed genes when fold change and adjusted p-value thresholds were applied.

4.3.1.1 IL12RB1 data set

The numbers of genes within the enriched IL12RB1 data set can be found in Table 4.1.

Table 4.1. Numbers of genes within the enriched data set in PBMCs from IL12RB1 deficient patients compared to healthy control PBMCs. Results are from three biological replicates, each stimulated with the indicated treatments for 4 hours. Threshold cut-offs were a p-value < 0.05.

Treatment	Positive Fold Change	Negative Fold Change	Total DE Genes
Media alone	631	561	1192
IL12	708	640	1348
IL18	780	707	1487
IL12/IL18	1170	976	2146
IFNγ	830	681	1511
LPS	1381	1131	2512
IL12/LPS	1260	927	2187
IFNγ/LPS	1932	1297	3229

The numbers of genes potentially differentially expressed for each treatment in the IL12RB1 patient group compared to control group can be found in Table 4.2. Genes considered to be differentially expressed had fold changes greater than 1.5 or less than -1.5 with adjusted p-values less than 0.05. The gene list for each treatment are found in Appendices C.1-C.8.

Table 4.2. Differentially expressed genes in PBMCs from IL12RB1 deficient patients compared to healthy control PBMCs. Results are from three biological replicates, each stimulated with the indicated treatments for 4 hours. Threshold cut-offs were fold changes of ± 1.5 and adjusted p-values < 0.05.

Treatment	Upregulated Genes	Downregulated Genes	Total DE Genes
Media alone	10	18	28
IL12	12	32	44
IL18	14	22	36
IL12/IL18	44	128	172
IFNγ	10	18	28
LPS	37	52	89
IL12/LPS	34	42	76
IFNγ/LPS	68	103	171

4.3.1.2 IFNGR1 data set

The numbers of genes within the enriched IFNGR1 data set are found in Table 4.3.

Table 4.3. Numbers of genes within the enriched data set in PBMCs from IFNGR1 deficient patients compared to healthy control PBMCs. Results are from three biological replicates, each stimulated with the indicated treatments for 4 hours. Threshold cut-offs were a p-value < 0.05.

Treatment	Positive Fold Change	Negative Fold Change	Total DE Genes
Media alone	378	389	767
IL12	437	474	911
IL18	497	661	1158
IL12/IL18	342	443	785
IFNγ	412	462	874
LPS	309	360	669
IL12/LPS	348	427	775
IFNγ/LPS	449	452	901

The numbers of genes potentially differentially expressed for each treatment in the IFNGR1 patient group compared to control group can be found in Table 4.4. Genes considered to be differentially expressed are those with fold changes greater than 1.5 (or 1.2 in brackets) or less than -1.5 (or -1.2 in brackets) with adjusted p-values less than 0.05. The list of genes, fold change ± 1.2, for each treatment are found in Appendices C.9-C.16.

Table 4.4. Differentially expressed genes in PBMCs from IFNGR1 deficient patients compared to healthy control PBMCs. Results are from three biological replicates, each stimulated with the indicated treatments for 4 hours. Threshold cut-offs are fold change +/- 1.5 (± 1.2) and adjusted p-value < 0.05. A * denotes that genes similar to other treatments were observed within the fold change cut-offs but the mean difference between patient and control did not reach statistical difference.

Treatment	Upregulated Genes	Downregulated Genes	Total DE Genes
Media alone	0 (1)	1 (10)	1 (11)
IL12	0 (2)	3 (17)	3 (19)
IL18	0 (3)	10 (50)	10 (53)
IL12/IL18	0 (1)	0 (8)	0 (9)
IFNγ	0 (3)	0 (6)	0 (9)
LPS	0 (2)	0 (5)	0 (7)
IL12/LPS	0 (3)	0 (8)	0 (11)
IFNγ/LPS	0 (0)	0 (0)	0 (0) *

4.3.2 IL12RB1 analysis

The IL12RB1 untreated cell (media alone) comparison showed that the majority of differentially expressed genes with negative fold changes were natural killer (NK) cell receptors (*KLRB1*, *KIR2DL1*, *KIR2DL3*, *KIR2DL4*), as well as granzyme H (*GZMH*) and fibroblast growth factor binding protein 2 (*FGFBP2*). This would suggest that these patients have an inherent deficiency in NK cells. It has been found that IL12RB1 deficient patients have normal numbers of CD3-CD56+ NK cells but these are less responsive to stimulation in terms of IFNG production and surface expression of CD107 (210), which characterizes degranulating NK cells (211). However IL12RB1 deficient patients were found to have reduced $CD56^+$ T cells, which can also express NK-like receptors including KIR, CD94/NKG2A, and CD161 (210). Therefore the decrease in NK cell receptors of the IL12RB1 patients in this work, and over-representation of enriched genes in the 'natural killer cell mediated cytotoxicity' KEGG pathway is likely due to reduced $CD56^+$ T cells. Both $CD56^+$ T cells and NK cells produce IFNG in response to IL12/IL18 (212, 213). Other genes of note downregulated in the media alone control for IL12RB1 patients were acyloxyacyl hydrolase (*AOAH*), which deacylates LPS thus allowing robust host responses to Gram-negative re-infection (214), gap junction protein (*GJB2*), which as a connexin may be involved in immunological synapses between peripheral blood mononuclear cells, in addition to maintaining the epithelial barrier (215), and sphingosine-1-phosphate receptor 5 (*S1PR5*), which is expressed on NK cells upon maturation and is required for migration to sites of inflammation (216).

The IL12RB1 IFNG treatment condition led is similar to that of media alone, which is expected since cells of these patients can respond to IFNG. The IL12 and IL18 treatment comparisons are also similar to that of the media, except downregulated genes are *IL2RB*, *IL18R1*, *NOD2*, *LAG3*, *MAP3K8*, *IL18RAP*, *SIGLEC10* and *NFKB1*, *IL18R1*, *MAP3K8*, respectively. The LPS and IL12/LPS treatment comparisons are similar to each other, which suggests that the gene expression observed at the four hour timepoint is not affected by the release/production of secondary mediators. However there IFNG/LPS treatment has a lot more genes downregulated compared to IFNG, which, with the IL12/LPS and LPS expression data, points towards a defect in LPS recognition or LPS-induced signalling. There is either synergy between IFNG and LPS, as has been observed for IL12 production (206), since the number of differentially expressed genes is approximately double that of LPS alone or IL12/LPS, or gene expression for this treatment and timepoint is affected by the production of secondary mediators. The gene expression observed for the IL12/IL18 treatment confirms a downregulation of many genes involved in TLR4 to NF-κB signalling. Additionally RELA (NF-κB3) target genes are overrepresented (downregulated) in the IL12/IL18 treatment. Therefore the main treatment analyzed here will be the IL12/IL18 combination.

Most of the genes differentially expressed for IL12RB1 IL12/IL18 treatment comparison are downregulated, and involved in IL12-mediated signalling, JAK-STAT, cytokine-cytokine receptor interaction, TLR signalling and TNFR2 signalling pathways, determined by pathway overrepresentation analysis (Table 4.5). The upregulated genes are *CD79B*, *PPARG*, *VAV3*, *MAPK13*, *CD27*, sequestosome 1 (*SQSTM1*). The most highly downregulated genes are *DEFB1* (-5.43), *UBD* (-10.8), *IDO1* (-11.5), *CXCL9* (-18.0) and *IFNG* (-25.7). The analysis of transcription factor targets in the IL12RB1 IL12/IL18 enriched data set highlights deficiencies in genes with STAT (1, 2, 3, 5A, 5B, 6), IRF (1, 2, 4, 5, 8), CREB1, CEBPB, RELA and SP1 predicted binding sites (Appendix C.17).

Table 4.5. Pathway analysis of IL12RB1 differentially expressed genes following IL12/IL18 treatment. Thresholds for fold change (+/- 1.5) and adj p-value (< 0.05) were applied to determine pathways overrepresented in terms of numbers of differentially expressed genes.

Pathway Name	Upregulated Genes	Downregulated Genes
IL12-mediated signaling events	none	*IFNG* , *IL18R1* , *IL18RAP* , *IL2RB* , *NFKB1* , *SOCS1* , *JAK2* , *GADD45B* , *GZMA* , *GZMB* , *PRF1*
JAK-STAT pathway and regulation pathway	*MAPK13*	*CDK6* , *CSF2RB* , *IFNG* , *IL15* , *IL15RA* , *IL18R1* , *IL2RB* , *JAK2* , *LTA* , *MAP3K8* , *MAPK13* , *SOCS1* , *STAT2* , *TNFSF10* , *BCL2L1* ,
Cytokine-cytokine receptor interaction	*CD27*	*CXCL9* , *CSF2RB* , *IFNG* , *IL15* , *IL15RA* , *IL18R1* , *IL18RAP* , *IL2RB* , *LTA* , *TNFRSF18* , *TNFSF10* , *TNFSF13B*
TLR signaling pathway	*MAPK13*	*TLR4* , *TMED7* ;*TICAM2* , *IRF7* , *CXCL9* , *MAP3K8* , *NFKB1*
TNFR2 signaling pathway	none	*TANK* , *TRAF1* , *LTA*
Unassigned	*CD79B* , *PPARG* , *VAV3* , *MAPK13* , *CD27* , *SQSTM1*	*DEFB1* , *IDO1* , *NOD2* , *TNFAIP2* , *TNFAIP8* , *IFI35* , *LRRK2*

4.3.3 IFNGR1 analysis

The IFNGR1 treatment comparisons generally show fewer differentially expressed genes than those of the IL12RB1 comparisons. A possible explanation for this has to do with the genotypes of the patients analyzed. The *IL12RB1* mutations are homozygous for complete IL12RB1 deficiency in both patients, whereas the *IFNGR1* are heterozygous, which means one receptor allele functions normally. The mutated IFNGR1 has a functional ligand binding domain but is truncated just following the transmembrane domain (208). The lack of signal domain results in no signal transduction thus the mutated receptor is dominant negative. Additionally the receptor-recycling domain is also deleted, so eventually the cells of these patients have proportionally more of the non-functional receptor because the functional ones get recycled normally (208). Therefore the four hour time point for these experiments might not have been long enough to observe significant changes in gene expression in the treatment comparisons. For this reason, the fold change thresholds were reduced (+/- 1.2).

The IFNGR1 media alone comparison again showed a high proportion of NK receptors (*KIR2DL1*, *KIR2DL3*, *KIR2DL4*, *KIR3DL1*) downregulated. An NK or T cell deficiency, similar to IL12RB1 patients, has not been described for IFNGR1 patients. The most highly downregulated gene is *CXCL10*, which as an interferon-inducible gene may not be surprising

since any inherent defects *in vivo* will be reflected upon treatment *ex vivo*. The IL12/IL18 and LPS comparisons were similar to media alone. The IL12, IFNG, IL12/LPS comparisons were similar to media alone, except a modest downregulation of *CXCL9*, *IFNG*, *CXCL9*, respectively, was also observed. None of the changes in gene expression in the IFNG/LPS comparison reached statistical significance. The IL18 comparison showed the greatest number of differentially expressed genes for the IFNGR1 data sets, the majority of which are downregulated including *TNF*, *IFIT3*, *LRRK2*, *STAT1*, *TNFAIP2*, *RIPK2*, *IDO1* (-1.53), *UBD* (-1.63), *CCL8* (-1.65), *CXCL9* (-1.75) and *CXCL10* (-1.87). The analysis of transcription factor targets in the IFNGR1 comparison enriched data sets highlights general deficiencies in genes with STAT and IRF predicted binding sites (Appendix C.18).

4.4 Discussion

Microarray analysis is a powerful tool in the understanding of cell signalling on a global level, and as such can reveal inherent defects in this signalling or in cellular populations of people susceptible to disease. As a mixed cell population was used for this work, the results are limited to general signalling events that can not be assigned to a specific cell type. The focus of the analyses of patients with mutations in IL12RB1 and IFNGR1 has been on similarities in gene expression to further elucidate the mechanisms involved in the susceptibility to NTM, as well as the differences that might be responsible for an increased incidence of *Salmonella* infections in IL12RB1 patients compared to IFNGR1 patients (217, 218).

The IL12RB1 data confirms the IL12/IFNG axis as the major signalling defect likely involved in disease susceptibility. The reduced IFNG production of PBMCs from IL12RB1 patients (206), confirmed here for *IFNG* as the most downregulated of the differentially expressed genes, in response to IL12/IL18 treatment could be a consequence of reduced CD56+ T cells in these patients. The second largest discrepancy between patient and control for the IL12/IL18 treatment is *CXCL9* (discussed below). Other genes which may be important mediators of IL12RB1 disease susceptibility are *DEFB1*, *UBD* and *IDO1*. A defensin deficiency has not been previously described for this patient set, nor has the production of β-defensin 1 (DEFB1) following IL12/IL18 stimulation. Degradation of intracellular proteins by ubiquitin D (UBD) is necessary for antigen presentation, and activation of NF-κB signalling as well as negatively regulating signalling molecules, e.g. peroxisome proliferator-activated receptor gamma (PPARG) (219), which is consistent with many of the downregulated genes. PPARG has recently been shown to induce *DEFB1* expression in human colon epithelial cells, but a PPARG-

independent mechanism of *DEFB1* production was also implicated (220). DEFB1 is constitutively expressed but can be induced, and its promoter contains predicted transcription factor binding sites for NF-κB1 (p50/p105), activator protein-1 (AP-1) and CCACT enhancer binding protein-β (CEBP-β) (221). *CEBPD*, downregulated in this study, and *CEBPB* are induced in inflammation and their protein products can form heterodimers (222). *DEFB1* is expressed in a variety of epithelial cells but its expression in monocytes, monocyte-derived macrophages and monocyte-derived DCs has been described, as well as the ability of LPS and IFNG to increase this expression (223). Additionally lymphocytes have been found to express DEFB1, which may be CD3+ and CD8+ T-cells, but this needs further elucidation (224). Indoleamine 2,3-dioxygenase 1 (IDO1) produced by DCs inhibits T cell responses through contact with regulatory T cells, and as such is hypothesized to function in chronic infections as well as other situations of tolerance (225). These data suggest general state of immune hyporesponsiveness, which may be due insufficient T cell responses.

The IL12RB1 data also suggests that these patients have a defect in either the recognition of or responses to LPS that, based on the genes downregulated within the TLR signalling pathway, is hypothesized to be within the TLR4-Myeloid differentiation primary response (MyD)88-independent signalling pathway (226). This may explain why patients with IL12/23-component (IL12RB1, IL12 or IL23) deficiencies present with more *Salmonella* infections than patients with IFNG-component (IFNG, IFNGR1 or IFNGR1) deficiencies (217, 218). The authors argue that the IL12/IL23 axis is more important in controlling *Salmonella* infection, and that there are IFNG–independent pathways (217). The hypothesis of the involvement of TLR4-MyD88-independent signalling may not be directly related to the observation that these patients have fewer CD56+ T cells and less responsive NK cells. However the apparent *CXCL9* deficiency may provide this link, depending on which cell type is normally responsible for its production. Why has this apparent LPS-recognition/response not been recognized, if indeed it is true? The cytokine response pattern (Figure 4.2) shows a defect in IFNG production in patients in response to IL12/LPS (206). This is most likely a secondary effect due to the inability of IL12RB1 deficient cells to utilize the IL18 produced in response to LPS in combination with the IL12 supplied. Neither patients nor control cells produce significant amounts of IFNG upon LPS treatment, as efficient IL12 or IL23 production requires a second signal from IFNG (206). Patient and controls cells show similar IL12A/B and TNF production in response to LPS or IFNG/LPS (206). However the differences observed in the array data could impact how the overall response to LPS or bacteria by these patients is mediated.

The direct comparison of differentially expressed genes between IL12RB1 and IFNGR1 patient cells needs to be approached cautiously until validation studies are performed, considering that cell populations within these two patient sets may be different, in addition to the consequences of complete versus partial receptor deficiency. However it appears that these two patient sets may lack or have reduced levels of Th1-attracting chemokine(s).

The interferon-inducible chemokines C-X-C ligand (CXCL)9-11 are produced by dendritic cells (DCs), B cells and macrophages. Additionally CD56+ T cells can induce *CXCL9* expression (227). CXCL9 deficiency following NTM infection has not been associated with either IL12RB1 or IFNGR1 deficient patients, but IFNGR KO mice have decreased *CXCL9-11* expression after infection with *Mycobacterium avium,* which may contribute to granuloma necrosis (228). Mice immunized with *Mycobacterium tuberculosis* vaccines and then challenged with *M. tuberculosis* had increased levels of *IFNG* and *CXCL9-11* expression in the lung, which was correlated with significantly lower bacterial counts in the lung and spleen (229). CXCL9 recruits CD4+ Th1 cells (230), which are necessary for *M. tuberculosis* clearance (231). Similar to CXCL9, CXCL10 also appears to be important in protection against *M. tuberculosis*. A single nucleotide polymorphism in *CXCL10*, which may increase the susceptibility to infection with *M. tuberculosis*, has recently been described (232). Synergy in CXCL10 production by a human monocyte-like cell line (THP-1) by IFNG and TNF (233). Patients with tuberculosis produce more CXCL10, CCL2 (MCP-1), and IFNG in response to mycobacterial-specific antigens than healthy control individuals (234). CXCL10 and IL2 are produced in significant amounts from whole blood treated with *M. tuberculosis*–specific antigens of children at high risk for tuberculosis compared to those at low risk (235). CXCL10 also seems to play a role in autoimmune disorders and viral infections (236, 237), but has not been associated with either IL12RB1 or IFNGR1 deficiencies.

CXCL9-11 bind CXCR3, which is expressed on NK cells, dendritic cells, B cells and in particular Th1 activated/memory T cells (230). CXCL9-11 are therefore involved in the trafficking of Th1 cells. The absence or reduced expression of CXCL9 or CXCL10 in IL12RB1 and IFNGR1 patients may be a contributing factor in their inability to control the normally non-pathogenic NTM and BCG vaccine strains, through defective recruitment of Th1 cells. Additionally, this work here has provided a rational explanation for the apparent discrepancies observed between clinical phenotype of IL12RB1 and IFNGR1 deficiencies despite their both being susceptible to NTM infections.

5 DISCUSSION AND CONCLUSIONS

The aim of this research was originally to study the immune function of defensins through *in vitro* cell culture assays and the generation of an α-defensin deficient mouse. This was in order to better understand their biological role *in vivo*, both during normal immune development and infection challenge. In the field there has been significant research in elucidating the mechanisms of action of defensins but there still is not a clear picture of how they contribute to the overall function of the immune system. In the intestinal tract, there is a delicate balance between commensal and pathogenic organisms, and the ability of one α-defensin, DEFA5, to regulate this balance has recently been described (96). Whether all α-defensins, especially in the mouse, function in the same manner is unknown. The gene expression of α-defensins within Paneth cell granules is generally thought to be constitutive, whereas upon degranulation, their release into the intestinal lumen in inducible. Paneth cells degranulation can be stimulated by cholinergic receptor engagement, Gram-positive and Gram-negative bacteria, and the bacterial products LPS, lipotechoic acid, lipid A and muramyl dipeptide (35). In terms of the mouse, it was known, in strains other than the reference C57BL/6J, that there is a discrepancy between genomic content and peptide repertoire for the α-defensins, but the reason for this discrepancy was unknown. Murine α-defensins have not been as well characterized due to several issues which all reflect their high degree of similarity. Targeted deletion of α-defensin genes is near impossible, study of individual genes expressed is only possible for a few genes and this is even suspect, changes in relative gene expression under various conditions is therefore not been established, and the problems with the genomic assembly have hindered complete phylogenic analyses. This work has highlighted these problems, and also provided some solutions. There is now a clear, reliable α-defensin gene set for the reference mouse strain. This work has provided a novel method for the quantification of murine α-defensin gene expression, which is free from the biases associated with conventional bacteria cloning methodologies, provides full-length transcript sequences, and potentiates unannotated gene discovery. As a result, this work has also provided insights into murine α-defensin gene expression, especially with respect to copy number variation. The significance of the regulation of the expression of gene duplications in humans is not clear, despite defensin CNV being implicated in several human diseases. There is now the ability to study the regulation of α-defensin gene expression in greater detail following infection or other intestinal assault. In these circumstances subtle differences may not be apparent using current detection technologies.

Studies involving the DEFA5 transgenic mouse suggest that the role of this defensin involves the regulation of intestinal bacterial populations, in addition to original finding that these mice are more resistant to *Salmonella* infection (43, 96). These two observations are most likely linked as increasing evidence suggests the normal flora plays a significant role in protecting against infection (238, 239). If intestinal defensins do indeed regulate the composition of the normal flora, this warrants further investigation when considering defensins or other host defence peptides as novel therapeutics against infection. The role that defensins play in immune homeostasis may be vastly different than the role, if any, they play upon active defence against an invading pathogen. This beneficial role of DEFA5 in controlling bacterial infection is contrasted by the detrimental role CRS4C peptides are hypothesized to play in SAMP1/YitFc mice, which are predisposed to ileitis and correlatively have increased levels of CRS4C mRNA and peptides (240). Due to the differences in the α- and CRS-defensin mature peptide sequences, these two families of peptides likely have distinctive functions but this needs further elucidation, in addition to determining their relative levels in disease. Under conditions of disease, there is likely a plethora of molecules within the inflammatory milieu, which may dilute the functional activity of defensins or conversely there may be synergy with some of these other factors, e.g. cytokines. Synergy has been demonstrated for IL1B and colony-stimulating factor 2 (granulocyte-macrophage) (CSF2, GM-CSF) with the host defense peptide, LL-37 (241, 242). Therefore the mouse model of conditional, intestinal-specific IL8 expression, and Il10 expression once generated, can be used to study the role of defensins in inflammation and immunosuppression, respectively. There are many immune activating and suppressing cytokines, some with apparent redundant functions, and even though IL8 and Il10 are just two of them, these models will provide vital information regarding defensin function under these conditions, and also in the presence of infection.

The IL8 knock-in mouse can also be used to study the *in vivo* role of IL8 in infection, which to-date has not been well characterized. The neutrophil chemoattractant property of IL8 is its most notable function, but how this contributes *in vivo* is not clear, as well as the downstream benefits or consequences of neutrophil recruitment. A novel function of neutrophils recently discovered is their ability to form a matrix upon degranulation. These matrices have been termed neutrophil extracellular traps (NETs), and consist of chromatin complexed with cytosolic and granular components (243). NET formation is dependent on neutrophil activation; various stimuli induce NET formation including IL8, phorbol myristate acetate (PMA) (protein kinase C activator), LPS, bacteria, fungi and activated platelets (243, 244). Infection of neutrophils with

either Staphylococcus aureus or *S. flexneri* induces NET formation and subsequent bacterial killing *in vitro* (243). Additionally NETs kill *Streptococcus pyogenes*, and *Bacillus anthracis*, *C. albicans*, *Leishmania amazonensis* and can trap but not kill *Streptococcus pneumoniae*, Group A Streptococci (GAS) and *Mycobacterium tuberculosis in vitro* (245). This suggests differing sensitivity to killing by NETs and indeed some microbes have developed mechanisms to evade killing by NETs (245). *In vivo* NET formation occurs in rabbits infected with Shigella and in humans with appendicitis (243), however the *in vivo* physiological concentration of stimuli needed for NET formation has not been well characterized and the outcome of infection following IL8-mediated NET formation *in vivo* is unknown. As neutrophils can be detrimental to the host and cause tissue damage associated with excessive inflammation, the timing of neutrophil recruitment is obviously crucial to this outcome. This mouse model can be used to control neutrophil recruitment and determine whether NET formation occurs in infection *in vivo* in the presence of IL8. If NET formation does occur *in vivo* then their role in protecting host tissue from damage due to neutrophil degranulation in response to infection can be assessed.

It is tempting to speculate that the absence of neutrophils in patients with typhoid fever is due to decreased levels of IL8 in response to *S. typhi*. However in patients with typhoid fever, neither the level of IL8 protein nor the presence of polymorphisms within the IL8 gene has been associated with disease severity. *S. typhimurium* elicits flagella-mediated IL8 production from a polarized human intestinal epithelial cell line, T84 (83, 246), and additionally deletion of the viaB locus, including all genes necessary for the regulation, synthesis and export of the Vi antigen, of *S. typhi* leads to increased IL8 production from T84 cells, a macrophage-like cell line (THP-1) and human biopsy colonic tissue (83), suggesting the Vi antigen functions in immune evasion (247). However further research revealed that the viaB locus regulator, TviA, also reduces *S. typhi* flagellin expression and it is this reduction that leads to decreased IL8 production (248). Unfortunately the host-restricted nature of *S. typhi* excludes the study of its pathogenicity in a mouse infection model. However it may be possible to prime the intestinal mucosa through IL8 production and subsequent neutrophil infiltration, which may increase the colonization ability of *S. typhi* similar to the administration of antibiotics in the mouse enterocolitis model of *S. typhimurium* infection (82). The conditional IL8 mouse model generated in this work could be used to model host-pathogen interactions in the dissemination of *S. typhi* from the mucosal barrier and the resulting systemic disease, in addition to determining the kinetics of neutrophil recruitment in *S. typhimurium* infection and how it contributes to the resulting disease in mice.

In additional to functional significance, there are unanswered questions regarding the evolution of mammalian defensins and cytokines. Why has there been such a rapid expansion of α-defensins in mice? Why is IL8, and indeed another chemokine, Cxcl6, within the same genomic region, deleted in mice? Even though murine granulocyte chemotactic protein (Gcp)-2 was referred to as Cxcl6 (61), it has now been annotated as Cxcl5 in mouse; human GCP-2 is CXCL6 in both (61) and Ensembl (Pubmed). Looking at the chromosomal position of IL8, CXCL6, CXCL5 and Cxcl5 in Ensembl would suggest that Cxcl6 is also deleted in mouse. IL8 is at 74.6 Mb on Chr4 and the region surrounding this position 52.7-89.0 Mb holds synteny with mouse Chr5 73.9-105.0 Mb, so the deletion does not appear to be within a chromosome rearrangement breakpoint. CXCL6 is also within the human genomic region but Cxcl6 is missing from syntenic region in mouse. CXCL5 and CXCL6 share 79% identity and are predicted as paralogous, CXCL5 may have undergone a duplication event compared to mouse having deleted Cxcl5. Interestingly, Cxcl5, CXCL5 and CXCL6 are all ELR(+) chemokines, and therefore a deletion in IL8 and Cxcl5 could limit neutrophil infiltration in the mouse intestinal tract in response to bacterial invasion. Both the human and mice genomes contain CXCL1-3, which are also ELR(+) chemokines, within these same regions. The guinea pig IL8 protein sequence is the most divergent of the known IL8 mammalian proteins which suggests two methods to prevent damage due to potent IL8 activity. Either accumulate mutations affecting neutrophil receptor binding, or delete the gene altogether (249).

There are a large proportion of duplications in olfactory genes in mouse, with much fewer in human (124); a keen sense of smell is far more beneficial for mice than humans. Duplication of murine intestinal α-defensin genes could have undergone a similar process since the gastrointestinal tract of mice is subject to a constant barrage of microbes. This same reasoning can be employed for why mice have deleted IL8 (and Cxcl6). Too much inflammation can be a bad thing, especially in the gut where there is a barrage of potential pathogens and may be why there has been such an expansion of other genes, e.g. α-defensins, in order to deal with the threat of pathogens more subtly. If the innate immune system can control bacterial growth at the epithelial surface of the intestinal tract then there is less chance that they will be able to invade the epithelial barrier, translocate and cause massive inflammation and cellular infiltration, which ultimately leads to the destruction of the pathogenic barrier.

Due to many structural and functional similarities between defensins and chemokines, and the fact that they can induce each other's expression/release indicates that they should both be considered in the design of novel therapeutics for the treatment of infectious or chronic

inflammatory diseases. The contribution of CXCL9 and CXCL10 to IL12- and IFNG-mediated immunity may be such an example. CXCL9 and CXCL10 are important for protection against mycobacterial infection through their recruitment of CD4+ Th1 cells, but antimicrobial activity has also been described for these chemokines (250). Similar to defensins, high salt inhibits this function however the concentrations of CXCL9 and CXCL10 induced by IFNG may be high enough to overcome this inhibition (250). Also similar to defensins, CXCL9-11 have been implicated in inflammatory bowel disease as blocking CXCL10 or CXCR3 function reduces colitis (251, 252).

Overall this research has contributed to the understanding of the murine α-defensin genomic repertoire. A next-generation sequencing method has been developed for the quantitative analysis of α-defensin transcriptional expression, which is applicable for other highly similar gene families. In addition this work has led to the generation of novel murine models of inflammation and (with anticipation) immunosuppression, which can be used to study potential therapeutics in addition to the *in vivo* role of human interleukin 8. Finally this work has led to new hypotheses for the involvement of CXCL9 and CXCL10 in IL12RB1- and IFNGR1-mediated susceptibility to mycobacteria, and for impaired TLR4 responses in the increased frequency of *Salmonella* infection for IL12RB1 deficiency. As defensin induction can be mediated by NF-κB signalling, it is also tempting to speculate that decreased DEFB1 levels in IL12RB1 deficient patients may also contribute to disease susceptibility.

REFERENCES

1. Powers, J.P.S., and Hancock, R.E.W. 2003. The relationship between peptide structure and antibacterial activity. *Peptides* 24:1681-1691.

2. Bowdish, D.M., Davidson, D.J., and Hancock, R.E.W. 2005. A re-evaluation of the role of host defence peptides in mammalian immunity. *Curr Protein Pept Sci* 6:35-51.

3. Yoshimura, A., Naka, T., and Kubo, M. 2007. SOCS proteins, cytokine signalling and immune regulation. *Nat Rev Immunol* 7:454-465.

4. Hoover, D.M., Boulegue, C., Yang, D., Oppenheim, J.J., Tucker, K., Lu, W., and Lubkowski, J. 2002. The structure of human macrophage inflammatory protein-3alpha /CCL20. Linking antimicrobial and CC chemokine receptor-6-binding activities with human beta-defensins. *J Biol Chem* 277:37647-37654.

5. Power, C.A., Church, D.J., Meyer, A., Alouani, S., Proudfoot, A.E.I., Clark-Lewis, I., Sozzani, S., Mantovani, A., and Wells, T.N.C. 1997. Cloning and characterization of a specific receptor for the novel CC chemokine MIP-3alpha from lung dendritic cells. *J Exp Med* 186:825-835.

6. Liao, F., Alderson, R., Su, J., Ullrich, S.J., Kreider, B.L., and Farber, J.M. 1997. STRL22 is a receptor for the CC chemokine MIP-3alpha. *Biochem Biophys Res Commun* 236:212.

7. Yang, D., Chertov, O., Bykovskaia, S.N., Chen, Q., Buffo, M.J., Shogan, J., Anderson, M., Schroder, J.M., Wang, J.M., Howard, O.M., and Oppenheim, J.J. 1999. Beta-defensins: linking innate and adaptive immunity through dendritic and T cell CCR6. *Science* 286:525-528.

8. Bachmann, M.F., Kopf, M., and Marsland, B.J. 2006. Chemokines: more than just road signs. *Nat Rev Immunol* 6:159.

9. Biragyn, A., Surenhu, M., Yang, D., Ruffini, P.A., Haines, B.A., Klyushnenkova, E., Oppenheim, J.J., and Kwak, L.W. 2001. Mediators of innate immunity that target immature, but not mature, dendritic cells induce antitumor immunity when genetically fused with nonimmunogenic tumor antigens. *J Immunol* 167:6644-6653.

10. Ganz, T., Selsted, M., and Lehrer, R. 1990. Defensins. *Eur J Haematol* 44:1-8.

11. Zou, J., Mercier, C., Koussounadis, A., and Secombes, C. 2007. Discovery of multiple beta-defensin like homologues in teleost fish. *Mol Immunol* 44:638-647.

12. Patil, A., Hughes, A.L., and Zhang, G. 2004. Rapid evolution and diversification of mammalian alpha-defensins as revealed by comparative analysis of rodent and primate genes. *Physiol Genomics* 20:1-11.

13. Semple, C.A., Gautier, P., Taylor, K., and Dorin, J.R. 2006. The changing of the guard: molecular diversity and rapid evolution of beta-defensins. *Mol Divers* 10:575-584.

14. Radhakrishnan, Y., Fares, M.A., French, F.S., and Hall, S.H. 2007. Comparative genomic analysis of a mammalian beta-defensin gene cluster. *Physiol Genomics* 30:213-222.

15. Lynn, D.J., Lloyd, A.T., Fares, M.A., and O'Farrelly, C. 2004. Evidence of positively selected sites in mammalian alpha-defensins. *Mol Biol Evol* 21:819-827.

16. Ganz, T. 2003. Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol* 3:710-720.

17. Lehrer, R.I., and Ganz, T. 2002. Defensins of vertebrate animals. *Curr Opin Immunol* 14:96-102.

18. Ouellette, A.J. 2004. Defensin-mediated innate immunity in the small intestine. *Best Pract Res Clin Gastroenterol* 18:405-419.

19. Nguyen, T.X., Cole, A.M., and Lehrer, R.I. 2003. Evolution of primate theta-defensins: a serpentine path to a sweet tooth. *Peptides* 24:1647-1654.

20. Ouellette, A.J., Greco, R.M., James, M., Frederick, D., Naftilan, J., and Fallon, J.T. 1989. Developmental regulation of cryptdin, a corticostatin/defensin precursor mRNA in mouse small intestinal crypt epithelium. *J Cell Biol* 108:1687-1695.

21. Ouellette, A.J., and Lualdi, J.C. 1990. A novel mouse gene family coding for cationic, cysteine-rich peptides. Regulation in small intestine and cells of myeloid origin. *J Biol Chem* 265:9831-9837.

22. Ouellette, A.J., and Lauldi, J.C. 1994. Erratum - A novel gene family coding for cationic, cysteine-rich peptides. Regulation in mouse small intestine and cells of myeloid origin. *J Biol Chem* 269:18702.

23. Ouellette, A.J., Pravtcheva, D., Ruddle, F.H., and James, M. 1989. Localization of the cryptdin locus on mouse chromosome 8. *Genomics* 5:233-239.

24. Ouellette, A.J., Pravtcheva, D., Ruddle, F.H., and James, M. 1992. Erratum - Localization of the cryptdin locus on mouse chromosome 8. *Genomics* 12:626.

25. Lin, M.Y., Munshi, I.A., and Ouellette, A.J. 1992. The defensin-related murine CRS1C gene: expression in paneth cells and linkage to Defcr, the cryptdin locus. *Genomics* 14:363-368.

26. Huttner, K.M., Selsted, M.E., and Ouellette, A.J. 1994. Structure and diversity of the murine cryptdin gene family. *Genomics* 19:448-453.

27. Huttner, K.M., Selsted, M.E., and Ouellette, A.J. 1995. Erratum - Structure and diversity of the murine cryptdin gene family. *Genomics* 25:762.

28. Ouellette, A.J., Miller, S.I., Henschen, A.H., and Selsted, M.E. 1992. Purification and primary structure of murine cryptdin-1, a Paneth cell defensin. *FEBS Lett* 304:146-148.

29. Selsted, M.E., Miller, S.I., Henschen, A.H., and Ouellette, A.J. 1992. Enteric defensins: antibiotic peptide components of intestinal host defense. *J Cell Biol* 118:929-936.

30. Mestas, J., and Hughes, C.C.W. 2004. Of mice and not men: differences between mouse and human immunology. *J Immunol* 172:2731-2738.

31. Lehrer, R.I. 2007. Multispecific myeloid defensins. *Curr Opin Hematol* 14:16-21.

32. Wehkamp, J., Salzman, N.H., Porter, E., Nuding, S., Weichenthal, M., Petras, R.E., Shen, B., Schaeffeler, E., Schwab, M., Linzmeier, R., Feathers, R.W., Chu, H., Lima, H., Jr., Fellermann, K., Ganz, T., Stange, E.F., and Bevins, C.L. 2005. Reduced Paneth cell alpha-defensins in ileal Crohn's disease. *PNAS* 102:18129-18134.

33. Ganz, T., Metcalf, J.A., Gallin, J.I., Boxer, L.A., and Lehrer, R.I. 1988. Microbicidal/cytotoxic proteins of neutrophils are deficient in two disorders: Chediak-Higashi syndrome and "specific" granule deficiency. *J Clin Invest* 82:552-556.

34. Eisenhauer, P.B., and Lehrer, R.I. 1992. Mouse neutrophils lack defensins. *Infect Immun* 60:3446-3447.

35. Ayabe, T., Satchell, D.P., Wilson, C.L., Parks, W.C., Selsted, M.E., and Ouellette, A.J. 2000. Secretion of microbicidal alpha-defensins by intestinal Paneth cells in response to bacteria. *Nat Immunol* 1:113.

36. Mastroianni, J.R., and Ouellette, A.J. 2009. Alpha-defensins in enteric innate immunity: functional Paneth cell alpha-defensins in mouse colonic lumen. *J Biol Chem* 284:27848-27856.

37. Koslowski, M.J., Beisner, J., Stange, E.F., and Wehkamp, J. 2010. Innate antimicrobial host defense in small intestinal Crohn's disease. *Int J Med Microbiol* 300:34-40.

38. Loose, C., Jensen, K., Rigoutsos, I., and Stephanopoulos, G. 2006. A linguistic model for the rational design of antimicrobial peptides. *Nature* 443:867-869.

39. Scott, M.G., Dullaghan, E., Mookherjee, N., Glavas, N., Waldbrook, M., Thompson, A., Wang, A., Lee, K., Doria, S., Hamill, P., Yu, J.J., Li, Y., Donini, O., Guarna, M.M., Finlay, B.B., North, J.R., and Hancock, R.E.W. 2007. An anti-infective peptide that selectively modulates the innate immune response. *Nat Biotech* 25:465-472.

40. Fjell, C.D., Jenssen, H., Hilpert, K., Cheung, W.A., Pante, N., Hancock, R.E.W., and Cherkasov, A. 2009. Identification of novel antibacterial peptides by chemoinformatics and machine learning. *J Med Chem* 52:2006-2015.

41. Selsted, M.E., and Ouellette, A.J. 2005. Mammalian defensins in the antimicrobial immune response. *Nat Immunol* 6:551-557.

42. Rehaume, L.M., and Hancock, R.E.W. 2008. Neutrophil-derived defensins as modulators of innate immune function. *Crit Rev Immunol* 28:185-200.

43. Salzman, N.H., Ghosh, D., Huttner, K.M., Paterson, Y., and Bevins, C.L. 2003. Protection against enteric salmonellosis in transgenic mice expressing a human intestinal defensin. *Nature* 422:522-526.

44. Scott, M.G., Davidson, D.J., Gold, M.R., Bowdish, D., and Hancock, R.E.W. 2002. The human antimicrobial peptide LL-37 is a multifunctional modulator of innate immune responses. *J Immunol* 169:3883-3891.

45. Remick, D.G. 2005. Interleukin-8. *Crit Care Med* 33:S466-S467.

46. Waugh, D.J., and Wilson, C. 2008. The interleukin-8 pathway in cancer. *Clin Cancer Res* 14:6735-6741.

47. Radkowski, M., Bednarska, A., Horban, A., Stanczak, J., Wilkinson, J., Adair, D.M., Nowicki, M., Rakela, J., and Laskus, T. 2004. Infection of primary human macrophages with hepatitis C virus in vitro: induction of tumour necrosis factor-alpha and interleukin 8. *J Gen Virol* 85:47-59.

48. Donnarumma, G., Paoletti, I., Buommino, E., Iovene, M.R., Tudisco, L., Cozza, V., and Tufano, M.A. 2007. Anti-inflammatory effects of moxifloxacin and human beta-defensin 2 association in human lung epithelial cell line (A549) stimulated with lipopolysaccharide. *Peptides* 28:2286-2292.

49. Nijnik, A., Madera, L., Ma, S., Waldbrook, M., Elliott, M.R., Easton, D.M., Mayer, M.L., Mullaly, S.C., Kindrachuk, J., Jenssen, H., and Hancock, R.E.W. 2010. Synthetic cationic peptide IDR-1002 provides protection against bacterial infections through chemokine induction and enhanced leukocyte recruitment. *J Immunol* 184:2539-2550.

50. Murphy, P., Baggiolini, M., Charo, I., Hébert, C., Horuk, R., Matsushima, K., Miller, L., Oppenheim, J., and Power, C. 2000. International union of pharmacology. XXII. Nomenclature for chemokine receptors. *Pharmacol Rev* 52:145-176.

51. Kobayashi, Y. 2008. The role of chemokines in neutrophil biology. *Front Biosci* 1:2400-2407.

52. Stillie, R., Farooq, S.M., Gordon, J.R., and Stadnyk, A.W. 2009. The functional significance behind expressing two IL-8 receptor types on PMN. *J Leukoc Biol* 86:529-543.

53. Soehnlein, O. 2009. An elegant defense: how neutrophils shape the immune response. *Trends Immunol* 30:511-512.

54. Pruijt, J., Verzaal, P., van Os, R., de Kruijf, E., van Schie, M., Mantovani, A., Vecchi, A., Lindley, I., Willemze, R., Starckx, S., Opdenakker, G., and Fibbe, W. 2002. Neutrophils are indispensable for hematopoietic stem cell mobilization induced by interleukin-8 in mice. *Proc Natl Acad Sci USA* 99:6228-6233.

55. Hol, J., Wilhelmsen, L., and Haraldsen, G. 2010. The murine IL-8 homologues KC, MIP-2, and LIX are found in endothelial cytoplasmic granules but not in Weibel-Palade bodies. *J Leukoc Biol* 87:501-508.

56. Harada, A., Kuno, K., Nomura, H., Mukaida, N., Murakami, S., and Matsushima, K. 1994. Cloning of a cDNA encoding a mouse homolog of the interleukin-8 receptor. *Gene* 142:297-300.

57. Cerretti, D.P., Nelson, N., Kozlosky, C.J., Morrissey, P.J., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Dosik, J.K., and Mock, B.A. 1993. The murine homologue of the human interleukin-8 receptor type B maps near the Ity-Lsh-Bcg disease resistance locus. *Genomics* 18:410-413.

58. Bozic, C.R., Gerard, N.P., von Uexkull-Guldenband, C., Kolakowski LF, J., Conklyn, M.J., Breslow, R., Showell, H.J., and Gerard, C. 1994. The murine interleukin 8 type B receptor homologue and its ligands. Expression and biological characterization. *J Biol Chem* 269:29355-29358.

59. Fu, W., Zhang, Y., Zhang, J., and Chen, W.F. 2005. Cloning and characterization of mouse homolog of the CXC chemokine receptor CXCR1. *Cytokine* 31:9-17.

60. Moepps, B., Nuesseler, E., Braun, M., and Gierschik, P. 2006. A homolog of the human chemokine receptor CXCR1 is expressed in the mouse. *Mol Immunol* 43:897-914.

61. Fan, X., Patera, A.C., Pong-Kennedy, A., Deno, G., Gonsiorek, W., Manfra, D.J., Vassileva, G., Zeng, M., Jackson, C., Sullivan, L., Sharif-Rodriguez, W., Opdenakker, G., Van Damme, J., Hedrick, J., Lundell, D., Lira, S.A., and Hipkin, R.W. 2007. Murine CXCR1 is a functional receptor for GCP-2/CXCL6 and interleukin-8/CXCL8. *J Biol Chem* 282:11658-11666.

62. Hall, D.A., Beresford, I.J.M., Browning, C., and Giles, H. 1999. Signalling by CXC-chemokine receptors 1 and 2 expressed in CHO cells: a comparison of calcium mobilization, inhibition of adenylyl cyclase and stimulation of GTPgammaS binding induced by IL-8 and GROalpha. *Br J Pharmacol* 126:810-818.

63. Moore, K.W., de Waal Malefyt, R., Coffman, R.L., and O'Garra, A. 2001. Interleukin-10 and the interleukin-10 receptor. *Annu Rev Immunol* 19:683-765.

64. O'Garra, A., and Vieira, P. 2007. T(H)1 cells control themselves by producing interleukin-10. *Nat Rev Immunol* 7:425-428.

65. Mosser, D.M., and Zhang, X. 2008. Interleukin-10: new perspectives on an old cytokine. *Immunol Rev* 226:205-218.

66. Saraiva, M., and O'Garra, A. 2010. The regulation of IL-10 production by immune cells. *Nat Rev Immunol* 10:170-181.

67. Fiorentino, D.F., Bond, M.W., and Mosmann, T.R. 1989. Two types of mouse T helper cell. IV. Th2 clones secrete a factor that inhibits cytokine production by Th1 clones. *J Exp Med* 170:2081-2095.

68. Monteleone, I., Platt, A.M., Jaensson, E., Agace, W.W., and Mowat, Allan M. 2008. IL-10-dependent partial refractoriness to Toll-like receptor stimulation modulates gut mucosal dendritic cell function. *Eur J Immunol* 38:1533-1547.

69. Boniotto, M., Jordan, W.J., Eskdale, J., Tossi, A., Antcheva, N., Crovella, S., Connell, N.D., and Gallagher, G. 2006. Human beta-defensin 2 induces a vigorous cytokine

response in peripheral blood mononuclear cells. *Antimicrob Agents Chemother* 50:1433-1441.

70. World Health Organization [http://www.who.int/en/]

71. Girard, M.P., Steele, D., Chaignat, C.L., and Kieny, M.P. 2006. A review of vaccine research and development: human enteric infections. *Vaccine* 24:2732-2750.

72. Shanahan, F., and Bernstein, C.N. 2009. The evolving epidemiology of inflammatory bowel disease. *Curr Opin Gastroenterol* 25:301-305.

73. Center, M.M., Jemal, A., Smith, R.A., and Ward, E. 2009. Worldwide variations in colorectal cancer. *CA Cancer J Clin* 59:366-378.

74. Chin, A.C., and Parkos, C.A. 2007. Pathobiology of neutrophil transepithelial migration: implications in mediating epithelial injury. *Annu Rev Pathol* 2:111-143.

75. Smith, A.J.P., and Humphries, S.E. 2009. Cytokine and cytokine receptor gene polymorphisms and their functionality. *Cytokine Growth Factor Rev* 20:43-59.

76. Sansonetti, P., Kopecko, D., and Formal, S. 1982. Involvement of a plasmid in the invasive ability of Shigella flexneri. *Infect Immun* 35:852-860.

77. Pédron, T., Thibault, C., and Sansonetti, P.J. 2003. The invasive phenotype of Shigella flexneri directs a distinct gene expression pattern in the human intestinal epithelial cell line Caco-2. *J Biol Chem* 278:33878-33886.

78. Phalipon, A., and Sansonetti, P.J. 2007. Shigella's ways of manipulating the host intestinal innate and adaptive immune system: a tool box for survival? *Immunol Cell Biol* 85:119-129.

79. Singer, M., and Sansonetti, P.J. 2004. IL-8 is a key chemokine regulating neutrophil recruitment in a new mouse model of Shigella-induced colitis. *J Immunol* 173:4197-4206.

80. Phalipon, A., Kaufmann, M., Michetti, P., Cavaillon, J., Huerre, M., Sansonetti, P., and Kraehenbuhl, J. 1995. Monoclonal immunoglobulin A antibody directed against serotype-specific epitope of Shigella flexneri lipopolysaccharide protects against murine experimental shigellosis. *J Exp Med* 182:769-778.

81. Rehaume, L.M., Jouault, T., and Chamaillard, M. 2010. Lessons from the inflammasome: a molecular sentry linking Candida and Crohn's disease. *Trends Immunol* 31:171-175.

82. Valdez, Y., Ferreira, R., and Finlay, B. 2009. Molecular mechanisms of Salmonella virulence and host resistance. *Curr Top Microbiol Immunol* 337:93-127.

83. Raffatellu, M., Chessa, D., Wilson, R.P., Dusold, R., Rubino, S., and Baumler, A.J. 2005. The Vi capsular antigen of Salmonella enterica serotype Typhi reduces Toll-like receptor-dependent interleukin-8 expression in the intestinal mucosa. *Infect Immun* 73:3367-3374.

84. Pickard, D., Wain, J., Baker, S., Line, A., Chohan, S., Fookes, M., Barron, A., O Gaora, P., Chabalgoity, J.A., Thanky, N., Scholes, C., Thomson, N., Quail, M., Parkhill, J., and Dougan, G. 2003. Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding Salmonella enterica pathogenicity island SPI-7. *J Bacteriol* 185:5055-5065.

85. Chamekh, M., Phalipon, A., Quertainmont, R., Salmon, I., Sansonetti, P., and Allaoui, A. 2008. Delivery of biologically active anti-inflammatory cytokines IL-10 and IL-1ra in vivo by the Shigella type III secretion apparatus. *J Immunol* 180:4292-4298.

86. Yang, D., Chen, Q., Chertov, O., and Oppenheim, J.J. 2000. Human neutrophil defensins selectively chemoattract naive T and immature dendritic cells. *J Leukoc Biol* 68:9-14.

87. Grigat, J., Soruri, A., Forssmann, U., Riggert, J., and Zwirner, J. 2007. Chemoattraction of macrophages, T lymphocytes, and mast cells is evolutionarily conserved within the human alpha-defensin family. *J Immunol* 179:3958-3965.

88. Ouellette, A., Hsieh, M., Nosek, M., Cano-Gauci, D., Huttner, K., Buick, R., and Selsted, M. 1994. Mouse Paneth cell defensins: primary structures and antibacterial activities of numerous cryptdin isoforms. *Infect Immun* 62:5040-5047.

89. Morrison, G., Kilanowski, F., Davidson, D., and Dorin, J. 2002. Characterization of the mouse beta-defensin 1, Defb1, mutant mouse model. *Infect Immun* 70:3053-3060.

90. Moser, C., Weiner, D.J., Lysenko, E., Bals, R., Weiser, J.N., and Wilson, J.M. 2002. Beta-defensin 1 contributes to pulmonary innate immunity in mice. *Infect Immun* 70:3068-3072.

91. Welling, M.M., Hiemstra, P.S., van den Barselaar, M.T., Paulusma-Annema, A., Nibbering, P.H., Pauwels, E.K.J., and Calame, W. 1998. Antibacterial activity of human neutrophil defensins in experimental infections in mice is accompanied by increased leukocyte accumulation. *J Clin Invest* 102:1583-1590.

92. Ericksen, B., Wu, Z., Lu, W., and Lehrer, R.I. 2005. Antibacterial activity and specificity of the six human alpha-defensins. *Antimicrob Agents Chemother* 49:269-275.

93. Niyonsaba, F., Iwabuchi, K., Matsuda, H., Ogawa, H., and Nagaoka, I. 2002. Epithelial cell-derived human beta-defensin-2 acts as a chemotaxin for mast cells through a pertussis toxin-sensitive and phospholipase C-dependent pathway. *Int Immunol* 14:421-426.

94. Territo, M.C., Ganz, T., Selsted, M.E., and Lehrer, R. 1989. Monocyte-chemotactic activity of defensins from human neutrophils. *J Clin Invest* 84:2017-2020.

95. Jan, M.S., Huang, Y.H., Shieh, B., Teng, R.H., Yan, Y.P., Lee, Y.T., Liao, K.K., and Li, C. 2006. CC chemokines induce neutrophils to chemotaxis, degranulation and alpha-defensin release. *J Acquir Immune Defic Syndr* 41:6-16.

96. Salzman, N., Hung, K., Haribhai, D., Chu, H., Karlsson-Sjoberg, J., Amir, E., Teggatz, P., Barman, M., Hayward, M., Eastwood, D., Stoel, M., Zhou, Y., Sodergren, E.,

Weinstock, G., Bevins, C., Williams, C., and Bos, N. 2010. Enteric defensins are essential regulators of intestinal microbial ecology. *Nature Immunol* 11:76-82.

97. Gaire, M., Magbanua, Z., McDonnell, S., McNei, l.L., Lovett, D.H., and Matrisian, L.M. 1994. Structure and expression of the human gene for matrix metalloproteinase matrilysin. *J Biol Chem* 21:2032-2040.

98. Lu, P.C., Ye, H.M., M., and Azar, D.T. 1999. Immunolocalization and gene expression of matrilysin during corneal wound healing. *Invest Ophthal Vis Sci* 40:20-27.

99. Li, Q., Park, P.W., Wilson, C.L., and Parks, W.C. 2002. Matrilysin shedding of syndecan-1 regulates chemokine mobilization and transepithelial efflux of neutrophils in acute lung injury. *Cell* 111:635-646.

100. Bevins, C.L. 2006. Paneth cell defensins: key effector molecules of innate immunity. *Biochem Soc Trans* 34:263-266.

101. Hornef, M.W., Putsep, K., Karlsson, J., Refai, E., and Andersson, M. 2004. Increased diversity of intestinal antimicrobial peptides by covalent dimer formation. *Nat Immunol* 5:836-843.

102. Darmoul, D., Brown, D., Selsted, M.E., and Ouellette, A.J. 1997. Cryptdin gene expression in developing mouse small intestine. *Am J Physiol Gastrointest Liver Physiol* 272:G197-206.

103. Darmoul, D., and Ouellette, A.J. 1996. Positional specificity of defensin gene expression reveals Paneth cell hetergeneity in mouse small intestine. *Am J Physiol Gastrointest Liver Physiol* 271:G68-74.

104. Potter, S.C., Clarke, L., Curwen, V., Keenan, S., Mongin, E., Searle, S.M.J., Stabenau, A., Storey, R., and Clamp, M. 2004. The Ensembl analysis pipeline. *Genome Res* 14:934-941.

105. Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27:573-580.

106. Mott, R. 1997. EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. *Comput Appl Biosci* 13:477-478.

107. Birney, E., Clamp, M., and Durbin, R. 2004. GeneWise and Genomewise. *Genome Res* 14:988-995.

108. Burge, C., and Karlin, S. 1997. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 268:78-94.

109. Lowe, T.M., and Eddy, S.R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955-964.

110. Down, T.A., and Hubbard, T.J.P. 2002. Computational detection and location of transcription start sites in mammalian genomic DNA. *Genome Res* 12:458-461.

111. Searle, S.M.J., Gilbert, J., Iyer, V., and Clamp, M. 2004. The Otter annotation system. *Genome Res* 14:963-970.

112. Sonnhammer, E., and Wootton, J. 2001. Integrated graphical analysis of protein sequence features predicted from sequence composition. *Proteins* 45:262-273.

113. Vega Genome Browser [http://vega.sanger.ac.uk/index.html]

114. Genome Reference Consortium [http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/index.shtml]

115. RestrictionMapper [www.restrictionmapper.org]

116. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948.

117. GeneDoc: a tool for editing and annotating multiple sequence alignments [http://www.psc.edu/biomed/genedoc]

118. Suyama, M., Torrents, D., and Bork, P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucl Acids Res* 34:W609-612.

119. Kibbe, W.A. 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* 35:W43-46.

120. Thomas, M., Hesse, S., McKie, A., and Farzaneh, F. 1993. Sequencing of cDNA using anchored oligo dT primers. *Nucleic Acids Res* 21:3915-3916.

121. Nam, D., Lee, S., Zhou, G., Cao, X., Wang, C., Clark, T., Chen, J., Rowley, J., and SM, W. 2002. Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. *Proc Natl Acad Sci USA* 99:6152-6156.

122. Wu, Z., Powell, R., and Lu, W. 2003. Productive folding of human neutrophil alpha-defensins in vitro without the pro-peptide. *J Am Chem Soc* 125:2402-2403.

123. Darveau, R.P., and Hancock, R.E. 1983. Procedure for isolation of bacterial lipopolysaccharides from both smooth and rough Pseudomonas aeruginosa and Salmonella typhimurium strains. *J Bacteriol* 155:831-838.

124. Church, D.M., Goodstadt, L., Hillier, L.W., Zody, M.C., Goldstein, S., She, X., Bult, C.J., Agarwala, R., Cherry, J.L., DiCuccio, M., Hlavina, W., Kapustin, Y., Meric, P., Maglott, D., Birtle, Z., Marques, A.C., Graves, T., Zhou, S., Teague, B., Potamousis, K., Churas, C., Place, M., Herschleb, J., Runnheim, R., Forrest, D., Amos-Landgraf, J., Schwartz, D.C., Cheng, Z., Lindblad-Toh, K., Eichler, E.E., Ponting, C.P., and The Mouse Genome Sequencing Consortium. 2009. Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol* 7:e1000112.

125. Hall, S.H., Yenugu, S., Radhakrishnan, Y., Avellar, M.C.W., Petrusz, P., and French, F.S. 2007. Characterization and functions of beta-defensins in the epididymis. *Asian J Androl* 9:453-462.

126. Patil, A.A., Cai, Y., Sang, Y., Blecha, F., and Zhang, G. 2005. Cross-species analysis of the mammalian beta-defensin gene family: presence of syntenic gene clusters and preferential expression in the male reproductive tract. *Physiol Genomics* 23:5-17.

127. Huttner, K.M., and Ouellette, A.J. 1994. A family of defensin-like genes codes for diverse cysteine-rich peptides in mouse Paneth cells. *Genomics* 24:99-109.

128. Jalkanen, J., Huhtaniemi, I., and Poutanen, M. 2005. Discovery and characterization of new epididymis-specific beta-defensins in mice. *Biochim Biophys Acta* 1730:22-30.

129. Tsutsumi-Ishii, Y., Hasebe, T., and Nagaoka, I. 2000. Role of CCAAT/enhancer-binding protein site in transcription of human neutrophil peptide-1 and -3 defensin genes. *J Immunol* 164:3264-3273.

130. Landry, C.R., Lemos, B., Rifkin, S.A., Dickinson, W.J., and Hartl, D.L. 2007. Genetic properties influencing the evolvability of gene expression. *Science* 317:118-121.

131. Zheng, D., Frankish, A., Baertsch, R., Kapranov, P., Reymond, A., Choo, S.W., Lu, Y., Denoeud, F., Antonarakis, S.E., Snyder, M., Ruan, Y., Wei, C.L., Gingeras, T.R., Guigo, R., Harrow, J., and Gerstein, M.B. 2007. Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. *Genome Res* 17:839-851.

132. Piehler, A., Hellum, M., Wenzel, J., Kaminski, E., Haug, K., Kierulf, P., and Kaminski, W. 2008. The human ABC transporter pseudogene family: evidence for transcription and gene-pseudogene interference. *BMC Genomics* 9:165.

133. Blake, J.A., Bult, C.J., Eppig, J.T., Kadin, J.A., Richardson, J.E., and the Mouse Genome Database, G. 2008. The Mouse Genome Database genotypes::phenotypes. *Nucleic Acids Res* 37:D712-719.

134. Jing, W., Hunter, H.N., Tanabe, H., Ouellette, A.J., and Vogel, H.J. 2004. Solution structure of cryptdin-4, a mouse Paneth cell alpha-defensin. *Biochemistry* 43:15759-15766.

135. Shirafuji, Y., Tanabe, H., Satchell, D.P., Henschen-Edman, A., Wilson, C.L., and Ouellette, A.J. 2002. Structural determinants of procryptdin recognition and cleavage by matrix metalloproteinase-7. *J Biol Chem* 278:7910-7919.

136. UCSC Genome Browser [http://genome.ucsc.edu/cgi-bin/hgGateway]

137. Mouse Chromosome 8 Linkage Map [http://www.informatics.jax.org/searches/linkmap_form.shtml]

138. Rat Genome Sequencing Project Consortium. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428:493-521.

139. Twigger, S.N., Pruitt, K.D., Fernandez-Suarez, X.M., Karolchik, D., Worley, K.C., Maglott, D.R., Brown, G., Weinstock, G., Gibbs, R.A., Kent, J., Birney, E., and Jacob, H.J. 2008. What everybody should know about the rat genome and its online resources. *Nat Genet* 40:523-527.

140. Human Genome Sequencing Consortium International. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431:931-945.

141. Rudd, M.K., and Willard, H.F. 2004. Analysis of the centromeric regions of the human genome assembly. *Trends Genet* 20:529-533.

142. Kyte, J., and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157:105-132.

143. Peptide Calculator [http://www.sigma-genosys.com/calc/pepCalc.asp]

144. Schutte, B.C., Mitros, J.P., Bartlett, J.A., Walters, J.D., Jia, H.P., Welsh, M.J., Casavant, T.L., and McCray, P.B. 2002. Erratum - Discovery of five conserved beta-defensin gene clusters using a computational search strategy. *PNAS* 99:14611.

145. Hollox, E.J., Armour, J.A., and Barber, J.C. 2003. Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. *Am J Hum Genet* 72:591-600.

146. Abu Bakar, S., Hollox, E.J., and Armour, J.A.L. 2009. Allelic recombination between distinct genomic locations generates copy number diversity in human beta-defensins. *Proc Natl Acad Sci USA* 106:853-858.

147. Linzmeier, R.M., and Ganz, T. 2005. Human defensin gene copy number polymorphisms: comprehensive analysis of independent variation in alpha- and beta-defensin regions at 8p22-p23. *Genomics* 86:423-430.

148. Freeman, J.L., Perry, G.H., Feuk, L., Redon, R., McCarroll, S.A., Altshuler, D.M., Aburatani, H., Jones, K.W., Tyler-Smith, C., Hurles, M.E., Carter, N.P., Scherer, S.W., and Lee, C. 2006. Copy number variation: new insights in genome diversity. *Genome Res* 16:949-961.

149. Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T.C., Trask, B., Patterson, N., Zetterberg, A., and Wigler, M. 2004. Large-scale copy number polymorphism in the human genome. *Science* 305:525-528.

150. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. 2004. Detection of large-scale variation in the human genome. *Nat Genet* 36:949-951.

151. Barber, J.C., Joyce, C.A., Collinson, M.N., Nicholson, J.C., Willatt, L.R., Dyson, H.M., Bateman, M.S., Green, A.J., Yates, J.R., and Dennis, N.R. 1998. Duplication of 8p23.1: a cytogenetic anomaly with no established clinical significance. *J Med Genet* 35:491-496.

152. O'Malley, D., and Storto, P. 1999. Confirmation of the chromosome 8p23.1 euchromatic duplication as a variant with no clinical manifestations. *Prenat Diagn* 19:183-184.

153. Aldred, P.M.R., Hollox, E.J., and Armour, J.A.L. 2005. Copy number polymorphism and expression level variation of the human alpha-defensin genes DEFA1 and DEFA3. *Hum Mol Genet* 14:2045-2052.

154. Ballana, E., Gonzalez, J.R., Bosch, N., and Estivill, X. 2007. Inter-population variability of DEFA3 gene absence: correlation with haplotype structure and population variability. *BMC Genomics* 8:1-10.

155. Adams, D.J., Biggs, P.J., Cox, T., Davies, R., van der Weyden, L., Jonkers, J., Smith, J., Plumb, B., Taylor, R., Nishijima, I., Yu, Y., Rogers, J., and Bradley, A. 2004. Mutagenic insertion and chromosome engineering resource (MICER). *Nature Genet* 36:867-871.

156. Ouellette, A.J., Darmoul, D., Tran, D., Huttner, K.M., Yuan, J., and Selsted, M.E. 1999. Peptide localization and gene structure of cryptdin 4, a differentially expressed mouse Paneth cell alpha-defensin. *Infect Immun* 67:6643-6651.

157. Sakamoto, N., Mukae, H., Fujii, T., Ishii, H., Yoshioka, S., Kakugawa, T., Sugiyama, K., Mizuta, Y., Kadota, J.-i., Nakazato, M., and Kohno, S. 2005. Differential effects of alpha- and beta-defensin on cytokine production by cultured human bronchial epithelial cells. *Am J Physiol Lung Cell Mol Physiol* 288:L508-513.

158. Finlay, B.B., and Hancock, R.E.W. 2004. Can innate immunity be enhanced to treat microbial infections? *Nat Rev Micro* 2:497-504.

159. Fleckenstein, J.M., and Kopecko, D.J. 2001. Breaching the mucosal barrier by stealth: an emerging pathogenic mechanism for enteroadherent bacterial pathogens. *J Clin Invest* 107:27-30.

160. Dharmani, P., and Chadee, K. 2008. Biologic therapies against inflammatory bowel disease: a dysregulated immune system and the cross talk with gastrointestinal mucosa hold the key. *Curr Mol Pharmacol* 3:195-212.

161. Padilla-Carlin, D.J., McMurray, D.N., and Hickey, A.J. 2008. The guinea pig as a model of infectious diseases. *Comp Med* 58:324-340.

162. Austin, C.P., Battey, J.F., Bradley, A., Bucan, M., Capecchi, M., Collins, F.S., Dove, W.F., Duyk, G., Dymecki, S., Eppig, J.T., Grieder, F.B., Heintz, N., Hicks, G., Insel, T.R., Joyner, A., Koller, B.H., Lloyd, K.C., Magnuson, T., Moore, M.W., Nagy, A., Pollock, J.D., Roses, A.D., Sands, A.T., Seed, B., Skarnes, W.C., Snoddy, J., Soriano, P., Stewart, D.J., Stewart, F., Stillman, B., Varmus, H., Varticovski, L., Verma, I.M., Vogt, T.F., von Melchner, H., Witkowski, J., Woychik, R.P., Wurst, W., Yancopoulos, G.D., Young, S.G., and Zambrowicz, B. 2004. The knockout mouse project. *Nat Genet* 36:921-924.

163. Simonet, W.S., Hughes, T.M., Nguyen, H.Q., Trebasky, L.D., Danilenko, D.M., and Medlock, E.S. 1994. Long-term impaired neutrophil migration in mice overexpressing human interleukin-8. *J Clin Invest* 94:1310-1319.

164. Kucharzik, T., Hudson, J.T., Lügering, A., Abbas, J., Bettini, M., Lake, J.G., Evans, M.E., Ziegler, T.R., Merlin, D., Madara, J.L., and Williams, I.R. 2005. Acute induction

of human IL-8 production by intestinal epithelium triggers neutrophil infiltration without mucosal injury. *Gut* 54:1565-1572.

165. Simonet, W., Bucay, N., Lauer, S., and Taylor, J. 1993. A far-downstream hepatocyte-specific control region directs expression of the linked human apolipoprotein E and C-I genes in transgenic mice. *J Biol Chem* 268:8221-8229.

166. Matthaei, K. 2007. Genetically manipulated mice: a powerful tool with unsuspected caveats. *J Physiol* 582:481-488.

167. Mouse Genome Informatics [http://www.informatics.jax.org/]

168. Vieira, P., and O'Garra, A. 2007. Regula'ten' the gut. *Nat Immunol* 8:905-907.

169. Davidson, N.J., Leach, M.W., Fort, M.M., Thompson-Snipes, L., Kuhn, R., Muller, W., Berg, D.J., and Rennick, D.M. 1996. T helper cell 1-type CD4+ T cells, but not B cells, mediate colitis in interleukin 10-deficient mice. *J Exp Med* 184:241-251.

170. O'Garra, A., Vieira, P.L., Vieira, P., and Goldfeld, A.E. 2004. IL-10-producing and naturally occurring CD4+ Tregs: limiting collateral damage. *J Clin Invest* 114:1372-1378.

171. Sellon, R.K., Tonkonogy, S., Schultz, M., Dieleman, L.A., Grenther, W., Balish, E., Rennick, D.M., and Sartor, R.B. 1998. Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. *Infect Immun* 66:5224-5231.

172. Battaglia, M., Stabilini, A., Draghici, E., Migliavacca, B., Gregori, S., Bonifacio, E., and Roncarolo, M.G. 2006. Induction of tolerance in type 1 diabetes via both CD4+CD25+ T regulatory cells and T regulatory type 1 cells. *Diabetes* 55:1571-1580.

173. Maynard, C.L., Harrington, L.E., Janowski, K.M., Oliver, J.R., Zindl, C.L., Rudensky, A.Y., and Weaver, C.T. 2007. Regulatory T cells expressing interleukin 10 develop from Foxp3+ and Foxp3- precursor cells in the absence of interleukin 10. *Nat Immunol* 8:931-941.

174. Kamanaka, M., Kim, S.T., Wan, Y.Y., Sutterwala, F.S., Lara-Tejero, M., Galán, J.E., Harhaj, E., and Flavell, R.A. 2006. Expression of interleukin-10 in intestinal lymphocytes detected by an interleukin-10 reporter knockin tiger mouse. *Immunity* 25:941-952.

175. De Winter, H., Elewaut, D., Turovskaya, O., Huflejt, M., Shimeld, C., Hagenbaugh, A., Binder, S., Takahashi, I., Kronenberg, M., and Cheroutre, H. 2002. Regulation of mucosal immune responses by recombinant interleukin 10 produced by intestinal epithelial cells in mice. *Gastroenterology* 122:1829-1841.

176. Rajan, S., Vyas, D., Clark, A., Woolsey, C., Clark, J., Hotchkiss, R., Buchman, T., and Coopersmith, C. 2008. Intestine-specific overexpression of IL-10 improves survival in polymicrobial sepsis. *Shock* 29:483-489.

177. Samuel, M.S., Suzuki, H., Buchert, M., Putoczki, T.L., Tebbutt, N.C., Lundgren-May, T., Christou, A., Inglese, M., Toyota, M., Heath, J.K., Ward, R.L., Waring, P.M., and Ernst,

M. 2009. Elevated Dnmt3a activity promotes polyposis in ApcMin mice by relaxing extracellular restraints on Wnt signaling. *Gastroenterology* 137:902-913.e911.

178. Ensembl Genome Browser [http://www.ensembl.org/index.html]

179. Abud, H.E., Johnstone, C.N., Tebbutt, N.C., and Heath, J.K. 2000. The murine A33 antigen is expressed at two distinct sites during development, the ICM of the blastocyst and the intestinal epithelium. *Mech Dev* 98:111-114.

180. Johnstone, C.N., Tebbutt, N.C., Abud, H.E., White, S.J., Stenvers, K.L., Hall, N.E., Cody, S.H., Whitehead, R.H., Catimel, B., Nice, E.C., Burgess, A.W., and Heath, J.K. 2000. Characterization of mouse A33 antigen, a definitive marker for basolateral surfaces of intestinal epithelial cells. *Am J Physiol Gastrointest Liver Physiol* 279:G500-510.

181. Johnstone, C.N., White, S.J., Tebbutt, N.C., Clay, F.J., Ernst, M., Biggs, W.H., Viars, C.S., Czekay, S., Arden, K.C., and Heath, J.K. 2002. Analysis of the regulation of the A33 antigen gene reveals intestine-specific mechanisms of gene expression. *J Biol Chem* 277:34531-34539.

182. Pereira-Fantini, P.M., Judd, L.M., Kalantzis, A., Peterson, A., Ernst, M., Heath, J.K., and Giraud, A.S. 2009. A33 antigen-deficient mice have defective colonic mucosal repair. *Inflamm Bowel Dis* 16:604-612.

183. Hopp, T.P., Prickett, K.S., Price, V.L., Libby, R.T., March, C.J., Pat Cerretti, D., Urdal, D.L., and Conlon, P.J. 1988. A short polypeptide marker sequence useful for recombinant protein identification and purification. *Nat Biotech* 6:1204-1210.

184. NCBI Primer-BLAST [http://www.ncbi.nlm.nih.gov/tools/primer-blast/]

185. Bonfield, J.K., Smith, K.F., and Staden, R. 1995. A new DNA sequence assembly program. *Nucleic Acids Res* 23:4992-4999.

186. Orner, G.A., Dashwood, W.-M., Blum, C.A., Díaz, G.D., Li, Q., Al-Fageeh, M., Tebbutt, N., Heath, J.K., Ernst, M., and Dashwood, R.H. 2002. Response of Apc(min) and A33 (delta N beta-cat) mutant mice to treatment with tea, sulindac, and 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP). *Mutat Res* 506-507:121-127.

187. Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., III, Burland, V., Mau, B., Petrosino, J.F., Qin, X., Muzny, D.M., Ayele, M., Gibbs, R.A., Csorgo, B., Posfai, G., Weinstock, G.M., and Blattner, F.R. 2008. The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. *J Bacteriol* 190:2597-2606.

188. Vincze, T., Posfai, J., and Roberts, R.J. 2003. NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res* 31:3688-3691.

189. Pettitt, S.J., Liang, Q., Rairdan, X.Y., Moran, J.L., Prosser, H.M., Beier, D.R., Lloyd, K.C., Bradley, A., and Skarnes, W.C. 2009. Agouti C57BL/6N embryonic stem cells for mouse genetic resources. *Nat Meth* 6:493-495.

190. Denning, T.L., Campbell, N.A., Song, F., Garofalo, R.P., Klimpel, G.R., Reyes, V.E., and Ernst, P.B. 2000. Expression of IL-10 receptors on epithelial cells from the murine small and large intestine. *Int Immunol* 12:133-139.

191. Friedrich, G., and Soriano, P. 1991. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* 5:1513-1523.

192. Zambrowicz, B.P., Imamoto, A., Fiering, S., Herzenberg, L.A., Kerr, W.G., and Soriano, P. 1997. Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. *Proc Natl Acad Sci USA* 94:3789-3794.

193. Marjou, F.E., Janssen, K.-P., Chang, B.H.-J., Li, M., Hindie, V., Chan, L., Louvard, D., Chambon, P., Metzger, D., and Robine, S. 2004. Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* 39:186-193.

194. Briones, G., Hofreuter, D., and Galan, J.E. 2006. Cre reporter system to monitor the translocation of type III secreted proteins into host cells. *Infect Immun* 74:1084-1090.

195. Salzman, N.H., Chou, M.M., de Jong, H., Liu, L., Porter, E.M., and Paterson, Y. 2003. Enteric salmonella infection inhibits Paneth cell antimicrobial peptide expression. *Infect Immun* 71:1109-1115.

196. Gunn, J.S., and Miller, S.I. 1996. PhoP-PhoQ activates transcription of pmrAB, encoding a two-component regulatory system involved in Salmonella typhimurium antimicrobial peptide resistance. *J Bacteriol* 178:6857-6864.

197. Eswarappa, S.M., Panguluri, K.K., Hensel, M., and Chakravortty, D. 2008. The yejABEF operon of Salmonella confers resistance to antimicrobial peptides and contributes to its virulence. *Microbiology* 154:666-678.

198. Crouch, M.L., Becker, L.A., Bang, I.-S., Tanabe, H., Ouellette, A.J., and Fang, F.C. 2005. The alternative sigma factor sigma(E) is required for resistance of Salmonella enterica serovar Typhimurium to anti-microbial peptides. *Mol Microbiol* 56:789-799.

199. Tani, K., Murphy, W.J., Chertov, O., Salcedo, R., Koh, C.Y., Utsunomiya, I., Funakoshi, S., Asai, O., Herrmann, S.H., Wang, J.M., Kwak, L.W., and Oppenheim, J.J. 2000. Defensins act as potent adjuvants that promote cellular and humoral immune responses in mice to a lymphoma idiotype and carrier antigens. *Int Immunol* 12:691-700.

200. Soehnlein, O., Kai-Larsen, Y., Frithiof, R., Sorensen, O.E., Kenne, E., Scharffetter-Kochanek, K., Eriksson, E.E., Herwald, H., Agerberth, B., and Lindbom, L. 2008. Neutrophil primary granule proteins HBP and HNP1-3 boost bacterial phagocytosis by human and murine macrophages. *J Clin Invest* 118:3491-3502.

201. Yang, Y., Jiang, Y., Yin, Q., Liang, H., and She, R. 2010. Chicken intestine defensins activated murine peripheral blood mononuclear cells through the TLR4-NF-kappaB pathway. *Vet Immunol Immunopathol* 133:59-65.

202. Naoko, K., and Shinichi, W. 2008. IL-12, IL-23, and IL-27 enhance human beta-defensin-2 production in human keratinocytes. *Eur J Immunol* 38:1287-1296.

203. Joly, S., Organ, C.C., Johnson, G.K., McCray, J.P.B., and Guthmiller, J.M. 2005. Correlation between beta-defensin expression and induction profiles in gingival keratinocytes. *Mol Immunol* 42:1073-1084.

204. Simmons, C.P., Goncalves, N.S., Ghaem-Maghami, M., Bajaj-Elliott, M., Clare, S., Neves, B., Frankel, G., Dougan, G., and MacDonald, T.T. 2002. Impaired resistance and enhanced pathology during infection with a noninvasive, attaching-effacing enteric bacterial pathogen, Citrobacter rodentium, in mice Lacking IL-12 or IFN-gamma. *J Immunol* 168:1804-1812.

205. Döffinger, R., Dupuis, S., Picard, C., Fieschi, C., Feinberg, J., Barcenas-Morales, G., and Casanova, J.L. 2002. Inherited disorders of IL-12- and IFN-gamma-mediated immunity: a molecular genetics update. *Mol Immunol* 38:903-909.

206. Patel, S., Doffinger, R., Barcenas-Morales, G., and Kumararatne, D. 2008. Genetically determined susceptibility to mycobacterial infection. *J Clin Pathol* 61:1006-10112.

207. Altare, F., Durandy, A., Lammas, D., Emile, J., Lamhamedi, S., Le Deist, F., Drysdale, P., Jouanguy, E., Doffinger, R., Bernaudin, F., Jeppsson, O., Gollob, J., Meinl, E., Segal, A., Fischer, A., Kumararatne, D., and Casanova, J.L. 1998. Impairment of Mycobacterial immunity in human interleukin-12 receptor deficiency. *Science* 280:1432-1435.

208. Jouanguy, E., Lamhamedi-Cherradi, S., Lammas, D., Dorman, S.E., Fondaneche, M.C., Dupuis, S., Doffinger, R., Altare, F., Girdlestone, J., Emile, J.F., Ducoulombier, H., Edgar, D., Clarke, J., Oxelius, V.-A., Brai, M., Novelli, V., Heyne, K., Fischer, A., Holland, S.M., Kumararatne, D.S., Schreiber, R.D., and Casanova, J.L. 1999. A human IFNGR1 small deletion hotspot associated with dominant susceptibility to mycobacterial infection. *Nat Genet* 21:370-378.

209. Fieschi, C., Dupuis, S., Catherinot, E., Feinberg, J., Bustamante, J., Breiman, A., Altare, F., Baretto, R., Le Deist, F., Kayal, S., Koch, H., Richter, D., Brezina, M., Aksu, G., Wood, P., Al-Jumaah, S., Raspall, M., Da Silva Duarte, A.J., Tuerlinckx, D., Virelizier, J.L., Fischer, A., Enright, A., Bernhöft, J., Cleary, A.M., Vermylen, C., Rodriguez-Gallego, C., Davies, G., Blütters-Sawatzki, R., Siegrist, C.A., Ehlayel, M.S., Novelli, V., Haas, W.H., Levy, J., Freihorst, J., Al-Hajjar, S., Nadal, D., De Moraes Vasconcelos, D., Jeppsson, O., Kutukculer, N., Frecerova, K., Caragol, I., Lammas, D., Kumararatne, D.S., Abel, L., and Casanova, J.L. 2003. Low penetrance, broad resistance, and favorable outcome of interleukin 12 receptor beta1 deficiency: medical and immunological implications. *J Exp Med* 197:527-535.

210. Guia, S., Cognet, C., de Beaucoudrey, L., Tessmer, M.S., Jouanguy, E., Berger, C., Filipe-Santos, O., Feinberg, J., Camcioglu, Y., Levy, J., Al Jumaah, S., Al-Hajjar, S., Stephan, J.L., Fieschi, C., Abel, L., Brossay, L., Casanova, J.L., and Vivier, E. 2008. A role for interleukin-12/23 in the maturation of human natural killer and CD56+ T cells in vivo. *Blood* 111:5008-5016.

211. Uhrberg, M. 2005. The CD107 mobilization assay: viable isolation and immunotherapeutic potential of tumor-cytolytic NK cells. *Leukemia* 19:707-709.

212. Stewart, C., Walzer, T., Robbins, S., Malissen, B., Vivier, E., and Prinz, I. 2007. Germ-line and rearranged Tcrd transcription distinguish bona fide NK cells and NK-like gammadelta T cells. *Eur J Immunol* 37:1442-1452.

213. Rosenblum, J.M., Shimoda, N., Schenk, A.D., Zhang, H., Kish, D.D., Keslar, K., Farber, J.M., and Fairchild, R.L. 2010. CXC Chemokine Ligand (CXCL) 9 and CXCL10 Are antagonistic costimulation molecules during the priming of alloreactive T cell effectors. *J Immunol* 184:3450-3460.

214. Lu, M., Varley, A.W., Ohta, S., Hardwick, J., and Munford, R.S. 2008. Host inactivation of bacterial lipopolysaccharide prevents prolonged tolerance following Gram-negative bacterial infection. *Cell Host Microbe* 4:293-302.

215. Lee, J.R., and White, T.W. 2009. Connexin-26 mutations in deafness and skin disease. *Expert Rev Mol Med* 11:e35.

216. Walzer, T., Chiossone, L., Chaix, J., Calver, A., Carozzo, C., Garrigue-Antar, L., Jacques, Y., Baratin, M., Tomasello, E., and Vivier, E. 2007. Natural killer cell trafficking in vivo requires a dedicated sphingosine 1-phosphate receptor. *Nat Immunol* 8:1337-1344.

217. MacLennan, C., Fieschi, C., Lammas, David A., Picard, C., Dorman, Susan E., Sanal, O., MacLennan, Jenny M., Holland, Steven M., Ottenhoff, Tom H.M., Casanova, J.L., and Kumararatne, D.S. 2004. Interleukin-12 and IL23 are key cytokines for immunity against Salmonella in humans. *J Infect Dis* 190:1755-1757.

218. Bustamante, J., Boisson-Dupuis, S., Jouanguy, E., Picard, C., Puel, A., Abel, L., and Casanova, J.L. 2008. Novel primary immunodeficiencies revealed by the investigation of paediatric infectious diseases. *Curr Opin Immunol* 20:39-48.

219. Blanquart, C., Barbier, O., Fruchart, J.C., Staels, B., and Glineur, C. 2003. Peroxisome proliferator-activated receptors: regulation of transcriptional activities and roles in inflammation. *J Steroid Biochem Mol Biol* 85:267-273.

220. Peyrin-Biroulet, L., Beisner, J., Wang, G., Nuding, S., Oommen, S.T., Kelly, D., Parmentier-Decrucq, E., Dessein, R., Merour, E., Chavatte, P., Grandjean, T., Bressenot, A., Desreumaux, P., Colombel, J.-F., Desvergne, B., Stange, E.F., Wehkamp, J., and Chamaillard, M. 2010. Peroxisome proliferator-activated receptor gamma activation is required for maintenance of innate antimicrobial immunity in the colon. *Proc Natl Acad Sci USA* 107:8772-8777.

221. Prado-Montes de Oca, E. 2010. Human beta-defensin 1: A restless warrior against allergies, infections and cancer. *Int J Biochem Cell Biol* 42:800-804.

222. Pennini, M.E., Liu, Y., Yang, J., Croniger, C.M., Boom, W.H., and Harding, C.V. 2007. CCAAT/enhancer-binding protein beta and delta binding to CIITA promoters is

associated with the inhibition of CIITA expression in response to Mycobacterium tuberculosis 19-kDa lipoprotein. *J Immunol* 179:6910-6918.

223. Duits, L.A., Ravensbergen, B., Rademaker, M., Hiemstra, P.S., and Nibbering, P.H. 2002. Expression of beta-defensin 1 and 2 mRNA by human monocytes, macrophages and dendritic cells. *Immunology* 106:517-525.

224. Wah, J., Wellek, A., Frankenberger, M., Unterberger, P., Welsch, U., and Bals, R. 2006. Antimicrobial peptides are present in immune and host defense cells of the human respiratory and gastroinstestinal tracts. *Cell Tissue Res* 324:449-456.

225. Puccetti, P., and Grohmann, U. 2007. IDO and regulatory T cells: a role for reverse signalling and non-canonical NF-kB activation. *Nat Rev Immunol* 7:817-823.

226. Jenkins, K.A., and Mansell, A. 2010. TIR-containing adaptors in Toll-like receptor signalling. *Cytokine* 49:237-244.

227. Ye, L., Wang, X., Wang, S., Wang, Y., Song, L., Hou, W., Zhou, L., Li, H., and Ho, W. 2009. CD56+ T cells inhibit hepatitis C virus replication in human hepatocytes. *Hepatology* 49:753-762.

228. Aly, S., Laskay, T., Mages, J., Malzan, A., Lang, R., and Ehlers, S. 2007. Interferon-gamma-dependent mechanisms of mycobacteria-induced pulmonary immunopathology: the role of angiostasis and CXCR3-targeted chemokines for granuloma necrosis. *J Pathol* 212:295-305.

229. Lim, J., Derrick, S.C., Kolibab, K., Yang, A.L., Porcelli, S., Jacobs, W.R., and Morris, S.L. 2009. Early pulmonary cytokine and chemokine responses in mice immunized with three different vaccines against Mycobacterium tuberculosis determined by PCR array. *Clin Vaccine Immunol* 16:122-126.

230. Lacotte, S., Brun, S., Muller, S., and Dumortier, H. 2009. CXCR3, inflammation, and autoimmune diseases. *Ann NY Acad Sci* 1173:310-317.

231. Caccamo, N., Guggino, G., Joosten, S.A., Gelsomino, G., Carlo, P.D., Titone, L., Galati, D., Bocchino, M., Matarese, A., Salerno, A., Sanduzzi, A., Franken, W.P.J., Ottenhoff, T.H.M., and Dieli, F. 2010. Multifunctional CD4+T cells correlate with active Mycobacterium tuberculosis infection. *Eur J Immunol* Jun 10 [Epub ahead of print].

232. Tang, N.L.-S., Fan, H.P.Y., Chang, K.C., Ching, J.K.L., Kong, K.P.S., Yew, W.W., Kam, K.M., Leung, C.C., Tam, C.M., Blackwell, J., and Chan, C.Y. 2009. Genetic association between a chemokine gene CXCL-10 (IP-10, interferon gamma inducible protein 10) and susceptibility to tuberculosis. *Clinica Chimica Acta* 406:98-102.

233. Xu-Feng, Q., Dong-Heui, K., Yang-Suk, Y., Dan, J., Xue-Zhu, H., Jian-Hong, L., Young-Kun, D., and Kyu-Jae, L. 2009. Essential involvement of cross-talk between IFN-gamma and TNF-alpha in CXCL10 production in human THP-1 monocytes. *J Cell Physiol* 220:690-697.

234. Ruhwald, M., Bodmer, T., Maier, C., Jepsen, M., Haaland, M.B., Eugen-Olsen, J., Ravn, P., and on behalf of, T. 2008. Evaluating the potential of IP-10 and MCP-2 as biomarkers for the diagnosis of tuberculosis. *Eur Respir J* 32:1607-1615.

235. Ruhwald, M., Petersen, J., Kofoed, K., Nakaoka, H., Cuevas, L.E., Lawson, L., Squire, S.B., Eugen-Olsen, J., and Ravn, P. 2008. Improving T cell assays for the diagnosis of latent TB infection: potential of a diagnostic test based on IP-10. *PLoS ONE* 3:e2858.

236. De Paepe, B., Creus, K.K., and De Bleecker, J.L. 2009. Role of cytokines and chemokines in idiopathic inflammatory myopathies. *Curr Opin Rheumatol* 21:610-616.

237. Ip, P.P., and Liao, F. 2010. Resistance to dengue virus infection in mice is potentiated by CXCL10 and is independent of CXCL10-mediated leukocyte recruitment. *J Immunol* 184:5705-5714.

238. Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., and Medzhitov, R. 2004. Recognition of commensal microflora by Toll-like receptors is required for intestinal homeostasis. *Cell* 118:229-241.

239. Lavelle, E.C., Murphy, C., O'Neill, L.A.J., and Creagh, E.M. 2010. The role of TLRs, NLRs, and RLRs in mucosal innate immunity and homeostasis. *Mucosal Immunol* 3:17-28.

240. Shanahan, M., Vidrich, A., Shirafuji, Y., Dubois, C., Henschen-Edman, A., Hagen, S., Cohn, S., and Ouellette, A. 2010. Elevated expression of Paneth cell CRS4C in ileitis-prone SAM1/YitFc mice: regional distribution, subcellular localization, and mechanism of action. *J Biol Chem* 5:7493-7504.

241. Bowdish, D.M.E., Davidson, D.J., Speert, D.P., and Hancock, R.E.W. 2004. The human cationic peptide LL-37 induces activation of the extracellular signal-regulated kinase and p38 kinase pathways in primary human monocytes. *J Immunol* 172:3758-3765.

242. Yu, J., Mookherjee, N., Wee, K., Bowdish, D.M.E., Pistolic, J., Li, Y., Rehaume, L., and Hancock, R.E.W. 2007. Host defence peptide LL-37, in synergy with inflammatory mediator IL-1beta, augments immune responses by multiple pathways. *J Immunol* 179:7684-7691.

243. Brinkmann, V., Reichard, U., Goosmann, C., Fauler, B., Uhlemann, Y., Weiss, D.S., Weinrauch, Y., and Zychlinsky, A. 2004. Neutrophil extracellular traps kill bacteria. *Science* 303:1532-1535.

244. Brinkmann, V., and Zychlinsky, A. 2007. Beneficial suicide: why neutrophils die to make NETs. *Nat Rev Micro* 5:577-582.

245. Papayannopoulos, V., and Zychlinsky, A. 2009. NETs: a new strategy for using old weapons. *Trends Immunol* 30:513-521.

246. Zeng, H., Carlson, A.Q., Guo, Y., Yu, Y., Collier-Hyams, L.S., Madara, J.L., Gewirtz, A.T., and Neish, A.S. 2003. Flagellin is the major proinflammatory determinant of enteropathogenic Salmonella. *J Immunol* 171:3668-3674.

247. Raffatellu, M., Wilson, R., Winter, S., and Bäumler, A. 2008. Clinical pathogenesis of typhoid fever. *J Infect Dev Ctries* 2:260-266.

248. Winter, S.E., Raffatellu, M., Wilson, R.P., Rüssmann, H., and Bäumler, A.J. 2008. The Salmonella enterica serotype Typhi regulator TviA reduces interleukin-8 production in intestinal epithelial cells by repressing flagellin secretion. *Cell Microbiol* 10:247-261.

249. Modi, W.S., and Yoshimura, T. 1999. Isolation of novel GRO genes and a phylogenetic analysis of the CXC chemokine subfamily in mammals. *Mol Biol Evol* 16:180-193.

250. Cole, A.M., Ganz, T., Liese, A.M., Burdick, M.D., Liu, L., and Strieter, R.M. 2001. Cutting edge: IFN-inducible ELR- CXC chemokines display defensin-like antimicrobial activity. *J Immunol* 167:623-627.

251. Hyun, J.G., Lee, G., Brown, J.B., Grimm, G.R., Tang, Y., Mittal, N., Dirisina, R., Zhang, Z., Fryer, J.P., Weinstock, J.V., Luster, A.D., and Barrett, T.A. 2005. Anti-interferon-inducible chemokine, CXCL10, reduces colitis by impairing T helper-1 induction and recruitment in mice. *Inflamm Bowel Dis* 11:799-805.

252. Singh, U., Venkataraman, C., Singh, R., and Lillard, J.J. 2007. CXCR3 axis: role in inflammatory bowel disease and its therapeutic implication. *Endocr Metab Immune Disord Drug Targets* 7:111-123.

APPENDICES

A. SUPPLEMENTARY INFORMATION FOR CHAPTER 2

A.1. Virtual digest and optical map alignment

Feb. 22, 2006

Comparison of the Chromosome 8:18,508,450-24,203,501 bp optical map, generated using the restriction enzyme SwaI, to the SwaI-virtual digest of the corresponding sequence downloaded from the Vega Genome Browser

-Optical Map generated by Steve Goldstein (obtained through Darren Grafham, WTSI)
-contiguous DNA restriction digested using SwaI

-Vitual Map generated using RestrictionMapper version 3 (www.restrictionmapper.org) on sequence exported from Vega v23 (Feb.2007) for same bp region
-the N's in the 5 gaps in this region were deleted and replaced by a space because the program can not handle non-base letters, but numbers and spaces are okay
-digested with SwaI and sites mapped

Virtual Digest

SwaI Cut Sites (bp)	8870	23495	57111	61037	72396	87776	121550	150747	170583	177143	200883	215217	293418	323786	333268
Fragment Sizes (bp)	8870	14625	33616	3926	11359	15380	33774	29197	19836	6560	23740	14334	78201	30368	9482

Optical Map

u:4e17acd1-59bc-4dd0-8e53-ce0681057d56:contig749.trimmed

Fragment Sizes (kb)	9.433	14.632	33.562	3.802	11.238	15.667	32.304	28.867	20.085	6.65	23.63	14.567	78.794	30.097	9.65
(bp)	9433	14632	33562	3802	11238	15667	32304	28867	20085	6650	23630	14567	78794	30097	9650
Position (bp)	9433	24065	57627	61429	72667	88334	120638	149505	169590	176240	199870	214437	293231	323328	332978

Alignment of Virtual Digest and Optical Map

Position (bp)	8870	23495	57111	61037	72396	87776	121550	150747	170583	177143	200883	215217	293418	323786	333268
Virtual Digest	8870	14625	33616	3926	11359	15380	33774	29197	19836	6560	23740	14334	78201	30368	9482
Optical Map	9433	14632	33562	3802	11238	15667	32304	28867	20085	6650	23630	14567	78794	30097	9650

fragments approximately the same size (same restriction site)
gap in sequence alignment
inverted contig AC152164.14
inserted fragments (small so most likely minor changes in restriction sites)
different restriction sites (same bp size overall)

357160 357687 358035 362231 375530 381168 445828 454073 473840 501501 518490 528311 610233 610268 610951 618624 661489

23892 527 348 4196 13299 5638 64660 8245 19767 27661 16989 9821 81922 35 683 7673 42865

23.923 4.676 12.855 5.658 64.303 7.778 19.165 27.14 16.775 9.4 78.456 7.618 42.281 17.36 2.603 23.532 14.88
23923 4676 12855 5658 64303 7778 19165 27140 16775 9400 78456 7618 42281 17360 2603 23532 14880

356901 361577 374432 380090 444393 452171 471336 498476 515251 524651 603107 610725 653006 670366 672969 696501 711381

357687, 358035 610268, 610951
357160 362231 375530 381168 445828 454073 473840 501501 518490 528311 610233 618624 661489 678447 681094 705087 720386
527, 348 35, 683
23892 4196 13299 5638 64660 8245 19767 27661 16989 9821 81922 7673 42865 60 2647 23993 15299
23923 4676 12855 5658 64303 7778 19165 27140 16775 9400 78456 7618 42281 17360 2603 23532 14880

678447 681094 705087 720386 762662 810777 824512 896337 922028 932331 933020 944530 944746 946284 947445 960006 986435

16958 2647 23993 15299 42276 48115 13735 71825 25691 10303 689 11510 216 1538 1161 12561 26429

41.527 48.798 13.881 70.446 26.118 10.472 12.209 1.798 13.512 27.105 1.934 23.513 5.04 35.052 60.77 50.477 17.076
41527 48798 13881 70446 26118 10472 12209 1798 13512 27105 1934 23513 5040 35052 60770 50477 17076

752908 801706 815587 886033 912151 922623 934832 936630 950142 977247 979181 1002694 1007734 1042786 1103556 1154033 1171109

1111792
762662 810777 824512 896337 922028 932331 988214 1011426 1016470 1050912 1111552 1161231 1177631
240
42276 48115 13735 71825 25691 10303 1779 23212 5044 34442 60640 49439 16400
41527 48798 13881 70446 26118 10472 12209 1798 13512 27105 1934 23513 5040 35052 60770 50477 17076

933020 944530 944746 946284 947445 960006 986435

689 11510 216 1538 1161 12561 26429

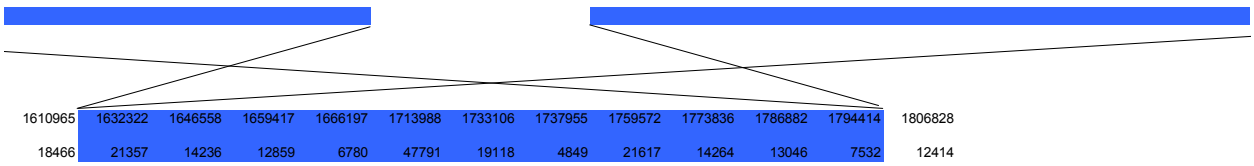
988214	1011426	1016470	1050912	1111552	1111792	1161231	1177631	1198668	1203855	1218992	1222969	1266933	1279793	1335786	1357296	1367943
1779	23212	5044	34442	60640	240	49439	16400	21037	5187	15137	3977	43964	12860	55993	21510	10647
20.947	5.313	15.594	3.893	43.413	12.438	18.217	20.691	5.349	19.589	82.532	7.768	12.533	13.511	21.745	5.195	18.267
20947	5313	15594	3893	43413	12438	18217	20691	5349	19589	82532	7768	12533	13511	21745	5195	18267
1192056	1197369	1212963	1216856	1260269	1272707	1290924	1311615	1316964	1336553	1419085	1426853	1439386	1452897	1474642	1479837	1498104

1198668	1203855	1218992	1222969	1266933	1279793											
21037	5187	15137	3977	43964	12860											
20947	5313	15594	3893	43413	12438	18217	20691	5349	19589	82532	7768	12533	13511	21745	5195	18267

							1335786	1357296	1367943	1383105	1391546	1401346	1415378	1432270	1571900	1592499
							55993	21510	10647	15162	8441	9800	14032	16892	139630	20599

1383105	1391546	1401346	1415378	1432270	1571900	1592499	1610965	1632322	1646558	1659417	1666197	1713988	1733106	1737955	1759572	1773836
15162	8441	9800	14032	16892	139630	20599	18466	21357	14236	12859	6780	47791	19118	4849	21617	14264
51.94	7.505	12.943	13.588	21.128	4.981	19.498	66.108	7.269	12.789	13.972	20.757	5.004	18.622	0.4	47.437	7.305
51940	7505	12943	13588	21128	4981	19498	66108	7269	12789	13972	20757	5004	18622	400	47437	7305
1550044	1557549	1570492	1584080	1605208	1610189	1629687	1695795	1703064	1715853	1729825	1750582	1755586	1774208	1774608	1822045	1829350

51940	7505	12943	13588	21128	4981	19498	66108	7269	12789	13972	20757	5004	18622	400	47437	7305



1610965	1632322	1646558	1659417	1666197	1713988	1733106	1737955	1759572	1773836	1786882	1794414	1806828	
18466	21357	14236	12859	6780	47791	19118	4849	21617	14264	13046	7532	12414	

1786882	1794414	1806828	1813225	1828373	1842396	1856341	1856383	1857948	1875694	1894553	1895852	1895894	1919958	1931894	1948475	1981398
13046	7532	12414	6397	15148	14023	13945	42	1565	17746	18859	1299	42	24064	11936	16581	32923
11.849	17.529	21.436	5.365	18.924	56.75	7.935	13.038	17.057	21.505	5.324	14.969	104.187	10.156	2.474	15.558	31.285
11849	17529	21436	5365	18924	56750	7935	13038	17057	21505	5324	14969	104187	10156	2474	15558	31285
1841199	1858728	1880164	1885529	1904453	1961203	1969138	1982176	1999233	2020738	2026062	2041031	2145218	2155374	2157848	2173406	2204691
11849	17529	21436	5365	18924	56750	7935	13038	17057	21505	5324	14969	104187	10156	2474	15558	31285

759879
bp size of gap - 12414 fragment at end of tandem repeat

759879

1993071	1993113	1994681	2012227	2031089	2032388	2032430	2063315	2068079	2078321	2126917	2128218	2128260	2153457	2171945	2190498	2203783
11673	42	1568	17546	18862	1299	42	30885	4764	10242	48596	1301	42	25197	18488	18553	13285
35.426	22.898	7.856	10.126	2.388	14.973	21.427	24.06	59.629	7.676	13.405	20.851	21.645	5.119	19.766	21.252	117.733
35426	22898	7856	10126	2388	14973	21427	24060	59629	7676	13405	20851	21645	5119	19766	21252	117733
2240117	2263015	2270871	2280997	2283385	2298358	2319785	2343845	2403474	2411150	2424555	2445406	2467051	2472170	2491936	2513188	2630921
35426	22898	7856	10126	2388	14973	21427	24060	59629	7676	13405	20851	21645	5119	19766	21252	117733

2213156	2223401	2249239	2260890	2260932	2267415	2280702	2280744	2304378	2315699	2317000	2317042	2328458	2341746	2351103	2367474	2416071
9373	10245	25838	11651	42	6483	13287	42	23634	11321	1301	42	11416	13288	9357	16371	48597
3.114	18.422	6.689	15.913	14.435	14.119	1.385	18.071	19.514	0.95	24.19	12.011	16.753	34.275	11.901	1.55	18.36
3114	18422	6689	15913	14435	14119	1385	18071	19514	950	24190	12011	16753	34275	11901	1550	18360
2634035	2652457	2659146	2675059	2689494	2703613	2704998	2723069	2742583	2743533	2767723	2779734	2796487	2830762	2842663	2844213	2862573
						1856383				1895894					1993113	
		1813225	1828373	1842396	1856341	1857948	1875694	1894553	1895852	1919958	1931894	1948475	1981398	1993071	1994681	2012227
						42				42					42	
		6397	15148	14023	13945	1565	17746	18859	1299	24064	11936	16581	32923	11673	1568	17546
3114	18422	6689	15913	14435	14119	1385	18071	19514	950	24190	12011	16753	34275	11901	1550	18360

2417372	2417414	2453714	2460230	2507993	2544709	2555049	2556917	2563307	2564174	2602317	2629275	2669944	2683142	2692356	2736403	2769347
1301	42	36300	6516	47763	36716	10340	1868	6390	867	38143	26958	40669	13198	9214	44047	32944
19.223	1.12	31.569	5.332	10.846	49.875	1.146	25.2	18.809	26.195	12.238	6.784	6.851	36.321	1.181	11.73	13.544
19223	1120	31569	5332	10846	49875	1146	25200	18809	26195	12238	6784	6851	36321	1181	11730	13544
2881796	2882916	2914485	2919817	2930663	2980538	2981684	3006884	3025693	3051888	3064126	3070910	3077761	3114082	3115263	3126993	3140537
		2032430					2128260									
2031089	2032388	2063315	2068079	2078321	2126917	2128218	2153457	2171945								2203783
		42					42									
18862	1299	30885	4764	10242	48596	1301	25197	18488								13285
19223	1120	31569	5332	10846	49875	1146	25200	18809	26195	12238	6784	6851	36321	1181	11730	13544

2190498

18553

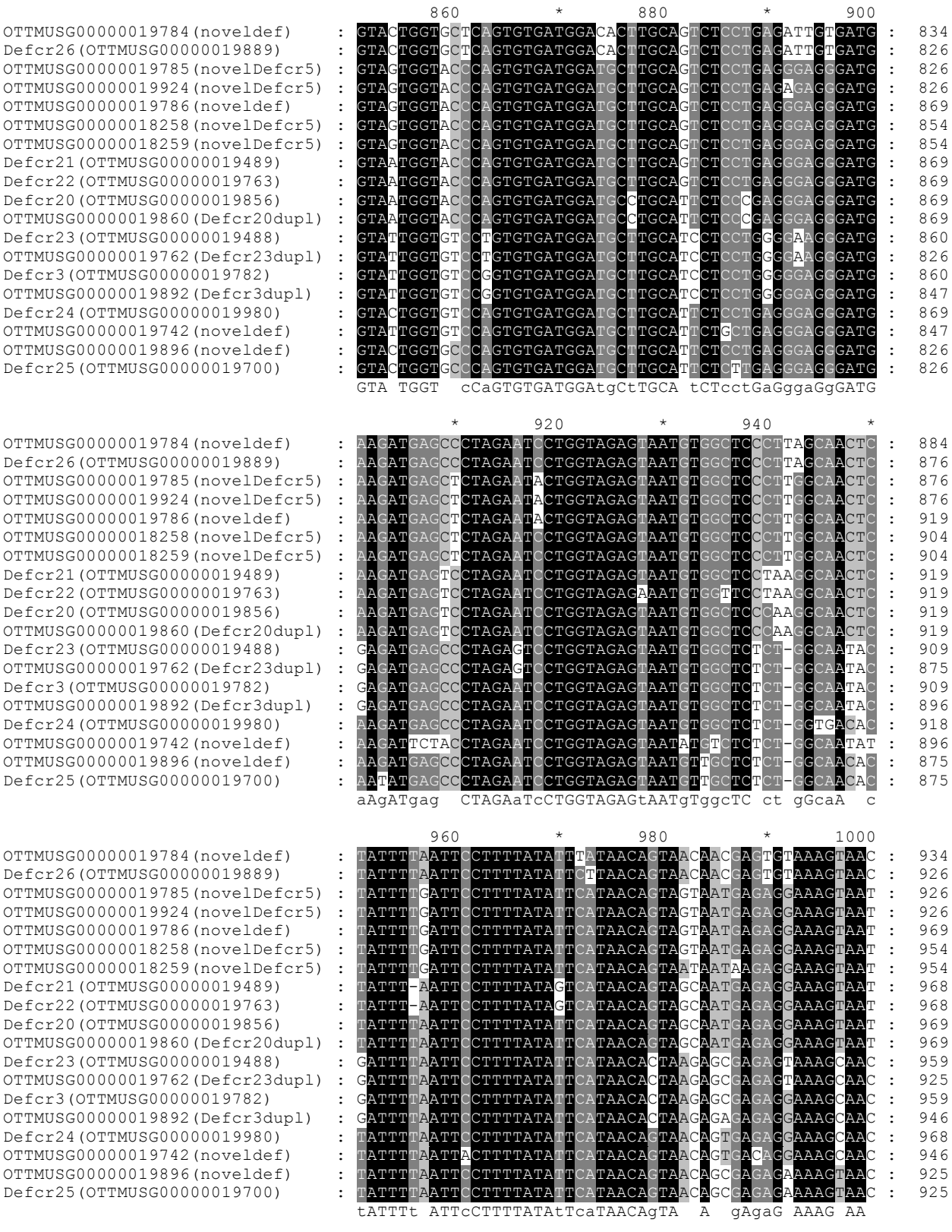
2831539	2900826	2903138	2954859	2955907	2962960	2967132	2975141	2985848	2992761	2994594	2995176	2997153	3018392	3027837	3080823	3091042
62192	69287	2312	51721	1048	7053	4172	8009	10707	6913	1833	582	1977	21239	9445	52986	10219
10.01	10.763	26.973	11.964	6.706	13.992	24.953	12.089	11.57	13.733	9.374	17.273	51.579	1.154	39.076	7.217	50.336
10010	10763	26973	11964	6706	13992	24953	12089	11570	13733	9374	17273	51579	1154	39076	7217	50336
3150547	3161310	3188283	3200247	3206953	3220945	3245898	3257987	3269557	3283290	3292664	3309937	3361516	3362670	3401746	3408963	3459299
				2260932		2280744		2317000, 2317042						2417414		
2213156	2223401	2249239	2260890	2267415	2280702	2304378	2315699	2328458	2341746	2351103	2367474	2416071	2417372	2453714	2460230	2507993
				42		42		1301, 42						42		
9373	10245	25838	11651	6483	13287	23634	11321	11416	13288	9357	16371	48597	1301	36300	6516	47763
10010	10763	26973	11964	6706	13992	24953	12089	11570	13733	9374	17273	51579	1154	39076	7217	50336

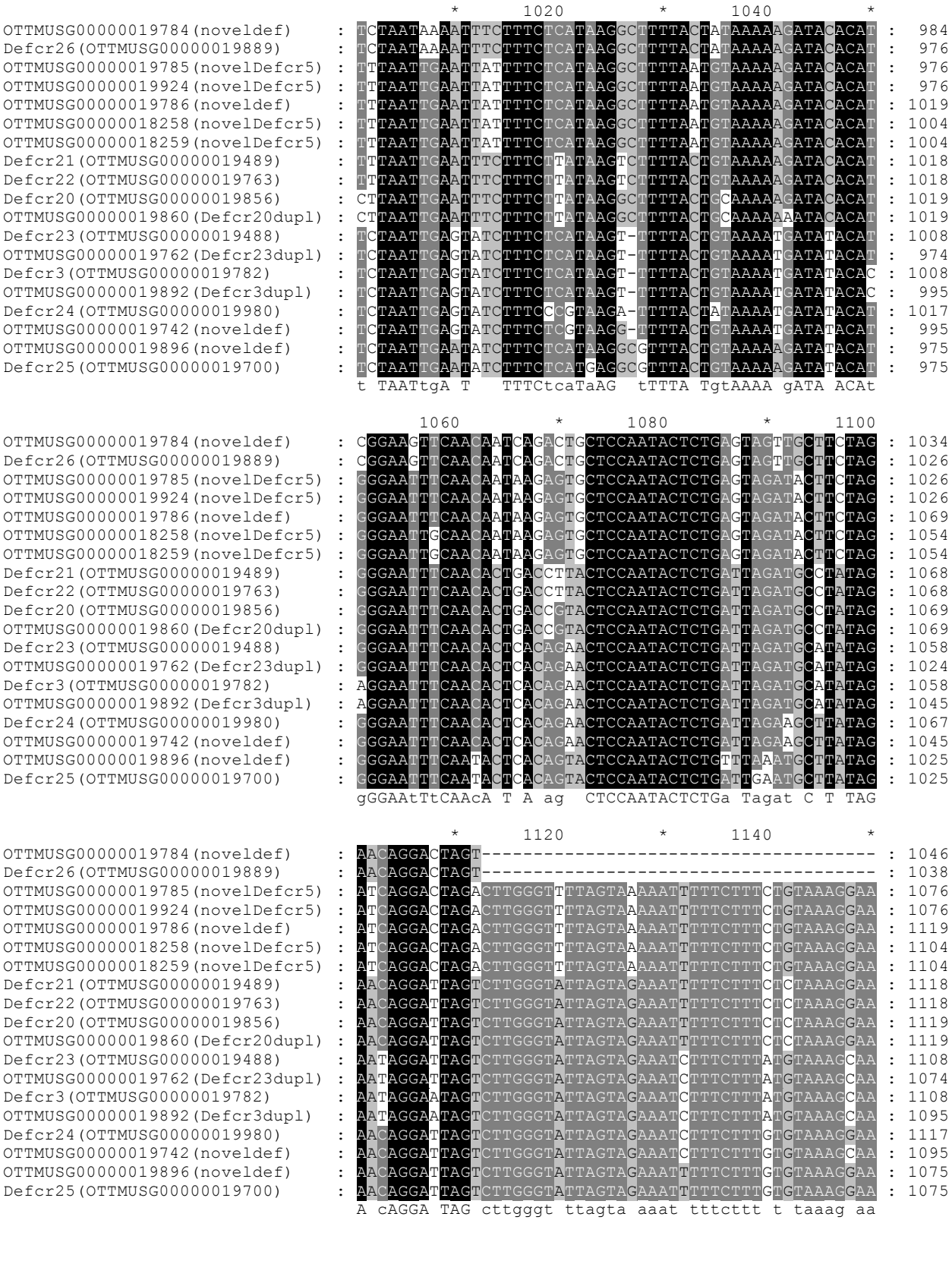
3096574	3097463	3169378	3189190	3202710	3210536	3213517	3226298	3236049	3237578	3268875	3331083	3337073	3345873	3424721	3466223	
5532	889	71915	19812	13520	7826	2981	12781	9751	1529	31297	62208	5990	8800	78848	41502	
39.099	11.819	7.609	0.692	37.888	26.869	41.697	13.429	9.263	44.966	33.784	64.645	70.321	1.918	52.532	0.921	7.518
39099	11819	7609	692	37888	26869	41697	13429	9263	44966	33784	64645	70321	1918	52532	921	7518
3498398	3510217	3517826	3518518	3556406	3583275	3624972	3638401	3647664	3692630	3726414	3791059	3861380	3863298	3915830	3916751	3924269
		2556917														
2544709	2555049	2563307	2564174	2602317	2629275	2669944	2683142	2692356	2736403	2769347	2831539	2900826	2903138	2954859	2955907	2962960
		1868														
36716	10340	6390	867	38143	26958	40669	13198	9214	44047	32944	62192	69287	2312	51721	1048	7053
39099	11819	7609	692	37888	26869	41697	13429	9263	44966	33784	64645	70321	1918	52532	921	7518

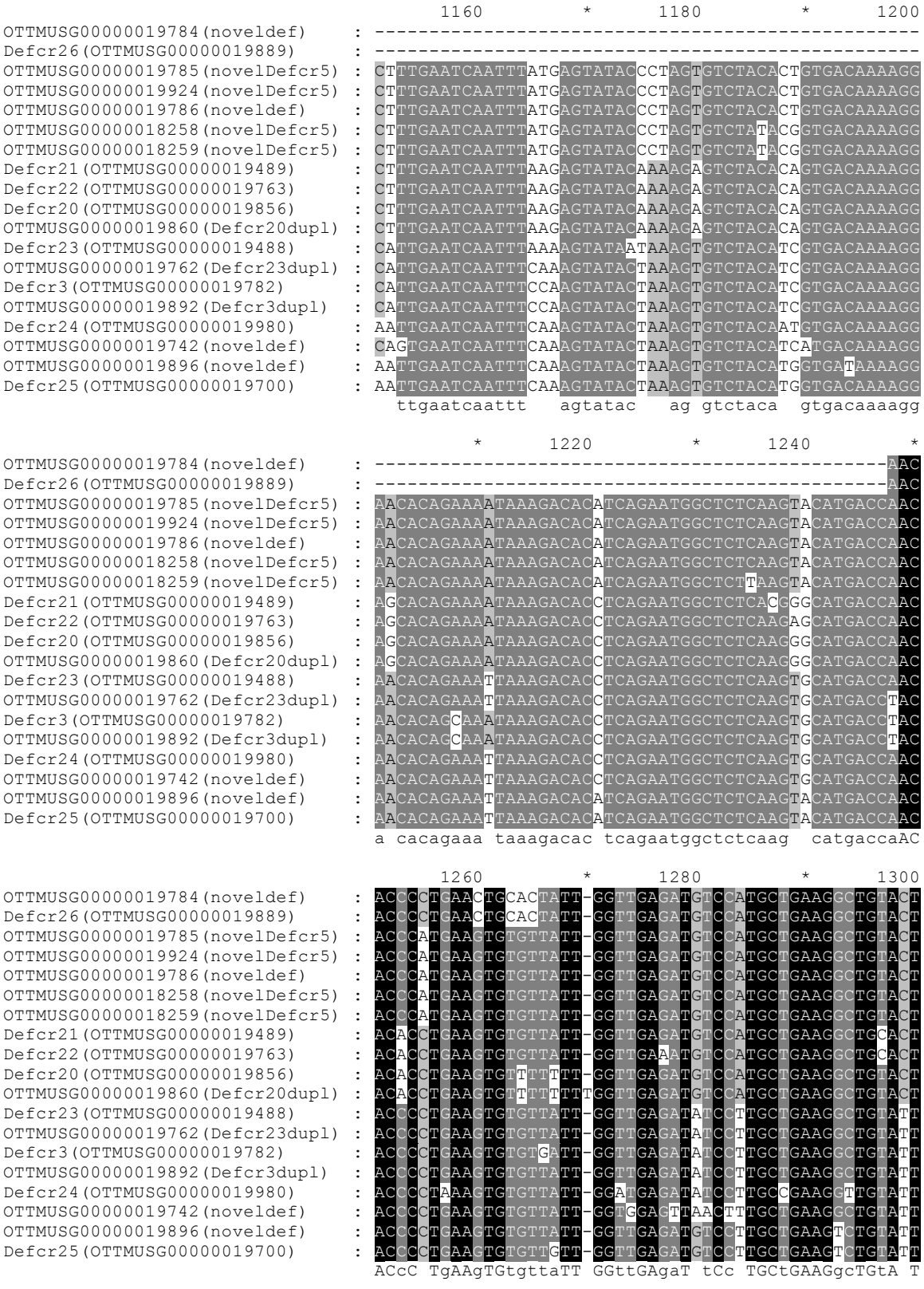
4.178	8.302	10.571	7.355	0.4	2.574	21.644	8.958	51.032	10.147	5.522	0.705	70.17	18.85	13.016	7.408	2.479
4178	8302	10571	7355	400	2574	21644	8958	51032	10147	5522	705	70170	18850	13016	7408	2479
3928447	3936749	3947320	3954675	3955075	3957649	3979293	3988251	4039283	4049430	4054952	4055657	4125827	4144677	4157693	4165101	4167580
				2994594												
2967132	2975141	2985848	2992761	2995176	2997153	3018392	3027837	3080823	3091042	3096574	3097463	3169378	3189190	3202710	3210536	3213517
				1833												
4172	8009	10707	6913	582	1977	21239	9445	52986	10219	5532	889	71915	19812	13520	7826	2981
4178	8302	10571	7355	400	2574	21644	8958	51032	10147	5522	705	70170	18850	13016	7408	2479

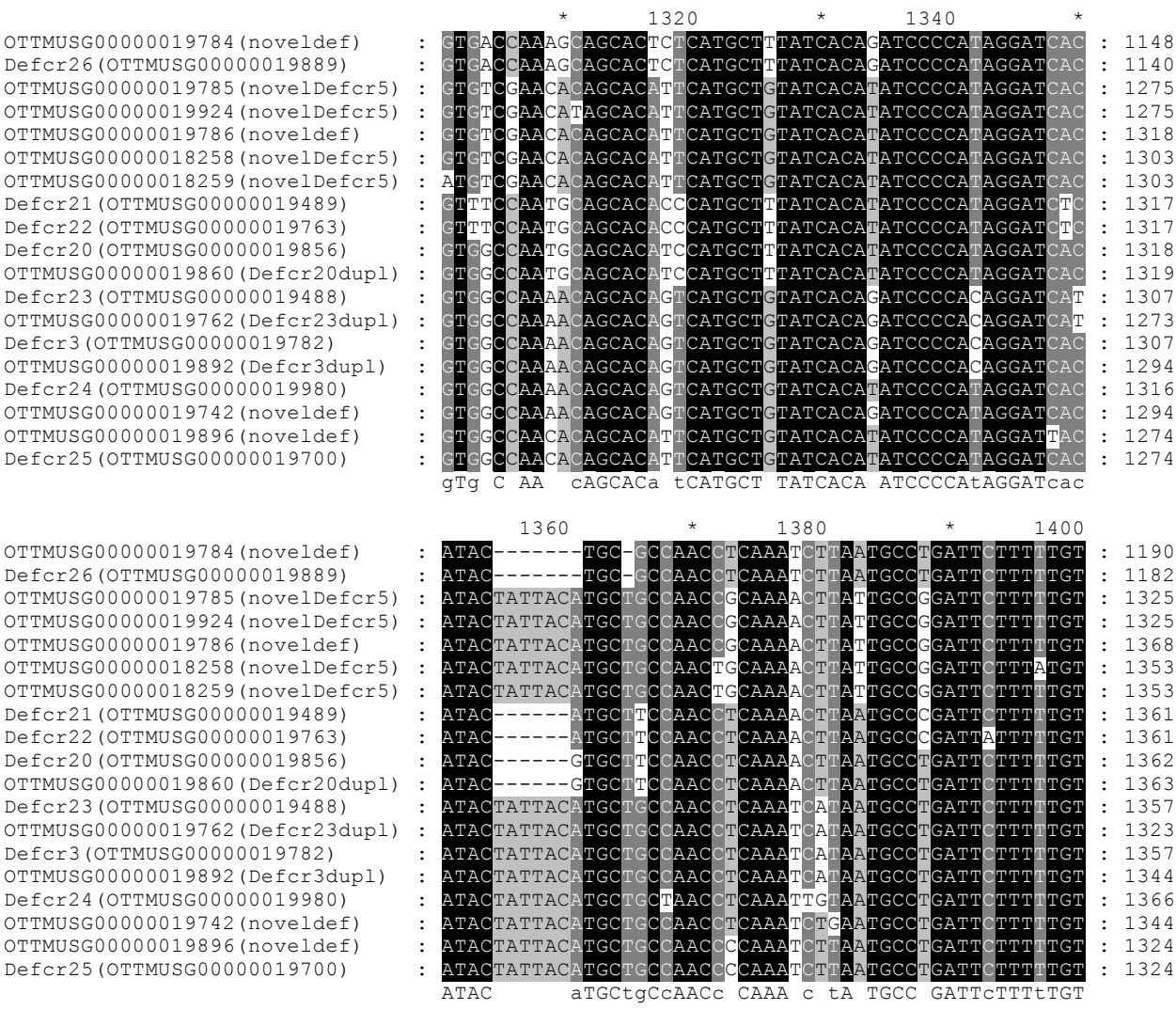
12.696	9.727	1.382	31.089	57.538	5.382	8.662	74.898	42.114	23.554
12696	9727	1382	31089	57538	5382	8662	74898	42114	23554
4180276	4190003	4191385	4222474	4280012	4285394	4294056	4368954	4411068	4434622
3226298	3236049	3237578	3268875	3331083	3337073	3345873	3424721	3466223	
12781	9751	1529	31297	62208	5990	8800	78848	41502	
12696	9727	1382	31089	57538	5382	8662	74898	42114	23554

A.2. Nucleotide alignment of α-defensin intron 1-2 genomic sequence

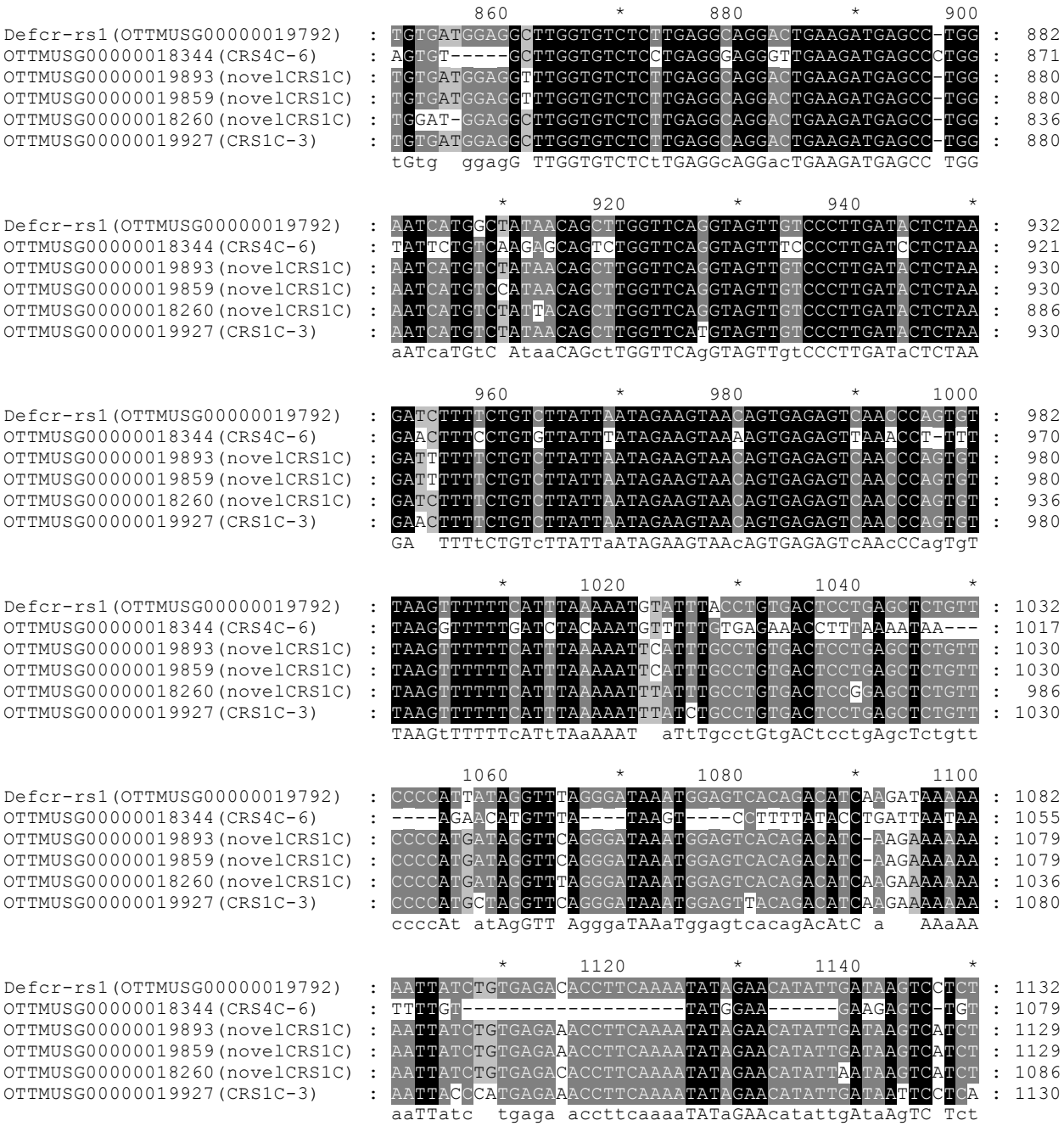



OTTMUSG00000019784(noveldef)
Defcr26(OTTMUSG00000019889)
OTTMUSG00000019785(novelDefcr5)
OTTMUSG00000019924(novelDefcr5)
OTTMUSG00000019786(noveldef)
OTTMUSG00000018258(novelDefcr5)
OTTMUSG00000018259(novelDefcr5)
Defcr21(OTTMUSG00000019489)
Defcr22(OTTMUSG00000019763)
Defcr20(OTTMUSG00000019856)
OTTMUSG00000019860(Defcr20dupl)
Defcr23(OTTMUSG00000019488)
OTTMUSG00000019762(Defcr23dupl)
Defcr3(OTTMUSG00000019782)
OTTMUSG00000019892(Defcr3dupl)
Defcr24(OTTMUSG00000019980)
OTTMUSG00000019742(noveldef)
OTTMUSG00000019896(noveldef)
Defcr25(OTTMUSG00000019700)
1020
1040
1060
1080
1100
1120
1140


1160 * 1180 * 1200
OTTMUSG00000019784(noveldef)
Defcr26(OTTMUSG00000019889)
OTTMUSG00000019785(novelDefcr5)
OTTMUSG00000019924(novelDefcr5)
OTTMUSG00000019786(noveldef)
OTTMUSG00000018258(novelDefcr5)
OTTMUSG00000018259(novelDefcr5)
Defcr21(OTTMUSG00000019489)
Defcr22(OTTMUSG00000019763)
Defcr20(OTTMUSG00000019856)
OTTMUSG00000019860(Defcr20dupl)
Defcr23(OTTMUSG00000019488)
OTTMUSG00000019762(Defcr23dupl)
Defcr3(OTTMUSG00000019782)
OTTMUSG00000019892(Defcr3dupl)
Defcr24(OTTMUSG00000019980)
OTTMUSG00000019742(noveldef)
OTTMUSG00000019896(noveldef)
Defcr25(OTTMUSG00000019700)
ttgaatcaattt agtatac ag gtctaca gtgacaaaagg
* 1220 * 1240 *
a cacagaaa taaagacac tcagaatggctctcaag catgaccaAC
1260 * 1280 * 1300
ACcC TgAAgTGtgttaTT GGttGAgaT tCc TGCtGAAGgcTGtA T

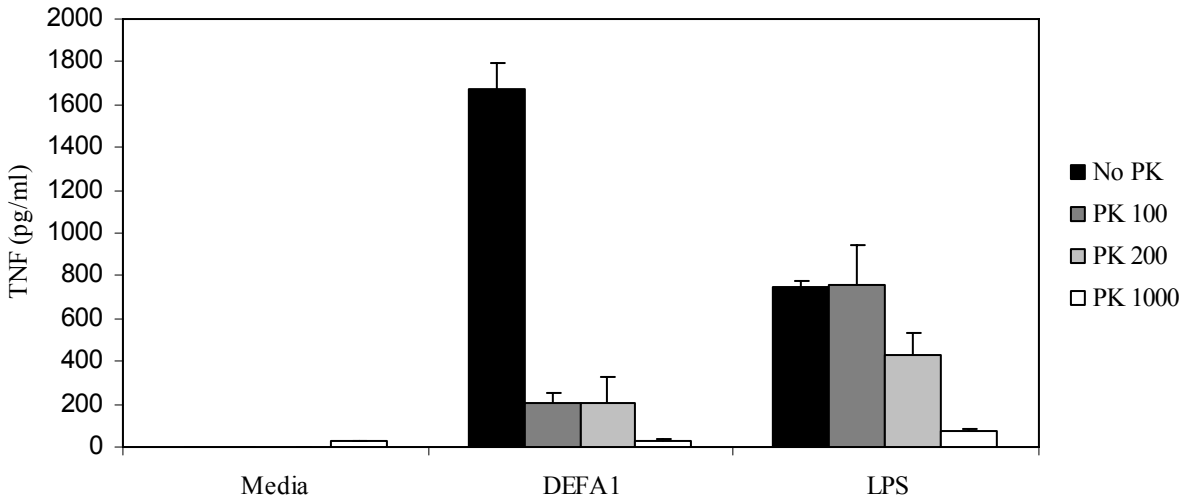

1320
1340
OTTMUSG00000019784(noveldef) : GTGACCAAAGCAGCACTCTCATGCTTTATCACAGATCCCCATAGGATCAC : 1148
Defcr26(OTTMUSG00000019889) : GTGACCAAAGCAGCACTCTCATGCTTTATCACAGATCCCCATAGGATCAC : 1140
OTTMUSG00000019785(novelDefcr5) : GTCTCGAACACAGCACATTCATGCTGTATCACATATCCCCATAGGATCAC : 1275
OTTMUSG00000019924(novelDefcr5) : GTCTCGAACATAGCACATTCATGCTGTATCACATATCCCCATAGGATCAC : 1275
OTTMUSG00000019786(noveldef) : GTGTCGAACACAGCACATTCATGCTGTATCACATATCCCCATAGGATCAC : 1318
OTTMUSG00000018258(novelDefcr5) : GTGTCGAACACAGCACATTCATGCTGTATCACATATCCCCATAGGATCAC : 1303
OTTMUSG00000018259(novelDefcr5) : ATGTCGAACACAGCACATTCATGCTGTATCACATATCCCCATAGGATCAC : 1303
Defcr21(OTTMUSG00000019489) : GTTTCCAATGCAGCACACCCATGCTTTATCACATATCCCCATAGGATCTC : 1317
Defcr22(OTTMUSG00000019763) : GTTTCCAATGCAGCACACCCATGCTTTATCACATATCCCCATAGGATCTC : 1317
Defcr20(OTTMUSG00000019856) : GTGGCCAATGCAGCACATCCATGCTTTATCACATATCCCCATAGGATCAC : 1318
OTTMUSG00000019860(Defcr20dupl) : GTGGCCAATGCAGCACATCCATGCTTTATCACATATCCCCATAGGATCAC : 1319
Defcr23(OTTMUSG00000019488) : GTCGCCAAAACAGCACAGTCATGCTGTATCACAGATCCCCACAGGATCAT : 1307
OTTMUSG00000019762(Defcr23dupl) : GTGGCCAAAACAGCACAGTCATGCTGTATCACAGATCCCCACAGGATCAT : 1273
Defcr3(OTTMUSG00000019782) : GTGGCCAAAACAGCACAGTCATGCTGTATCACAGATCCCCACAGGATCAC : 1307
OTTMUSG00000019892(Defcr3dupl) : GTGGCCAAAACAGCACAGTCATGCTGTATCACAGATCCCCACAGGATCAC : 1294
Defcr24(OTTMUSG00000019980) : GTGGCCAAAACAGCACAGTCATGCTGTATCACATATCCCCATAGGATCAC : 1316
OTTMUSG00000019742(noveldef) : GTGGCCAAAACAGCACAGTCATGCTGTATCACAGATCCCCATAGGATCAC : 1294
OTTMUSG00000019896(noveldef) : GTGGCCAACACAGCACATTCATGCTGTATCACATATCCCCATAGGATTAC : 1274
Defcr25(OTTMUSG00000019700) : GTGGCCAACACAGCACATTCATGCTGTATCACATATCCCCATAGGATCAC : 1274
gTg C AA cAGCACa tCATGCT TATCACA ATCCCCAtAGGATcac
1360
1380
1400
OTTMUSG00000019784(noveldef) : ATAC-------TGC-GCCAACCTCAAATCTTAATGCCTGATTCTTTTTGT : 1190
Defcr26(OTTMUSG00000019889) : ATAC-------TGC-GCCAACCTCAAATCTTAATGCCTGATTCTTTTTGT : 1182
OTTMUSG00000019785(novelDefcr5) : ATACTATTACATGCTGCCAACCGCAAAACTTATTGCCGGATTCTTTTTGT : 1325
OTTMUSG00000019924(novelDefcr5) : ATACTATTACATGCTGCCAACCGCAAAACTTATTGCCGGATTCTTTTTGT : 1325
OTTMUSG00000019786(noveldef) : ATACTATTACATGCTGCCAACCGCAAAACTTATTGCCGGATTCTTTTTGT : 1368
OTTMUSG00000018258(novelDefcr5) : ATACTATTACATGCTGCCAACTGCAAAACTTATTGCCGGATTCTTTATGT : 1353
OTTMUSG00000018259(novelDefcr5) : ATACTATTACATGCTGCCAACTGCAAAACTTATTGCCGGATTCTTTTTGT : 1353
Defcr21(OTTMUSG00000019489) : ATAC------ATGCTTCCAACCTCAAAACTTAATGCCCGATTCTTTTTGT : 1361
Defcr22(OTTMUSG00000019763) : ATAC------ATGCTTCCAACCTCAAAACTTAATGCCCGATTATTTTTGT : 1361
Defcr20(OTTMUSG00000019856) : ATAC------GTGCTTCCAACCTCAAAACTTAATGCCTGATTCTTTTTGT : 1362
OTTMUSG00000019860(Defcr20dupl) : ATAC------GTGCTTCCAACCTCAAAACTTAATGCCTGATTCTTTTTGT : 1363
Defcr23(OTTMUSG00000019488) : ATACTATTACATGCTGCCAACCTCAAATCATAATGCCTGATTCTTTTTGT : 1357
OTTMUSG00000019762(Defcr23dupl) : ATACTATTACATGCTGCCAACCTCAAATCATAATGCCTGATTCTTTTTGT : 1323
Defcr3(OTTMUSG00000019782) : ATACTATTACATGCTGCCAACCTCAAATCATAATGCCTGATTCTTTTTGT : 1357
OTTMUSG00000019892(Defcr3dupl) : ATACTATTACATGCTGCCAACCTCAAATCATAATGCCTGATTCTTTTTGT : 1344
Defcr24(OTTMUSG00000019980) : ATACTATTACATGCTGCTAACCTCAAATTGTAATGCCTGATTCTTTTTGT : 1366
OTTMUSG00000019742(noveldef) : ATACTATTACATGCTGCCAACCTCAAATCTGAATGCCTGATTCTTTTTGT : 1344
OTTMUSG00000019896(noveldef) : ATACTATTACATGCTGCCAACCCCAAATCTTAATGCCTGATTCTTTTTGT : 1324
Defcr25(OTTMUSG00000019700) : ATACTATTACATGCTGCCAACCCCAAATCTTAATGCCTGATTCTTTTTGT : 1324
ATAC aTGCtgCcAACc CAAA c tA TGCC GATTcTTTtTGT

A.3. Nucleotide alignment of CRS-defensin genes intron 1-2 genomic sequence



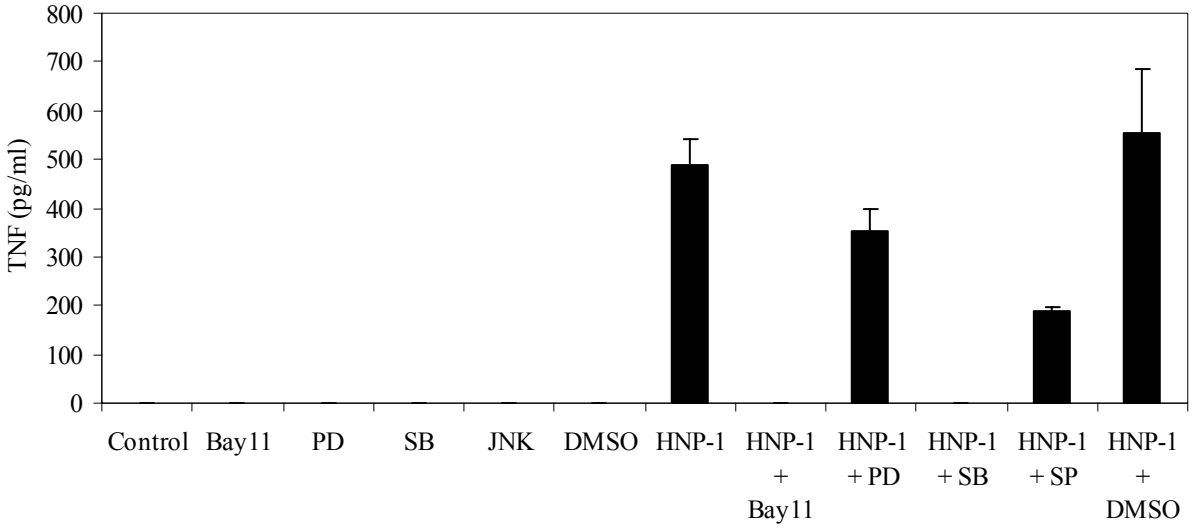
A.4. Proteinase K treatment of DEFA1 to determine functional specificity

Due to their disulphide bonds, defensins are very resistant to proteolysis, the following modified protocol was used. DEFA1 (10 μg/ml) or LPS (50 ng/ml) as the control were boiled with 5% β-ME for 5 minutes, then digested with 100, 200, 1000 μg/ml proteinase K at 50°C for 60 minutes. PBMCs were treated with DEFA1 +/- β-ME/PK for 24 hours and TNF released into the supernatant measured by ELISA. Results are from one of two experiments with separate donors. No PK, no proteinase K; PK 100, proteinase K 100 μg/ml; PK 200, proteinase K 200 μg/ml; PK 1000, proteinase K 1000 μg/ml.



A.5. DEFA1 release of TNF occurred through NF-κB and MAPK14 activation

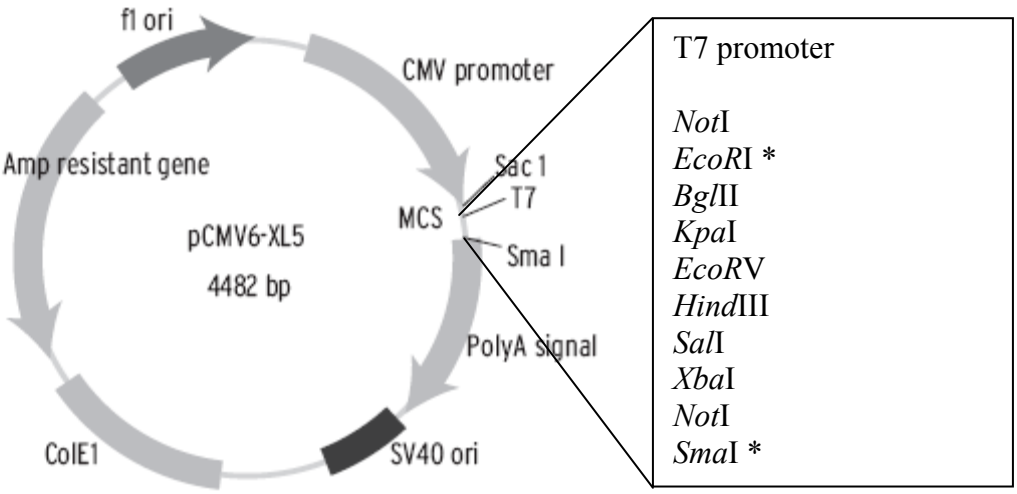
Pretreatment of human PBMCs with 10 μM of Iκ-Bα phosphorylation inhibitor Bay 11-7085 (Bay11) or MAPK14 (p38) inhibitor SB208530 (SB), each dissolved in dimethyl sulphoxide (DMSO), for 1 hour prior to stimulation with 10 ng/ml DEFA1 reduces the amount of TNF released into the supernatant after 24 hours, determined by ELISA. Pretreatment with 10 μM of MAP2K2 (MEK) inhibitor PD98059 (PD) or MAPK8 (JNK) inhibitor SP600125 (SP), each dissolved in DMSO, partially abrogates the release of TNF under the same conditions. Results are from one of two experiments with separate donors.



B. SUPPLEMENTARY INFORMATION FOR CHAPTER 3

B.1. IL8 cDNA vector

a. Origene vector map of pCMV6-XL5 (catalogue number TC119807) containing human *IL8* cDNA (permission obtained)



* human *IL8* (NM_000584.2) cloned between *EcoR*I and *Sal*I of MCS by Origene

b. Multiple cloning site (MCS) of pCMV6-XL5

Vector Primer v1.5 >

TTTGGCACCAAAATCAACG**GGACTTTCCAAAATGTCG**TAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACG

AAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCATTATTGGGGCGGGGCAACTGCGTTTACCCGCCATCCGCACATGC

T7 promoter >

Sac 1 (Not for sequencing) Not I EcoRI

GTGGGAGGTCTATATAAGCA**GAGCTC**GTTTAGTGAACCGTCAGAATTTT**GTAATACGACTCACTATAGGGCGGCCGCGAATT**

CACCCTCCAGATATATTCGT**CTCGAG**CAAATCACTTGGCAGTCTTAAAACATTATGCTGAGTGATATCC**CGCCGGCGCTTAA**

Xho I/Sal I Xba 1 Not I Sma I

C----TrueClone_Insert-----**CTCGACTCTAGATTGCGGCCGC**GGTCATAGCTGTTTCCTGAACATGTGAT**CCCGGGT**

G----TrueClone_Insert-----**GAGCTGAGATCTAACGCCGGCG**CCAGTATCGACAAAGGACTTGTACACTA**GGGCCCA**

GGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTGCCCACCAGCCTTGTCCTAATAAA

CCGTAGGGACACTGGGGAGGGGTCACGGAGAGGACCGGGACCTTCAACGGTGAGGTCAC**GGGTGGTCGGAACAGGATTA**TTT

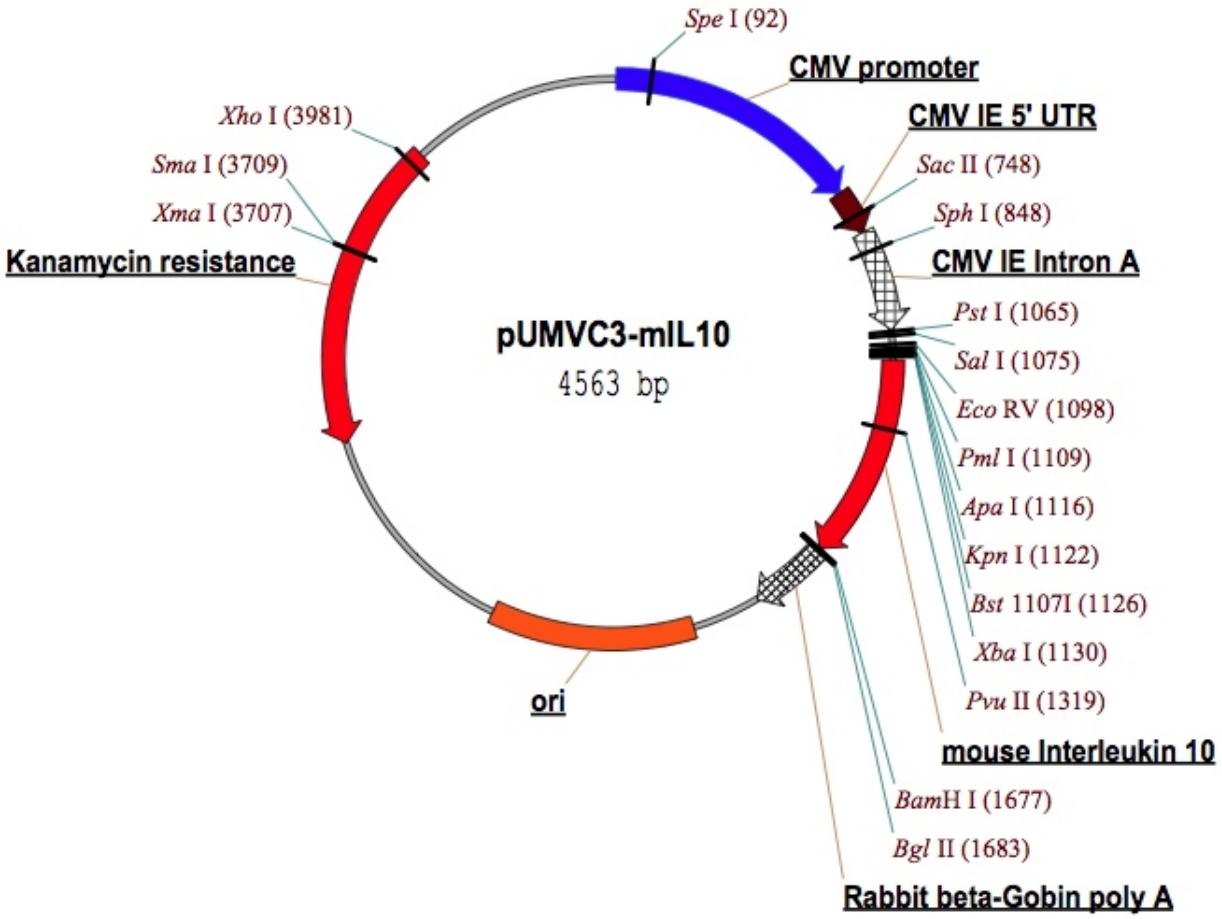
< Vector Primer XL39

c. NCBI reference sequence: NM_000584.2 *Homo sapiens* interleukin 8 (*IL8*), mRNA sequence (coding sequence capitalized and underlined)

```
   1 ctccataagg cacaaacttt cagagacagc agagcacaca agcttctagg acaagagcca
  61 ggaagaaacc accggaagga accatctcac tgtgtgtaaa cATGACTTCC AAGCTGGCCG
 121 TGGCTCTCTT GGCAGCCTTC CTGATTTCTG CAGCTCTGTG TGAAGGTGCA GTTTTGCCAA
 181 GGAGTGCTAA AGAACTTAGA TGTCAGTGCA TAAAGACATA CTCCAAACCT TTCCACCCCA
 241 AATTTATCAA AGAACTGAGA GTGATTGAGA GTGGACCACA CTGCGCCAAC ACAGAAATTA
 301 TTGTAAAGCT TTCTGATGGA AGAGAGCTCT GTCTGGACCC CAAGGAAAAC TGGGTGCAGA
 361 GGGTTGTGGA GAAGTTTTTG AAGAGGGCTG AGAATTCATA Aaaaaattca ttctctgtgg
 421 tatccaagaa tcagtgaaga tgccagtgaa acttcaagca aatctacttc aacacttcat
 481 gtattgtgtg ggtctgttgt agggttgcca gatgcaatac aagattcctg gttaaatttg
 541 aatttcagta aacaatgaat agtttttcat tgtaccatga aatatccaga acatacttat
 601 atgtaaagta ttatttattt gaatctacaa aaaacaacaa ataattttta aatataagga
 661 ttttcctaga tattgcacgg gagaatatac aaatagcaaa attgaggcca agggccaaga
 721 gaatatccga actttaattt caggaattga atgggtttgc tagaatgtga tatttgaagc
 781 atcacataaa aatgatggga caataaattt tgccataaag tcaaatttag ctggaaatcc
 841 tggatttttt tctgttaaat ctggcaaccc tagtctgcta gccaggatcc acaagtcctt
 901 gttccactgt gccttggttt ctcctttatt tctaagtgga aaaagtatta gccaccatct
 961 tacctcacag tgatgttgtg aggacatgtg gaagcacttt aagttttttc atcataacat
1021 aaattatttt caagtgtaac ttattaacct atttattatt tatgtattta tttaagcatc
1081 aaatatttgt gcaagaattt ggaaaaatag aagatgaatc attgattgaa tagttataaa
1141 gatgttatag taaatttatt ttattttaga tattaaatga tgttttatta gataaatttc
1201 aatcagggtt tttagattaa acaaacaaac aattgggtac ccagttaaat tttcatttca
1261 gataaacaac aaataatttt ttagtataag tacattattg tttatctgaa attttaattg
1321 aactaacaat cctagtttga tactcccagt cttgtcattg ccagctgtgt tggtagtgct
1381 gtgttgaatt acggaataat gagttagaac tattaaaaca gccaaaactc cacagtcaat
1441 attagtaatt tcttgctggt tgaaacttgt ttattatgta caaatagatt cttataatat
1501 tatttaaatg actgcatttt taaatacaag gctttatatt tttaacttta agatgttttt
1561 atgtgctctc caaatttttt ttactgtttc tgattgtatg gaaatataaa agtaaatatg
1621 aaacatttaa aatataattt gttgtcaaag taaaaaaaaa aaaaaa
```

B.2. Il10 cDNA vector

a. Aldevron vector map of pUMVC3-mIL10 (catalogue number 4032) containing mouse *Il10* cDNA (permission obtained)

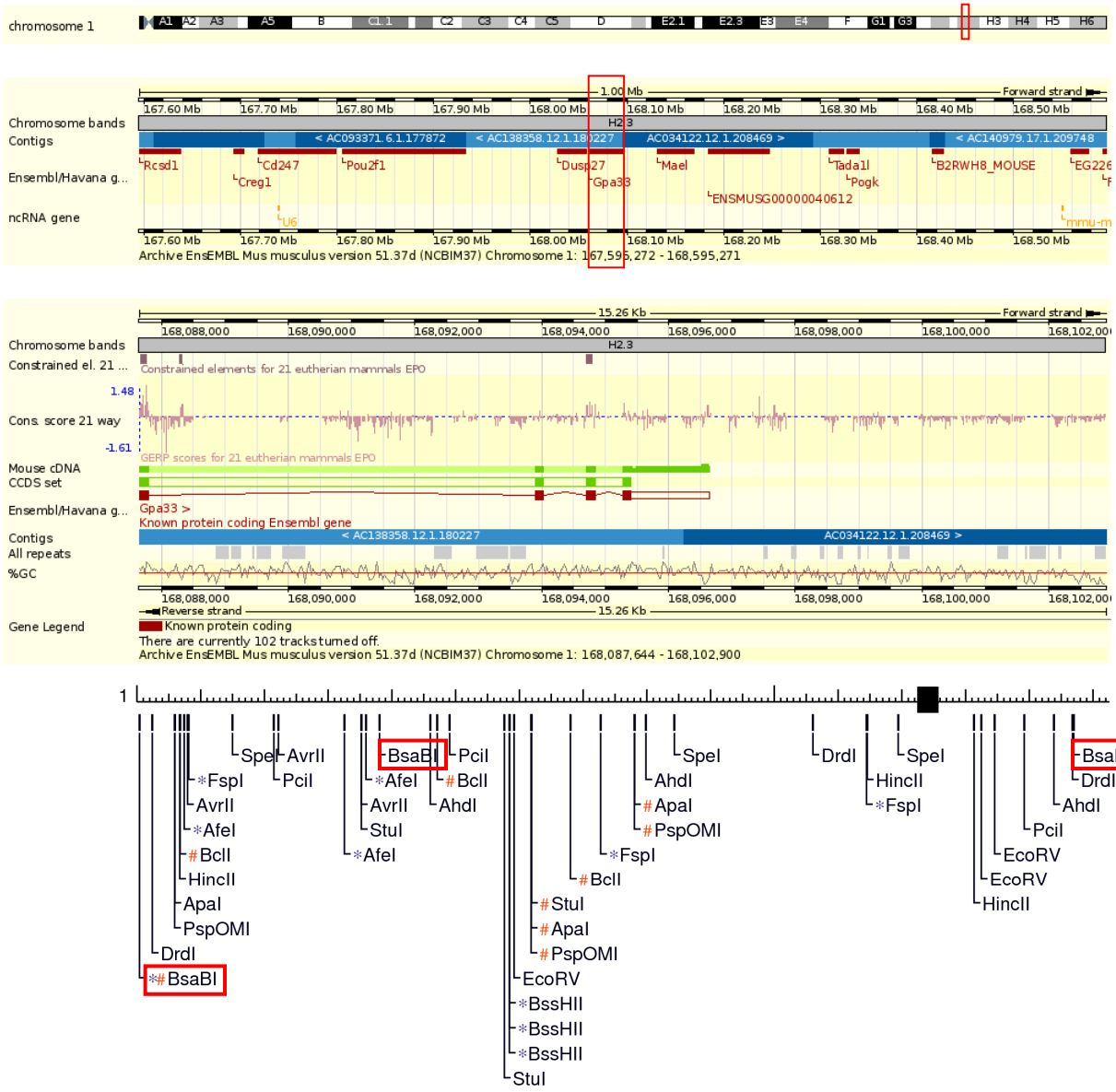


b. NCBI Reference Sequence: NM_010548.1 *Mus musculus* interleukin 10 (*Il10*), mRNA sequence (coding sequence capitalized)

```
   1 gggggggggg atttagagac ttgctcttgc actaccaaag ccacaaagca gccttgcaga
  61 aaagagagct ccatcATGCC TGGCTCAGCA CTGCTATGCT GCCTGCTCTT ACTGACTGGC
 121 ATGAGGATCA GCAGGGGCCA GTACAGCCGG GAAGACAATA ACTGCACCCA CTTCCCAGTC
 181 GGCCAGAGCC ACATGCTCCT AGAGCTGCGG ACTGCCTTCA GCCAGGTGAA GACTTTCTTT
 241 CAAACAAAGG ACCAGCTGGA CAACATACTG CTAACCGACT CCTTAATGCA GGACTTTAAG
 301 GGTTACTTGG GTTGCCAAGC CTTATCGGAA ATGATCCAGT TTTACCTGGT AGAAGTGATG
 361 CCCCAGGCAG AGAAGCATGG CCCAGAAATC AAGGAGCATT TGAATTCCCT GGGTGAGAAG
 421 CTGAAGACCC TCAGGATGCG GCTGAGGCGC TGTCATCGAT TTCTCCCCTG TGAAAATAAG
 481 AGCAAGGCAG TGGAGCAGGT GAAGAGTGAT TTTAATAAGC TCCAAGACCA AGGTGTCTAC
 541 AAGGCCATGA ATGAATTTGA CATCTTCATC AACTGCATAG AAGCATACAT GATGATCAAA
 601 ATGAAAAGCT AAaacacctg cagtgtgtat tgagtctgct ggactccagg acctagacag
 661 agctctctaa atctgatcca gggatcttag ctaacggaaa caactccttg gaaaacctcg
 721 tttgtacctc tctccgaaat atttattacc tctgatacct cagttcccat tctatttatt
 781 cactgagctt ctctgtgaac tatttagaaa gaagcccaat attataattt tacagtattt
 841 attattttta acctgtgttt aagctgtttc cattggggac actttatagt atttaaaggg
 901 agattatatt atatgatggg aggggttctt ccttgggaag caattgaagc ttctattcta
 961 aggctggcca cacttgagag ctgcagggcc ctttgctatg gtgtcctttc aattgctctc
1021 atccctgagt tcagagctcc taagagagtt gtgaagaaac tcatgggtct tgggaagaga
1081 aaccagggag atcctttgat gatcattcct gcagcagctc agagggttcc cctactgtca
1141 tcccccagcc gcttcatccc tgaaaactgt ggccagtttg ttatttataa ccacctaaaa
1201 ttagttctaa tagaactcat ttttaactag aagtaatgca attcctctgg gaatggtgta
1261 ttgtttgtct gcctttgtag cagcatctaa ttttgaataa atggatctta ttcg
```

B.3. Genomic locus of homologous recombination

A screenshot of mouse chromosome 1: 168,087,644-168,102,900 showing the genomic locus of homologous recombination from exon 4 of Gpa33 to downstream of 3' arm of homology as depicted with Ensembl (v.51, NCBIM37), and the corresponding NEBcutter *in silico* digest for Southern blot restriction enzyme and probe region determination.



*BsaB*I-digested WT allele = 10.9 kb
*BsaB*I-digested hIL8:c allele = 13.1 kb
*BsaB*I-digested mIL10c: allele = 13.3 kb

Approximate location of Southern blot probe hybridization in non-repeat region is indicated by a black box.

B.4. Genomic sequence of homologous recombination

Murine genomic sequence of Gpa33 from exon 4 to downstream of the 3'UTR, 3' arm of homology and 3' *BsaB*I cut site (Ensembl v.51, NCBIM37) for probe design and Southern blot restriction digest enzyme determination

>1 dna:chromosome chromosome:NCBIM37:1:168087644:168102900:1
TTCCACCCTCCAAGCCAGACTGCAGCATCCAAGGGGAGATGGTGATCGGGAACAACATCC
AGCTGACATGCCACTCTGCAGAGGGCTCTCCGAGCCCACAATACAGCTGGAAGAGTTACA
ACGCCCAGAACCAGCAGCGGCCCCTCACCCAGCCAGGTGAGGGGGCAGCAGGAAGGCCGT
GTATCCCACATAACGGGGCACTTGACTGCGTGGGGTTGCTAGAGAGACAAAGGCTGACCA
ACCAGTCCTACTAACACAGCCCCGCCCTTGAGCTACCTCGGACCGTCCCTCCGAGTATTG
GTTTTCCATACAAAAGCCTGTACACCCCACTTCACAGCTGTCAAGGACAGCAGTGGTTCC
CAGAAGCGTAGAGCAGGCGATGGTAGAGGTCACGTGACTTCTGTTCTTCCCACCCTCCAT
CCTCTGGAGGAAGTGAATCCCGCTAAGCATTCTGTGCAAAACACCGGGGCTTGTTCTTTT
AAGCTTTATTATAGCCTAAAACCCACTTTCGAGGTGGGGAATTTCCTAGCCCCACTTCCC
ATCCACACCAGTGCGTCCCAGACACCACCATGAACACAACATGGGCTGTTAGGGGCCCCG
TCCTCAGGCAGGCTACGGCCGATTTAGCTGTGGTTGCCAAGAGACCCATTTTGTACTGGT
CACCTAATAGTTGACCTTTGATCAGGAAATAAACAAGATACAGAAGGAGGTTAGGGGAGA
AAGAAGAGGAGAGAGGGGCAGCGCTGGGACCATGTAAACACTCAGCTGCCATTATGGCTG
CTGGAGTGGAAATCGCCTAGGCTCTCAGTGTGCGCACCAGCTGCTGCGGCTGCACCAGTG
AACCGAGGTGTGTCCATCAGTTTTACCTTTGAGGTCGTTCAGTGCTGCGTGTGTAGAAAT
GGGATAAGGAGGTGAGAGCAAAGCCTGTGGGTGAGGCTGATCTGGGTGAGGGAGGAGCAT
CGGGTTCCCTGGGGAAGCCCCGCGAAACATCCAGTGCCCTGGGAAGCCCCACGGATGAAC
ACCCCACACCCATGGGCTTCCATCAAAGCCGAGTGGATTATCACGTTGCCTCTCAGTCAT
CTGGAGGCCAAACAATAACTTCTTTGAAAAACCCAAAAAGTTAGTAGAATTCACTCGACT
TAAAGTAAAAGGGACGATGTACTAGATATGGATTCCCACAAGCGTCCAGTTTTGCTGGTA
TGTTTCTGGTTTACCAGTTTTATTTTTAAAAGGTTTCTTTATTTATATTATGCATGTGAG
TACACAGTTGTTCTCTTCAAAGACACCAGAAGAGGGCATCGGATCCCATTACGGATGGTC
GTGAGCCACCATGTGGTTGCTGGGAACTGAACTCATCTCTGGCAGAGCAGTCATTGAGTG
CTCTTAACCACGGAGCCATCCCTCCAGCCCGGATGGCCAAGTTAAAACACGTTTCTAAAA
CTGCCTTTGTGAGGTGTGTATGTATGTGGGAGAACTGCCCAGTGAGAGGCCAGAGGACAA
CTAGTTGGGAGCCTCCTCTCTCCACCACGTGAGCCCCAGGGCTTGAACTCACTTATCAGG
CTTGGCACCAAGTGCCTTCACCCACAGGGTCATCTCACCAGCTCCCGGCCTGCCAGTGTT
TGAATGTCTGAGATTTCTGTTGAGCTCACCTGGCAGCTGTGCTTAAAACCTGTGGTGTTC
TGTCTACCCGCTGTTGCCTGCTCCCGATCTAAGAACACAGTGATTTCGTGACTTTCCTCT
TGCAAAGTTTGCACCGTGGGGGCCAGGGCTGAAGGTCACGTGGTCATGAGCTTCGGTGGG
GGGTGGGGGGCTGACACAGTGATGTGATGCAGACTGTGTGAGGGATCCTGGAGGCTTTCT
TCACAAGAGGGATGTTGGAGTTGCAGCAAGACCAGGTGCTGAAGAGTAGCCACGGGCTTT
AATCCTAGCTGTGCTCCTTGCCAGCAAGTAAGGCAAGATACTTAGCCTGTCTCTGCCTTG
GTCTGTACATCTGTAGAATGGGCACAACACCAGTGTTCATCTCATGCTGTTACGGGAAAG
GGTGAACCAAAGTATGTAAAACACGGAGATTAGTGCCTGGCCAGCAGACCCATCAGAACG
GGAAAACAGCATGAGAGTCATGATGTCCGTGTGGCTGAGGGGTAGGGTACATGTCCATGA
TGGCCAATGTTGTTCATGGGAAGGACAATGGGGGAAGATGGAGAGGGGCCTGTGTACCAG
CCTAGGAGGTTGGACTTTTCCCTGTAAGCACTGGGCTCCTCACAGTTTAAGAGCACAGTC
CTCCAAGGTGGGGGAGCTGTGGCAGCTGCAGCTCTCAATGCCCCTGGCCACTGGAAAAGG
GGGGGGGCAGGGAGAGAGGGAGGGGAGAGGGAGAGAGAGGGAGAGAGGGAGAGGGAGAAA
GGGAGAGGGAGACACACACAGAGAGACAGAGAGACAGACAGACACACACAGAGGGATAAA
GAGAGGGGGAGAGGAGAGACACAGAGAGAGACACACAGACAGACAACACAGAAAGAAACA
GACAGACAGACAGACAGACACACACAGAGCCAGAGAGAGACAGAGAGACAGAGAGAGAGA
GGAGGGAGGGGGCAGGCATAGAGGGAGGAATTCTGGTGCACAACTCACTTTCTACTTTTT
ATTCAGTCCAATATTCCACTCCAAGGGATGGTGCAGTCTACAGCTGGGCTGGGGGGATGG
GGGATGGGGCATGGGGGAGGGTCTTCCCACCTTAATTAACCTAATCTAGACTGCCCCACC
CAACCACTCCCAAGGTGGATTTGATCTGGAATTCCACTAGGCGTGCCCAGAGGCTTCTCC
TGTAGGTGATTTCGATTCAGACTGACAGTATTAACTCTGGAAGCGATGCGGTTTTGTGTC
TGGATGTTGTAATGAGGTAGGGGTGCTGTTCTCAGACTCTTGCTCCCAGAGTTCCCTGCA
GAGCCGGGGAGTCCTATTTTAACTTTGTAGCTGTAAGTCACTTACAAGCTCCCTGGTCCT
TCAAAAACCACAATAAAGGCTGAGATGGTGCTAATGTTACCTTCCTTGAAGTAAGAGGCA
GTTTCCCATAGGCGACCTGTGAAAATGCCGGTGAGCTCAGGAGTGGACCGAACAGGAATA

CTTCAGGGGTCCGCCCCAGAGCCTGGGGTGAGACGTGGCAGCCAGAGCTTACTGACCGGC
GTTGGTGGGAGGGGACAGAAAGAGCGGATCCTGTTCATTCAACATCAATATCTTGGGGTT
ATCAGGGCTTCTTAGCGCTTCGGCGGGACAGATGACAGCCATATGCATACCCTTTCCTGA
CCACAAGATGGCGCCACAACAACGGGAAACCGGGTTTCAGACAGTCGTGGGCTCTTCCAG
AAGGAAAGGAGGCTGAGATTTCTCAGGTCAATCAAAGTCCATCAGTGACGTGGGATGAGC
CCACTTTTATCCCAGATAAGTTTCTAGAACATATCGTTTGCTTTAGAGACTAGGGGAAGG
ATGGTGCAGCAAGAACCCTCATGATCCCACCCCTAAAGGAGGCCTAGGAGCTGCCCTGGG
GTGCTGTGGACACTTGACTCTCAGTACACTCTCAGGTGGGGGCTCTGATGCCAGCGCTAG
GCGTCCTAAACACACCTGTACCTGGAGACAGAGAGCATTCTGTAAAGTGTGTTCCTTTAA
AATATTCTTAAGATTAACCAAATCCGGTTGGGTTAGAAAATGGGACTCTGAATCCCCATC
ATGGTTTCCACTGGCATTATCAAAACACCCGTAGGCCACGCCCCCAAAGTAGGCCACGCC
CCTGAAGCTTCATCACCATAGTGATTTACATCCTTGCTTTGCCTGTGGTCCCCTTTCCAT
TATCATATACGAATACCACCTGAAAATGGTATACCATGAAACACTCAGGTGTGGGGTACT
AGAATCCAAGGCTGCCACATTTTACAAAAAAATGTGCTTTGAGTTTTAGGGAAAGAATGT
GAAGAAAGTAGCATAATTATGATAAGAAGAGAAGAAAATAAAATCAATGTAGTGTGACTA
AGGTGATTTTATTTTATTTTCTCTTCTTTTAGAGGCTCCTTCTGGCCCTCAGACAAGAGT
AAACCCCTTCACCATGAGTTCAGATACAACCCCACTTTGAACCGTAACTGCTCATGAAGT
TACCACTGTTTCATGTGTGTCTTCCTTCATAGATTGGAAATGGGAGGAAGACAGGCGCTC
TCCTCGTTTTGTTTGTAGCTGTGATTGGAATGGAGATGGTTTTAGTTTGAAGGGAGTTGG
TGTGATAGCTTTATGTCATACATAAATTTTTTATCATAGGCATACATTGCTTTTATAGTG
CAAAGGTATTTTGCTTTTTAAACATTTATTTACTTATTATTAGTGTGTGTGTGCATGTGT
GCACCCGCGTGCACAGATACCAACATAAAACTCAGAGGACTCTCCTGTGAAACTGGAGTT
GGGAACACGTGGTCCACCTTCCATACAAGAATGTTGGAAGGATGCAACACCATTGAGGAT
ACTGAGAGACCAACAACCCCAGAGTGTGGAGGCAGGAGCTCGTGTGGACTGTTCCCTGGT
TTGGCACATGGGGAAGGAGCCGGACATAAGGCTGGTGTGACTGCAGGTCCAGACTGCAGT
TTTGTCACTGTATTAAAAGCCCAGAACGCAGACCTTAGGGGGGTGGTTCAATCAGCAAAC
TGCTTGGCATGAAATTGTGAGGACCTTGGTTTGATCAGGAGCAACCCACACCTGCCCGTA
ATTGCAGCACCGGGCGTGGGCGATGGGCAGAGACAGGAGGACCCCAGGGCTCACCGTGCA
TCCTCTGCACAGCAGAATCAGTGAGTTCCAGAGTCATCGAGAGACTCCACCTCAAAAAGC
AAGAGAGAAAAGAGACCGAAACACCTGAAGCTAACTTCTGGTCTCCACATGTGTACACAT
ATGTGCATGTACACACACACACACACACACACACACACACACACACATACACACACACTGAACT
GCCTATTGAAGATTAAAGATAGCTTCTGAAATTTCTAAAGACTGCAACAGGTCCCTCTTG
ATAACTTGGAGCGTGTCCTTGTGACAGATTGTGAAACCTGGTGGCACAAGGTTCTGATGA
GGCTGGCGCTGTAGTTAGGGGATGTGGGTTTCTTCTCTTGAAATAGCAGTGATGATGGAG
CATCTCTGAGAACAGGACATACACATGGGCTGTCATGGGTTCAGCCACTGCTGAAGGACT
GGATGGAGGAAGAGTTTAAACTGACACGTCCCTCAACTCACAAACATTCTCCCATAGCGT
ATCCATACCATACACCACACACATCACATATGTACATTTCCTGTTCTGTATATAGGACAA
GTTTCAGCCCAGGACCACCAAGACCAAATTTTCAGCACAAAGGAGATTTATCTGCCCCAG
AGGGACAAAGGGCAGGGATAAGGGGCCAGAACAGGAGATAGAGGAAGAGGGAGAAGGGGA
AGGAGGAGGGAGGGAGGGAGAGGAGAAAGTGGTATTTGTCCTGCAAGGGAACAATGACTG
TCTCTGGATAGAAAGGAGACAGATGTGGCCAATAGGCAAATGGCTGTTTATAAAGGTAAA
TGGGGAACCTGTGTTAGGATGCAGTGCTTAATTTTAACTGGCAGTGAACCCAAAAGGGCT
TTTGATTGCTGGACTTTAGTACTTTGAGAGCTGGCCCTTGGAAATAAGCCTTAGGAGGAG
GTAGCAGCCAAAGAAAAGGATAGACCCTGTGGCTAGCTTTGGGAAGAACTGGGGAGAAGG
GCAAGGCCTGCCAGAGTCATGTTCGCCATGCTTGAGCTGACAAGAGTCCCTGTGTGTGTG
TGTGTGTGTGTGTGTGTGTGTGCGCGCGCGCGTGCGTGTTTCCCAAGTACTTACTGCA
ATCAGGCATTCTACTGGACACTTATATTTATTATTGATATCTCATCTTAGACTCACAAGT
TCCTTGTCATACAGGTACCATCATTTCTAGTTATAGATGAAGAAATAGATTCCATAGGAC
AAGTTATTTGCAAAACACATCAAATAAGTAGGCAAAGCCAGAGTTGGCTCCCAGGGGTTT
TGCTGTGTTTGAAGCTCCTCACTGTTGTCCCATCTGCCATGTTGATCTCCCAGCATCCTT
AGAGGTGATTTTCCTTCCATCACCGATGGACCGGCTCAGAGAGGCTCAGAGTTTGCAGCT
CCCAGGGCCCAGGCCTCCCAGCCCTGGGCTGTGCACACTCCACCAAATGCTGTCTTTTCA
CCATGCTGTTCTGCGTTCAGTCTCAGGCGAGCCATTGTTGCTGAAGAACATCTCCACGGA
GACAGCGGGTTACTACATCTGCACCTCCAGCAATGACGTGGGGATAGAGTCCTGCAACAT
CACCGTGGCACCCAGACCTCGTAAGAGCAGGGTGGGGAGGGGAGGGCTACAAGGTCCAGA
GATGGCTGTTTGAGGATACTGGAGGTGAATGAGTGGCCGCTGTGACCCTGGAAAGATGAC
TTCTGGCCCCGTTTCTCTCACCCTGAGAAGCCCAGATCCTCTGGTTCTTTGGAAATAGAG
CCTGGGACATATTTCTGGAGACCTCTTCAGCCATGCCTGCCTTGTGGCCACTCAGAGAGT
GTGACAGGAGTTTCTTGAACCAGTTGTCTTGGTTTAAACCCTATTCTCTTTTGGTGGGGA
GAGAGAGACCATAGCATTAGCCAGAAGCAGGCGAAGCTAACACGCGGTGTGTTTGGAGAG

CCCAGGGAGCCCCGAGGCAGGATGCTCTGGTCTGTCTCTCCTTGTCTGCGCTTCTAACCT
AAAGAATGTCCTCAGATCTGCCTGATCACAATGCTCCAGTGCTTGACTGTTGTAGTGTGT
TACCAGGAGCTTCCAGCCATGGCTCCCTGGTCAGGCAGAAGAACTGAAGTTGCCCAGAGG
CTGGGATGGGATTAGCCTGCTGCGCCTTGTACCAGAACAATGGTTCCCCACCTCCTGCCT
GACTTTCTATTCTATACCCGTGATTGACTTGGAGGGGGGGAAGGGTAGAGCTGGGGAGAG
GAGTAGAGTTGGGCATCGGAAAGTACGGCTGTTCCATCCTCTTACAGCCTCCATGAACAT
TGCGCTGTATGCGGGCATTGCGGGGAGTGTCTTTGTGGCCCTCATCATCATCGGTGTCAT
TGTCTACTGCTGCTGCTGCCGGGAAAAGGATGACAAAGATCAAGACAGGGAGGATGCGCG
GCCGTAAGTGTGCTGGGATGGTTGATCCCTTGGAGGAGGGCAGATAGAGATGTATAGCCC
AGCCCCTCGGGCATGCGCATTAGCACCTTGGGAGACACCAACGAGTTTCAATCTTTCCCA
GTTGCAGTTGAGGAATGAATGACGCATCCAAGTCAAGGGTGGACTGAGGTCCTGCCTTTG
TCCTCTGGGCCTGTGGTTTGCAGCCTCTCTTCTCATGGCCAGCCGAGGATCAGACAGAGA
CTGGAACACACGCACACTCTTTCCACCTCGATCTCCTCAGCTCACGAGTTGGGAGATCTC
ACTGTGGGCTGGGTAACCAGCTTGCAGCAGAGTCCCTTGGCAGTTTACGAGTGGAGATGG
GAGTAGGAACAGTTCACCCCCCTGCCCCCTCACGAGGGAGGGCTAACTTTCTTTGTTTCT
TTATTGCCCTATAGGAACCGAGCAGCTTACCAAGTGCCTAAAAAGGAGCAGAAAGAAATT
TCCAGAGGGCGGGAAGACGAAGATGACCACAGACATGAGGATCGGTGGAGCTCGGGGCGT
AGCACTCCAGACCAGCCTTTCCAATGAAGGAGCCTTCTGCCCAGCGGGGAGAGGGGGGAG
GTCTAGGGCCCAGGCTCCTGCTCTGGCCACCCTTCTCCCCAGCCCCATAGTCCTGCTCCT
CCCCCCAGGCATTGGTGGAGCACTTCTTTCTGTGTATCTGCCTCAGGGGCGCATCACTGG
CCTGGCTACTGACGTCCTACCTGATGACCACTGAAAGGACCCTTCCTCATTGCTAAATCT
GACTCAGGGTCTGGCCTTCTCCCTAGACCAGACAGCATCCTCTGCCTCTCTTACGGCTGG
CAGGGCCTCCTAGAATCTCGGCACGATGCAGGGCTGCTTCTGTTCCAAGCTCGCCATACA
GGCAAGCGCCAGCTTCAGCACCAGTTTCAGAAGGTGGGAGTCTGCCCTGCCTCACCATGC
CCGACACTCAGAGGAGAACCAGTGATGGGCCTGCACGCAATATTTTGTTGAATGAAGGGA
AGAAATCAACGATAAGGTGTAGTTTCTGTGATTTACCCAGCTTAATTGCGCCCCCCCCCC
CCCCATGCTGAGGGGCCACGGTGGCCTCCAGCGTCCCCCTCTGCTCCATCTCCCACCATG
GCGTTTCTGTTCCTGGACACACACCTTCTTAGGACTTCTCCAAACACAGTTCCATCTCTG
CCTGTCAAGGGCCTGCCTGTCCTTAGAACTCACTAGTCTTGGATCCCCAAATCTTTCTTA
ATCCTGTGACTGCTAACACACCAGGAGGCAATGAAGTCACTTTGTCCCACACCTAGACCT
GGGCTGCATGCAAGGCCATCAGAGCTACACTTCCTGCTGTGTTAGAGTAGGTGCCCTGGA
GCAAGGTCTGGATCTTGGAGCCTGGCAGCCTCAACTGGGGACTTGCTCTCATGGACAGTG
CTTTGCCTCCTGGGTGTGACTTGGGACACCTGGTGACTCCACCCAAGCCCCGTGTTAGTC
TGGCTAGGTGGGCAGACTGATTTACAATGAACACAGACAAAAGCTTACTTCCCTCGCACC
TGCCAGCTGCCCTGCCTGCCGCAGCCTCCCCTTTCCATTACTATCCAGGACTCAGGCCAG
ATGGGCTGCTCCAAATGTGTTTTACTACTGGATAAACCTTTCTGGATCACCAGAAAAGCA
ACTGTATCAATCCCTTCCCTCCTCTGATTTTATGTACACATATATTTTATATACTTGGCT
AAATTCTTTGTTTCAATCAAAATGTTTCTTTATCAATAAAGTTACTGGTGGAAATTCACT
TTTCTTTGTTCACCCATCTTGTCTTTCCTTCCTGCCGAGTACACCTGTGCCTGGCGGGTC
GGTGAGGCAAGGTGACCAGCTAGAGAATCTTGGGGTTCTATTACTTGTGGCCAAGGTAAG
GGTAGGCAGTGACTCTCAAAGTTGTTGGGGACAGGACAGAGGTCAGCCTGATGGTGGAGG
TATATTCCGAGAGCCTACAGTGTCAGAAGACCTGGCAGGCGTGGTAGTCATCCATGTGCT
TCCAAGATGGTGACAGACACTCCCTTGCCTGTGATACCTACCCTACAGTGCCCCTTCTGA
GGAGATGGGCAGATCATCCTGTTACAGAGAACCATGGCACCTTTCTCCCAAAGCAGCTTG
GATCCGTGGTCGGTGCCCAGCTTGAACACAGACTCGTGAGACATTTTTAAGGTCATCTCC
TTCATGTCCATTCCTGTGAAGCTCTGTGTTTCTCCTGCCTCAGCATTTGCCATATGGTGC
CACCATGACCCGGGTGTACACATTCCCCTCTGCCTCTGCCTCCGAGAGGCATGAGCGTCG
GAACTGGCTCACGGGCATTTGGTAAGGGTTAGTTGCGTAGATAATGCTTACTGTCACAGT
GTCTTGCTCAAGGGCTAGCGCCTAGCTCCCTGGCTACTTAGAGATGCCACTATCCCACAT
GGTTCACGGCTCCACAGGCATGTCTGGAGGCCACTCCATCCTTGTTGGCTCCCATCTTAG
CTTGGGTTTAGGGAAGAGCACAGGAGAAAACAGGCTCAACTTCAGCAGGGAATGTTGTGA
GCACCTGGTACTGAAGAGCTAGTATCTAGGGCAGGAAGCAGACACAGGAACAGGTGCATG
AGACAGAAGCACTGGGGCAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
TGTGTGTGACACAAGCCTGGGGATGATTCAAATGGGTGATATTTGGTTGCATTTTGACAA
AACAGAGCAGGTGCTGATGCTCACTTGTCCTCTAGCGCAGGGCTCTATTATTTGGGTGTC
AGCTCTCATTATCAGTTGGCAACACTACAGATTCCTGCTGAATTTATCACAGTCTAGAGG
CTGGGGCAGTTAGATTGCTGTAGGTTTCCATACAGCTTCAGCTAGCCCAGCATGGATGCT
CGGAGCCTCTGGGCGGGTGCTCAGACAGGCCCCTGAAAGATGAGATGCAGCTTGTAAATT
AGCTACTCATCAGGGCTTCTAATGAAGTTGCTTTTGTGTCACAGTGTATTGGGTCTCAAA
GGCAGAGGCATGTACCAACGGCCAGTAGTAAGGCTGTACATACTGTGGACCAGCTACGGG

TCTAACCATTTCCCACATCTCTCCTACAGTCCTCACGGGACCAGTCTTAGTCAGTGTTCT
ATTGCTGGGAAGAGGCAGCCATGACCGAGGCAATTCTTATAAAGGAAAACATTTAATTGG
AGCTGGATTACAGTTTCAGAGGCTTAATCCATTATCATCATGGCAGGGAACATGGTAGCA
TGCAGGCAGACATGATGCTGGAGAGGTGGCTGAGAATTCTACATCAAGATCCGCAGGCAG
CAGGAAGAGAGAGACACTGGGCCTGGCTTGTGCTTTTTAGACCTCAAAGTCCACCCCCAG
TGACACACTTCTTACAATATGGTCACACCTACTTCCTCAAGGCCACACCCACTAACCTGT
CTAAAGTAAAGGCACTCCCTAATGGCTAAGCATTCACATCTATGAGCCTACGGGGGCTGG
TTCTATCAAAGTCACCAGATGCTTTTAGTATTCCCGTTTTACAGAGTGGAAGTTGAAGCA
TGGAAAGTCAAGCCCTTAAAACACAGGATGTCAAGTTGCTATTTCCAGCTATAAAGTCTT
TTTGTCGCGGTGACTAGGGACTATGTTGCTATCGAGTACGGAATCAGAATAAAAGAAGAT
GGCTGCCTTTCAGTAACCCTCATCATCTATAACAGTACTTACAAGCAGGCATTATATCGC
CAGTGAAGAGTCTAGAAGCTTCTGGGGTCTTGGTTCTATCCCTGTATGTTGTTTTTATTT
ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATATATGTCAGT
ACACTGTAGCTGTCTTAGATACTCCAGAAGAGGCATCCAATCTTGTTATGGATGGTTGTG
AGCCACCATGTGGTTGCTGGGATTTGAACTCAGGACCTTTGGAAGAGCAGCCGGTGCTCT
TAACCCCTGAGCCATCTCACCAGCCCCCCTGTATGTTGTTTTAAGGCTCTGGTCATCTTC
ACAAGGTGATCTTGCTGGGTCCTGTGGGCGGGGCCTCTTTGACCATGCCCCTATATTTGC
ATACTGGTGTTTTTTTTTTTTGTTTTTTTTTTTTTTTCTTTTCTGGTCTCAGTCAGGTAC
ATTTCCAGAGGCACCTAGCTGGTATCTCTGATGACCCTTGTTTGCTGCGCAGCCTCAGTT
AACCACGTTCGCTTCCTTCTCCCTTTCTTCCTATTTACACACTTTTTTTTCAGACTGGAG
AAAAACTTCTTCCGAGGTCAGCTTGACCTGGGAGCTGCGACTTATGGCTAGCTTTCTGCT
GTGATCTGGGATGGGTTGGTTTTCACACTCATCTCAAGTGGGCACGTTTACCCCTGATTG
CCAGGCTGTTGAGTGGATTCTTGATTCAACGAGGATAAACGAAACTGCTCCAGAGATGGA
CGTAAAAATCGTCATGACTGAAAAATCAGCTGACAATGCTTTCTTACACGTGCAAGTTCC
CCATCGCCAGTTTATCTTTTTGTTATTTATTTATTCATTTGGGTGGGTGTATTGTGGAGG
TTAGAGGACAAGTTTCAGGGGTTGGTTCTCTCCTTCCATCCATCCTGGGGATCCCGCTGA
GGCAGCCTCTGAGGCCCCCGTCTGTCTAGTGGGTGTATATCCCGCCCTCAGGGTGCTGTT
TGCCACTAGTCAGCAGAGGAGACGGAACACACAATAAACGTCCAGTGGTTTGTCTTTCTT
TCTTTCTTTCTTCCTTCCTTCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT
TCTTTCTTTCTTTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTTTCGAGACAGGGTTTCTCTGT
ATAGCCCTGGCAGCCCTGTAACTCATTTTGTAGACCAGGCTGGCCTCGAACTCAGAAATC
CACCTGCCTCTGCCTCCTGAATGCTGGAATTAAAGGCATGCGCCACTACGTGGGGCTCAG
TGGTTTCTTTCAACTTTGGTCCTGAGGTGACCACAGAGACTCTGCTTAGGCCACAGATGA
GTCACCATGTAACAACCATGTCAGTGGGTGGCAAGATGTCAGTCAGTGTGACCTTGACAT
GGGGAGAAGATGAATTGGCTGTTGGTCTCGGCCACCATCGCCTCCTTCCAGTCTCTGAGA
ATTCTCCATTTTCCGGAGGAGGCTGGGATGTCATTAGAGGCGACAAAGCCAGGAGGATGA
GCCTGAGACTCAGATCTGATGATGAACACCTTGTCCTTGTCACCTGGTCACATCTTCTCC
TGGTCCTGTGCACATGCAGAGGGTGGGAGCAGAGGAGGCTGAGTTAGGGCTGCTTGCTTT
GCATTTATATATATGCTACATAATGAGACCCTGTTCCGAAAGTAATCGACTTGTGTGATG
AGCTGAGGTCTTTGTCAGGCACACTTCATCTTCTTACACCCTTTGAAGGGACAGGGTGAC
TCATTCAGCTTTGCAAAAAAGCCAAAGCTTATTCATTAACATTTTCACTTGCCACAGTGT
AAGTCCCCTCAAACATGGGTCTCAGAAATGTTCAGATCCTATTTTGAGGGGATAAGTGTC
ATGTGTTGCCTTTCAAGATGCAGGAGGTCAGGGGAGGGATTCCTCAGCTAGCCACGCCTA
CAGCAGTGTGTCTAGGGACCAGAGAGGCTGGGGCAGGGAAAACCGACAGATATGCACAGC
GTGCGGGGAAGTGGCACAGTTCTGCAGGACATCTGTCCTGCACCTTTTAAATCTCTAACT
TAAAACAAGGTCCAGGTACAGTAGGAAGACCTGAATGGTGTGAACTCAGTTGTTGACACA
CTGTTCACTTTAGTATGATGCAACAGGAAGGGACGCCATCACCCCCTCATTACAGCATCA
TCCTCAGAACCCATCAAGGTCCACAAGATTCCCAGGAATGCTGGGGATATCAGCTGTGTT
CACTTGGTCAAGGGTGTGTGCAGGATTTCTCTGCTCACTTCCCATCGCTAACACTTGGAG
ACTTTGGAAGCATCTCAGAGAGTAGGTAAACCTTGTCTCCTCCAACCTTATTGACGTATT
TATGAAGTCTCTGCAGAATTCTGCGCGTCTGTTTCATTCCGCAGGTTCTGACCTATCGCT
GACCTCCCCCTCTTGATATCTGAATCACTTGCGGGTTGGCTAGGAAGCACCTTGGCCACT
GCCATGTCCCTACCATTCTATGTGTACCAAGTGCCTTCTCACCGTAAGCTTGCTTTGTAC
CATTTCTGTCCCAGCTCTGACGTCAGCTGTTTCTCCATGGACCATCTGCCCGGCTTAGGG
TAACACTTTTCAGAATGTGAGGTGTGCTCAATGTTGCTAAGGAGTAACTGCTCTCAGCCT
TCCTCAGGAGCAGACAGACTTAGTGAGTGAAAGACGCGCCTCTTGAACTTGCAACTACGT
TCTTGTGTTGAAGATAAACACTCACTCCAATGTCTCTGTTTGGCCTCCAACAAGCCAAGG
GTTCAGCTGAGTGTCTTTTGGCTGTGAGAGGCTCTAGAACCCCATTCTCTTTTATCTATT
TACCAATTTGAGCAGTTTCTCTGTGTTGTACTTTCCCATTTATGGTCTTTCCCTCTTTGC

TACATGTGACAGGGCTCTGCCTTCCCTACTGAAGAGCCTGGCTTGTGAAGACAGTCACTT
CCACTTCTTCTTCTTCTTTTTTTTTAACTGAAATATATATTTTTATATAAAACATATCAT
AAGATTTACCTAACGTACAAGAGGCACACAGAACACCAGAACACTGGACCAGAAAAGACA
TTCACTTTAGCAGATAATCAAAACACTAATGTACATAACAAAGAATGAATATTAAAAGCT
GCAAAGGAAAAAGAGAAGTTGCAGATAAAGACAGAGCCATTCAAATGACACCCGGCTTCG
CAATGGAGACCCTGAACAAGAGAAGGACCTGGCAGACGTTTTTATAAGCGCCAAGGGACC
GTAGATGCCAGTCCACAGTACTATTCCACTCTACGAGATAGGACCCATCCATACTTGAGA
CTGCTGAAGAGGCCAAGAACCTGAGACTAGCTAGGTTATAAGACTAGAGGTCCCACTTCC
ACTTCTAGAAGCGCGTGGCTTCTGCTTCCCTAAGCCCTTACCCTGCTCATGTCGGCTTCA
GGACGTAGTGTGAAAGGTTCAAACATAAGCTGAAGGTACTTTTGTCTACACACACACACA
CACACACACACACACACACACACACACACAGAGAGTTGTGTCAATACCACTTACTGAAAGGT
TTCTTTCTCCCACTGACCTACCTTGGTATCTTGGGGGATGCGGGTGGGGTGGTGGAAATC
AACTGAGGATCTAGGTGTGAGTCTATTTCTGGACTCTCTGGTCTTCTCCCTTGAGATGTT
TATCTGTCTTGATGCCAGTATTGTCCTCGATACTATGGTGCAGAGTATGCCTTGGTTGTG
ATAAAGCACTTAAAATCTCCTGAGTGCTTTTTAAGGAATGTTTCTGACCAGTCTAGATTC
TTTGATTTTTTTGTATCATAGGAGAGTTAGTTTTTATATCTATGAAATAAGCTCTTGGAA
TGTTAGTTGCAATGTCAAGCTGAGAAGGATCATCAGGAAGGCTGACTCCTTCTAGCTATG
AAGGAGCGTGCTTCTCTGCTGTTCAGATCTTTTGAACTCCTCACAGAGACACTAGGCAGT
TCTCAGTGCAGAGATTTTGCAGGTATTTTGTTCCATTTGAAAATGCTTGCTGTTTTTGAT
ACTATGATAATCGGAATGGGTTTTTACATCCCATTTACTTTGCTTAATTTATTTTGCATC
TCTATTCTTAGATACATACATTTAAAATAATTATAGCTTCCTGTTTAATTGGCCCTTTTG
TCAGTAGGAAGAAGTGTTCTTAATTAGCTATGTTAATACTCTATCTTGAAGTTTATATTA
ACTAATAAATGTTAAAC

Sequence Decorations Legend

A33 exons 4-7 underlined

GAATTCCACTAGGCGTGCCC end of 5' arm of homology

3' UTR

TGGAGAGGTGGCTGAGAATT end of 3' arm of homology

*BsaB*I cut sites (GATNNNNATC)

Probe

Repeat elements in genomic sequence

B.5. Microinjection observations – WTSI RSF barrier

a. Donor/Recipient Mice

The 3.5d embryos used for the microinjection are from albino C57BL/6N donor mice (7-8 embryos are flushed per donor mouse) so the coat colour of F0 chimera pups can be used to determine the percentage of microinjected embryonic stem cells (ESC) that integrate into the inner cell mast of the blastocyst, i.e. the darker the coat the greater chimera percentage. If the F0 chimeras are paired with albino mice the coat colour of F1 offspring can again be used to determine whether the mutation has gone germline i.e. black pups indicate germline transmission has occurred whereas white pups indicate that the mutation has not been transmitted. However the F0 hIL8 and mIL10 chimeras will be crossed with CBLN mice (black coat colour) so I will need to genotype all F1 offspring to determine whether germline transmission occurs.

The 2.5d pseudo-pregnant females used as recipients of injected blastocysts are induced into pseudo-pregnancy by mating with sterile males; mice are the only rodents that do this which makes it possible to time donor and recipient mice.

At 2.5d, the ovaries of pseudo-pregnant females have two red spots, hematoma, which indicates that hormone levels have increased and the ovary has released an egg, i.e. the uterus will be able receptive to blastocyst implantation.

b. Microinjection

On the morning of microinjection, the uterus is removed from donor mice and embryos flushed out with media. The ESC to be injected and blastocysts are put together onto a depression slide and covered with mineral oil. Approximately 10-15 ESC are injected into each blastocyst into the space below ICM and approximately 30 blastocysts are injected per ESC clone, depending on number obtained per donor mouse and how many clones there are to inject on that particular day. ESC sucked up a capillary needle as close to one another as possible to reduce the amount of media/volume to be injected. The blastocyst is positioned against the suction arm (microinjection arms controlled by air pressure) and then the needle is inserted in between the cells lining the blastocyst and into the inner space of the blastocyst, avoiding the ICM. It is important not to disturb the cells of the ICM because these are what divide to form the embryo and the injected ESC need to become part of the ICM. Loss of pressure upon injection causes the blastocyst to collapse, which increases the chance of the ESC being incorporated into the ICM. The injected blastocysts are incubated for 1-2 hours at 37°C with 5% CO_2; these blastocysts can be implanted right away but the incubation allows them to recover and then it is possible to make sure that implanted blastocysts are viable. The blastocysts need to be implanted on the same day but time of incubation isn't critical. Viable cells have an intact membrane and have rounded up again; it may be possible to see the ICM starting to divide!

c. Surgery

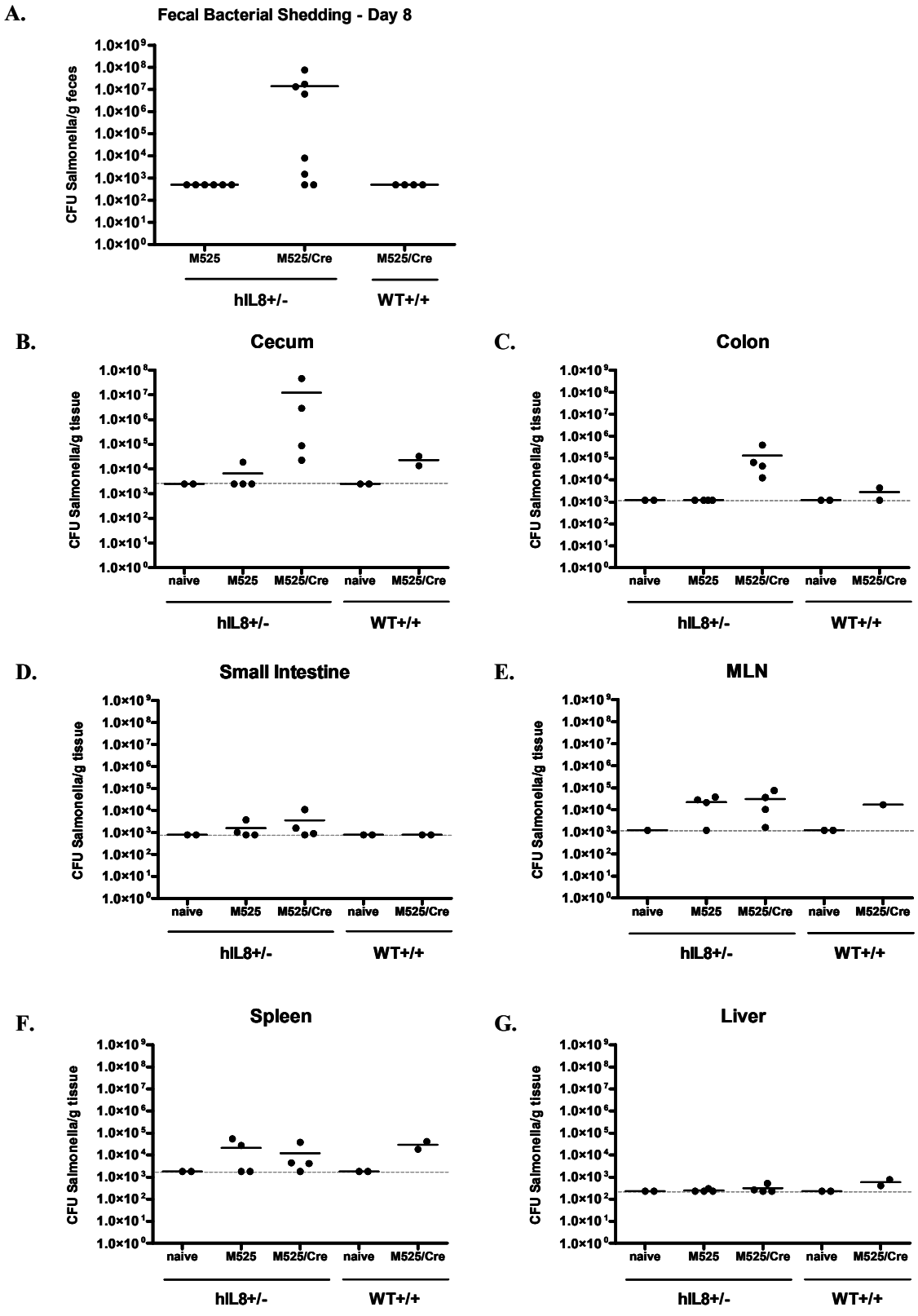
Recipient mice are anesthetized by intra peritoneal (IP) injection; RSF also has the capacity for gas anesthesia but it is more difficult to do large numbers and the mice have to remain in the same position throughout the procedure, which makes manipulation more difficult. It takes about five minutes for the anesthesia to kick in so everything is set up and the blastocysts retrieved from the incubator during this time. The eyes of the mouse are moistened with drops since the anesthesia doesn't allow them to blink and causes their eyes to bulge, and also don't want their eyes to dry out. A small incision is cut in the back of mouse in the abdomen region (sterilized with 70% EtOH) and small piece of underlying connective tissue removed. The fat pad connected to ovary is found and used to pull the ovary and uterus out of the incision. At this stage it is possible to tell whether the mouse is pseudo-pregnant (2.5d stage) because of the

hematoma on the ovaries. If the hematoma is not present then the blastocysts will not implant and the mouse can not be used as a recipient. Since surgery has commenced the mouse can not be used for any other procedure and therefore has to be culled. If the hematoma is present, a needle is used to make a hole in the uterus; a spot of blood will direct back to site of hole. The blastocysts are drawn up a Pasteur pipette with mouth pipetting creating bubble guides; blastocysts will be positioned between 2-3 bubbles for insertion into the uterus. 2.5d blastocysts are inserted into the top of the uterus to coincide with normal embryo placement at this stage; if 0.5d blastocysts are used, the blastocysts are implanted into the oviduct, which is a much more difficult procedure. Approximately ten blastocysts are implanted into each mouse so the number of recipients depends on total number of blastocysts injected. The blastocysts are mouth pipetted into the uterus and the small hole will repair itself so suturing is not necessary. The fat pad is used to direct the uterus and ovary back into the abdomen; the organs will sort their positioning out themselves. The incision is then wiped with 70% EtOH and held together to heal with two small clamps; the clamps will be taken off approximately ten days later and the skin will have healed, i.e. again, no suturing. The mice are then wrapped in kimwipes to keep warm and placed in a recovery chamber for 90-120 minutes and monitored throughout; the mice are very quick to recover from gas but take much longer following IP injection.

The mice are then put into the RSF barrier (clean side) until F0 chimera offspring are born. The staff of the RSF genotype the pups based on their coat colour, denoted as a percentage and normally females and any males below 40% will not be used in subsequent matings for F1 generation. Exceptions are made if no high percentage male chimeras are born but this is on a case by case basis.

B.6. Fecal and organ bacterial counts from hIL8:c/- mice infected with *Salmonella typhimurium* expressing Cre recombinase

Homozygous hIL8:c/- knock-in and C57BL/6 wild type mice were infected orally with 2 x 10^7 colony forming units (CFU) of *S. typhimurium* alone (M525) or *S. typhimurium* transformed with pSopE-Cre (M525/Cre). Uninfected hIL8:c/- or WT mice were used as naïve controls. Fecal shedding of the bacteria was monitored every 2-3 days by plating serial dilutions of feces resuspended at 100 mg/ml in PBS onto LB plates supplemented with 100 mg/ml ampicillin or 25 mg/ml chloramphenicol. CFU of Salmonella per gram of feces observed on Day 8 for each of the infected groups shows a higher geometric mean for the hIL8:c/- mice infected with the *S. typhimurium* expressing Cre (A). The four hIL8:c/- mice with the highest fecal bacterial counts showed signs of sickness (ruffled fur, hunched posture, weight loss and less movement) on Day 10 and had to be culled in accordance with ethical protocol. Organs from these mice and others from each of the control groups were taken for bacterial counts, RNA extraction, histology, immunohistochemistry and electron microscopy. Much higher CFU of Salmonella were observed for the hIL8:c/- mice infected with the *S. typhimurium* expressing Cre in the cecum (B) and colon (C) and a modest increase in the small intestine (D), however bacterial numbers within the mesenteric lymph nodes (MLN) (E) and spleen (F) did not show significant differences between the infected groups of mice, and bacterial numbers within the liver (G) did not show significant differences between all (naïve and infected) groups of mice. Limits of detection for bacterial enumeration are indicated with a dashed line. The original pSopE-Cre plasmid was a gift from Jorge Galan. This plasmid contained an ampicillin resistance gene, however the *Salmonella enterica* serovar Typhimurium (*S. typhimurium*) M525 strain is also ampicillin resistant. Therefore the antibiotic cassette was exchanged for chloramphenicol (backbone plasmid obtained from Sidney Kushner). The final plasmid used for transformation into M525 contained the SopE-Cre fusion under the control of the SopE promoter with a chloramphenicol resistance gene.


A.
Fecal Bacterial Shedding - Day 8
CFU Salmonella/g feces
M525
M525/Cre
M525/Cre
hIL8+/-
WT+/+
B.
Cecum
C.
Colon
D.
Small Intestine
E.
MLN
F.
Spleen
G.
Liver
CFU Salmonella/g tissue
naive
M525
M525/Cre
naive
M525/Cre
hIL8+/-
WT+/+

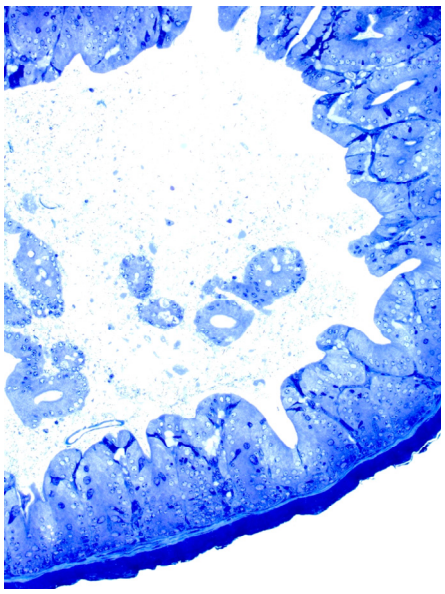
B.7. Cecal micrographs from hIL8:c/- mice infected with *Salmonella typhimurium* expressing Cre recombinase

Toluidine blue staining of sections of cecum, hIL8:c/- naïve (A), hIL8:c/- M525 (B), WT+/+ naïve (C), WT+/+ M525/Cre (D), hIL8:c/- M525/Cre-1 (E), hIL8:c/- M525/Cre-2 (F), hIL8:c/- M525/Cre-2 (G), taken at Day 10 of infection with *S. typhimurium* alone (M525) or *S. typhimurium* expressing Cre (M525/Cre) fixed for transmission electron microscopy and viewed under light microscopy. The infection conditions are described in Appendix B.6. Longer crypts, thicker submucosa, and neutrophils in the mucosa and lumen are visible in E-F, and G shows increased detail of the neutrophils within the lumen. All sections are 20X magnification except for G, which is 40X magnification.

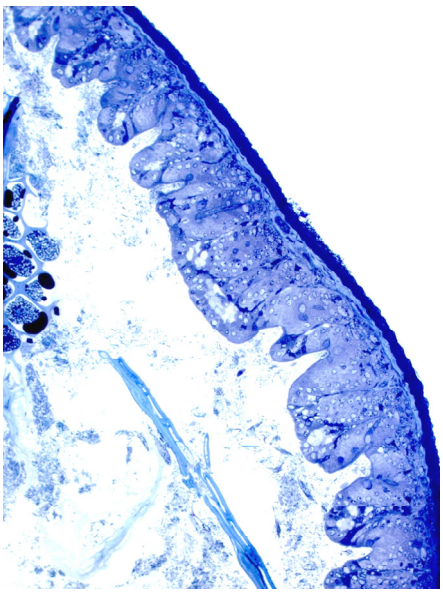
Summary of hIL8:c/- infection with Cre-expressing S. typhimurium:
High M525/Cre bacterial shedding was observed in the IL8+/- mice. The IL8+/- mice infected with M525/Cre had similar bacterial counts within the liver and spleen compared to the other infection groups. This suggests that intestinal pathology, not systemic infection, caused the sickness, which correlates with the high cecum and colon bacterial colonization in the IL8+/- mice infected with M525/Cre. The hypothesis from this experiment is that IL8 recruits neutrophils leading to inflammatory neutrophil damage reducing the effectiveness of host defences and increasing *Salmonella* infection.

Sections of small intestine, including the Peyer's Patch, cecum and colon tissue were taken from each of the mice culled at Day 10 for RNA extraction, histology, immunohistochemistry and electron microscopy. IL8 expression will be tested in these tissues by qPCR and immunohistochemistry. Tissue morphology will be analyzed by histology and electron microscopy.

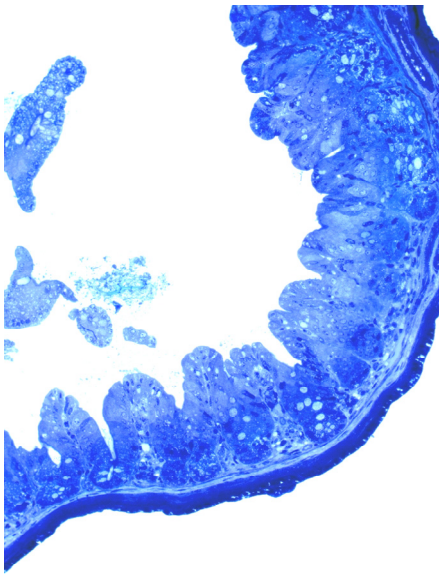
A.



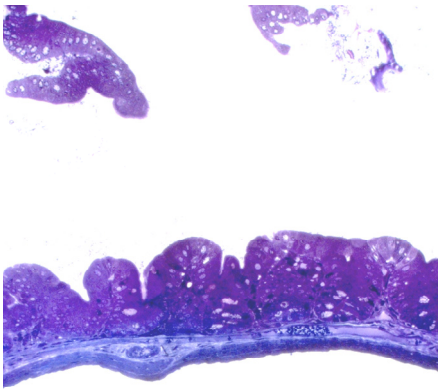
B.



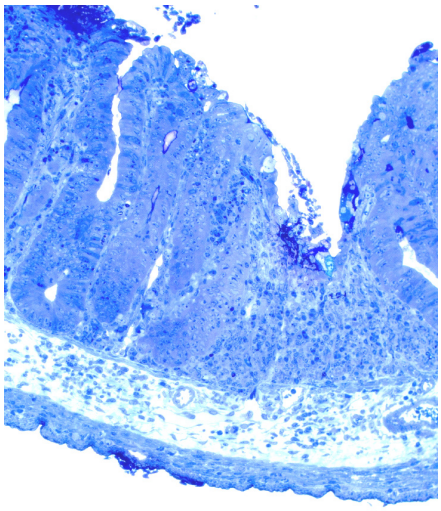
C.



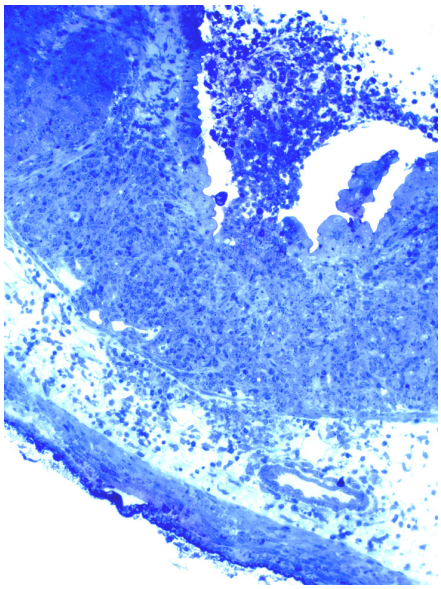
D.



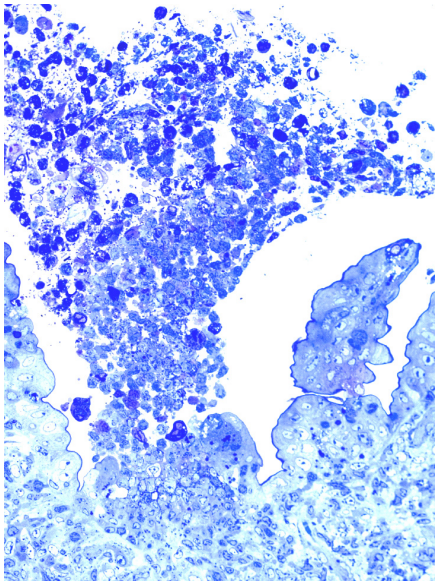
E.



F.



G.



C. SUPPLEMENTARY INFORMATION FOR CHAPTER 4

C.1. IL12RB1 patient data set DE genes – media alone treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
CHURC1	91612	1.51	2.84	2.84	6.29E-09	3.40E-05
ATP8B4	79895	1.31	2.48	2.48	2.75E-06	4.17E-03
NT5C3L	115024	1.17	2.25	2.25	2.85E-13	5.39E-09
PASK	23178	1.15	2.22	2.22	3.51E-05	2.76E-02
IGLL1	3543	0.97	1.96	1.96	9.32E-06	9.53E-03
RPL23AP7	118433	0.97	1.95	1.95	1.02E-08	4.30E-05
WARS2	10352	0.92	1.90	1.90	6.22E-12	7.85E-08
RNASE6	6039	0.75	1.69	1.69	3.33E-05	2.68E-02
TSHZ2	128553	0.73	1.66	1.66	1.01E-10	6.34E-07
ARL17P1	51326	0.65	1.57	1.57	4.18E-07	9.88E-04
CLDND2	125875	-0.66	0.63	-1.58	1.48E-08	5.51E-05
ADRB2	154	-0.66	0.63	-1.58	2.19E-05	1.92E-02
APOBEC3H	164668	-0.78	0.58	-1.72	2.02E-08	6.36E-05
AOAH	313	-0.82	0.57	-1.76	2.90E-06	4.21E-03
KLRB1	3820	-0.88	0.54	-1.84	1.55E-11	1.46E-07
GPR114	221188	-0.96	0.51	-1.95	3.62E-11	2.74E-07
GJB2	2706	-0.96	0.51	-1.95	6.69E-05	4.04E-02
ATP13A1	57130	-0.99	0.50	-1.98	8.36E-15	3.16E-10
RPS23	6228	-1.07	0.48	-2.10	7.18E-06	8.76E-03
GSTM1	2944	-1.10	0.47	-2.14	9.25E-06	9.53E-03
GSTM2	2946	-1.29	0.41	-2.44	4.14E-05	3.07E-02
S1PR5	53637	-1.30	0.41	-2.47	6.24E-05	3.91E-02
KIR2DL1	3802	-1.51	0.35	-2.85	2.46E-08	7.17E-05
KIR2DL4	3805	-1.77	0.29	-3.42	1.20E-06	2.16E-03
GZMH	2999	-1.88	0.27	-3.68	6.90E-07	1.31E-03
ZNF683	257101	-1.88	0.27	-3.68	9.11E-06	9.53E-03
FGFBP2	83888	-1.90	0.27	-3.74	2.89E-07	7.81E-04
KIR2DL3	3804	-2.05	0.24	-4.14	5.66E-07	1.13E-03

C.2. IL12RB1 patient data set DE genes – IL12 treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
CHURC1	91612	1.51	2.84	2.84	6.35E-09	2.40E-05
PASK	23178	1.17	2.24	2.24	2.85E-05	2.17E-02
NT5C3L	115024	1.08	2.12	2.12	2.30E-12	2.18E-08
RPL23AP7	118433	1.04	2.06	2.06	1.94E-09	8.17E-06
IGLL1	3543	1.04	2.06	2.06	2.80E-06	3.53E-03
WARS2	10352	0.96	1.94	1.94	2.31E-12	2.18E-08
OLIG2	10215	0.86	1.82	1.82	3.89E-05	2.71E-02
COL18A1	80781	0.65	1.57	1.57	7.56E-08	1.91E-04
ARL17P1	51326	0.65	1.57	1.57	4.56E-07	8.21E-04
VAV3	10451	0.61	1.53	1.53	1.65E-06	2.23E-03
TSHZ2	128553	0.60	1.51	1.51	1.21E-08	4.15E-05
EEF1G	1937	0.59	1.50	1.50	1.06E-05	9.54E-03
IL2RB	3560	-0.66	0.63	-1.58	3.53E-07	7.04E-04
SETBP1	26040	-0.67	0.63	-1.59	9.39E-06	9.02E-03
CCND2	894	-0.71	0.61	-1.63	1.00E-05	9.25E-03
ADRB2	154	-0.74	0.60	-1.67	3.50E-06	4.27E-03
CMAH	8418	-0.78	0.58	-1.71	3.95E-05	2.71E-02
CLDND2	125875	-0.78	0.58	-1.72	2.52E-10	1.86E-06
IL18R1	8809	-0.85	0.55	-1.81	1.01E-07	2.26E-04
APOBEC3F	200316	-0.87	0.55	-1.82	9.61E-05	4.72E-02
APOBEC3H	164668	-0.94	0.52	-1.92	2.94E-10	1.86E-06
CEBPD	1052	-0.97	0.51	-1.96	1.41E-05	1.19E-02
NOD2	64127	-0.98	0.51	-1.97	4.41E-06	5.06E-03
SIGLEC10	89790	-0.98	0.51	-1.97	3.12E-05	2.29E-02
GSTM1	2944	-0.98	0.51	-1.97	4.81E-05	3.03E-02
KIR3DL3	115653	-1.01	0.50	-2.02	9.05E-05	4.58E-02
ATP13A1	57130	-1.03	0.49	-2.04	2.80E-15	1.06E-10
RPS23	6228	-1.03	0.49	-2.04	1.40E-05	1.19E-02
KLRB1	3820	-1.05	0.48	-2.08	1.04E-13	1.98E-09
LAG3	3902	-1.07	0.47	-2.11	3.24E-05	2.31E-02
KIR3DL1	3811	-1.28	0.41	-2.43	4.02E-05	2.71E-02
KIR2DS5	3810	-1.35	0.39	-2.55	9.28E-05	4.62E-02
MAP3K8	1326	-1.47	0.36	-2.77	6.41E-07	1.01E-03
GZMA	3001	-1.50	0.35	-2.82	6.92E-05	3.85E-02
ZNF683	257101	-1.62	0.32	-3.08	7.64E-05	4.07E-02
FGFBP2	83888	-1.65	0.32	-3.14	3.61E-06	4.27E-03
KIR2DL1	3802	-1.74	0.30	-3.34	1.05E-09	4.96E-06
FES	2242	-1.79	0.29	-3.45	2.02E-08	6.36E-05
IL18RAP	8807	-1.82	0.28	-3.52	4.33E-07	8.19E-04
GZMH	2999	-1.84	0.28	-3.58	1.00E-06	1.41E-03
G0S2	50486	-1.85	0.28	-3.61	2.22E-05	1.79E-02
KIR2DL4	3805	-2.06	0.24	-4.17	6.94E-08	1.88E-04
PRF1	5551	-2.13	0.23	-4.39	5.46E-07	9.39E-04
KIR2DL3	3804	-2.36	0.20	-5.12	3.59E-08	1.05E-04

C.3. IL12RB1 patient data set DE genes – IL18 treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
CHURC1	91612	1.47	2.77	2.77	1.10E-08	3.78E-05
ATP8B4	79895	1.36	2.57	2.57	1.37E-06	2.22E-03
PASK	23178	1.20	2.30	2.30	1.88E-05	1.47E-02
NT5C3L	115024	1.16	2.24	2.24	3.43E-13	1.06E-08
RPL23AP7	118433	1.13	2.19	2.19	2.66E-10	1.67E-06
IGLL1	3543	1.07	2.10	2.10	1.68E-06	2.44E-03
WARS2	10352	1.01	2.01	2.01	5.62E-13	1.06E-08
RNASE6	6039	0.78	1.72	1.72	1.84E-05	1.47E-02
ARL17P1	51326	0.72	1.65	1.65	6.05E-08	1.64E-04
PPARG	5468	0.70	1.62	1.62	1.01E-06	1.82E-03
EEF1D	1936	0.62	1.54	1.54	1.38E-11	1.14E-07
VAV3	10451	0.62	1.54	1.54	1.41E-06	2.22E-03
COL18A1	80781	0.60	1.52	1.52	3.30E-07	7.82E-04
TSHZ2	128553	0.58	1.50	1.50	2.08E-08	6.23E-05
NFKB1	4790	-0.58	0.67	-1.50	4.47E-05	2.82E-02
RSPO3	84870	-0.64	0.64	-1.56	1.67E-06	2.44E-03
PRR5L	79899	-0.65	0.64	-1.57	4.27E-05	2.79E-02
AOAH	313	-0.68	0.63	-1.60	5.20E-05	3.12E-02
IL18R1	8809	-0.71	0.61	-1.63	3.41E-06	4.47E-03
GPR114	221188	-0.76	0.59	-1.69	1.01E-08	3.78E-05
SLC15A4	121260	-0.77	0.58	-1.71	8.06E-05	4.12E-02
ATP13A1	57130	-0.83	0.56	-1.78	1.20E-12	1.52E-08
APOBEC3H	164668	-0.86	0.55	-1.82	2.48E-09	1.19E-05
KLRB1	3820	-0.88	0.54	-1.84	1.51E-11	1.14E-07
APOL3	80833	-0.93	0.52	-1.91	1.01E-06	1.82E-03
DRAM1	55332	-0.99	0.50	-1.99	2.15E-05	1.54E-02
GSTM1	2944	-1.04	0.49	-2.05	2.13E-05	1.54E-02
MAP3K8	1326	-1.30	0.41	-2.46	5.37E-06	6.16E-03
S1PR5	53637	-1.46	0.36	-2.74	1.25E-05	1.20E-02
FGFBP2	83888	-1.50	0.35	-2.82	1.73E-05	1.42E-02
KIR2DL1	3802	-1.58	0.33	-3.00	8.99E-09	3.78E-05
GZMH	2999	-1.65	0.32	-3.15	6.20E-06	6.71E-03
KIR2DL4	3805	-1.77	0.29	-3.40	1.31E-06	2.22E-03
ZNF683	257101	-1.77	0.29	-3.42	2.20E-05	1.54E-02
G0S2	50486	-1.90	0.27	-3.74	1.51E-05	1.33E-02
KIR2DL3	3804	-2.06	0.24	-4.16	5.34E-07	1.06E-03

C.4. IL12RB1 patient data set DE genes – IL12/IL18 treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
CHURC1	91612	1.57	2.97	2.97	2.46E-09	5.09E-06
C5AR1	728	1.34	2.52	2.52	6.37E-05	1.35E-02
RPL14	9045	1.31	2.48	2.48	1.26E-04	2.11E-02
NRIP3	56675	1.24	2.35	2.35	2.72E-05	7.79E-03
PASK	23178	1.23	2.34	2.34	1.32E-05	4.60E-03
CD79B	974	1.22	2.33	2.33	2.82E-04	3.73E-02
ATP8B4	79895	1.20	2.30	2.30	1.11E-05	3.99E-03
CHST7	56548	1.15	2.23	2.23	1.92E-05	6.05E-03
RPL23AP7	118433	1.13	2.19	2.19	2.74E-10	9.42E-07
VCAN	1462	1.13	2.19	2.19	1.13E-04	1.97E-02
WARS2	10352	1.08	2.12	2.12	7.78E-14	7.67E-10
NT5C3L	115024	1.06	2.09	2.09	3.77E-12	2.38E-08
IGLL1	3543	1.01	2.02	2.02	4.60E-06	2.26E-03
PPARG	5468	1.01	2.01	2.01	3.87E-10	1.13E-06
SPRY2	10253	0.97	1.96	1.96	1.16E-06	7.32E-04
OLIG2	10215	0.94	1.92	1.92	1.05E-05	3.89E-03
TM4SF19	116211	0.89	1.85	1.85	2.96E-04	3.84E-02
VAV3	10451	0.85	1.81	1.81	2.03E-09	4.51E-06
MTSS1	9788	0.84	1.79	1.79	5.98E-05	1.33E-02
MAPK13	5603	0.80	1.74	1.74	2.80E-05	7.79E-03
CD27	939	0.78	1.71	1.71	2.78E-05	7.79E-03
RNASE6	6039	0.76	1.70	1.70	2.68E-05	7.74E-03
SLC24A6	80024	0.74	1.67	1.67	2.32E-04	3.20E-02
ARL17P1	51326	0.73	1.66	1.66	3.95E-08	4.76E-05
GLT25D1	79709	0.73	1.66	1.66	1.31E-04	2.17E-02
SQLE	6713	0.72	1.65	1.65	2.27E-05	6.86E-03
HSPBAP1	79663	0.72	1.65	1.65	7.49E-05	1.51E-02
TSHZ2	128553	0.72	1.64	1.64	1.56E-10	6.54E-07
POLR1D	51082	0.70	1.63	1.63	3.25E-06	1.74E-03
NMB	4828	0.71	1.63	1.63	4.57E-06	2.26E-03
GLIPR1	11010	0.69	1.62	1.62	2.96E-06	1.67E-03
FOSL1	8061	0.68	1.60	1.60	7.53E-05	1.51E-02
SQSTM1	8878	0.68	1.60	1.60	2.97E-04	3.84E-02
KIAA1598	57698	0.66	1.59	1.59	5.27E-08	5.87E-05
HPCAL1	3241	0.65	1.57	1.57	1.81E-04	2.77E-02
COL18A1	80781	0.64	1.56	1.56	1.10E-07	1.12E-04
BCAS4	55653	0.64	1.55	1.55	1.11E-04	1.95E-02
EEF1G	1937	0.63	1.54	1.54	3.59E-06	1.84E-03
ARRDC5	645432	0.61	1.52	1.52	9.73E-07	6.46E-04
GEMIN4	50628	0.60	1.52	1.52	1.48E-05	5.04E-03
GALK1	2584	0.61	1.52	1.52	4.07E-05	1.06E-02
RPL13A	23521	0.60	1.51	1.51	4.29E-05	1.10E-02
MERTK	10461	0.59	1.51	1.51	1.59E-04	2.52E-02
NOP16	51491	0.60	1.51	1.51	1.66E-04	2.62E-02
DUSP10	11221	-0.58	0.67	-1.50	1.56E-07	1.37E-04
CFLAR	8837	-0.59	0.67	-1.50	5.99E-06	2.80E-03
LOC643031	643031	-0.60	0.66	-1.51	1.22E-04	2.07E-02
RBCK1	10616	-0.59	0.66	-1.51	1.67E-04	2.62E-02
NFKB1	4790	-0.61	0.66	-1.52	2.57E-05	7.60E-03
PIGV	55650	-0.60	0.66	-1.52	2.57E-05	7.60E-03
MACF1	23499	-0.61	0.65	-1.53	6.95E-08	7.52E-05
IGF2R	3482	-0.61	0.65	-1.53	1.93E-04	2.85E-02

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
SLC37A1	54020	-0.62	0.65	-1.54	3.21E-06	1.74E-03
ATP1B2	482	-0.63	0.65	-1.55	2.28E-05	6.86E-03
TANK	10010	-0.63	0.65	-1.55	4.88E-05	1.15E-02
PRR5L	79899	-0.63	0.65	-1.55	6.86E-05	1.41E-02
TWSG1	57045	-0.65	0.64	-1.57	1.88E-04	2.84E-02
SLC6A12	6539	-0.67	0.63	-1.59	8.63E-06	3.44E-03
NCF4	4689	-0.68	0.63	-1.60	1.50E-07	1.35E-04
IL15RA	3601	-0.68	0.63	-1.60	7.47E-06	3.18E-03
TSHZ3	57616	-0.68	0.63	-1.60	8.87E-06	3.46E-03
ASPHD2	57168	-0.67	0.63	-1.60	8.49E-05	1.61E-02
LOC80154	80154	-0.70	0.62	-1.63	4.03E-08	4.76E-05
SETBP1	26040	-0.71	0.61	-1.63	3.55E-06	1.84E-03
OPTN	10133	-0.70	0.61	-1.63	1.23E-05	4.35E-03
RARRES3	5920	-0.73	0.60	-1.65	2.25E-05	6.86E-03
NOD2	64127	-0.73	0.60	-1.66	2.84E-04	3.73E-02
HSPA5	3309	-0.74	0.60	-1.67	8.45E-06	3.44E-03
STAT2	6773	-0.74	0.60	-1.67	1.93E-04	2.85E-02
RALB	5899	-0.76	0.59	-1.69	3.40E-06	1.78E-03
CSF2RB	1439	-0.77	0.59	-1.70	5.29E-07	3.85E-04
LRRK2	120892	-0.77	0.59	-1.70	4.53E-05	1.12E-02
TNFRSF18	8784	-0.76	0.59	-1.70	1.93E-04	2.85E-02
TRAF1	7185	-0.77	0.59	-1.71	9.23E-06	3.53E-03
RNF24	11237	-0.78	0.58	-1.72	7.10E-05	1.44E-02
TNFAIP8	25816	-0.78	0.58	-1.72	2.80E-04	3.73E-02
TLR4	7099	-0.79	0.58	-1.73	2.64E-05	7.73E-03
LIMK2	3985	-0.80	0.57	-1.74	2.18E-04	3.08E-02
SP110	3431	-0.81	0.57	-1.75	8.47E-05	1.61E-02
IL2RB	3560	-0.83	0.56	-1.78	2.56E-09	5.09E-06
RBMS1	5937	-0.84	0.56	-1.79	2.93E-07	2.36E-04
TAP2	6891	-0.85	0.55	-1.80	2.77E-05	7.79E-03
P2RY14	9934	-0.85	0.56	-1.80	1.01E-04	1.82E-02
LMNB1	4001	-0.85	0.56	-1.80	2.88E-04	3.76E-02
ATG3	64422	-0.85	0.55	-1.81	5.79E-05	1.30E-02
OASL	8638	-0.87	0.55	-1.83	8.42E-05	1.61E-02
RPS23	6228	-0.88	0.54	-1.84	1.25E-04	2.10E-02
BATF	10538	-0.88	0.54	-1.84	2.27E-04	3.15E-02
CD97	976	-0.89	0.54	-1.85	1.95E-04	2.87E-02
LAMP3	27074	-0.90	0.53	-1.87	2.24E-04	3.13E-02
ATP13A1	57130	-0.91	0.53	-1.88	8.10E-14	7.67E-10
GNG2	54331	-0.92	0.53	-1.89	1.85E-04	2.82E-02
APOBEC3H	164668	-0.93	0.53	-1.90	4.12E-10	1.13E-06
MLKL	197259	-0.93	0.53	-1.90	8.72E-05	1.64E-02
GSTM1	2944	-0.93	0.53	-1.90	1.01E-04	1.82E-02
KLRB1	3820	-0.93	0.52	-1.91	2.98E-12	2.26E-08
CTSZ	1522	-0.93	0.52	-1.91	3.05E-06	1.70E-03
OAS2	4939	-0.93	0.52	-1.91	1.78E-04	2.74E-02
RNF144B	255488	-0.96	0.51	-1.95	5.23E-05	1.22E-02
TDRD7	23424	-0.97	0.51	-1.96	6.31E-05	1.34E-02
TAX1BP3	30851	-0.98	0.51	-1.97	8.27E-06	3.40E-03
TRIM22	10346	-1.00	0.50	-1.99	1.79E-07	1.54E-04
CEBPD	1052	-0.99	0.50	-1.99	1.01E-05	3.79E-03
SERPING1	710	-1.00	0.50	-2.00	1.01E-04	1.82E-02

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
BRSK1	84446	-1.00	0.50	-2.00	2.69E-04	3.64E-02
BCL2L1	598	-1.00	0.50	-2.00	3.84E-04	4.72E-02
OAS1	4938	-1.02	0.49	-2.03	6.77E-05	1.41E-02
IL15	3600	-1.04	0.49	-2.06	1.45E-04	2.34E-02
CCND2	894	-1.06	0.48	-2.08	4.31E-09	7.41E-06
ENPP2	5168	-1.08	0.47	-2.11	1.76E-05	5.73E-03
SOCS1	8651	-1.08	0.47	-2.11	5.85E-05	1.31E-02
TICAM2	353376	-1.09	0.47	-2.12	6.96E-06	3.05E-03
IL18R1	8809	-1.09	0.47	-2.13	4.60E-10	1.13E-06
KIAA1618	57714	-1.09	0.47	-2.13	4.78E-09	7.86E-06
TMEM140	55281	-1.12	0.46	-2.18	4.67E-05	1.13E-02
S1PR5	53637	-1.13	0.46	-2.18	3.85E-04	4.72E-02
NDFIP2	54602	-1.13	0.46	-2.19	8.75E-15	1.66E-10
GNG10	2790	-1.15	0.45	-2.21	8.56E-06	3.44E-03
LAG3	3902	-1.15	0.45	-2.21	1.26E-05	4.41E-03
APOBEC3F	200316	-1.15	0.45	-2.22	1.27E-06	7.89E-04
PLSCR1	5359	-1.15	0.45	-2.23	2.31E-04	3.20E-02
PTAFR	5724	-1.19	0.44	-2.28	5.43E-05	1.25E-02
IRF7	3665	-1.20	0.44	-2.29	3.16E-05	8.57E-03
NBN	4683	-1.20	0.44	-2.30	2.98E-04	3.84E-02
KIR3DL1	3811	-1.21	0.43	-2.31	8.89E-05	1.66E-02
FPR2	2358	-1.21	0.43	-2.31	2.82E-04	3.73E-02
PARP14	54625	-1.24	0.42	-2.36	2.03E-04	2.95E-02
FEZ1	9638	-1.26	0.42	-2.40	2.66E-05	7.73E-03
KIR2DS5	3810	-1.26	0.42	-2.40	2.12E-04	3.05E-02
STX11	8676	-1.27	0.41	-2.42	6.23E-05	1.33E-02
CMAH	8418	-1.29	0.41	-2.44	3.53E-09	6.68E-06
BHLHE40	8553	-1.33	0.40	-2.51	6.67E-06	3.05E-03
SAMD9L	219285	-1.33	0.40	-2.51	3.14E-05	8.57E-03
IFI35	3430	-1.37	0.39	-2.59	2.84E-06	1.63E-03
CDK6	1021	-1.38	0.38	-2.60	8.90E-09	1.30E-05
PRIC285	85441	-1.39	0.38	-2.62	4.75E-05	1.13E-02
GCH1	2643	-1.39	0.38	-2.63	4.01E-06	2.02E-03
XAF1	54739	-1.43	0.37	-2.70	1.46E-07	1.35E-04
DRAM1	55332	-1.44	0.37	-2.71	2.59E-08	3.50E-05
GADD45B	4616	-1.49	0.36	-2.80	2.70E-07	2.22E-04
JAK2	3717	-1.50	0.35	-2.82	2.24E-08	3.14E-05
ASCL2	430	-1.49	0.36	-2.82	2.82E-06	1.63E-03
APOL3	80833	-1.51	0.35	-2.85	2.33E-11	1.26E-07
RSPO3	84870	-1.57	0.34	-2.97	4.53E-16	1.72E-11
FGFBP2	83888	-1.57	0.34	-2.97	8.26E-06	3.40E-03
LTA	4049	-1.59	0.33	-3.02	3.77E-09	6.79E-06
RGS16	6004	-1.60	0.33	-3.02	6.83E-07	4.70E-04
KIR2DL1	3802	-1.60	0.33	-3.03	7.29E-09	1.10E-05
ZNF683	257101	-1.60	0.33	-3.03	9.30E-05	1.72E-02
MAP3K8	1326	-1.61	0.33	-3.05	1.15E-07	1.15E-04
GBP4	115361	-1.66	0.32	-3.16	8.43E-05	1.61E-02
VAMP5	10791	-1.69	0.31	-3.22	5.65E-09	8.91E-06
TNFAIP2	7127	-1.69	0.31	-3.22	4.58E-05	1.12E-02
IL18RAP	8807	-1.73	0.30	-3.31	1.07E-06	6.86E-04
GZMA	3001	-1.75	0.30	-3.36	6.99E-06	3.05E-03
G0S2	50486	-1.76	0.30	-3.39	4.67E-05	1.13E-02

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
GZMH	2999	-1.78	0.29	-3.44	1.73E-06	1.06E-03
TNFSF13B	10673	-1.79	0.29	-3.46	5.12E-06	2.42E-03
GK	2710	-1.80	0.29	-3.49	1.20E-05	4.30E-03
KIR2DL4	3805	-1.85	0.28	-3.61	5.62E-07	4.01E-04
TNFSF10	8743	-2.04	0.24	-4.11	2.77E-05	7.79E-03
RSAD2	91543	-2.10	0.23	-4.30	1.08E-04	1.92E-02
FES	2242	-2.11	0.23	-4.33	4.48E-10	1.13E-06
PRF1	5551	-2.16	0.22	-4.47	4.28E-07	3.31E-04
KIR2DL3	3804	-2.21	0.22	-4.62	1.37E-07	1.30E-04
GZMB	3002	-2.33	0.20	-5.04	2.75E-04	3.71E-02
KCNJ2	3759	-2.39	0.19	-5.24	2.24E-04	3.13E-02
DEFB1	1672	-2.44	0.18	-5.43	6.33E-07	4.44E-04
UBD	10537	-3.43	0.09	-10.80	1.18E-07	1.15E-04
IDO1	3620	-3.53	0.09	-11.50	3.42E-07	2.70E-04
CXCL9	4283	-4.17	0.06	-18.00	1.01E-04	1.82E-02
IFNG	3458	-4.68	0.04	-25.70	3.78E-08	4.76E-05

C.5. IL12RB1 patient data set DE genes – IFNG treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
CHURC1	91612	1.41	2.65	2.65	2.89E-08	9.95E-05
PASK	23178	1.19	2.28	2.28	2.04E-05	1.98E-02
NT5C3L	115024	0.98	1.97	1.97	3.43E-11	4.33E-07
RPL23AP7	118433	0.93	1.91	1.91	2.24E-08	8.48E-05
ARL17P1	51326	0.93	1.90	1.90	1.97E-10	1.24E-06
WARS2	10352	0.83	1.78	1.78	9.45E-11	7.56E-07
TSHZ2	128553	0.80	1.74	1.74	9.15E-12	1.73E-07
CARM1	10498	0.71	1.64	1.64	4.95E-07	1.25E-03
BOLA2	552900	0.59	1.51	1.51	6.67E-05	4.68E-02
COL18A1	80781	0.58	1.50	1.50	6.22E-07	1.39E-03
LOC643031	643031	-0.62	0.65	-1.54	7.04E-05	4.84E-02
NPTX1	4884	-0.66	0.63	-1.58	8.06E-08	2.54E-04
ADRB2	154	-0.67	0.63	-1.59	1.82E-05	1.90E-02
CLDND2	125875	-0.68	0.63	-1.60	7.85E-09	3.30E-05
APOBEC3H	164668	-0.72	0.61	-1.64	1.27E-07	3.71E-04
AOAH	313	-0.79	0.58	-1.73	4.90E-06	7.42E-03
GPR114	221188	-0.80	0.58	-1.74	3.36E-09	1.59E-05
KLRB1	3820	-0.82	0.57	-1.76	9.98E-11	7.56E-07
ATP13A1	57130	-0.89	0.54	-1.85	1.76E-13	6.68E-09
RPS23	6228	-0.97	0.51	-1.97	2.99E-05	2.64E-02
GSTM1	2944	-1.04	0.49	-2.06	2.03E-05	1.98E-02
GSTM2	2946	-1.27	0.41	-2.42	4.84E-05	3.74E-02
KIR2DL1	3802	-1.68	0.31	-3.19	2.55E-09	1.38E-05
GZMH	2999	-1.69	0.31	-3.23	4.27E-06	6.73E-03
FGFBP2	83888	-1.78	0.29	-3.43	1.00E-06	1.89E-03
KIR2DL4	3805	-1.87	0.27	-3.65	4.68E-07	1.25E-03
ZNF683	257101	-1.94	0.26	-3.83	5.51E-06	7.72E-03
KIR2DL3	3804	-2.00	0.25	-4.00	8.80E-07	1.84E-03

C.6. IL12RB1 patient data set DE genes – LPS treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
WARS2	10352	1.44	2.71	2.71	1.51E-17	5.71E-13
PASK	23178	1.42	2.67	2.67	1.23E-06	1.04E-03
CHURC1	91612	1.39	2.62	2.62	3.56E-08	8.42E-05
FASN	2194	1.05	2.08	2.08	1.43E-04	2.78E-02
IGLL1	3543	1.03	2.04	2.04	3.64E-06	2.29E-03
RPL23AP7	118433	0.98	1.98	1.98	7.07E-09	2.43E-05
NT5C3L	115024	0.97	1.95	1.95	4.66E-11	2.94E-07
WDR18	57418	0.94	1.92	1.92	2.39E-09	1.00E-05
GALR2	8811	0.92	1.89	1.89	4.73E-07	4.72E-04
DEAF1	10522	0.84	1.79	1.79	2.37E-07	2.89E-04
REG4	83998	0.82	1.76	1.76	2.21E-11	1.73E-07
MCM6	4175	0.81	1.76	1.76	1.24E-07	1.87E-04
CD7	924	0.79	1.73	1.73	6.94E-06	3.41E-03
ITGA1	3672	0.78	1.71	1.71	1.95E-06	1.48E-03
CD5	921	0.77	1.71	1.71	2.35E-04	3.77E-02
DCXR	51181	0.75	1.68	1.68	1.33E-04	2.60E-02
CES4	51716	0.73	1.66	1.66	2.94E-05	9.52E-03
TSHZ2	128553	0.72	1.64	1.64	1.53E-10	7.26E-07
FAAH2	158584	0.70	1.63	1.63	4.68E-08	9.34E-05
BOLA2	552900	0.71	1.63	1.63	4.71E-06	2.74E-03
COL18A1	80781	0.69	1.61	1.61	2.30E-08	6.22E-05
CDR2	1039	0.68	1.60	1.60	9.16E-05	2.08E-02
EZR	7430	0.67	1.59	1.59	1.62E-05	6.12E-03
IMP4	92856	0.62	1.54	1.54	1.37E-07	2.00E-04
RANGAP1	5905	0.62	1.54	1.54	5.67E-07	5.50E-04
PDCL3	79031	0.62	1.54	1.54	7.02E-05	1.70E-02
SOX8	30812	0.62	1.54	1.54	2.47E-04	3.92E-02
EEF1D	1936	0.61	1.53	1.53	1.86E-11	1.73E-07
EEF1G	1937	0.61	1.53	1.53	4.89E-06	2.74E-03
IGFBP4	3487	0.61	1.53	1.53	2.78E-04	4.19E-02
RPL23AP45	729764	0.61	1.53	1.53	3.41E-04	4.83E-02
ARL17P1	51326	0.58	1.50	1.50	3.00E-06	2.02E-03
MYB	4602	0.58	1.50	1.50	6.67E-06	3.34E-03
ADSL	158	0.58	1.50	1.50	1.42E-05	5.69E-03
CARM1	10498	0.58	1.50	1.50	1.49E-05	5.86E-03
GTPBP6	8225	0.59	1.50	1.50	8.15E-05	1.88E-02
OSGEP	55644	0.58	1.50	1.50	2.38E-04	3.80E-02
KIR2DS1	3806	-0.58	0.67	-1.50	2.30E-04	3.72E-02
KIAA1671	85379	-0.60	0.66	-1.51	1.23E-07	1.87E-04
CTSC	1075	-0.60	0.66	-1.52	1.72E-07	2.25E-04
FCER1G	2207	-0.61	0.66	-1.52	1.07E-04	2.23E-02
RASSF4	83937	-0.62	0.65	-1.53	5.82E-05	1.51E-02
KLRB1	3820	-0.62	0.65	-1.54	5.19E-08	9.35E-05
LOC644090	644090	-0.63	0.64	-1.55	4.99E-06	2.74E-03
PTGER2	5732	-0.64	0.64	-1.56	2.07E-04	3.55E-02
SERPINA1	5265	-0.65	0.64	-1.57	3.47E-04	4.89E-02
LOC643031	643031	-0.66	0.63	-1.58	2.77E-05	9.19E-03
APOBEC3H	164668	-0.67	0.63	-1.59	4.67E-07	4.72E-04
MS4A6A	64231	-0.67	0.63	-1.59	3.39E-05	1.04E-02
FAM186B	84070	-0.68	0.62	-1.60	2.67E-06	1.87E-03
CSTA	1475	-0.69	0.62	-1.61	3.32E-04	4.77E-02
AQP9	366	-0.70	0.62	-1.62	9.29E-05	2.09E-02

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
TSHZ3	57616	-0.72	0.61	-1.65	3.39E-06	2.25E-03
RAB32	10981	-0.72	0.61	-1.65	5.93E-05	1.52E-02
DAPK1	1612	-0.76	0.59	-1.70	8.89E-05	2.03E-02
TNFRSF1A	7132	-0.78	0.58	-1.71	3.59E-05	1.08E-02
C12orf75	387882	-0.81	0.57	-1.76	9.86E-05	2.14E-02
RPS23	6228	-0.87	0.55	-1.83	1.31E-04	2.60E-02
RNF165	494470	-0.88	0.54	-1.84	2.35E-05	8.32E-03
ATP13A1	57130	-0.89	0.54	-1.85	1.66E-13	3.15E-09
PPAP2B	8613	-0.89	0.54	-1.85	2.33E-06	1.70E-03
AMY1C	278	-0.89	0.54	-1.85	2.48E-04	3.92E-02
HMOX1	3162	-0.90	0.54	-1.87	2.11E-04	3.58E-02
DSE	29940	-0.93	0.53	-1.90	9.64E-07	8.40E-04
CST3	1471	-0.93	0.53	-1.90	2.46E-04	3.91E-02
TYROBP	7305	-0.97	0.51	-1.96	2.35E-04	3.77E-02
ZEB2	9839	-0.98	0.51	-1.97	6.31E-05	1.59E-02
SPI1	6688	-0.99	0.50	-1.98	2.29E-05	8.19E-03
AOAH	313	-1.03	0.49	-2.04	3.88E-08	8.64E-05
MAP3K4	4216	-1.08	0.47	-2.11	1.59E-07	2.15E-04
FCN1	2219	-1.12	0.46	-2.17	4.69E-08	9.34E-05
KIR3DL1	3811	-1.14	0.45	-2.20	1.89E-04	3.36E-02
GSTM1	2944	-1.18	0.44	-2.26	2.78E-06	1.91E-03
SIRPA	140885	-1.18	0.44	-2.27	2.13E-04	3.58E-02
FES	2242	-1.24	0.42	-2.36	1.60E-05	6.12E-03
S1PR5	53637	-1.24	0.42	-2.37	1.15E-04	2.37E-02
CEBPD	1052	-1.26	0.42	-2.39	1.52E-07	2.12E-04
GSTM2	2946	-1.34	0.39	-2.54	2.29E-05	8.19E-03
FPR2	2358	-1.37	0.39	-2.58	5.70E-05	1.49E-02
LILRB2	10288	-1.44	0.37	-2.72	7.18E-05	1.73E-02
LILRA3	11026	-1.51	0.35	-2.85	9.77E-07	8.40E-04
FGFBP2	83888	-1.54	0.34	-2.91	1.11E-05	4.68E-03
AKR1C3	8644	-1.54	0.34	-2.91	4.41E-05	1.25E-02
KIR2DL1	3802	-1.58	0.34	-2.98	9.93E-09	3.13E-05
SMPDL3A	10924	-1.72	0.30	-3.30	7.18E-11	3.88E-07
CSF1R	1436	-1.73	0.30	-3.32	9.52E-05	2.11E-02
KIR2DL4	3805	-1.82	0.28	-3.53	7.70E-07	7.28E-04
KIR2DL3	3804	-1.94	0.26	-3.84	1.50E-06	1.17E-03
GJB2	2706	-2.04	0.24	-4.11	2.28E-11	1.73E-07

C.7. IL12RB1 patient data set DE genes – IL12/LPS treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
PASK	23178	1.51	2.85	2.85	3.63E-07	3.81E-04
CHURC1	91612	1.34	2.54	2.54	7.34E-08	9.26E-05
WARS2	10352	1.14	2.20	2.20	1.69E-14	3.21E-10
FASN	2194	1.11	2.16	2.16	7.52E-05	2.26E-02
GALR2	8811	1.03	2.04	2.04	4.79E-08	6.97E-05
RPL23AP7	118433	0.93	1.90	1.90	2.57E-08	4.86E-05
WDR18	57418	0.92	1.89	1.89	3.77E-09	1.02E-05
NT5C3L	115024	0.91	1.88	1.88	1.87E-10	1.35E-06
NELL2	4753	0.90	1.87	1.87	1.39E-04	3.44E-02
CD5	921	0.88	1.84	1.84	4.06E-05	1.61E-02
IGLL1	3543	0.87	1.83	1.83	4.55E-05	1.74E-02
RPL23AP45	729764	0.86	1.82	1.82	2.40E-06	1.69E-03
REG4	83998	0.84	1.79	1.79	1.14E-11	1.44E-07
SESN3	143686	0.83	1.78	1.78	1.05E-04	2.78E-02
ITGA1	3672	0.81	1.75	1.75	9.97E-07	8.20E-04
DEAF1	10522	0.80	1.74	1.74	5.56E-07	5.18E-04
MCM6	4175	0.79	1.72	1.72	2.54E-07	2.74E-04
COL18A1	80781	0.76	1.69	1.69	2.60E-09	7.56E-06
CD7	924	0.75	1.68	1.68	1.77E-05	7.79E-03
SQSTM1	8878	0.72	1.64	1.64	1.56E-04	3.62E-02
TSHZ2	128553	0.69	1.61	1.61	4.37E-10	2.36E-06
IGFBP4	3487	0.69	1.61	1.61	6.95E-05	2.23E-02
KLF2	10365	0.68	1.60	1.60	6.42E-05	2.17E-02
SGSM2	9905	0.66	1.58	1.58	1.41E-04	3.46E-02
FAAH2	158584	0.65	1.57	1.57	2.34E-07	2.61E-04
MYB	4602	0.66	1.57	1.57	8.71E-07	7.32E-04
EZR	7430	0.65	1.57	1.57	2.56E-05	1.07E-02
EEF1G	1937	0.64	1.56	1.56	2.30E-06	1.64E-03
CES4	51716	0.65	1.56	1.56	1.59E-04	3.65E-02
BOLA2	552900	0.63	1.55	1.55	3.09E-05	1.26E-02
WDR46	9277	0.60	1.52	1.52	4.10E-05	1.61E-02
PDCL3	79031	0.60	1.52	1.52	1.04E-04	2.78E-02
RBKS	64080	0.59	1.50	1.50	7.51E-09	1.58E-05
ADSL	158	0.59	1.50	1.50	1.28E-05	6.03E-03
FCER1G	2207	-0.60	0.66	-1.51	1.25E-04	3.19E-02
CFD	1675	-0.62	0.65	-1.54	2.08E-05	8.93E-03
KIR2DS1	3806	-0.64	0.64	-1.56	7.19E-05	2.23E-02
PPAP2B	8613	-0.68	0.63	-1.60	1.34E-04	3.36E-02
KLRB1	3820	-0.73	0.60	-1.65	1.69E-09	5.80E-06
DSE	29940	-0.73	0.60	-1.66	4.51E-05	1.74E-02
FAM186B	84070	-0.75	0.59	-1.68	4.34E-07	4.33E-04
APOBEC3H	164668	-0.76	0.59	-1.70	3.53E-08	5.56E-05
CLEC12A	160364	-0.77	0.59	-1.70	6.75E-05	2.22E-02
RNF165	494470	-0.79	0.58	-1.73	1.04E-04	2.78E-02
CMAH	8418	-0.79	0.58	-1.74	2.75E-05	1.13E-02
IFI30	10437	-0.81	0.57	-1.76	1.81E-04	3.96E-02
C12orf75	387882	-0.84	0.56	-1.79	6.34E-05	2.16E-02
CISH	1154	-0.84	0.56	-1.79	7.09E-05	2.23E-02
TNFRSF1A	7132	-0.85	0.56	-1.80	9.72E-06	4.91E-03
FCGRT	2217	-0.86	0.55	-1.81	1.34E-04	3.36E-02
TSHZ3	57616	-0.92	0.53	-1.89	3.25E-08	5.56E-05
HMOX1	3162	-0.92	0.53	-1.90	1.53E-04	3.60E-02

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
KIR3DL3	115653	-0.94	0.52	-1.92	2.22E-04	4.56E-02
MAP3K4	4216	-0.95	0.52	-1.93	1.76E-06	1.31E-03
ATP13A1	57130	-0.97	0.51	-1.96	1.48E-14	3.21E-10
AOAH	313	-1.00	0.50	-2.00	7.12E-08	9.26E-05
AMY1C	278	-1.07	0.48	-2.10	2.18E-05	9.27E-03
SPI1	6688	-1.09	0.47	-2.13	4.93E-06	2.96E-03
GSTM1	2944	-1.10	0.47	-2.14	8.76E-06	4.48E-03
FCN1	2219	-1.14	0.45	-2.20	3.45E-08	5.56E-05
KIR3DL1	3811	-1.18	0.44	-2.27	1.16E-04	3.02E-02
GSTM2	2946	-1.25	0.42	-2.39	6.03E-05	2.09E-02
FPR2	2358	-1.27	0.42	-2.41	1.56E-04	3.62E-02
LILRB2	10288	-1.32	0.40	-2.50	2.25E-04	4.59E-02
CEBPD	1052	-1.34	0.39	-2.53	3.97E-08	6.00E-05
LILRA3	11026	-1.34	0.39	-2.54	7.16E-06	3.87E-03
FES	2242	-1.39	0.38	-2.61	2.74E-06	1.82E-03
GZMH	2999	-1.42	0.37	-2.68	6.00E-05	2.09E-02
GZMA	3001	-1.49	0.35	-2.82	7.14E-05	2.23E-02
SMPDL3A	10924	-1.53	0.35	-2.89	1.22E-09	5.14E-06
FGFBP2	83888	-1.54	0.34	-2.92	1.07E-05	5.28E-03
KIR2DL1	3802	-1.60	0.33	-3.04	6.98E-09	1.58E-05
AKR1C3	8644	-1.75	0.30	-3.37	6.25E-06	3.53E-03
KIR2DL4	3805	-1.85	0.28	-3.60	5.66E-07	5.18E-04
GJB2	2706	-1.88	0.27	-3.69	1.72E-10	1.35E-06
KIR2DL3	3804	-1.90	0.27	-3.74	2.14E-06	1.56E-03

C.8. IL12RB1 patient data set DE genes – IFNG /LPS treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
PASK	23178	1.55	2.93	2.93	2.27E-07	1.87E-04
CHURC1	91612	1.38	2.61	2.61	4.14E-08	5.85E-05
FASN	2194	1.32	2.50	2.50	5.61E-06	1.84E-03
WARS2	10352	1.23	2.35	2.35	1.70E-15	6.44E-11
RPL23AP7	118433	1.11	2.16	2.16	3.98E-10	1.51E-06
SH3TC1	54436	1.00	1.99	1.99	9.60E-05	1.29E-02
ATIC	471	0.95	1.93	1.93	4.42E-05	7.40E-03
NELL2	4753	0.95	1.93	1.93	7.22E-05	1.08E-02
NT5C3L	115024	0.94	1.92	1.92	8.53E-11	4.61E-07
LOC391811	391811	0.93	1.90	1.90	1.06E-04	1.38E-02
REG4	83998	0.91	1.87	1.87	1.44E-12	1.82E-08
CD5	921	0.90	1.86	1.86	3.15E-05	5.86E-03
WDR18	57418	0.88	1.84	1.84	9.97E-09	2.20E-05
MCM6	4175	0.86	1.82	1.82	3.70E-08	5.83E-05
ITGA1	3672	0.87	1.82	1.82	2.72E-07	2.14E-04
CARM1	10498	0.86	1.81	1.81	1.18E-08	2.35E-05
BOLA2	552900	0.85	1.81	1.81	1.70E-07	1.53E-04
TSHZ2	128553	0.84	1.79	1.79	2.33E-12	2.21E-08
PITPNC1	26207	0.84	1.79	1.79	3.42E-05	6.19E-03
ARL17P1	51326	0.81	1.75	1.75	5.29E-09	1.33E-05
PLEKHG2	64857	0.80	1.75	1.75	4.23E-04	3.47E-02
COL18A1	80781	0.78	1.72	1.72	1.41E-09	4.85E-06
DEAF1	10522	0.78	1.72	1.72	9.67E-07	5.01E-04
DCTPP1	79077	0.78	1.72	1.72	1.84E-04	2.01E-02
WDR6	11180	0.78	1.72	1.72	5.80E-04	4.20E-02
IGLL1	3543	0.77	1.71	1.71	2.19E-04	2.31E-02
GALR2	8811	0.76	1.70	1.70	1.10E-05	2.96E-03
MYB	4602	0.76	1.69	1.69	5.11E-08	6.90E-05
IMP4	92856	0.74	1.67	1.67	3.44E-09	9.30E-06
AKTIP	64400	0.74	1.67	1.67	9.95E-05	1.32E-02
DCXR	51181	0.73	1.66	1.66	1.83E-04	2.01E-02
SESN3	143686	0.73	1.66	1.66	4.59E-04	3.66E-02
PSAT1	29968	0.73	1.65	1.65	1.29E-04	1.56E-02
AARS	16	0.72	1.65	1.65	2.57E-04	2.56E-02
EEF1G	1937	0.70	1.63	1.63	4.94E-07	3.21E-04
CD7	924	0.71	1.63	1.63	3.83E-05	6.69E-03
SGSM2	9905	0.70	1.63	1.63	6.61E-05	1.01E-02
ADSL	158	0.69	1.61	1.61	9.52E-07	5.01E-04
EZR	7430	0.68	1.61	1.61	1.20E-05	3.13E-03
PDCL3	79031	0.69	1.61	1.61	1.56E-05	3.80E-03
RPP21	79897	0.69	1.61	1.61	1.78E-04	1.99E-02
USP24	23358	0.68	1.60	1.60	2.15E-07	1.85E-04
EXOSC7	23016	0.68	1.60	1.60	1.70E-04	1.93E-02
WDR46	9277	0.67	1.59	1.59	7.32E-06	2.22E-03
NHP2L1	4809	0.66	1.58	1.58	5.23E-07	3.30E-04
POLR1D	51082	0.66	1.58	1.58	8.32E-06	2.39E-03
RPUSD2	27079	0.65	1.57	1.57	8.89E-08	1.03E-04
SLC2A5	6518	0.65	1.57	1.57	4.93E-04	3.85E-02
IMP3	55272	0.64	1.56	1.56	2.14E-06	9.12E-04
RPL23AP45	729764	0.64	1.56	1.56	2.03E-04	2.16E-02
NDUFB8	4714	0.63	1.55	1.55	1.68E-05	4.01E-03
RPP40	10799	0.63	1.55	1.55	5.06E-04	3.91E-02

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
MMS19	64210	0.61	1.53	1.53	2.91E-05	5.62E-03
FAAH2	158584	0.60	1.52	1.52	9.79E-07	5.01E-04
LOC729008	729008	0.60	1.52	1.52	5.76E-05	9.20E-03
GPX7	2882	0.60	1.52	1.52	6.14E-05	9.60E-03
CMTM7	112616	0.61	1.52	1.52	5.38E-04	4.04E-02
MRPL20	55052	0.61	1.52	1.52	5.80E-04	4.20E-02
B4GALT3	8703	0.59	1.51	1.51	1.36E-05	3.38E-03
GEMIN4	50628	0.59	1.51	1.51	2.03E-05	4.55E-03
STOML2	30968	0.60	1.51	1.51	7.35E-04	4.77E-02
RPSAP12	387867	0.59	1.51	1.51	7.80E-04	4.95E-02
BEX2	84707	0.58	1.50	1.50	9.99E-08	1.11E-04
RANGAP1	5905	0.58	1.50	1.50	1.80E-06	7.82E-04
DDX24	57062	0.58	1.50	1.50	3.17E-06	1.22E-03
TXNDC12	51060	0.58	1.50	1.50	1.20E-04	1.50E-02
NAT9	26151	0.58	1.50	1.50	1.28E-04	1.55E-02
TPRG1L	127262	0.58	1.50	1.50	1.50E-04	1.74E-02
KLRB1	3820	-0.58	0.67	-1.50	1.84E-07	1.62E-04
CUX1	1523	-0.58	0.67	-1.50	8.33E-07	4.57E-04
FAM186B	84070	-0.58	0.67	-1.50	2.96E-05	5.65E-03
PVRL2	5819	-0.59	0.67	-1.50	6.28E-04	4.34E-02
RTN4	57142	-0.60	0.66	-1.51	1.49E-06	6.96E-04
EFHD2	79180	-0.59	0.66	-1.51	2.15E-05	4.75E-03
MTMR11	10903	-0.60	0.66	-1.51	4.85E-04	3.80E-02
IL15RA	3601	-0.60	0.66	-1.52	4.09E-05	7.06E-03
SBK1	388228	-0.61	0.65	-1.53	1.10E-06	5.51E-04
IFNGR2	3460	-0.61	0.65	-1.53	4.48E-04	3.61E-02
CCND3	896	-0.62	0.65	-1.54	2.89E-06	1.13E-03
KIAA1671	85379	-0.63	0.65	-1.55	4.17E-08	5.85E-05
TICAM1	148022	-0.63	0.65	-1.55	1.98E-05	4.54E-03
MARCKS	4082	-0.63	0.64	-1.55	9.92E-05	1.32E-02
TNIP1	10318	-0.64	0.64	-1.55	2.60E-04	2.57E-02
RAB21	23011	-0.63	0.65	-1.55	3.09E-04	2.83E-02
PPAP2B	8613	-0.63	0.65	-1.55	3.23E-04	2.94E-02
SLC43A3	29015	-0.64	0.64	-1.56	1.97E-05	4.54E-03
FHL3	2275	-0.64	0.64	-1.56	1.07E-04	1.38E-02
CTSZ	1522	-0.64	0.64	-1.56	5.15E-04	3.95E-02
MS4A6A	64231	-0.65	0.64	-1.57	5.20E-05	8.42E-03
TNFRSF1A	7132	-0.65	0.64	-1.57	3.76E-04	3.26E-02
OPTN	10133	-0.66	0.63	-1.58	3.06E-05	5.81E-03
LILRB4	11006	-0.66	0.63	-1.58	2.79E-04	2.68E-02
PTGER2	5732	-0.67	0.63	-1.59	1.23E-04	1.53E-02
ABCA1	19	-0.67	0.63	-1.59	5.68E-04	4.13E-02
APOBEC3H	164668	-0.69	0.62	-1.61	2.88E-07	2.18E-04
FCER1G	2207	-0.69	0.62	-1.61	1.79E-05	4.20E-03
RAB32	10981	-0.69	0.62	-1.61	1.17E-04	1.48E-02
MPP1	4354	-0.69	0.62	-1.61	3.02E-04	2.80E-02
CARD17	440068	-0.72	0.61	-1.64	1.52E-08	2.74E-05
TSHZ3	57616	-0.71	0.61	-1.64	3.80E-06	1.38E-03
FCGR2A	2212	-0.71	0.61	-1.64	6.34E-05	9.83E-03
SLC15A3	51296	-0.71	0.61	-1.64	7.99E-05	1.14E-02
SDCBP	6386	-0.72	0.61	-1.65	7.56E-04	4.90E-02
CSF3	1440	-0.73	0.60	-1.66	3.66E-05	6.48E-03

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
PICALM	8301	-0.74	0.60	-1.67	4.15E-05	7.14E-03
FXYD6	53826	-0.75	0.59	-1.68	6.11E-04	4.28E-02
CMAH	8418	-0.78	0.58	-1.72	3.34E-05	6.07E-03
CLEC12A	160364	-0.78	0.58	-1.72	4.76E-05	7.86E-03
CISH	1154	-0.78	0.58	-1.72	1.98E-04	2.14E-02
PSAP	5660	-0.78	0.58	-1.72	6.03E-04	4.27E-02
IRAK2	3656	-0.79	0.58	-1.73	4.66E-05	7.74E-03
SLC2A6	11182	-0.81	0.57	-1.75	2.54E-04	2.56E-02
FGR	2268	-0.82	0.57	-1.76	1.64E-04	1.88E-02
ATP13A1	57130	-0.83	0.56	-1.77	1.40E-12	1.82E-08
CFD	1675	-0.82	0.57	-1.77	1.59E-07	1.47E-04
CAMK1G	57172	-0.82	0.57	-1.77	4.53E-04	3.63E-02
C12orf75	387882	-0.84	0.56	-1.78	6.77E-05	1.03E-02
FCGRT	2217	-0.84	0.56	-1.79	1.83E-04	2.01E-02
GNG2	54331	-0.84	0.56	-1.79	5.44E-04	4.04E-02
DSE	29940	-0.85	0.56	-1.80	4.76E-06	1.65E-03
RNF165	494470	-0.85	0.56	-1.80	4.29E-05	7.24E-03
ITGB8	3696	-0.85	0.55	-1.80	5.56E-05	8.92E-03
KIR3DL3	115653	-0.86	0.55	-1.82	6.35E-04	4.36E-02
ANO6	196527	-0.88	0.54	-1.84	7.26E-05	1.08E-02
TYROBP	7305	-0.88	0.54	-1.84	6.86E-04	4.57E-02
SMPDL3A	10924	-0.90	0.53	-1.87	2.82E-05	5.62E-03
AQP9	366	-0.91	0.53	-1.88	1.58E-06	7.28E-04
DRAM1	55332	-0.93	0.53	-1.90	5.35E-05	8.62E-03
KLF4	9314	-0.94	0.52	-1.91	6.55E-04	4.43E-02
EMILIN2	84034	-0.94	0.52	-1.92	3.97E-07	2.69E-04
STARD8	9754	-0.94	0.52	-1.92	5.78E-06	1.88E-03
IVNS1ABP	10625	-0.94	0.52	-1.92	3.00E-04	2.80E-02
CST3	1471	-0.95	0.52	-1.94	1.73E-04	1.96E-02
CPVL	54504	-0.97	0.51	-1.96	6.82E-05	1.04E-02
EXT1	2131	-0.97	0.51	-1.96	1.24E-04	1.53E-02
ATF5	22809	-1.04	0.49	-2.06	6.25E-04	4.34E-02
KIR3DL1	3811	-1.07	0.47	-2.11	3.68E-04	3.23E-02
SERPINA1	5265	-1.09	0.47	-2.13	1.06E-07	1.15E-04
AOAH	313	-1.11	0.46	-2.16	7.01E-09	1.66E-05
PRDM8	56978	-1.11	0.46	-2.16	1.19E-05	3.13E-03
EPB41L3	23136	-1.11	0.46	-2.16	5.66E-04	4.12E-02
MAP3K4	4216	-1.12	0.46	-2.18	6.89E-08	8.70E-05
SPI1	6688	-1.13	0.46	-2.18	2.89E-06	1.13E-03
MAFB	9935	-1.13	0.46	-2.18	4.73E-06	1.65E-03
CEBPD	1052	-1.14	0.45	-2.20	9.67E-07	5.01E-04
ASCL2	430	-1.14	0.45	-2.20	1.57E-04	1.82E-02
TLR4	7099	-1.16	0.45	-2.23	2.89E-08	4.97E-05
S1PR5	53637	-1.20	0.43	-2.30	1.81E-04	2.00E-02
GSTM1	2944	-1.21	0.43	-2.32	1.76E-06	7.74E-04
ADAP1	11033	-1.22	0.43	-2.33	3.78E-04	3.26E-02
FPR2	2358	-1.22	0.43	-2.34	2.37E-04	2.46E-02
AMY1C	278	-1.25	0.42	-2.39	1.64E-06	7.37E-04
LILRB2	10288	-1.26	0.42	-2.40	3.72E-04	3.24E-02
TREM1	54210	-1.27	0.42	-2.41	3.48E-05	6.25E-03
MAP3K8	1326	-1.28	0.41	-2.44	6.60E-06	2.07E-03
FAM148B	388125	-1.30	0.41	-2.47	8.98E-06	2.54E-03

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
GSTM2	2946	-1.32	0.40	-2.50	2.95E-05	5.65E-03
SIRPA	140885	-1.33	0.40	-2.52	4.64E-05	7.73E-03
LILRA3	11026	-1.35	0.39	-2.54	6.98E-06	2.17E-03
FES	2242	-1.39	0.38	-2.61	2.73E-06	1.10E-03
GZMH	2999	-1.47	0.36	-2.76	3.83E-05	6.69E-03
FCN1	2219	-1.48	0.36	-2.78	8.21E-11	4.61E-07
AKR1C3	8644	-1.48	0.36	-2.78	7.90E-05	1.14E-02
FGFBP2	83888	-1.56	0.34	-2.94	9.45E-06	2.65E-03
KIR2DL1	3802	-1.57	0.34	-2.98	1.04E-08	2.20E-05
RIN2	54453	-1.58	0.33	-2.99	6.46E-04	4.40E-02
CFB	629	-1.63	0.32	-3.10	3.12E-05	5.86E-03
G0S2	50486	-1.71	0.31	-3.27	7.00E-05	1.06E-02
KIR2DL3	3804	-1.78	0.29	-3.43	6.55E-06	2.07E-03
KIR2DL4	3805	-1.84	0.28	-3.59	5.94E-07	3.57E-04
GJB2	2706	-2.20	0.22	-4.61	2.94E-12	2.22E-08

C.9. IFNGR1 patient data set DE genes – media alone treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
VAV3	10451	0.54	1.46	1.46	6.19E-10	7.55E-06
KIR3DL1	3811	-0.27	0.83	-1.20	1.04E-07	7.28E-04
KCNJ2	3759	-0.26	0.83	-1.20	3.21E-07	1.74E-03
KIR2DL1	3802	-0.27	0.83	-1.21	5.93E-08	4.82E-04
GBP5	115362	-0.27	0.83	-1.21	2.55E-07	1.56E-03
KIR2DL4	3805	-0.27	0.83	-1.21	4.42E-06	1.08E-02
KIR2DL3	3804	-0.27	0.83	-1.21	7.67E-06	1.63E-02
FCGR1B	2210	-0.30	0.81	-1.23	2.67E-06	8.34E-03
GBP1	2633	-0.36	0.78	-1.28	7.99E-09	7.80E-05
ANKRD22	118932	-0.44	0.74	-1.35	4.63E-11	1.13E-06
CXCL10	3627	-0.73	0.60	-1.66	4.01E-12	1.96E-07

C.10. IFNGR1 patient data set DE genes – IL12 treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
VAV3	10451	0.51	1.43	1.43	2.56E-09	1.13E-05
D4S234E	27065	0.28	1.22	1.22	8.58E-06	9.30E-03
CD274	29126	-0.26	0.84	-1.20	5.63E-08	1.61E-04
KIR2DL3	3804	-0.26	0.84	-1.20	1.53E-05	1.52E-02
KIR2DL1	3802	-0.27	0.83	-1.21	6.90E-08	1.77E-04
KCNJ2	3759	-0.28	0.82	-1.21	8.42E-08	2.05E-04
LOC728744	728744	-0.28	0.82	-1.21	2.87E-07	5.85E-04
KIR2DL4	3805	-0.27	0.83	-1.21	4.40E-06	5.24E-03
P2RX7	5027	-0.28	0.82	-1.21	2.29E-05	2.11E-02
RIPK2	8767	-0.30	0.81	-1.23	2.00E-08	6.97E-05
WARS	7453	-0.32	0.80	-1.24	8.12E-10	4.80E-06
CXCL9	4283	-0.43	0.74	-1.34	2.42E-06	3.40E-03
GBP5	115362	-0.47	0.72	-1.38	1.16E-12	1.14E-08
FCGR1A	2209	-0.47	0.72	-1.38	7.03E-12	4.90E-08
GBP1	2633	-0.50	0.71	-1.41	3.45E-12	2.80E-08
IRG1	730249	-0.54	0.69	-1.45	5.16E-07	8.99E-04
FCGR1B	2210	-0.60	0.66	-1.51	6.94E-13	8.47E-09
ANKRD22	118932	-0.69	0.62	-1.61	2.21E-16	1.08E-11
CXCL10	3627	-0.82	0.57	-1.76	2.04E-13	3.32E-09

C.11. IFNGR1 patient data set DE genes – IL18 treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
VAV3	10451	0.52	1.43	1.43	2.18E-09	3.43E-06
CES1	1066	0.34	1.27	1.27	1.07E-04	3.73E-02
OLIG2	10215	0.29	1.22	1.22	4.31E-06	2.36E-03
TAP1	6890	-0.26	0.84	-1.20	3.13E-10	7.27E-07
TNF	7124	-0.27	0.83	-1.20	6.75E-10	1.37E-06
RAB20	55647	-0.26	0.83	-1.20	2.10E-07	1.71E-04
DUSP5	1847	-0.26	0.84	-1.20	2.20E-07	1.76E-04
RHOBTB3	22836	-0.26	0.83	-1.20	2.90E-07	2.18E-04
GK	2710	-0.26	0.84	-1.20	3.87E-07	2.82E-04
RSPO3	84870	-0.26	0.84	-1.20	8.93E-07	6.05E-04
IFIT3	3437	-0.26	0.83	-1.20	5.56E-06	2.95E-03
LOC728744	728744	-0.27	0.83	-1.21	4.68E-07	3.36E-04
FBXO6	26270	-0.27	0.83	-1.21	3.84E-05	1.59E-02
STEAP4	79689	-0.28	0.82	-1.21	9.86E-05	3.46E-02
PSTPIP2	9050	-0.29	0.82	-1.22	9.45E-08	8.71E-05
GCH1	2643	-0.28	0.82	-1.22	3.46E-05	1.47E-02
EPSTI1	94240	-0.29	0.82	-1.23	4.32E-09	6.20E-06
LRRK2	120892	-0.30	0.81	-1.23	5.52E-09	7.69E-06
KIR2DS5	3810	-0.30	0.81	-1.23	1.57E-07	1.33E-04
SNX10	29887	-0.30	0.81	-1.23	3.04E-07	2.25E-04
MYOF	26509	-0.30	0.81	-1.23	1.67E-05	7.97E-03
BATF2	116071	-0.30	0.81	-1.23	7.91E-05	2.96E-02
STAT1	6772	-0.31	0.81	-1.24	2.12E-10	5.18E-07
KIR3DL1	3811	-0.30	0.81	-1.24	6.52E-09	8.16E-06
EDN1	1906	-0.32	0.80	-1.25	3.11E-05	1.37E-02
GBP4	115361	-0.34	0.79	-1.26	7.50E-10	1.41E-06
KIR2DL1	3802	-0.33	0.79	-1.26	8.70E-10	1.52E-06
APOL6	80830	-0.35	0.79	-1.27	1.65E-06	1.07E-03
C11orf75	56935	-0.35	0.79	-1.27	6.84E-05	2.61E-02
TNFAIP2	7127	-0.36	0.78	-1.28	4.21E-10	9.34E-07
RIPK2	8767	-0.36	0.78	-1.28	4.57E-10	9.70E-07
NIACR1	338442	-0.35	0.78	-1.28	6.52E-09	8.16E-06
KIR2DL3	3804	-0.36	0.78	-1.28	4.39E-08	4.66E-05
C1QB	713	-0.35	0.78	-1.28	6.00E-06	3.15E-03
KIR2DL4	3805	-0.36	0.78	-1.29	1.83E-08	2.23E-05
KCNJ2	3759	-0.38	0.77	-1.30	1.05E-10	2.86E-07
P2RX7	5027	-0.38	0.77	-1.30	1.28E-07	1.14E-04
WARS	7453	-0.39	0.76	-1.31	4.18E-12	1.57E-08
LAP3	51056	-0.40	0.76	-1.32	1.77E-11	5.76E-08
CD274	29126	-0.41	0.75	-1.33	1.17E-12	7.11E-09
SLAMF8	56833	-0.45	0.73	-1.36	5.71E-09	7.74E-06
LOC400759	400759	-0.50	0.71	-1.42	5.84E-07	4.07E-04
FCGR1A	2209	-0.54	0.69	-1.46	1.58E-13	1.10E-09
GBP1	2633	-0.60	0.66	-1.52	2.12E-14	1.72E-10
GBP5	115362	-0.62	0.65	-1.53	5.36E-16	8.71E-12
IRG1	730249	-0.62	0.65	-1.53	3.87E-08	4.20E-05
IDO1	3620	-0.61	0.65	-1.53	6.48E-08	6.46E-05
FCGR1B	2210	-0.70	0.62	-1.62	9.17E-15	1.12E-10
UBD	10537	-0.71	0.61	-1.63	1.14E-09	1.85E-06
CCL8	6355	-0.73	0.60	-1.65	1.97E-05	8.97E-03
CXCL9	4283	-0.81	0.57	-1.75	2.42E-12	1.26E-08
ANKRD22	118932	-0.87	0.55	-1.82	2.17E-19	1.06E-14
CXCL10	3627	-0.90	0.54	-1.87	1.42E-14	1.39E-10

C.12. IFNGR1 patient data set DE genes – IL12/IL18 treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
VAV3	10451	0.41	1.33	1.33	1.90E-06	8.61E-03
FGFBP2	83888	-0.26	0.83	-1.20	2.04E-05	4.98E-02
KIR3DL1	3811	-0.28	0.83	-1.21	4.23E-07	3.98E-03
RSPO3	84870	-0.28	0.83	-1.21	1.94E-06	8.61E-03
KIR2DS5	3810	-0.27	0.83	-1.21	5.69E-06	1.98E-02
KIR2DL1	3802	-0.35	0.79	-1.27	4.17E-09	1.17E-04
KIR2DL4	3805	-0.34	0.79	-1.27	5.51E-07	3.98E-03
KIR2DL3	3804	-0.34	0.79	-1.27	1.06E-06	6.49E-03
RSAD2	91543	-0.35	0.78	-1.28	1.27E-06	6.88E-03

C.13. IFNGR1 patient data set DE genes – IFNG treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
VAV3	10451	0.46	1.38	1.38	2.30E-08	2.80E-04
VCAN	1462	0.41	1.33	1.33	7.10E-08	6.93E-04
FPR3	2359	0.31	1.24	1.24	1.04E-09	2.53E-05
RSAD2	91543	-0.26	0.83	-1.20	2.65E-05	3.81E-02
KIR2DL4	3805	-0.27	0.83	-1.21	4.08E-06	9.96E-03
KIR2DL3	3804	-0.27	0.83	-1.21	6.41E-06	1.36E-02
CCNA1	8900	-0.31	0.81	-1.24	7.49E-07	2.81E-03
IFNG	3458	-0.44	0.74	-1.36	2.62E-05	3.81E-02
RSPO3	84870	-0.45	0.73	-1.37	3.73E-12	1.82E-07

C.14. IFNGR1 patient data set DE genes – LPS treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
VAV3	10451	0.42	1.34	1.34	1.66E-07	8.11E-04
CES1	1066	0.40	1.32	1.32	1.33E-05	3.25E-02
KIR3DL1	3811	-0.31	0.81	-1.24	3.89E-09	4.75E-05
KIR2DS5	3810	-0.31	0.81	-1.24	5.93E-08	4.14E-04
KIR2DL3	3804	-0.33	0.80	-1.25	3.41E-07	1.51E-03
KIR2DL1	3802	-0.33	0.79	-1.26	8.72E-10	1.50E-05
KIR2DL4	3805	-0.35	0.79	-1.27	5.01E-08	4.08E-04

C.15. IFNGR1 patient data set DE genes – IL12/LPS treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
CES1	1066	0.40	1.32	1.32	1.23E-05	2.73E-02
VAV3	10451	0.39	1.31	1.31	6.62E-07	2.29E-03
CES4	51716	0.27	1.21	1.21	4.69E-08	2.54E-04
FCGR1B	2210	-0.29	0.82	-1.23	3.90E-06	1.12E-02
KIR2DS5	3810	-0.34	0.79	-1.26	1.20E-08	7.33E-05
KIR2DL4	3805	-0.33	0.79	-1.26	1.15E-07	5.11E-04
KIR2DL3	3804	-0.33	0.80	-1.26	2.91E-07	1.18E-03
KIR3DL1	3811	-0.37	0.77	-1.29	7.34E-11	1.19E-06
KIR2DL1	3802	-0.38	0.77	-1.30	5.28E-11	1.19E-06
ANKRD22	118932	-0.42	0.75	-1.34	1.30E-10	1.27E-06
CXCL9	4283	-0.46	0.73	-1.37	7.02E-07	2.29E-03

C.16. IFNGR1 patient data set DE genes – IFNG /LPS treatment

Gene	Entrez Gene ID	logFC (log2)	Fold Change	Fold Change (transformed)	p-value	p-value (adjusted)
VAV3	10451	0.48	1.40	1.40	1.11E-04	5.04E-01
CES1	1066	0.42	1.34	1.34	3.84E-03	9.91E-01
GJB2	2706	0.40	1.32	1.32	9.94E-04	7.35E-01
HAMP	57817	0.38	1.30	1.30	1.98E-02	1.00E+00
CSF2	1437	0.30	1.24	1.24	1.01E-02	1.00E+00
TOMM5	401505	0.29	1.22	1.22	6.42E-03	1.00E+00
CSNK1G2	1455	0.26	1.20	1.20	4.34E-02	1.00E+00
KIR3DL2	3812	-0.27	0.83	-1.20	1.00E-04	5.04E-01
MYOM2	9172	-0.27	0.83	-1.20	9.04E-03	1.00E+00
C1QB	713	-0.28	0.82	-1.21	1.99E-02	1.00E+00
KIR2DS5	3810	-0.29	0.82	-1.22	8.65E-04	7.03E-01
GZMH	2999	-0.28	0.82	-1.22	1.53E-02	1.00E+00
KIR2DL1	3802	-0.30	0.81	-1.23	1.11E-04	5.04E-01
MMP7	4316	-0.30	0.81	-1.23	1.25E-02	1.00E+00
PLA2G16	11145	-0.30	0.81	-1.23	2.58E-02	1.00E+00
KIR3DL1	3811	-0.31	0.81	-1.24	7.87E-05	5.04E-01
CADM1	23705	-0.32	0.80	-1.24	1.40E-02	1.00E+00
CXCL9	4283	-0.32	0.80	-1.24	2.11E-02	1.00E+00
KRT86	3892	-0.31	0.81	-1.24	3.29E-02	1.00E+00
FGFBP2	83888	-0.33	0.80	-1.25	3.24E-04	6.68E-01
KRT17	3872	-0.35	0.79	-1.27	4.33E-05	5.04E-01
FCGR1A	2209	-0.34	0.79	-1.27	1.45E-04	5.04E-01
FCGR1B	2210	-0.35	0.78	-1.28	5.35E-04	6.68E-01
UBD	10537	-0.35	0.78	-1.28	2.24E-02	1.00E+00
NPTX2	4885	-0.39	0.76	-1.31	2.88E-02	1.00E+00
KLRC3	3823	-0.46	0.73	-1.38	1.32E-02	1.00E+00

C.17. Transcription factor targets in the IL12RB1 enriched data sets

Media	IL12	IL18	IL12/IL18	IFNg	LPS	IL12/LPS	IFNg/LPS
RUNX3	SP1	CREB1	SP1	none sig.	SP1	SP1	SP1
ELF1	STAT1	STAT1	STAT1		SPI1	SPI1	SPI1
ETS1	SP3	ATF1	IRF1		ELF1	IRF1	CREB1
EGR1	CREB1	IRF1	SP3		GATA1	USF1	IRF1
RBPJ	STAT3	RELA	STAT3		IRF1	TFAP2A	CEBPB
CEBPD	STAT5A	GABPA	STAT5A		ETS1	CREB1	JUN
SP4	IRF1	SPI1	RELA		CEBPB	JUN	RELA
SMAD4	EGR1	USF1	STAT5B		NFYB	ELF1	STAT1
SMAD3	RELA	RUNX3	IRF2		TFAP2A	USF2	ELF1
MECOM	JUN	ELF1	SPI1		CREB1	GATA1	IRF2
DNM3	ATF1	IRF2	HIF1A		GATA2	HNF4A	STAT3
SPIB	ATF2	ATF2	ATF2		YY1	CEBPB	ATF2
MTA1	ETS1	BCL6	TP53		NR1H3	GATA2	ETS1
TOPORS	USF1	TFAP2B	CREB1		GATA3	GABPA	FOS
PARP1	TFAP2A	CEBPE	USF1		ETS2	ATF2	HNF4A
HNF4G	RUNX3	RBPJ	CEBPB		IRF2	YY1	TFAP2A
KLF13	STAT4	IRF7	STAT4		GABPA	ETS1	GATA2
FOXO1	STAT5B	CEBPD	GATA3		USF2	EGR1	YY1
BCL6	FOXO3	IRF3	ELF1		GFI1	RELA	ETV4
YBX1	IRF2	GABPB1	GATA2		IRF8	FOS	GABPA
CEBPE	HIF1A	GATA3	ATF1		RUNX3	GABPB1	USF2
ELF3	ESR1	GATA2	ARNT		YBX1	GATA3	GATA1
E2F4	YY1	ETS2	ETS1		TFAP2B	ETS2	HOXA10
GLI1	HNF1A	MECOM	EGR1		TFAP2C	IRF2	RBPJ
IRF7	HNF4A	SPIB	RUNX3		NR1H2	HNF4G	PBX1
RARG	BCL6	NCOA3	RBPJ		HOXA10	RUNX3	NFATC1
IRF3	TFAP2C	TOPORS	IRF4		JAM2	TFAP2B	GABPB1
STAT4	NR1H2	PARP1	IRF7		RBPJ	TFAP2C	STAT5B
	RBPJ	EP300	IRF3		IRF4	HOXA10	ZNF148
	NFATC1	HNF4G	STAT6		PBX1	ATF3	NFE2L2
	SP4	KLF13	GABPA		CEBPD	RBPJ	NR3C1
	NR1H3	SATB1	JUND		ELK1	AR	GATA3
	RUNX2	C20orf166	E2F1		GABPB1	PBX1	ETS2
	SMAD4	CUX1	IRF8		ZNF148	CEBPD	HNF4G
	ELF1	IRF8	BCL6		NFE2L2	NFATC1	IRF8
	ETS2	NCOA1	HOXA10		STAT6	NR1H3	RUNX3
	SMAD3	HES1	MAFG		DDIT3	ELK1	TFAP2B
	GABPA	LEF1	PBX1		ETV1	SREBF2	TFAP2C
	JUND	TFAP2C	TP63		MECOM	ZNF148	KLF4
	SPI1	CREM	NFATC1		DNM3	SMAD4	JAM2
	E2F1	HOXA10	REL		HMGB2	DDIT3	IRF4
	NFAT5	MAFG	CTCF		HOXB7	EGR3	IRF7
	EGR3	PRDM1	GABPB1		INSM1	ETV1	NR5A1
	MECOM	SP2	NFE2L2		MEIS2	ATF6	RORA
	ATF6	TCF7L2	FOXO3		PFDN5	DNM3	IRF3
	FOXM1	GLI1	ZBTB33		SPIB	FOXN2	STAT4
	LOC729991-N	IRF4	DDB1		TGIF1	FOXK2	REL
	SPIB	PBX1	DDB2		XPA	PAX3	NR1H3
	NCOA3	RARG	EGR3		PRDM2	POU6F1	ELK1

C.18. Transcription factor targets in the IFNGR1 enriched data sets

Media	IL12	IL18	IL12/IL18	IFNg	LPS	IL12/LPS	IFNg/LPS
STAT1	STAT1	STAT1	STAT1	STAT3	STAT1	STAT1	none sig.
RELA	STAT3	IRF1	RXRA	ETS1	STAT3	STAT3	
CEBPB	IRF1	STAT3	CEBPB	STAT1	STAT5A	STAT5A	
STAT5A	RELA	CEBPB	STAT3	STAT4	RUNX3	HNF4G	
STAT3	CEBPB	RELA	FOS	STAT5A			
	IRF2	STAT5A	NFYA	STAT5B			
	NFYB	STAT5B	CREB1	GABPA			
	NFYA	IRF2	STAT6	IRF1			
	STAT5A	USF1	FOXO3	USF1			
	NFYC	STAT4	RARA	CEBPB			
	STAT5B	TP53	ATF2	NFYB			
	ZNF148	REL	NFYC	TFAP2A			
		YY1	ESR1	NFYA			
		ETS1	IRF1	CREB1			
		EGR1	ZBTB33	NFATC1			
		HNF4A	FOXM1	GABPB1			
		HDAC1	HINFP	ETV4			
		HOXA10	ZRANB2	ETS2			
		RBPJ	CDX1	NFYC			
		IRF4	VEZF1	EGR3			
		PBX1	RFX5	PITX3			
		TP63	RFXAP	FOXJ2			
		ZNF148	RFXANK	NCOA3			
		STAT6	RUNX3	HNF4G			
		NR3C1	HES1	NR4A2			
		ETV4	TFAP2B	VEZF1			
		FOXO3	TFAP2C	IRF5			
		GATA2	CEBPE	RFX5			
		NR2F1	GATA6	RFXAP			
		GBX2	RBPJ	RFXANK			
		LYL1	NFE2	RUNX3			
		RAD9A	RUNX1	NCOA1			
		ZBTB17	CDX2	HES1			
		ZRANB2	NFATC2	TFAP2B			
		ZNF281	NR1H3	CEBPE			
		GFI1	CTCF	ELF3			
		HNF4G	STAT5A	RBPJ			
		NF1	ZNF148	NFE2			
		RB1	RUNX2	RUNX1			
		SATB1	NR3C1	NR1I3			
		VEZF1	ETV4	NR1I2			
		TDRD9	GATA3	NR1H3			
		CUX1	NR2F2	CTCF			
		IRF8		ZNF148			
		IRF5		STAT6			
		RFX5		NR3C1			
		RFXAP		PPARG			
		RFXANK		GATA3			
		RUNX3		MYC			