ROLE OF AUTOIMMUNE REGULATOR (AIRE) IN THE PATHOGENESIS OF NEUROLOGICAL DISORDERS ASSOCIATED WITH AUTOIMMUNE POLYGLANDULAR SYNDROME TYPE 1

Fahad Albalawi

Submitted for the degree of Doctor of Philosophy

The University of Sheffield

Faculty of Medicine, Dentistry and Health

Department of Infection, Immunity and Cardiovascular Disease

March 2022

ABSTRACT

Autoimmune polyglandular syndrome type 1 (APS-1) or autoimmune polyendocrinopathy-candidiasis–ectodermal dystrophy/dysplasia (APECED) is a monogenic autosomal recessive disease that occur due to mutations in autoimmune regulator gene (AIRE). AIRE is a transcriptional regulator; it is expressed mainly in medullary thymic epithelial cells (mTEC) and plays an essential role in Tissue Specific Antigens (TSAs) expression, leading to self-reactive thymocytes deletion (negative selection). AIRE is also expressed in different organs such as the brain. It is common that patients with APS-1 show high level of autoantibodies against some proteins of the brain such as SOX9, SOX10, TH, TPH, GH, GHRH and GAD65. According to recent studies, some APS-1 patients show neurological manifestation such as stiff person syndrome, cerebellar ataxia, and memory loss. Besides such learning and memory deficiencies, a recent study observed a reduction in total brain and cerebellum volumes in APS-1 patients and correlated this reduction to the presence of autoantibodies against GAD+ and TH+ neurons (Meloni et al., 2019). We therefore tested the expression of genes encoding brain antigens (BAs) in Aire$^{-/-}$ mouse (C57BL/6) and in vitro using medullary thymic epithelial cells (4D6). Ex vivo and in vitro experiments in this study revealed that AIRE upregulates the expression of *Th/TH* in the thymus of C57BL/6 mice and in an immortal cell line for human mTEC (4D6 cell line). However, *SOX9* was only upregulated in 4D6 cells in the presence of AIRE, and *Sox10* was only upregulated in mice. Therefore, upregulation by AIRE appears to be variable between mice and immortal model cell line for human mTEC. In addition, we investigated the brain of C57BL/6 *Aire*$^{-/-}$ mice but Pathological changes to the anatomy of the brain reported in APS-1 patients were not observed in Aire$^{-/-}$ mice. It may be because of the genetic background as it is strongly influencing the autoimmune phenotype or the age of mice as it corresponds primarily to the duration of the disease.

In addition, Different APS-1 mutations were introduced into an *AIRE* expression construct in 4D6 cell line to test the effect of the most common AIRE mutations on the expression of *SOX9* and *TH*. Three common APS-1 mutations Y85C, R139X, R257X and G228W were introduced to AIRE. All tested AIRE mutations adversely affected the transactivation potential of AIRE, except Y85C. The promoters of *SOX9* and *TH*

were activated with no significant difference between cells expressing wild-type AIRE and those expressing the AIRE mutant Y85C.

Our ChIP-seq data conflicted with previous reports by Kumar et al. (2001) and Purohit et al. (2005) showing AIRE binding on ATTGGTTA or TTATTA motifs. However, HOMER software was used to identify AIRE consensus sequences from ChIP data. HOMER de novo motif results showed that AIRE bound to three motifs, mostly the TCTGCAAGTGGA motif. Our data showed that AIRE is enrich in the intergenic region in pericentromeric region of many chromosomes where cohesins were also enriched. AIRE may bind directly to the identified motifs in the pericentromeric region or indirectly by binding to cohesins such as RAD21 and SMC3 that already found to be among putative Aire-associated proteins.

ACKNOWLEDGMENTS

Firstly, I would like to thank my god Almighty Allah, the most gracious, the most merciful, for blessing, protecting me and giving me this opportunity to make my dream real.

I would like to thank Dr. Rachid and Prof. Mimoun Azzouz for their constant encouragement, patience, and guidance. The knowledge I gained from you during the past three years contributed greatly to my research and personal development.

Special thanks to my parents. They are always available to support and motivate. My deepest gratitude to my wife who supported me during my study despite the difficult situation she faced when we lost our baby. She was always strong and supportive. I would like to thank my daughters Rital, Farah and Lulwa for their love and thank you for making me always smile.

I would like to thank my brothers and sisters for being available all the time and ready for help.

My gratitude to my friends Abdulrahman Abdullah, Abdulaziz Alharbi and Musab Alhithili for being in contact with me all the time and continued to inspire me significantly.

My endless appreciation goes to my friends in the department, Md Miraj K Chowdhury and Bader Alabdelaly and I would like to thank all the technicians in the department especially Jon Kilby, Mark Ariaans who have managed to order me lab materials and Fiona wright for processing all the tissues samples collected from mice.

My final greetings go to my funding bodies, a scholarship funded by Ministry of education, Saudi Arabia.

Table of Contents

ABSTRACT ii
ACKNOWLEDGMENTS iv
List of abbreviations 1
Chapter One 5
1.1.1 The prevalence of APS-1 6
1.1.2 Diagnosis and clinical features of APS-1 8
1.2 The AIRE gene 9
1.2.1 The role of AIRE in the thymus 9
1.2.2 Human AIRE expression 12
1.2.3 Structure of AIRE 13
1.2.4 Molecular mechanisms of AIRE function 15
1.2.5 Effect of AIRE on antigens 19
1.2.6 Mutations of *AIRE* 22
1.3 Experimental APS-1 models 23
1.4 APS-1 autoantibodies 30
1.5 APS-1 brain autoantigens 31
1.6 APS-1 and neurological disorders 32
1.7 Hypothesis and aims 35
Chapter Two 37
2. Material and method 38
2.1 Mouse genotyping 38
2.2 Histological analysis 39
2.2.1 Tissue collection 39
2.2.2 Immunofluorescence 39
2.2.3 H&E staining 40
2.3 Nucleic acid techniques 40
2.3.1 Total RNA extraction and cDNA synthesis 40
2.3.2 PCR of cDNA 41
2.3.3 Gel electrophoresis 43
2.3.4 RT-qPCR 43
2.4 Protein analysis 47
2.4.1 Protein extraction 47
2.4.2 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) 48
2.5 Cloning of DNA fragments 50
2.5.1 Primer design and genomic DNA extraction 50

2.5.2 Purification of DNA from agarose gels 50
2.5.3 Restriction endonuclease digestion of DNA 51
2.5.4 Ligation of DNA sequence with a vector 51
2.5.5 Transformation of competent cells 52
2.5.6 Isolation of plasmid DNA 52
2.5.7 DNA sequence analysis 53
2.6 Cell culture 53
2.7 Dual luciferase assay 53
2.8 Site-directed mutagenesis 54
2.9 Chromatin immunoprecipitation (ChIP) assay 55
A- Crosslinking and cell pellet isolation 56
B- Lysis and MNase digestion 56
C- Immunoprecipitation 57
D- IP elution and DNA recovery 58
Chapter Three 59
3.1 Outline 60
3.2 Genotyping and AIRE expression 60
3.3 Brain and thymus size 63
3.4 Histological analysis of the thymus and brain in *Aire*-deficient mice 67
3.5 Expression of genes encoding BAs 72
3.6 Expression of *Aire* and the BA genes 73
3.7 Expression of SOX10 and TH at the protein level 82
3.8 Discussion 88
Chapter Four 95
The effect of AIRE and APS-1 mutations on the expression of BAs in 4D6 cells 95
4.1 Outline 96
4.2 Validating *AIRE* expression in vitro 96
4.3 Regulation by AIRE in vitro 98
4.4 Activation of the promoters of genes encoding BA by AIRE 101
4.4.1 Identifying and cloning the promoter region 101
4.4.2 Transfection and dual luciferase reporter assay 103
4.5 The effect of APS-1 mutations on AIRE transactivation potential 107
4.6 Discussion 114
Chapter Five 119
5.1 Outline 120
5.2 ChIP-seq assay 120

5.3 Next-generation sequencing (NGS) 121
5.4 Bioinformatic analysis of the ChIP-seq data 123
5.5 Discussion 132
Chapter Six 135
6.1 General discussion 136
6.2 Future plans 139
7- References 142
Chapter Eight 155
Appendices 155
8.1 Validation of RT-qPCR samples by gel electrophoresis 156
8.2 In silico sequences of promoters cloned in pGL3 158
8.4 In silico sequences of wild type and mutated AIRE in pBIND 173
8.5 Maps of plasmids 187

List of abbreviations

AD	Addison's disease
AIRE	Autoimmune regulator
Amp	Ampicillin
APC	Antigen presenting cell
APS	Ammonium Persulfate Solution
APS-1	Autoimmune polyendocrinopathy syndrome Type I
ATF7ip	Activating Transciption Factor 7- interacting protein
β-actin	Beta actin
BRD4	Bromodomain-containing protein 4
CARD	Caspase recruitment domain
CBP	Cyclic AMP response element-binding protein
CMC	Chronic mucocutaneous candidiasis
cTEC	Cortical thymic epithelial cell
DC	Dendritic cells
DNA-PK	DNA-dependent protein kinase
FEZF2	FEZ Family Zinc Finger 2
FOxN1	Forkhead box N1
FOxP3	Forkhead box P3
GAD65	Glutamic acid decarboxylase 65

GH	Growth Hormone
GHRH	Growth Hormone Releasing Hormone
H3K4	Histone 3 lysine 4
HLA	Human leukocyte antigen
IF	Immunofluorescence
Kan	Kanamycin
LAR II	Luciferase assay reagent II
LXXL	L is leucine and × is any amino acid
MHC	Major histocompatibility complex
mTEC	Medullary thymic epithelial cells
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NLS	Nuclear localisation signal
PAMPS	Pathogen associated molecular pattern
PBS	Phosphate-buffered saline
PHD	Plant homeodomain
PIAS1	Protein inhibitor of activated STAT 1
P-TEFb	Positive transcription elongation factor b
RA	Rheumatoid Arthritis
RANK	Receptor activator of nuclear factor K
SAND	Sp100, AIRE-1, NucP41/75, DEAF-1

SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SOX9	SRY-Box Transcription Factor 9
SOX10	SRY-Box Transcription Factor 10
TAE	Tris-acetic acid-EDTA
TBS	Tris-buffered saline (TBS)
TCR	T cell receptor
TEMED	Tetramethylethylenediamine
TH	Tyrosine Hydroxylase
TPH	Tryptophan Hydroxylase
TSAs	Tissue-specific antigens
TSS	Transcription start sites
Treg	Regulatory T cells
TOP2α	Topoisomerase II

Chapter One

Introduction

1.1.1 The prevalence of APS-1

APS-1 is a rare disease, and APS-1 patients from more than 20 different national cohorts have been published from 1992 to 2021 estimated that the total number of APS-1 patients to be approximately 1000 patients worldwide (Garelli et al., 2021). However, it is found to be more prevalent among certain populations with low genetic diversity, such as Sardinians (1:14,000; Rosatelli et al., 1998), Finns (1:25,000; Björses et al., 1996; Garelli et al., 2021) and Iranian Jews (1:6500 to 1:9000; Zlotogora and Shapiro, 1992; Garelli et al., 2021). In addition, APS-1 is found in Norway (1:90,000), Slovenia (1:43,000) and Poland/Ireland (1:130,000; Myhre et al., 2001; Garelli et al., 2021) and among the nine million people who live in the northwest region of France in Nord-Pas-De-Calais, Picardie, higher and lower Normandie (1:500,000; Proust-Lemoine et al., 2007). The prevalence of APS-1 in Japan (1:10,000,000) is low compared with that in the abovementioned countries (Kisand and Peterson, 2015; Kahaly and Frommer, 2018). AIRE mutations are more frequent in isolated populations and only specific mutations are maintained in these populations (Mathis and Benoist, 2009). This is mainly because of low genetic heterogeneity and high level of consanguinity. There is no evidence that there is a heterozygous advantage in case of AIRE mutations because the individuals heterozygous for AIRE mutations are extremely rare versus inexistent in healthy controls. The difference in the frequencies of AIRE mutations suggests the existence of a founder effect in several populations. This is because there are specific mutations in specific populations (*see section 1.2.6*; Björses et al., 2000., Wolff et al., 2007). It is also due to the fact that there are common haplotypes spanning AIRE locus on chromosome 21, suggesting the existence of pressure of selection on the whole chromosomal region (Pearce et al., 1998., Jin et al., 1999). It was reported that common Sardinian R139X mutation, is responsible for 92% of the disease alleles in the population of Sardinia (Rosatelli et al., 1998).

However, R257X mutation was found in 89% of the Finnish disease alleles with 72% of patients from Finland being homozygous for R257X mutation (Björses et al., 2000). Moreover, a study that includes 104 APS-1 patients from twelve different countries reported the frequencies of 13 different *AIRE* mutations (Halonen et al., 2002). Fifty-six of the patients were Finnish; 13 were Norwegian; 10, Swedish; 1, Australian; 3, English; 1, French; 4, German; 2, Dutch; 5, Iranian Jewish; 5, Italian; 1, Swiss; and 3, North American (Halonen et al., 2002). Thirteen *AIRE* mutations were identified, and the frequencies of AIRE mutation show that R257X is the most common mutation which was observed in 56.7% of the examined alleles. The second most common mutation is 967–979del13 bp which was observed in 18.3 of the examined alleles (Table 1.1; Halonen et al., 2002). The Y85C mutation was found in all Iranian Jewish patients in homozygous form (Halonen et al., 2002).

Table 1.1: The frequencies of different *AIRE* mutations (Halonen et al., 2002).

Mutation	**Number of alleles**	**Percentage**	**Number of homozygous**	**Number of compound heterozygous**
R257X	118	56.7	52	13
967–979del13bp (C322fsX372)	38	18.3	14	8
Y85C	10	4.8	5	0
A21V	4	1.9	1	2
R203X	1	0.5	0	1
C311Y	1	0.5	0	1
C311fsX376	1	0.5	0	0
M388fsX422	4	1.9	0	4
L397fsX478	1	0.5	0	1
P398fsX478	2	1	1	0
H415fsX422	1	0.5	0	1
L417fsX422	5	2.4	2	1
X546C+59aa	4	1.9	0	4

1.1.2 Diagnosis and clinical features of APS-1

Diagnosis of APS-1 at an early stage can be difficult due to the clinical variability of the disease, and patients can present the subsequently associated diseases differently. APS-1 patients can be diagnosed clinically based on the presence of at least two out of three conditions, namely chronic mucocutaneous candidiasis, chronic hypoparathyroidism and autoimmune adrenal insufficiency (Smith-Anttila et al., 2017). Usually, children who have APS-1 onset before the age of five years develop chronic mucocutaneous candidiasis caused by *Candida albicans,* which affects approximately five percent of the body surface. In subsequent years, the child develops hypoparathyroidism, which is typically the first sign of autoimmune endocrine disease. Usually, patients develop Addison's disease (adrenal insufficiency) as a second autoimmune endocrinopathy before the age of 15 years, and at this stage, the complete clinical picture is clear (Kahaly and Formmer, 2018). Besides these diseases, patients with APS-1 may develop other manifestations due to the formation of autoantibodies against the endocrine glands, such as type 1 diabetes (1–18%), hypogonadism (24–60%), pituitary failure (7%), and hypergonadotropic and autoimmune thyroid disease (4–36%). In addition, non-endocrine manifestations such as ectodermal dystrophy, which includes punctate nail dystrophy, keratoconjunctivitis and dental enamel hypoplasia, may occur. Patients with APS-1 can develop gastrointestinal and skin diseases such as malabsorption, alopecia areata and vitiligo (Paterson et al., 2004; Proust-Lemoine et al., 2012).

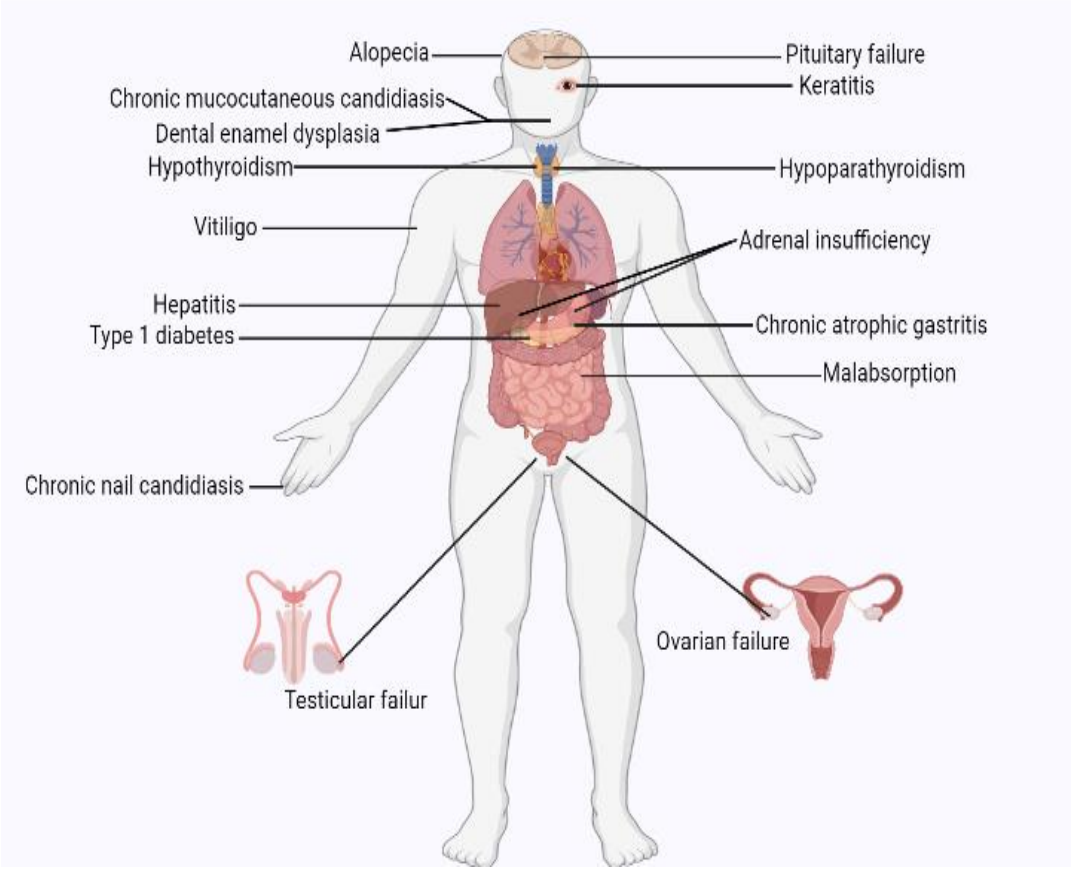

Figure 1.1: APS-1 clinical manifestations. Representative diagram of the manifestations of APS-1 and parts of the body affected. Figure created using BioRender.

1.2 The AIRE gene

The human AIRE gene is located on chromosome 21q22.3 (Aaltonen et al., 1997; Nagamine at al., 1997; Zhao et al., 2018). AIRE is a transcriptional regulator expressed mainly in the thymus in stromal cells called medullary thymic epithelial cells (mTEC; Aaltonen et al., 1997; Nagamine et al., 1997; Fetissov et al., 2009; Perniola et al., 2021). AIRE controls the expression of thymic ectopic expression of thousands of genes that are usually expressed only in one or few specific tissues. This is so-called ectopic gene expression (Org et al., 2009; Meredith et al., 2015).

1.2.1 The role of AIRE in the thymus

Experiments on *Aire*$^{-/-}$ mice have demonstrated the role of AIRE in central tolerance (Hubert et al., 2009). *Aire*$^{-/-}$ mice show organ-specific autoimmunity similar to the manifestations of APS-1, although it is milder (Proust-Lemoine et al., 2012). These

mice display abolished or reduced ectopic RNA transcript expression in mTEC and disruption of the thymic negative selection of autoreactive T cells (Anderson et al., 2002; Ramsey et al., 2002b; Jiang et al., 2005; Kuroda et al., 2005; Hubert et al., 2009) (APS-1 mouse model discussed in detail in Section 1.3). Central tolerance occurs by two mechanisms, namely positive and negative selection (Kyewski et al., 2006; Wang et al., 2012). Developing T cells migrate to thymic cortex and are stimulated to develop T-cell receptor (TCR) expressing CD4+CD8+ double positive T cells (Passos et al., 2015). This occurs by rearrangements of variable (V), diversity (D), and joining (J) gene segments leads to the highly diverse repertoire of TCR. The rearrangements of V(D)J gene segments almost occur in a random fashion and is guided by conserved noncoding DNA sequences, named recombination signal sequences (RSS; Kondo et al., 2019; Chi et al., 2020). V(D)J recombination is mediated by V(D)J recombinase, which is a diverse collection of enzymes used in common DNA repair mechanisms that is initiated by recombination activating genes (RAG1 and RAG2) that is specific to T cells (Chi et al., 2020). Positive selection occurs in the cortex of the thymus, where T cells are selected and avoid deletion if double positive T cells are reactive to self-antigens presented by major histocompatibility complex class I (MHC-I) or MHC class II (MHC-II) molecules on cortical thymic epithelial cells (cTECs) with low affinity. At CD4+CD8+ double-positive stage, T cells that do not interact because of the absence of T cell receptors (TCRs) or the inability of their TCRs to interact with self-peptide-MHC ligands are deleted by neglect (Stritesky et al., 2012; Wang et al., 2012). T cells that recognize MHC-II become CD4+ helper cells, while T cells that recognize MHC-I become CD8+ killer cells (Stritesky et al., 2012). Positively selected T cells (CD4+ helper or CD8+ killer cells) subsequently travel to the medulla, where negative selection occurs. In the medulla, T cells must interact with antigen presenting cells

(APC), including dendritic cells and mTEC, in order to delete autoreactive T cells that interact with a very high affinity to self-antigens (Org et al., 2009; Meredith et al., 2015). In the medulla, a distinct subpopulation of mTEC express AIRE that characterized by the expression of high levels of major histocompatibility complex class II (MHC-II)^{Hi} and cluster of designation 80 (CD80)^{Hi}, hence called mTEC^{Hi} (Org et al., 2009). mTEC^{Hi} differentiate from immature MHCII^{lo} CD80^{lo} AIRE- mTECs, known as mTEC^{lo} (Gray et al.,2007). Although mTEC^{Hi} had been thought to be non-proliferative (postmitotic) and undergo apoptosis after the expression of AIRE, it was shown that a fraction of mTEC^{Hi} differentiate to post-AIRE mTEC (Wang et al., 2012). The expression of AIRE, most AIRE-dependent genes, MHCII, CD80 molecules are lost in post-AIRE mTEC (Wang et al., 2012). Metzger et al. (2013) has been proposed that post-AIRE mTEC can convert to mTEC^{lo} and then differentiated into mTEC^{Hi} again.

AIRE plays an essential role in the negative selection process, regulating the expression of a wide array of tissue-specific antigens (TSAs) in mTEC, that is presented to developing T cells, leading to self-reactive thymocytes deletion (Org et al., 2009; Meredith et al., 2015). However, the affinity between TCRs and self-peptide-MHC defines the fate of T cells. Only T cells that interact with self-peptide-MHC with low affinity are allowed to survive and exit to the periphery. However, T cells that interact with self-peptide-MHC with high affinity are deleted but those that interact with self-peptide-MHC with intermediate affinity are differentiate into Foxp3+ regulatory T cells (T_{reg}; Stritesky et al., 2012; Kisand and Peterson, 2015). It has been reported that T_{reg} cells are found in insufficient numbers and malfunction in APS-1 patients (Kekalainen et al., 2007; Wolff et al., 2010). Although negative selection is crucial to delete autoreactive T cells, it cannot purge all reactive T cells and thus allows autoreactive T cells to escape the thymus reach the periphery (Kisand and Peterson,

2015). Therefore, the regulation of TSA expression by AIRE promotes immune tolerance by clonal deletion and the differentiation of T cells to T_{reg} cells.

1.2.2 Human AIRE expression

Two independent groups first described the expression of AIRE in humans in 1997 (Aaltonen et al., 1997; Nagamine at al., 1997). Expression of AIRE in the thymus, foetal liver and lymph node was detected using northern blot analysis (Nagamine et al., 1997). In addition, the expression of AIRE was detected in the spleen, pancreas and adrenal cortex (Aaltonen et al., 1997; Adamson et al., 2004; Heino et al., 1999). AIRE protein is mainly located to the nucleus; however, it can be translocated to the cytoplasm (Proust-Lemoine et al., 2012). Besides the expression of AIRE in the thymus, it can be detected in dendritic-like cells in the lymph nodes and tonsils (Kisand and Peterson, 2015; Zhao et al., 2018).

Immunohistochemical staining studies and real-time reverse transcription-polymerase chain reaction (RT-PCR) of samples from C57BL/6 and CD1 mice have shown that AIRE protein is expressed in other organs such as the spleen, heart, kidney, testis, brain, lung, gut and adrenal gland (Figure 1.2; Aaltonen et al., 1997; Nagamine at al., 1997; Adamson et al., 2004). Immunostaining has demonstrated the expression of AIRE protein in some parts of the brain, but high expression of Aire was detected only in Purkinje cells, many granular neurons of the cerebral cortex and the neurons of the hippocampus (Ramsey et al., 2002b; Adamson et al., 2004). It seems that AIRE highly expressed in specific cells of the cerebral cortex and hippocampus but not in whole brain, which may explain the low expression of Aire at RNA level since only a few specific cells in the brain express Aire. Therefore, there might be a dilution effect and RNA degradation when total RNA is extracted from the whole brain (*see section 3.2 & 3.8;* Figure 1.2).

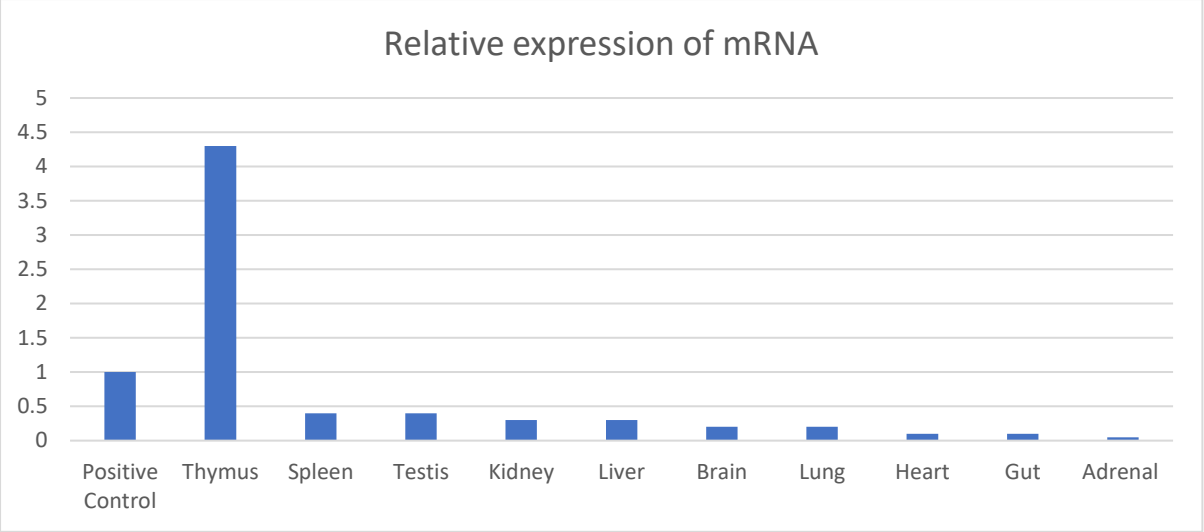

Figure 1.2: Distribution of *Aire* expression. The distribution of *Aire* expression in different organs of CD1 mice was assessed using real-time RT-PCR. The graph shows that *Aire* is mostly expressed in the thymus but also expressed in different organs including immunologically relevant tissues such as the spleen and extra-immunological tissues such as the brain and lung. Taken with permission from Adamson et al. (2004).

1.2.3 Structure of AIRE

The AIRE gene has 14 exons coding for 2445 base pairs and encodes a protein of 545 amino acids with a molecular mass of approximately 57.5 kDa (Aaltonen et al., 1997; Nagamine et al., 1997; Guo et al., 2018). AIRE consists of six domains, including the N-terminal caspase recruitment domain (CARD), also called the homogeneously staining region (HSR), nuclear localisation signal (NLS), SAND domain (Sp100, AIRE, NucP41/75, DEAF1), two plant homeodomain (PHD)-type zinc finger domains, proline-rich region (PRR) and four LXXLL motifs (Sparks et al., 2016; Guo et al., 2018).

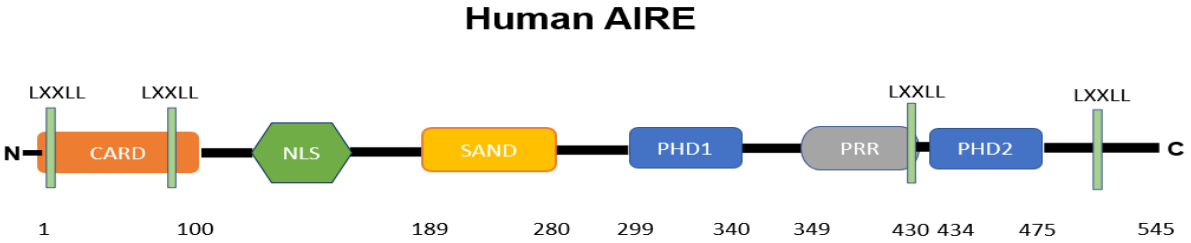

Figure 1.3: Functional domains of AIRE. The domains of AIRE protein are shown in the coloured boxes including CARD (aa 1-100), NLS (aa 110-167), SAND domain (aa 189-280), PHD1 (aa 299-340), PHD2 (aa 434-475), PRR (aa 349-430) and four LXXLL motifs that are scattered throughout the protein sequence. The position of four LXXLL motifs is marked above the diagram. The 545 amino acids of AIRE protein are labelled below the diagram.

The CARD domain is involved in AIRE homomultimerisation and mutations in the CARD domain affect the expression of AIRE-dependent genes (Peterson et al., 2008). Missense mutations in the CARD and/or SAND domains that have been identified in APS-1 patients can reduce AIRE's effectiveness in the activation of gene transcription (Halonen et al., 2004). The SAND domain does not have a distinct DNA- binding motif but is found to be involved in protein-protein interactions with activating transcription factor 7-integrating protein (ATF7IP) in the context of epigenetic processes (Peterson et al., 2008). NLS has an essential role in nuclear import (Sparks et al., 2016). PHD zinc finger 1 (PHD1) influences AIRE transcriptional activity by interacting with the N-terminal tail of unmethylated histone H3 molecules at lysine position 4 (H3K4) and with DNA-dependent protein kinase (DNA-PK; Org et al., 2008; Chignola et al., 2009). LXXLL motifs are involved in promoting gene transcription by facilitating the interaction of proteins with nuclear receptors (Heery et al., 1997; Sparks et al., 2016; Guo et al., 2018). The PRR, found between amino acid numbers 350 and 430, mediates protein-protein interactions (Guo et al., 2018).

There is no strong evidence showing that AIRE binds to DNA even if it has been shown by gel shift assay that AIRE can bind to two different consensus binding sequence motifs, namely ATTGGTTA and TTATTA (Kumar et al., 2001; Purohit et al., 2005). In addition, Purohit et al. (2005) demonstrated that the PHD domains in AIRE are responsible for binding to the ATTGGTTA sequence motif, while the TTATTA motif was shown to be bound by the SAND domain. However, later studies have contradicted these findings by showing that AIRE is not a conventional transcription factor since it recognises its dependent genes through recognition of the repressive epigenetic signature directly through its PHD1 domain or indirectly by binding to other

proteins (AIRE partners) instead of binding to DNA (*see section 1.2.4*; Org et al., 2008). In addition, it was found that AIRE binds weakly and non-specifically to DNA and no common motif has been identified for AIRE-dependent genes (Koh et al., 2008; Org et al., 2008). Moreover, The SAND domain in other proteins has been shown to be a DNA binding domain, however, the SNAD of AIRE lacks the motif responsible for DNA binding (KDWK; Waterfield et al., 2014).

1.2.4 Molecular mechanisms of AIRE function

It has been noted that the 3D medullary microenvironment and certain signalling were needed for the expression of AIRE and TSA genes and disruption of medullary microenvironment leads to AIRE expression reduction (Rossi et al., 2007; Irla et al., 2008). The expression of AIRE is triggered in $mTEC^{Hi}$ by members of the tumour necrosis factor (TNF) superfamily, including the receptor activator of nuclear factor kappa-B (RANK) and CD40 signalling pathways (Rossi et al., 2007; Irla et al., 2008). RANK is expressed by $mTEC^{Hi}$ and its related ligand, RANKL is expressed by CD4+ and CD8+ thymocytes (Desanti et al., 2012). Mice deficient for RANK or RANKL show a strong but not complete reduction of $AIRE^{+}$ $mTEC^{Hi}$ and TSA expression (Hikosaka et al., 2008). Additional studies have identified CD40 that induce the expression of AIRE in $mTEC^{Hi}$. CD40 is expressed by $mTEC^{Hi}$ and recognises its corresponding ligand CD40L that is expressed by CD4+ thymocytes (Irla et al., 2008).

Although the induction of AIRE expression by RANK and CD40 signalling pathway is necessary for inducing the expression of TSAs in $mTEC^{Hi}$, post-translational modifications of AIRE impact its activation. Sirtuin-1 (SIRT-1) deacetylases AIRE lysine residues located between AIRE's NLS and SAND domains, causing its activation and induce the expression of TSAs in $mTEC^{Hi}$ (Abramson and Husebye, 2016).

There are some factors to control the specificity of AIRE to recognise and target its dependent genes. AIRE is recruited to the silenced genes through recognition of the repressive epigenetic signature, including the unmethylated lysine 4 of histone 3 (H3K4me0; Org et al., 2008). AIRE binds to H3K4me0 directly through its PHD1 domain. However, some studies have shown that disruption of AIRE's PHD1 domain does not affect the expression of all genes induced by AIRE and H3K4me0 is abundant in the genome near the promoters of many AIRE-independent genes, leaving the possibility of other factors to control the specificity of AIRE. It was proposed by Waterfield et al. (2014) that the complex of Activating-transcription-factor-7-interacting protein (ATF7IP) and methyl-CpG-binding-domain protein 1 (MBD1) is essential for the recruitment of AIRE at its target genes. MBD1 is known to recognise the promoters of silenced genes by binding to unmethylated CpG dinucleotides and ATF7IP is a transcription factor that has been shown to control positively or negatively gene transcription. It was reported that AIRE interacts with ATF7IP by its SAND domain and knockdown of ATF7IP impairs the induction of many reporter genes depends on AIRE in HEK293 (Paterson et al., 2008; Waterfield et al., 2014). Moreover, it was found that AIRE is recruited to the silenced genes through recognition of the repressive epigenetic mark, including H3K27me3 (Sansom et al., 2014). AIRE may interact with H3K27me3 indirectly by interacting with chromodomain helicase DNA (CHD) 4 and 6 that have been found to be among the putative AIRE-associated proteins (Yang et al. 2013, Sansom et al., 2014). However, AIRE is found to bind and activate super-enhancers, the chromatin region that overloaded with transcriptional factors (Bansal et al., 2017). Aire is found to bind to super-enhancer by binding to histone 3 acetylated at lysine 27 (H3K27ac) and histone 3 lysine 4 aminomethylation (H3K4me1), active chromatin signatures that define the super-enhancer (Bansal et al., 2017).

AIRE has been found to bind and form complexes with several proteins known as AIRE partners (Yang et al., 2013). Mass spectrometry of co-immunoprecipitated proteins screening showed that Aire interact with several proteins that can be categorised into four groups based on function: nuclear transport, chromatin remodelling, transcription and mRNA processing (Yang et al., 2013). None of the identified AIRE partners were involved in transcriptional initiation, however, some of the identified AIRE partners are related to transcriptional elongation, indicating AIRE involvement in transcriptional elongation rather than transcriptional initiation (Giraud et al., 2014). In vitro and in vivo studies regarding AIRE have suggested its role in the activation of promiscuous gene expression of TSAs via the release of blocked RNA polymerase II (RNAP-II; Giraud et al., 2012). The RNAP-II stalling is characterised by the ability of RNAP-II to transcribe the first 30–60 base pairs (bp), which is then stopped by the action of transcription pausing components such as negative elongation factor (NELF) and 5,6-dichloro-1-beta-D-ribofuranosyl-benzimidazole (DRB) sensitivity-inducing factor (DSIF). To continue transcription, stalled RNAP-II is released by the positive transcription elongation factor b (P-TEFb). P-TEFb is cyclin dependent kinase that phosphorylates the transcription pausing components to terminates their inhibitory action. In addition, P-TEFb phosphorylates the serine-2 residues in the carboxy-terminal domain (CTD) of RNAP-II to recruit the chromatin modifying proteins. AIRE is found to interact to members of pTEFb complex such as the cyclin-dependent kinase 9 (CDK9), cyclin T2 (CCNT2), Hexamethylene bisacetamide-inducible protein 1 (HEXIM1) and heterogeneous nuclear ribonucleoprotein L (HNRNPL; Oven et al., 2007; Giraud et al., 2014). Later, it was demonstrated by Yoshida et al. (2015) that interactions between bromodomain-containing protein 4 (BRD4) and AIRE lead to the recruitment pTEFb resulting in the

release of the stalled RNAP-II. It was also shown by Kanno et al. (2014) that BRD4 recruit the pTEFb to stalled RNAP-II through its binding to acetylated histones H3/H4.

AIRE is found to interact with proteins associated with epigenetic changes and chromatin relaxation (Bansal et al., 2017). AIRE interacts with topoisomerase 2-alpha (TOP2α) via its PHD1 domain (Bansal et al., 2017). Although the exact sequential events occur due to the interaction between AIRE and TOP2α to release the blocked RNAP-II requires further investigations, TOP2α known to initiate breaks at the TSSs of silenced genes, which leads to relaxation of the chromatin superhelical tensions that formed by advancing RNAP-II during elongation stage of transcription, leading to more effective transcription by proceeding more easily through the unwinding helix (Abramson et al., 2010). DNA breaks initiated by TOP2α lead to the activation of DNA-PK and the other partners of AIRE associated with epigenetic changes and chromatin relaxation, such as the transcriptional co-activator, cyclic adenosine monophosphate response element-binding protein (CREB)-binding protein (CBP), which leads to chromatin relaxation through histone acetylation (Liiv et al., 2008; Abramson et al., 2010, Abramson and Husebye, 2016). CBP is the first protein identified to interact with AIRE and colocalise with AIRE in the nuclear speckles (Pitkänen et al., 2005). It has been shown by Pitkänen et al. 2005 that CBP promote the expression of AIRE-dependent reporter construct in HEK293 cells (Pitkänen et al., 2005). Moreover, Chromatin relaxation and the interaction between AIRE and BRD4 that leads to the recruitment pTEFb, facilitates the successful transcription of AIRE-dependent genes (Giraud et al., 2012; Yoshida et al., 2015).

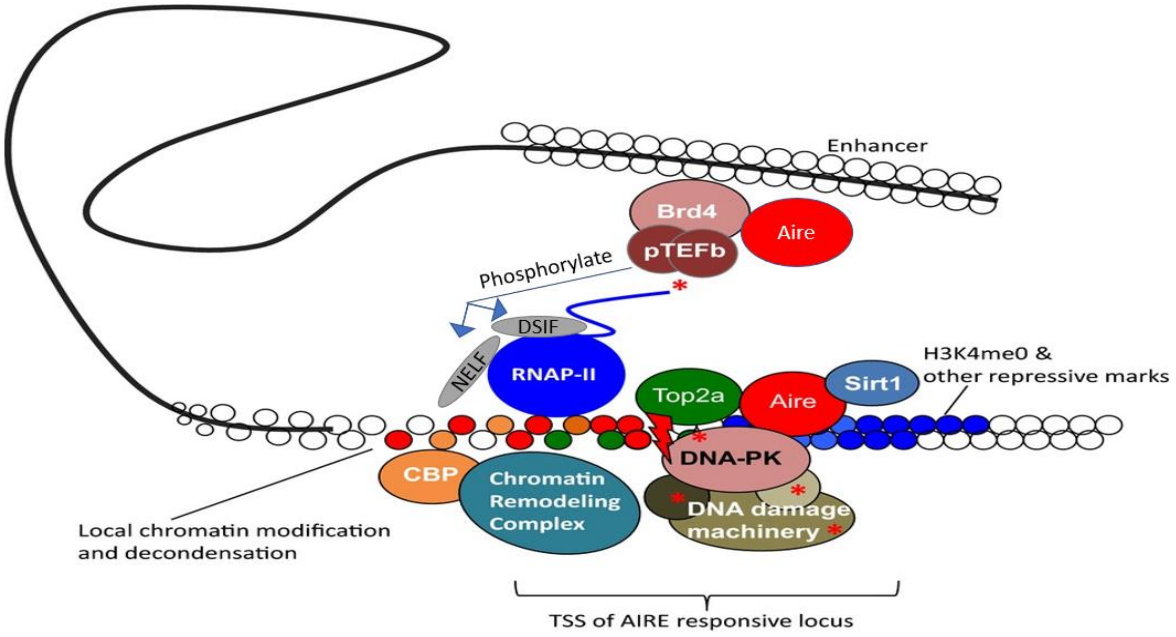

Figure 1.4: Hypothesised mechanism of AIRE-mediated gene activation. AIRE binds with several proteins to regulate the transcription of its dependent genes. SIRT-1 deacetylases AIRE, causing its activation. The interaction of AIRE with unmodified H3K4 leads to the formation of breaks at the TSSs of silenced genes by TOP2α. As a result, chromatin relaxation occurs through histone acetylation by CBP (red circles). These events lead to the recruitment of pTEFb by BRD4 to phosphorylate NELF and DSIF, which releases stalled RNAP-II and facilitates the successful transcription of AIRE-dependent genes. Taken with permission from Abramson and Goldfarb, 2016.

1.2.5 Effect of AIRE on antigens

It has been estimated that around 500 out of 2000–3000 organ-specific antigens, including APS-1-associated autoantigens, may be AIRE-dependent (Anderson et al., 2002; Gotter et al., 2004). Later, a study has revealed by microarray assays that AIRE can regulate thousands of genes in mTEC^{Hi} (Venanzi et al., 2008). However, more recently, the number of AIRE-dependent genes in mTEC^{Hi} has been re-evaluated by RNA-seq and these studies provided some evidence that AIRE is the main transcription regulator in mTEC^{Hi} by showing that AIRE induce the expression of thousands of TSA genes in mTEC^{Hi} (Sansom et al., 2014; St-Pierre et al., 2015). However, it was hypothesised that there are other transcription factors that can regulate the remaining thousands of genes that expressed independently from AIRE. As a result, FEZF2 has been identified to induce the expression of TSA genes in

mTEC^{Hi} but the number of genes induced by FEZF2 are much lower than AIRE-dependent genes (Takaba et al., 2015). The results of Sansom et al. 2014 and St-Pierre et al. 2015 revealed that mature mTEC expressing Aire can express 19,293 and 15,137 protein-coding genes, respectively. However, to determine genes regulated by Aire using RNA-seq, Sansome et al. 2014 identified 3980 genes upregulated in mouse mature mTEC expressing Aire comparing to mature Aire knockout mTEC while only 180 genes were downregulated. However, St-Pierre et al. 2015 identified 3272 genes upregulated in mouse mature mTEC expressing Aire comparing to Aire knockout mTEC while only 66 genes were downregulated. In addition, it was found by Sansom et al. 2014 that genes induced by Aire are strongly associated with H3K27me3 (*see section 1.2.4*; Sansom et al., 2014).

To identify the genes targeted by AIRE, our group generated a recombinant AIRE expression variant of the TEC 1A3 human cell line (TEC 1A3 AIRE^{hi}; Lovewell et al., 2018). Cell line models are suitable for exploring the molecular mechanism of AIRE function because there is no limitation in the number of AIRE^{+} cells, the cells can be chemically and genetically manipulated, and they represent the only way by which to study AIRE in a model system of human origin (Lovewell and Tazi-Ahnini, 2011). Although, cell line models provided key findings, they also have some limitations such as overexpression of AIRE in these cells induce much lower number of TSA genes comparing to the number of genes induced in vivo (Abramson et al., 2010, Giraud et al., 2014, Besnard et al., 2021). The number of TSA genes induced by AIRE is much lower in human cell line than in vivo because 3D medullary microenvironment and certain signalling is necessary for the expression of AIRE and TSA genes and disruption of medullary microenvironment leads to the reduction of AIRE and TSA genes expression (*see section 1.2.4*; Rossi et al., 2007; Irla et al., 2008).

Data from microarray analysis showed that 482 genes were significantly differentially expressed, with a p-value < 0.05. AIRE upregulated the expression of 353 genes and downregulated 129 genes, and these data were validated by quantitative PCR (qPCR), which confirmed the variable expression of the selected 12 genes known to be AIRE-dependent (Lovewell et al., 2018). *Aire* knockout mice have been used to confirm the reduced expression of some transcripts encoding peripheral TSAs, such as the *insulin2* (*Ins2*), fatty acid-binding protein (*Fabp2*), salivary gland protein 1 (*Spt1*) genes and casein- alpha (*Csna*; Anderson et al., 2002; Derbinski et al., 2005; Venanzi et al., 2008). However, AIRE can downregulate some genes, such as those encoding C-reactive protein (*CRP*) and glutamic acid decarboxylase 67 (*GAD67*; Derbinski et al., 2005). Moreover, our group examined the expression level of some candidate genes in the thymus of $Aire^{-/-}$ mice in comparison to $Aire^{+/+}$ mice using qPCR analysis (Almaghrabi et al., 2020). qPCR data revealed that the expression of chemokine (C-C motif) ligand 1 (*Ccl1*) and interleukin 3 (*Il3)* decreased four-fold in $Aire^{-/-}$ compared with wild-type mice, and the expression levels of *Fabp2*, *Ctrb1 and Csna* reduced seven-fold in *Aire* knockout mice. In addition, the level of high-density lipoprotein (*Hdl*) was reduced by almost three-fold in $Aire^{-/-}$ mice. Therefore, *Ccl1, Il3, Fabp2*, *Ctrb1, Csna* and *Hdl* are Aire-dependent genes in $Aire^{+/+}$ compared with $Aire^{-/-}$ mice (Almaghrabi et al., 2020). On the other hand, some genes seem to be downregulated in $Aire^{+/+}$ mice compared with $Aire^{-/-}$ mice, such as DNA methyltransferase 3-like (*Dnmt3l*), Monoamine oxidase A (*Maoa*), Phosphatidylinositol transfer protein cytoplasmic 1 (*Pitpnc1*) and Cyclin M2 (*Cnnm2*; Almaghrabi et al., 2020). Finally, some genes showed no changes regardless of the presence of AIRE, such as Coproporphyrinogen Oxidase (*Cpox)*, Nuclear Receptor Coactivator 6 (*Ncoa6)*,

Forkhead box N1 (*Foxn1)* and Peptidyl-prolyl cis-trans isomerase (*Pplb*; Almaghrabi et al., 2020).

1.2.6 Mutations of *AIRE*

Pathogenic mutations in *AIRE* are distributed throughout the coding region of the gene. To date, over 100 mutations in *AIRE* that lead to APS-1 have been reported, including deletions, substitutions and insertions (Guo et al., 2018; Garelli et al., 2021). The mutations can be nonsense or frameshift mutations which result in non-functional protein (Peterson et al., 2005). Nonsense mutations result in premature stop codons, resulting in the generation of a truncated polypeptide, while frameshift mutations due to the deletion of one or more base pairs from the gene lead to an unnatural reading frame from the mutation position to the end of the gene (Peterson et al., 2005). R257X in exon 6 and the deletion of a 13-base-pair segment (967–979del13 bp) in exon 8 are the most common mutations (Guo et al., 2018; Garelli et al., 2021). The 13-base-pair deletion is the most common mutation to affect the PHD1 domain and causes a shift in the reading frame, resulting in the formation of a truncated protein (Guo et al., 2018).

Some mutations are common in certain populations; for example, the R257X mutation is common in APS-1 patients of European origin and is the most common mutation in APS-1 patients from Finland (Björses et al., 2000, Gue et al., 2018). The most common mutation among APS-1 patients in the UK, Norway and North America is 967–979del13 bp (Wolff et al., 2007). In addition, the R139X mutation is common in Sardinian APS-1 patients and the Y85C mutation in Iranian Jews. R139X and R257X are nonsense mutations that result in truncated versions of AIRE, while Y85C is a missense mutation that changes a tyrosine to cysteine in the CARD domain (Björses et al., 2000; Ramsey et al., 2002a; Ilmarinen et al., 2005; De Martino et al., 2013). The

Y85C mutation is only detected in the Iranian Jewish population and has not been reported in any other ethnic group (Björses et al., 2000; Meloni et al 2012).

Almost all APS-1 mutations are inherited in an autosomal recessive manner. However, G228W mutation that acts in an autosomal dominant fashion has been reported in Italian kindred with APS-1 and Co-segregates with Hypothyroid Autoimmune Thyroiditis (Cetani et al., 2001). The G228W mutation results in an amino acid change from glycine to tryptophan in exon 6, which affects the SAND domain (Cetani et al., 2001). Later, it was shown by immunoprecipitation studies that G228W Aire multimerizes with WT Aire in 1C6 mTEC cell line (Su et al., 2008). Luciferase reporter assay showed that WT AIRE activated the transcription of a reporter vector containing the insulin (*INS*) promoter by 60-fold, however, G228W AIRE did not activate the transcription of *INS* promoter and inhibited the transactivation activity of WT AIRE when co-transfected (Su et al., 2008). In addition, it was suggested by Su et al. (2008) that G228W AIRE acts as a dominant negative by multimerizing to WT AIRE and prevent it to get access to the transcription site of AIRE-dependent genes. Chromatin immunoprecipitation experiments has been shown that WT AIRE enriched more than 6-fold at *Ins*2 promoter (AIRE-dependent gene) when transfected in 1C6 cells, however, co-transfection of G228W AIRE and WT AIRE prevented AIRE to bind to the *Ins*2 locus (Su et al., 2008).

1.3 Experimental APS-1 models

Since APS-1 is a monogenic condition arising from loss-of-function mutations in the *AIRE* gene, researchers such as Anderson et al. (2002) and Ramsey et al. (2002b) could soon develop an experimental in vivo model. Pereira et al. (2005) and Liston (2006) have since reported in vivo investigations in murine models leading to a major

increase in our understanding of the autoimmunity mechanisms underlying APS-1 cases.

Seven distinct types of *Aire*-deficient mice have been independently bred on a range of backgrounds by four different research groups, as published by Anderson et al. (2002), Ramsey et al. (2002b), Jiang et al. (2005) and Kuroda et al. (2005), and Hubert et al. (2009). Thus, as reported by Anderson et al. (2002), *Aire* knockout mice were independently generated on a C57BL/6 background by two groups in 2002. To model human APS-1, Peltonen et al. bred *Aire* knockout mice that harboured R257X, an *AIRE* mutation significant among Finns. In this example, rapid termination of all synthesised AIRE polypeptides was triggered by the insertion of a neomycin resistance (Neo)-cassette into exon 6 in a targeted disruption of murine *Aire* by homologous recombination. Meanwhile, *Aire* knockout mice were generated by Anderson et al. (2002) through deletion of exon 2. While both mouse models revealed typical levels of *Aire* gene transcripts but as expected the truncated versions were smaller than the wild type. However, sequence analysis demonstrated that the frameshifts in both *Aire* copies prevented the production of functional protein.

As described by Kuroda et al. (2005), an *Aire* knockout mouse model with Neo was subsequently generated using a gene-targeting vector produced by the deletion of exons 5−12. This large-scale deletion led to the absence of the key domains SAND, PHD-1 and PHD-2. Similarly, in the *Aire* knockout mouse model produced by Scott et al., the PHD-1 domain was disrupted at exon 8 to mirror the 13-bp deletion 967–979 in human APS-1 cases (Hubert et al., 2009). A comparison of the publications cited above indicates that functional AIRE protein expression in each mouse model was prevented due to abolished or restrained expression of premature, non-functional AIRE protein. However, the morphology of the thymus appears to be unaffected by

the absence of the *Aire* gene, as all the *Aire* knockout mice displayed healthy thymic compartment morphologies.

In terms of phenotype, all strains appear largely normal and are comparable to their wild-type siblings. Except for those produced by Matsumoto's group, the mice display increasing infertility with age, producing offspring only occasionally (Kuroda, 2005). Except for those on the non-obese diabetic (NOD) background, where rapid and severe onset of the disease leads to mortality by 15 weeks of age (Jiang et al., 2005), all *Aire* knockout mice survive to the typical wild-type age. NOD mouse model of type 1 diabetes (T1D) develops lymphocytic infiltration targeting insulin-producing pancreatic islets leading to insulitis (Serreze and Leither, 1994). There are different loci contributing to insulin-dependent diabetes known as *idd* that can affect the clonal deletion of autoreactive T cells in the thymus but the key genetic component responsible for T1D susceptibility in NOD mice is the strain's unique MHC-II haplotype known as $H2^{g7}$ (*idd1*; *Serreze* and Leither, 1994; Jiang et al., 2005). NOD mice not only develop T1D but also develop other autoimmune diseases such as autoimmune thyroiditis and autoimmune sialadenitis (Park et al., 2015). Therefore, the pre-existing autoimmune diseases that NOD mice experience may explain the severe $Aire^{-/-}$ phenotype on a NOD background.

Circulation of autoantibodies and spontaneous lymphocytic infiltration of many tissues have been observed in the abovementioned studies. Ramsey et al. (2002b) observed adrenal and ovarian atrophy in 42% of *Aire*-deficient mice and lymphocytic infiltration in 50% of the assessed liver sections. While this multi-organ autoimmune behaviour is comparable to the human APS-1 phenotype, these infiltrations never develop into the three standard diagnostic conditions of Addison's disease, hypoparathyroidism and chronic mucocutaneous candidiasis in mice.

The wide range of clinical symptoms displayed by APS-1 patients suggest that the characteristics of the disease are influenced by a variety of genetic or environmental elements. Jiang et al. (2005) examined the influence of genetic variation on the range of APS-1 symptoms in human patients by back-crossing mice on BALB/cJ, C57BL/6J, NOD/LtJ and SJL/J backgrounds with *Aire*$^{-/-}$ mice previously produced by Anderson et al. (2002). As in the earlier study, Jiang et al. observed lymphocytic infiltration and autoimmune endocrine damage in the *Aire*-deficient mice, with the different strains presenting moderately different pathologies. Their study also demonstrated the influence of genetic background on overall disease intensity, which varied from mild illness among C57BL/6 mice to more severe infiltration, pneumonitis and elevated autoantibody generation among *Aire* NOD/LtJ. For example, significantly higher levels of tissue infiltration targeting the stomach were observed in the BALB/c (100%) and NOD (87%) backgrounds compared with C57BL/6 (9%), while tissue infiltration targeting the pancreas was significant for NOD (100%) and SJL/J (71%). Thus, the minor autoimmune condition observed in the C57BL/6 background was transformed into a grave illness leading to mortality by 15 weeks of age in the NOD genetic background. Taken together, these studies demonstrate the presence of disease-altering loci in mice and further suggest the existence of such loci in human APS-1 patients as an explanation for the variations in clinical symptoms.

With its loss-of-function mutations and pathologies comparable to human APS-1, the overall influence of *Aire* on murine models is considered adequate to enable its direct comparison with the function of AIRE in humans. Nevertheless, the *Aire*-deficient APS-1 mouse model cannot replicate the onset of the three standard APS-1 symptoms of Addison's disease, candidiasis and hypoparathyroidism. Although studies by Pöntynen et al. (2006) and Kekalainen et al. (2007) provide inconsistent evidence for

the relevance of *Aire*$^{-/-}$ mice as a model for human APS-1, it should be noted that the *Aire*$^{-/-}$ mouse model used in Pöntynen's study had the C57BL/6 genetic background. As mentioned previously, this background has been shown to produce a milder immune phenotype than, for example, the NOD strain. Hence, the influence of genetic background could explain why Pöntynen's model did not mimic the human APS-1 autoantibody profile. Liston (2006) has also offered the potential explanation that while *Aire*-deficient mice display the same flaws in self-tolerance, the corresponding clinical symptoms affect distinct targets. The findings of Ramsey et al. (2002b) also suggest that the immune response in *Aire*$^{-/-}$ mice may be amplified by the significant influence of non-genetic (environmental) elements such as the acquisition of infections. Indeed, the absence of candidiasis infection in *Aire*-deficient mice may be a result of their being kept under sterile conditions.

In a study, Meredith et al. (2015) analysed microarray profiling data from the mTEC of two genetically identical wild-type mice to reveal notably distinct gene expression profiles. This result has major implications regarding individual tolerance variations within a given species. In a previous study by Venanzi et al. (2008), gene expression profiling of wild-type mTEC from one or two thymic lobes of the same mouse also revealed a notably higher variation in AIRE-dependent gene expression relative to AIRE-independent transcripts. While this variability may cause difficulties in interpreting some experimental results, Ahonen et al. (1990) and Ishii et al. (2000) have argued that it may help identify the source of variation in the clinical symptoms displayed by human APS-1 patients whose genetic backgrounds and consequent clinical profiles can similarly vary from individual to individual, even among siblings. However, studying the expression of AIRE-dependent genes at single-cell level revealed that TSA genes induced by AIRE are not expressed in each single mTEC^{Hi}

and the choice of genes induced by AIRE are set stochastically in every mTEC^{Hi} (Villaseñor et al., 2008, Derbinski et al., 2008, Pinto et al., 2013, Meredith et al., 2015). It was showed by Villaseñor et al. 2008 using single-cell RT-qPCR that some genes induced by AIRE in mTEC^{Hi} such as genes of S100 family are not expressed in all single mTEC^{Hi}. In addition, another group showed that the genes of the casein locus, whose expression is induced by AIRE in mTEC^{Hi}, are expressed in random arrangement in single mTEC^{Hi} while these genes are uniformly co-expressed in single cells of the mammary epithelial cells (Derbinski et al., 2008). After that, it was showed by Pinto et al. 2013 that minor subsets of mTEC^{Hi} expressing particular TSAs co-express distinct sets of AIRE-dependent genes. Three co-expression gene groups were identified, and it was suggested that the expression of AIRE-dependent genes at mTEC^{Hi} are expressed with some degree of cooperation as the co-expression groups contain overlapping and complementary gene sets, that mapped to specific chromosomes and intra-chromosomal gene clusters (Pinto et al., 2013). In addition, it was found by Meredith et al., 2015 that TSA genes induced by AIRE are expressed in only a minority of mTEC^{Hi} as small inter-chromosomal gene clusters triggered cooperatively in a fraction of mTEC^{Hi}. The clusters are not shared between two genetically identical wild-type mice but is stable across mTEC divisions (Meredith et al., 2015). Therefore, it was suggested my Meredith et al. 2015 that mTEC^{Hi} that share TSA clusters might be differentiated from the same epithelial cell progenitor that illustrate AIRE-dependent genes within a clone of mTEC is stable through cell division. Therefore, it is suggested that the co-expressed genes within single mTEC^{Hi} clustered in the genome in a distribution that seems to be organised at the DNA or epigenome by stochastic determinism (Meredith et al., 2015).

Aire$^{-/-}$ mice are crucial to study the role of AIRE in central tolerance and circulation of autoantibodies and spontaneous lymphocytic infiltration of many tissues have been observed in mice. However, these infiltrations never develop into the three standard diagnostic conditions of Addison's disease, hypoparathyroidism and chronic mucocutaneous candidiasis in mice. Therefore, a recently *Aire*-deficient rat (BN) was generated and shows significant autoimmune symptoms of APS-1, some of which have not been observed in *Aire*-deficient mice, such as nail dystrophy, alopecia and vitiligo (Ossart et al., 2018). In addition, sever autoimmune lesions were observed in a number of organs such as pancreas, liver, kidney and lung (Ossart et al., 2018). Moreover, autoantibodies against interferon type 1 and interleukin 17 (IL-17) were detected similar to APS-1 (Ossart et al., 2018). The parts of the body affected because of the absence of AIRE in different APS-1 models, whether mice or rats, are shown in Table 1.2. These models can complement each other to mimic APS-1.

Table 1.2: Summary of the organs affected in models of APS-1.

APS-1 model (group)	Animal	Design	Skin, nails and hair	Digestive system	Lung	Eye	Liver	Reproductive system	Mortality	Reference
C57BL/6 (Mathis)	Mouse	Cre-lox-mediated deletion of exon 2	-	✓	✓	✓	o	✓	-	Anderson et al., 2002
C57BL/6 (Peltonen)	Mouse	Neo cassette insertion in exon 6 emulating the R257X mutation	-	o	o	o	✓	✓	-	Ramsey et al., 2002b
C57BL/6 (Matsumoto)	Mouse	Neo cassette insertion replacing exons 5–12	-	✓	o	o	-	o	-	Kuroda et al., 2005
C57BL/6 (Scott)	Mouse	Cre-lox-mediated deletion of exon 8 emulating d1094-1106 del13	-	✓	o	o	-	o	-	Hubert et al., 2009
NOD/LtJ (Mathis)	Mouse	Backcrossed with *Aire*$^{-/-}$ C57BL/6	-	✓	✓	✓	✓	✓	✓	Jiang et al., 2005
BALB/cJ (Mathis)	Mouse	Backcrossed with *Aire*$^{-/-}$ C57BL/6	-	✓	✓	✓	✓	✓	-	Jiang et al., 2005
SJL/J (Mathis)	Mouse	Backcrossed with *Aire*$^{-/-}$ C57BL/6	-	✓	✓	-	✓	✓	-	Jiang et al., 2005
Brown Norway rat (Guillonneau)	Rat	ZFN to target exon 3 emulating the R139X mutation	✓	✓	✓		✓	✓	-	Ossart et al., 2018

✓ = affected; - = not affected; o = not stated.

1.4 APS-1 autoantibodies

APS-1 is characterised by the presence of high titres of autoantibodies against several TSAs, a feature that can help in diagnosing and treating the disease (Guo et al., 2018). APS-1 patients usually have different autoantibody types targeting and affecting different organs. For instance, patients with chronic mucocutaneus candidiasis caused by a mutation in *AIRE* have high titres of autoantibodies against IL-17A, IL-17F and IL-22, which are thought to be linked to the *Candida* infection (Depner et al., 2016). In addition, autoimmune disorders that affect a specific organ are often found to be linked with several organ-specific autoantibodies. For example, autoantibodies against NACHT leucine-rich-repeat protein 5 (NALP5) have been detected in more than half of all patients with hypoparathyroidism related to APS-1; however, no such autoantibodies were detected in patients without hypoparathyroidism (Alimohammadi et al., 2008). NALP5 expressed in chief cells of the parathyroid glands that subjected to high calcium concentration which may suggest that it is involved in the process of calcium sensing in chief cells (Alimohammadi et al., 2008). APS-1 can also affect the adrenal gland, resulting in Addison's disease (Uibo et al., 1994; Cihakova et al., 2001). Autoantibodies against P450 family members, 21-hydroxylase (CYPC21) and side-chain cleavage enzyme (CYPSCC) have also been found. However, autoantibodies against glutamic acid decarboxylase (GAD65), islet cell antigen (ICA) 512, zinc transporter 8 (ZnT8) and insulin are associated with type 1 diabetes (Alimohammadi et al., 2008; Fierabracci., 2016). Autoantibodies against CYPSCC and 17α hydroxylase (CYPC17) are markers for hypergonadotropic hypogonadism (Bensing et al., 2007; Perniola et al., 2021) . Some other autoantibodies can suggest Grave's disease, such as thyroglobulin antibody, thyroid peroxidase antibody and thyroid-stimulating immunoglobulins. In addition, autoimmune hepatitis can be predicted by detecting autoantibodies against cytochrome P450 1A2 (CYP1A2), aromatic L-amino

acid decarboxylase (AADC) and cytochrome P450 2A6 (CYP2A6; Björk et al. 1994; Fierabracci., 2016).

1.5 APS-1 brain autoantigens

Patients with APS-1 commonly show high levels of autoantibodies against proteins of the brain, which may aid diagnosis (Meloni et al., 2019). There are a number of autoantibodies against neurotransmitter enzymes involved in 5-hydroxytryptamine (5-TH) or catecholamine biosynthesis, such as tyrosine hydroxylase (TH), AADC and tryptophan hydroxylase (TPH) as well as GAD65, which synthesises gamma aminobutyric acid (GABA). In addition, brain antigens (BAs) that are associated with APS-1 include growth hormone (GH), growth hormone-releasing hormone (GHRH), SRY-Box Transcription Factor 10 (SOX10), SRY-Box Transcription Factor 9 (SOX9), Tudor domain-containing protein 6 (TDRD6) and luteinizing hormone (LH; Table 1.3). Moreover, a study published recently with the help of ProtoArray showed that patients with APS-1 display auto-reactivity against 100 self-proteins, indicated that many autoantigens remain unidentified (Meyer et al., 2016).

Table 1.3: APS-1 autoantigens in the brain. Different proteins targeted by autoantibodies associated with APS-1 are shown along with their function and location of expression.

Antigen	Function	Cells expressing the proteins	References
SOX9	Transcription factor; plays a role in neural crest cell specification and differentiation.	Neural crest cells, ectodermal, endodermal, and mesodermal derivatives and glial cells	(Scott et al., 2010; Haseeb and Lefebvre, 2019)
SOX10	Essential transcription factor in neuronal and glial differentiation; expressed in neural crest cells during embryonic development and in adult melanocytes.	Neural crest cells, glial cells and melanocytes	(Kim et al., 2003; Ferletta et al., 2007; Haseeb and Lefebvre, 2019)
Tyrosine hydroxylase (TH)	Converts tyrosine to dopamine to produce catecholamines.	Anterior pituitary cells	(Smith-Anttila et al., 2017)
Growth hormone (GH)	Stimulates tissue growth; plays a role in glucose and lipid homeostasis; involved in muscle mass function.	Anterior pituitary cells (somatotrophs)	(Cocco et al., 2012; Smith-Anttila et al., 2017)

Antigen	Function	Cells expressing the proteins	References
Growth hormone-releasing hormone (GHRH)	Plays an important role in pituitary somatotroph development and proliferation; stimulates the secretion of GH.	Hypothalamic GHRH neurons and pituitary somatotrophs	(Cocco et al., 2012; Kavinga et al., 2015; Novusbio.com, 2018)
Glutamic acid decarboxylase 65 (GAD65)	An enzyme that converts glutamate to γ-amino butyric acid (GABA).	Arcuate neurons in the intermediate lobe of the pituitary gland and GABA neurons	(Fetissov et al., 2009; Smith-Anttila et al., 2017)
Tryptophan hydroxylase (TPH)	An enzyme that functions as the rate-limiting factor for the synthesis of serotonin and catalyses the formation of 5-hydroxytryptophan (5-HTP) from the dietary precursor l-tryptophan.	Serotonergic neurons	(Fetissov et al., 2009; Quadros, Takahashi and Miczek, 2010)
Luteinizing hormone (LH)	During ovulation in women, it triggers ovaries to release the egg; in men, it produces testosterone from Leydig cells.	Gonadotropic cells in the anterior pituitary gland	(Cocco et al., 2012 ; Raju et al., 2013)
Tudor domain-containing protein 6 (TDRD6)	Spliceosome maturation and mRNA splicing in spermatocytes.	Anterior pituitary cells (median eminence dopaminergic nerve terminals and pituitary gonadotrophs)	(Akpınar et al., 2017; Bensing et al., 2007)
Endothelin-converting enzyme (ECE)-2	To convert large, inactive endothelins to the active form of potent vasoactive peptide endothelin-1 (ET-1)	The brain, with high levels in the cerebral cortex, cerebellum, hippocampus and hypothalamus	(Palmer et al., 2009; Smith-Anttila et al., 2017)
Aromatic L-amino acid decarboxylase (AADC)	An enzyme that converts L-dopa and 5-hydroxytryptophan to dopamine and serotonin.	Dopamine neurons in the arcuate nucleus (Arc) of the hypothalamus projecting to the median eminence (ME) and to the intermediate pituitary lobe (IPL) and 5-HT neurons of the DRN	Fetissov et al., 2009 ; Shih et al., 2013 ; Yuwen et al., 2013)

1.6 APS-1 and neurological disorders

Organ manifestations of APS-1 can occur due to autoantibody reactions against the organ's proteins, and despite the fact that many autoantibodies found in APS-1 patients target the proteins of the brain, neurological manifestations associated with APS-1 are rare (Husebye et al., 2009; Fetissov et al., 2009). APS-1 patients rarely develop pituitary manifestations (only 7% of cases have been reported) but some studies have suggested that hypopituitarism in patients with APS-1 may be underdiagnosed since these patients have complex polyendocrinopathies that affect development and growth (Betterle et al., 1998; Cocco et al., 2012; Smith-Anttila et al.,

2017). Most recently recorded APS-1 autoantigens are expressed in pituitary tissue, such as GH, LH, TH and GAD65 (Bensing et al., 2007; Cocco et al., 2005; Cocco et al., 2012; Smith-Anttila et al., 2017). APS-1 patients can develop either single or multiple pituitary hormone deficiencies, of which GH deficiency is the most reported (Ahonen et al., 1990; Betterle et al., 1998; Cocco et al., 2005).

Autoantibodies against endothelin-converting enzyme-2 (ECE-2) have been specifically detected in 46% (48 out of 104) of patients with APS-1, while they have not been detected in other autoimmune diseases, such as type 1 diabetes, Addison's disease, systemic lupus erythematosus, primary Sjogren's syndrome and lymphocytic hypophysitis, or in healthy controls (Smith-Anttila et al., 2017). ECE-2 is a zinc metalloprotease that belongs to the M13 family and is found in the trans-Golgi network and secretory vesicles. It converts large, inactive endothelins to the active form of potent vasoactive peptide endothelin-1 (ET-1; Palmer et al., 2009). It is expressed mostly in the pancreas but is also found in brain tissues, with high levels in the hippocampus, cerebral cortex, cerebellum, catecholaminergic nuclei and hypothalamus (Smith-Anttila et al., 2017). Mice can develop and live normally when ECE-2 is knocked out; however, they suffer minor learning and memory deficiencies (Rodriguiz et al., 2008; Cocco et al., 2012).

GAD65 is expressed in GABAergic neurons in the brain and pancreatic beta cells, and autoantibodies against GAD65 have been found to cause neurological manifestations such as stiff person syndrome, cerebellar ataxia, idiopathic epilepsy and limbic encephalitis (Fetissov et al., 2009). Studies in rats have shown that patients with APS-1 develop autoantibodies that react with the brains of rats, which contain AADC, TPH, GAD65 and/or TH (Cocco et al. 2005; Serguei et al., 2009). In another study, the sera of 12/14 APS-1 patients stained specific areas in the hypothalamic-pituitary axis with

neurons that have TH and GHRH. GH deficiency was detected in five of these patients and four of them had autoantibodies to one or more regions of the hypothalamic-pituitary axis. In addition, memory loss and/or anxiety disorders together with autoantibodies to GAD neurons were detected in two patients (Cocco et al., 2012).

Besides learning and memory deficiencies, a recent study observed a reduction in total brain and cerebellum volumes and correlated this reduction to the presence of autoantibodies against GAD+ and TH+ neurons (Meloni et al., 2019). This study observed and reported, for the first time, alterations in brain volume together with the presence of neuronal autoantibodies in 78.6% patients with APS-1 from Sardinia. Patients had smaller volumes of cerebellum and grey matter. They also had increased total cerebrospinal fluid in comparison with healthy people. In 11 out of 14 patients, the abnormalities of the brain were associated with autoantibodies against GAD+ and/or TH+ neurons that persisted for 10 years in seven out of the 11 patients. Interestingly, the decrease in the volume of the whole brain and cerebellum was associated with autoantibodies against GAD+ neurons, while autoantibodies against TH+ neurons were associated with total cerebrospinal fluid increase (Meloni et al., 2019). The authors suggested that the correlation between high titres of brain autoantibodies and reduction of total brain and cerebellum volumes form an immunological basis for brain abnormalities.

Finally, there are several indications that although neurological manifestations associated with APS-1 are not rare, they are underdiagnosed because the clinical manifestations vary greatly and affect different parts of the body. In addition, some clinical manifestations may not appear until the fifth decade of life (Ahonen et al., 1990; Carpino et al., 2021). Therefore, patients need lifelong follow-up to detect novel manifestations of the disease such as brain alterations.

1.7 Hypothesis and aims

APS-1 or APECED is a monogenic autosomal recessive disease that occurs due to mutations in *AIRE*. AIRE is a transcriptional regulator; it is expressed mainly in mTEC and plays an essential role in TSA expression, thereby contributing to autoreactive T cell deletion (negative selection). AIRE is also expressed in different organs such as the brain. Patients with APS-1 commonly have high levels of autoantibodies against proteins of the brain such as TPH, GH, GHRH, GAD65, SOX9, SOX10 and TH. According to recent studies, some APS-1 patients also show neurological manifestations such as stiff person syndrome, cerebellar ataxia and memory loss. Besides such learning and memory deficiencies, a recent study observed a reduction in total brain and cerebellum volumes in APS-1 patients and correlated this reduction to the presence of autoantibodies against GAD+ and TH+ neurons (Meloni et al., 2019). To support our hypothesis suggesting the presence of neurological modifications *Aire*$^{-/-}$ C57BL/6 mice observed in APS1 patients and the possible involvement of brain antigen in these modifications. We first aim to confirm that AIRE is involved in the expression of the genes encoding brain autoantigens (TPH, GH, GHRH, GAD65, SOX9, SOX10 and TH) then we will measure the total brain and cerebellum volumes and analysed brain tissue for possible lymphocytic infiltration. Thus, the broad aim of this project was to establish the role of AIRE in the pathogenesis of neurological disorders associated with APS-1, with the following specific objectives:

1. To measure the weights of the thymus and brain of wild-type (WT), heterozygous (Het) and knockout (KO) C57BL/6 mice at different developmental stages to identify any alterations in these parameters in *Aire*$^{-/-}$ mice.

2. To evaluate lymphocytic infiltration in the brain by haematoxylin and eosin (H&E) staining.

3. To validate the expression of AIRE and candidate genes in the thymus and brain of C57BL/6 mice (APS-1 model) by RT-PCR.

4. To test whether AIRE regulates the expression of genes encoding brain autoantigens using *Aire*$^{+/+}$ and *Aire*$^{-/-}$ C57BL/6 mice by RT-qPCR.

5. To test whether AIRE regulates the expression of genes encoding brain autoantigens using human mTEC (cell line 4D6) by RT-qPCR and dual luciferase assay.

6. To test the effect of common APS-1 mutations on the expression of AIRE-dependent genes by dual luciferase assay.

7. To define the role of AIRE as a DNA-binding element by ChIP-seq.

Chapter Two

Materials and Methods

2. Material and method

2.1 Mouse genotyping

Mice were genotyped to obtain 60 individuals with different *Aire* genotypes (20 *Aire*$^{+/+}$, 20 *Aire*$^{+/-}$ and 20 *Aire*$^{-/-}$) on a C57BL/6 background that generated by Mathis group (Anderson et al., 2002). Mice were grouped by age (5, 15, 30 and 38 weeks) and each age group contains 15 mice (5 *Aire*$^{+/+}$, 5 *Aire*$^{+/-}$ and 5 *Aire*$^{-/-}$) to test the expression of *Aire* and candidate genes in the thymus and brain in addition to investigating changes in brain morphology and volume in *Aire*$^{-/-}$ mice. Genotyping was done by PCR using the KAPA Mouse Genotyping Kit (KAPA Biosystems). DNA was extracted from ear clippings in a 100-µl solution containing 10 µl 10X KAPA Express Extract Buffer, 2 µl KAPA Express Extract Enzyme and 88 µl PCR-grade water incubated at 75 °C for 10 min and inactivated at 95 °C for 5 min. Three primers were used to amplify *Aire*, with reverse primers specific to either the mutant or wild-type gene (Table 2.1).

Table 2.1: Primers used to genotype *Aire* mice.

Primer	**Sequence**
***Aire* forward**	AGACTAGGTGTTCCCTCCCAACCTCAG
***Aire* wild-type reverse**	GGAGACTTGCCTATTCCTGTC
***Aire* mutant reverse**	CCGGCGGATTTGTCCTAC

For the PCR, 1 µl of the DNA template was separately mixed with 0.5 µM each of the forward and reverse (wild-type and mutant) primers and 12.5 µl 2X KAPA2G Fast (HotStart) Genotyping Mix with dye and diluted to a total volume of 25 µl with water. The PCR protocol was as follows: 120 s at 95 °C, 35 cycles of 15 s at 95 °C, 15 s at 60 °C and 20 s at 72 °C, and an additional 2 min at 72 °C followed by cooling to 4 °C

for permanent storage. The PCR product was subjected to electrophoresis at 120 V for 60 min on a 1.5% agarose gel with 0.5 µg/µl ethidium bromide. While the wild-type mice were expected to display one band at 195 bp and the homozygous transgenic mice were expected to display one band at 140 bp, the heterozygous mice were expected to display both bands.

2.2 Histological analysis

2.2.1 Tissue collection

Mice were euthanized by intraperitoneal injection of 20% w/v sodium pentobarbital solution (500 mg/kg, JML) and the tissues were collected immediately and fixed using 4% paraformaldehyde (PFA) at 4 °C overnight. After 24 h, the tissues from were transferred to phosphate buffered saline (PBS), then embedded in paraffin and sectioned at 5 µm and at least 10 sections (10 slides) were obtained from each tissue.

2.2.2 Immunofluorescence

For AIRE staining, thymus and brain sections from 5-week-old *Aire*$^{+/+}$ and *Aire*$^{-/-}$ C57BL/6 mice were incubated twice in xylene for 10 min, rehydrated in 100%, 95% and 70% ethanol, washed in water and incubated for 20 min in 3% hydrogen peroxide/methanol to block endogenous peroxidase. All slides were then subjected to heat-induced antigen retrieval by the pressure cooker method with citrate buffer solution (pH 6.0). The sections were then washed for 1 min in water and a further 5 min in PBS prior to incubation in PBS with 0.3% Triton X-100 for 10 min for permeabilization. Thereafter, the sections were incubated with protein block serum-free buffer (DAKO) at room temperature for 10 min to block any non-specific binding. This was followed by incubation with a 1:50 dilution of polyclonal goat anti-AIRE D-17 antibody (sc-017986, SantaCruz Biotechnology) in PBS with 0.15% Triton X-100 at 4 °C overnight. Then, the sections were washed with PBS three times (10 min per wash)

prior to the addition of the secondary antibody (Alexa Fluor 488 donkey anti-goat IgG antibody, Invitrogen) diluted 1:500 in PBS, incubated for 1 h at room temperature, and washed again with PBS three times (10 min per wash). The slides were mounted using VECTASHIELD Antifade Mounting Medium with 4′,6-diamidino-2-phenylindole (DAPI; Vector Labs) for nuclear staining and cured at room temperature overnight. A Nikon A1 confocal microscope was used for imaging.

2.2.3 H&E staining

APS-1 mouse model thymus and brain paraffin sections were incubated twice in xylene for 10 min and rehydrated in 100%, 95% and 70% ethanol and water (10 min each). Thereafter, the slides were incubated in haematoxylin for 3 min and rinsed with deionised water for 5 min. Then, the slides were incubated in eosin for 30 s, followed by 10 min in 95% ethanol, 10 min in 100% ethanol and 15 min in xylene. Finally, one drop of Permount mounting medium was placed on each slide and covered using a coverslip. Pictures were taken using a slide scanner.

2.3 Nucleic acid techniques

2.3.1 Total RNA extraction and cDNA synthesis

To test the ability of AIRE to regulate candidate gene expression, it was crucial to first validate the expression of candidate genes in the thymus and brain of mice and then investigate their expression levels by RT-qPCR. In addition, RNA was extracted from 4D6 cells to test the expression of the candidate genes in the presence and absence of AIRE in vitro. To perform RT-PCR and RT-qPCR, total RNA was extracted using TRI Reagent® (Sigma-Aldrich) in accordance with the manufacturer's instructions. TRI reagent is a monophasic solution of guanidinium isothiocyanate and phenol that lysis the cell and phase separation was achieved by adding 1-bromo-3-chloropropane to extract RNA. Mouse tissue was homogenized in liquid nitrogen using a mortar and

pestle, and 2 ml TRI Reagent was added to powdered brain tissue and 1 ml to one lobe powdered thymus tissue. 4D6 cells were washed with cold PBS and pelleted in 1.5-ml microcentrifuge tubes. Then, 200 µl TRI Reagent was added, and the mixture was incubated for 5 min at room temperature. Thereafter, 100 µl of 1-bromo-3-chloropropane was added per ml of TRI Reagent to the powdered tissues and cell pellets. The samples were then shaken vigorously for 15 s and incubated for 15 min at room temperature. After incubation, the samples were centrifuged at 12,000 × g for 15 min at 4 °C. After phase separation, the aqueous phase containing the RNA was transferred to a clean tube and the RNA precipitated by the addition of isopropanol (0.5 ml per ml of TRI Reagent). The RNA was then pelleted, washed with 75% ethanol, and resuspended in 30 µl RNase-free water. The RNA concentration was then measured using a spectrophotometer (NanoDrop 1000 Spectrophotometer version 3.8.1). Reverse transcription of RNA and synthesis of cDNA were performed using the SuperScript® IV First-Strand Synthesis System kit (Invitrogen) using random hexamers in accordance with the manufacturer's instructions. Reverse transcription was done using 700 ng of RNA and then 1 ul of cDNA was used to run the PCR of cDNA.

2.3.2 PCR of cDNA

Platinum® *Taq* DNA Polymerase High Fidelity (Life Technologies) was used in accordance with the manufacturer's protocol to amplify the desired genes (*Aire* and the candidate genes) via PCR with cDNA templates. Primers (Table 2.2) were designed using Primer3Plus (https://primer3plus.com/cgi-bin/dev/primer3plus.cgi) and chromosomal positions were identified using UCSC Genome Browser on Mouse (GRCm39/mm39; https://genome-euro.ucsc.edu/cgi-bin/hgGateway). 50−200 ng cDNA was mixed with 0.1 µl DNA polymerase (0.5 units), 5 µl 10x High Fidelity PCR

buffer, 1 µl dNTP mix (10 mM) and 1 µl of each primer (10 mM), and the mixture was made up to a total volume of 25−50 µl (depending on the template) using distilled water. After gentle mixing, the reaction mix was run in the Thermal Cycler. After an initial heating step at 95 °C for 2 min, the mixture was subjected to 35 cycles of 30 s at 95 °C, 30 s at 60 °C and 1 min at 72 °C, and an additional 1 min at 72 °C followed by cooling to 4 °C for permanent storage.

Table 2.2: cDNA primer sequences.

Gene	Primer sequence	Chromosomal positions	Recognised exon number	Size (bp)
Aire	Fwd 5' GTGGCCATAGACAGTGCCTT 3'	chr10:77879277-77879297	1	507
	Rev 5' GACAGAAGCTGCCATGGTCT 3'	chr10:77877421-77877441	6	
Gad65	Fwd 5' GAGTGGAGTAGAGAGGGCCA 3'	Chr2: 22563953-22563973	11	562
	Rev 5' GTGAGTTGCTGCAGGGTTTG 3'	chr2: 22580284-22580304	16	
Gh	Fwd 5' TGCTTGGCAATGGCTACAGA 3'	Chr11: 106192373-106192393	1	583
	Rev 5' TCTTGAAGCAGGAGAGCAGC 3'	chr11: 106191268-106191288	5	
Ghrh	Fwd 5' GAGCAGAACCTCAATCGGAG 3'	chr2: 157179297-157179317	1	403
	Rev 5' GGTACAGTTGTGTTTGGGGC 3'	chr2: 157171487-157171507	5	
Tph1	Fwd 5'AGAAGCCACCAAGACTCAGC 3'	Chr7: 46296204-46296224	11	591
	Rev 5'AGCCCTCTCTCTTACCCTGG 3'	Chr7: 46295633-46295653	11	

Gene	Primer sequence	Chromosomal positions	Recognised exon number	Size (bp)
Th	Fwd 5'GGACCACCAGCTTGCACTAT 3'	Chr7: 142453696-142453716	1	512
	Rev 5'GGAACCTTGTCCTCTCTGGC 3'	Chr7: 142450395-142450415	4	
Sox9	Fwd 5'CCAGCAAGAACAAGCCACAC 3'	Chr11: 112673702-112673722	1	545
	Rev 5'GCTCAGTTCACCGATGTCCA 3'	Chr11: 112675629-112675649	3	
Sox10	Fwd 5'CTACAAGAGTGCCCACCTGG 3'	Chr15: 79043394-79043414	3	522
	Rev 5'TAGGCGATCTGGGAAGTGGA 3'	Chr15: 79040399-79040419	3	

2.3.3 Gel electrophoresis

DNA was separated on 1−3 % (w/v) agarose gels (Bioline) using 1X TAE buffer (ThermoFisher Scientific) and 6X loading dye (New England BioLabs). The gels were run at 80-120 V. The DNA bands were visualised by adding 3.5 µl of 10 mg/ml ethidium bromide (Sigma-Aldrich) per 100 ml of gel. A 2-log DNA ladder (New England BioLabs) was run simultaneously as a size marker. The gels were imaged on a UGENIUS system ultraviolet transilluminator (Syngene).

2.3.4 RT-qPCR

PCR was performed in triplicate for each sample in a 384-well plate using the 7900HT Real-time PCR system (Applied Biosystems). Primers (Tables 2.3 and 2.4) were designed using the National Center for Biotechnology Information (NCBI) resource (https://www.ncbi.nlm.nih.gov/). Primer pairs sit on different exons, to avoid amplification of genomic DNA (Table 2.3 & 2.4). The chromosomal positions of the primers (Tables 2.3 and 2.4) were identified using UCSC Genome Browser on Mouse

(GRCm39/mm39; https://genome-euro.ucsc.edu/cgi-bin/hgGateway) or using UCSC Genome Browser on Human (GRCh38/hg38; https://genome-euro.ucsc.edu/cgi-bin/hgGateway). The reaction mixture consisted of SYBR™ Green PCR Master Mix (Applied Biosystems), cDNA and 600 nM forward and reverse primers made up to a total volume of 10 µl. After an initial heating step at 95 °C for 10 min, the mixture was subjected to 40 cycles of 30 s at 95 °C and 60 s at 60 °C, followed by a further 60 s at 95 °C and then the dissociation curve (30 s at 60 °C and 30 s at 95 °C). When the RT-qPCR run was completed, raw data were collected using the Applied Biosystems SDS2.4 software. The $\Delta\Delta C_t$ method was used to calculate changes in relative expression between wild-type and knockout mice or between cells expressing and lacking AIRE. β-Actin expression was used as a reference to normalize the C_t values for each RNA sample.

Table 2.3: Primer sequences for RT-qPCR of mouse samples.

Gene	Primer sequence	Chromosomal positions	Recognised exon number	Size (bp)
Aire	Fwd 5' AGTTCGAAGACCCCAGTGGC 3'	chr10: 77875842 - 77875862	6	199
	Rev 5' ACGGCACACTCATCCTCGTT 3'	chr10: 77873799-77873819	8	
Gad65	Fwd 5' CAGCTGGAACCACCGTGTAT 3'	Chr2: 22558286 - 22558306	10	135
	Rev 5' TCCACTTGTGTTTCCGGGAC 3'	chr2: 22559738-22559758	11	
Gh	Fwd 5' TGGCTGCTGACACCTACAAAGA 3'	Chr11: 106192225-106192247	2	191
	Rev 5' CAGCCATGACTGGATGAGCAG 3'	chr11: 106191704 - 106191725	3	

Gene	Primer sequence	Chromosomal positions	Recognised exon number	Size (bp)
Ghrh	Fwd 5' TCCTGAGCCAGCTGTATGCC 3'	chr2: 157175332 - 157175352	3	148
	Rev 5' TGCAAGATGCTCTCCAGGGT 3'	chr2: 157173680- 157173700	4	
Tph1	Fwd 5' CCCAGCAAGGACGGGATCAA 3'	Chr7: 46316703- 46316723	1	102
	Rev 5' TCCCTCTTTCGGAGGAATGGT 3'	Chr7: 46314667 - 46314688	2	
Th	Fwd 5' CAGGATACCAAGCAGGCCGA 3'	Chr7: 142453618 - 142453638	1	158
	Rev 5' ACCACAGCCTCCAATGGGTT 3'	Chr7: 142451799 - 142451819	2	
Sox9	Fwd 5' CCACATTCCTCCTCCGGCAT 3'	Chr11: 112674836 - 112674856	2	101
	Rev 5' ACGTCGGTTTTGGGAGTGGT 3'	Chr11: 112675526 - 112675546	3	
Sox10	Fwd 5' GTCAACGGTGCCAGCAAGAG 3'	Chr15: 79047617 - 79047637	2	182
	Rev 5' ATGAAGGGGCGCTTGTCACT 3'	Chr15: 79043570 - 79043590	3	
Ins2	Fwd 5' GCTTCTTCTACACACCCATGTC 3'	Chr7: 52253200 - 552253222	1	147
	Rev 5' AGCACTGATCTACAATGCCAC 3'	Chr7: 142232496 - 142232517	2	
Foxn1	Fwd 5' TTCCATCAGTACTCCCCGGGTGG 3'	Chr11: 78257805- 78257825	4	95
	Rev 5' GCGTTGGCCTGGGGTGCAAT 3'	Chr11: 78257737 - 78257757	5	
	Fwd 5' GGCTGTATTCCCCTCCATCG 3'	Chr5: 142891338 - 142891358	2	154

Gene	Primer sequence	Chromosomal positions	Recognised exon number	Size (bp)
Actb (β-actin)	Rev 5' CCAGTTGGTAACAATGCCATGT 3'	Chr5: 142891117 - 142891139	3	

Table 2.4: Primer sequences for RT-qPCR of 4D6 cells.

Gene	Primer sequence	Chromosomal positions	Recognised exon number	Size (bp)
AIRE	Fwd 5' TCGGGAACGGGATTCAGACC 3'	Chr21: 44288343-44288363	4	165
	Rev 5' TGTAGAACTCCCCGCCAACC 3'	Chr21: 44289678-44289698	6	
GAD65	Fwd 5' TCCCTCAAATGCTCTGGGGC 3'	Chr10: 26216486-26216506	1	146
	Rev 5' GCGTGTGTGTATGCGAGCTG 3'	Chr10: 26216612-26216632	2	
GH	Fwd 5' CCTCTGACAGCAACGTCTATGA 3'	Chr17: 63917802-63917824	4	152
	Rev 5' TGCGTCATCGTTGTGTGAGT 3'	chr17: 63917419-63917439	5	
GHRH	Fwd 5' GATGCGGCGGTATGCAGATG 3'	chr20: 37256478-37256498	3	154
	Rev 5' TACCTGACGACCAAGCCGTG 3'	chr20: 37254280-37254300	4	
TPH	Fwd 5' CTACCCAACCCATGCTTGCAG 3'	Chr11: 18029259-18029280	5	148
	Rev 5' AGCCACAGGACGGATGGAAA 3'	Chr11: 18026593-18026613	6	

Gene	Primer sequence	Chromosomal positions	Recognised exon number	Size (bp)
TH	Fwd 5' GTGCAGCCCTACCAAGACCA 3'	Chr11: 2165291-2165311	13	130
	Rev 5' ACGGGTCGAACTTCACGGAG 3'	Chr11: 2164342-2164362	14	
SOX9	Fwd 5' CAGTACCCGCACTTGCACAAC 3'	Chr17: 72121766-72121787	1	183
	Rev 5' CTGCCCGTTCTTCACCGACTT 3'	Chr17: 72122824-72122845	2	
SOX10	Fwd 5' CCGAGCAAGGTGGGACCG 3'	Chr22: 37977954-37977932	2	159
	Rev 5' CTGTCTTCGGGGTGGTTGGAG 3'	Chr22: 37974153-37974174	3	
INS	Fwd 5' TACACACCCAAGACCCGCC 3'	Chr11: 2160804-2160824	3	134
	Rev 5' TGTTCCACAATGCCACGCTTC 3'	Chr11: 2159903-2159924	4	
FOXN1	Fwd 5' AGTCTGACGTCACGCTGCC 3'	Chr17: 28523991-28523011	2	146
	Rev 5' AGGGCCGTCGGACACAAATG 3'	Chr17: 28524527-28524547	3	
ACTB (β-actin)	Fwd 5' CCTCGCCTTTGCCGATCC 3'	Chr7: 5530541-5530559	1	71
	Rev 5' GCGCGGCGATATCATCATCC 3'	Chr7: 5529635-5529655	2	

2.4 Protein analysis

2.4.1 Protein extraction

In this project, proteins were extracted from both 4D6 cells and mouse tissue. The brain and thymus tissues from *Aire* mice were homogenized in liquid nitrogen using a

mortar and pestle, while the cells were washed with cold PBS and pelleted in 1.5-ml microcentrifuge tubes prior to the addition of radio immunoprecipitation assay (RIPA) buffer (50 mM Tris-HCl (pH 7.4, Sigma-Aldrich), 150 mM NaCl (Sigma-Aldrich), 2mM ethylenediaminetetraacetic acid (EDTA), 1% Triton X-100 (Sigma-Aldrich), 0.5% sodium deoxycholate and 0.1% sodium dodecyl sulphate (SDS, Sigma-Aldrich) supplemented with 1% protease inhibitor cocktail (Sigma-Aldrich)). For lysis, the cell pellets and powdered tissues were suspended in 200−500 µl RIPA buffer and incubated at 4 °C with mixing for 30 min followed by 10 min of centrifugation at 13,000 × g at 4 °C. The supernatant containing the protein was then isolated and stored in a labelled 1.5-ml Eppendorf tube.

2.4.2 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

To perform SDS-PAGE, a 10% gel divided into two layers (4% stacking gel on top and 10% separating gel on the bottom) was made. To make the 10% separating gel, the solutions listed in Table 2.5 were mixed and 9 ml of the mixture was poured into a gel cassette and allowed to solidify. Thereafter, a stacking gel solution prepared by mixing the solutions listed in Table 2.6 was poured over the separating gel in the cassette.

Table 2.5: Separating gel composition.

Reagent	**Volume**
Water	4 ml
30% acrylamide (Protogel)	3.35 ml
Separating buffer (1.5 M Tris pH 8.8)	2.5 ml
10% Sodium dodecyl sulfate (SDS; Sigma)	100 µl
10% Ammonium Persulfate Solution (APS; Sigma)	50 µl
Tetramethylethylenediamine (TEMED; Sigma)	10 µl

Table 2.6: Stacking gel composition.

Reagent	**Volume**
Water	3 ml
30% acrylamide (Protogel)	650 µl
Stacking buffer (0.5 M Tris pH 6.5)	1.75 ml
10% SDS (Sigma)	50 µl
10% APS (Sigma)	50 µl
TEMED (Sigma)	5 µl

Protein samples were denatured by heating with Laemmli buffer (4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromphenol blue and 0.125 M Tris HCl, pH approx. 6.8; Sigma-Aldrich) at 95 °C for 5 min and loaded onto the 10% SDS-PAGE gel. The gel was then subjected to electrophoresis in running buffer (25 mM Tris base (Sigma), 250 mM glycine (Sigma) and 0.1% SDS (Sigma)) for 20 min at 80 V followed by 60–90 min at 100 V. The separated proteins were transferred onto a polyvinylidene difluoride (PVDF) membrane (Thermo Fisher) using an iBlot machine as per the manufacturer's protocol for immunodetection. Blocking was performed at room temperature for 60 min with 5% dried milk powder in 5 ml TBST buffer (50 mM Tris base (pH 7.5), 150 mM NaCl and 0.1% Tween) on a roller (Milton). Solutions of primary and secondary antibodies were prepared in 5% milk in 5 ml TBST blocking buffer. All primary antibodies were incubated with the samples at 4 °C overnight with agitation. Horseradish peroxidase (HRP)-conjugated secondary antibodies were added at a 1:5000 dilution and the samples were incubated at room temperature for 1 h on the roller. After each step, the membranes were subjected to three 10-min washes with TBST at room-temperature on the roller to remove any unbound antibodies. Protein chemiluminescence was detected via HRP using the ECL Plus chemiluminescence

detection kit (GE Healthcare) and visualised using the G-BOX image capture system (Syngene).

2.5 Cloning of DNA fragments

2.5.1 Primer design and genomic DNA extraction

Promoters of the candidate genes as well as a positive and negative control were identified using the Eukaryotic Promoter Database (EPD). Primers were designed using Primer3plus and DNA was extracted from human HEK293 cells to amplify the desired promoters using the Quick-gDNA mininPrep kit (ZYMO RESEARCH) as follows. Cells were centrifuged at 1500 rpm for 5 min, then 300 µl of genomic lysis buffer was added to the pellet, mixed briefly by vortexing, and let stand at room temperature for 7 min. Thereafter, the mixture was transferred to a Zymo-Spin Column in a collection tube and centrifuged for 1 min at 13,000 rpm. Then, 200 µl of pre-wash buffer was added to the spin column and centrifuged for 1 min at 13,000 rpm. To wash the DNA, 500 µl of g-DNA wash buffer was added and centrifuged for 1 min at 13,000 rpm. Finally, the spin column was transferred to a clean 1.5-mL Eppendorf tube, and 30 µl of water was added to the spin column and centrifuged for 1 min at 13,000 rpm to elute the DNA.

2.5.2 Purification of DNA from agarose gels

The candidate promoters were amplified by PCR (program in Section 2.2.1) and separated on agarose gels. DNA was extracted from the gel using the QIAquick™ Gel Extraction Kit (QIAGEN) in accordance with the manufacturer's instructions. Thus, Buffer QG 1 was added to the gel at a volume ratio of 3:1 and incubated for 10 min at 50 °C to allow dissolution of the gel slice. Isopropanol (Sigma-Aldrich) was then added to the gel solution in a 1:1 volume ratio and the solution was transferred to a 2-ml

collection tube, washed with Buffer PE and eluted from the column with 30 µl distilled water.

2.5.3 Restriction endonuclease digestion of DNA

All PCR products and the pGL3-Basic expression vector (Promega) were subjected to restriction digestion reactions (KpnI and HindIII or KpnI and NheI). Different digestion enzymes were used because some of the digestion enzymes are able to digest some of the promoters' sequences so choosing the digestion enzymes was depending on their ability to cut the inserts. To digest 30 µl of the PCR product, the DNA was mixed with 5 µl of 10X CutSmart™ buffer (New England BioLabs), 1 µl of each digestion enzyme and 13 µl water. The reaction was incubated at 37 °C for 3–4 h. 3 µl of the vector was digested with 1 µl of each restriction enzyme, 2.5 µl 10X CutSmart™ buffer (New England BioLabs) and 18.5 µl water.

2.5.4 Ligation of DNA sequence with a vector

The digested inserts and vector were ligated using T4 DNA ligase (Promega) by the reaction described in Table 2.7.

Table 2.7: Ligation of insert with Pgl3 vector.

Vector DNA	100 ng
Insert DNA	17 ng
Ligase 10X buffer	1 µl
T4 DNA ligase	0.5 µl
Water	Up to 10 µl

Table 2.8: List of plasmids used in this project.

Plasmid	Company	Catalogue number
pGL3-Basic vector	Promega	E1751
pBIND	Promega	E2440
AIRE-pBIND	Provided by the lab (*see appendix: figure 8.2*)	-
pCMV3-AIRE-CMYC	Sino Biological	HG17322-CM
Pcmv3-c-myc negative control vector	Sino Biological	CV014

2.5.5 Transformation of competent cells

One Shot® Mach1™ T1R NEB® 5-alpha Competent *E. coli* cells (ThermoFisher Scientific, New England BioLabs) were completely thawed on ice. A 50-µl aliquot was incubated with 4 µl transfected DNA solution for 30 min on ice, then subjected to heat shock in water at 42 °C for 30 s, followed by an additional 2 min on ice. 300 µl of super optimal broth (SOC) medium (Thermo Fisher) was added and the mixture incubated at 37 °C in a shaker at 225 rpm for 60 min. The mixture was then applied to pre-warmed lysogen broth (LB) agar plates containing 100 µg/ml ampicillin and incubated at 37 °C overnight to permit the growth of antibiotic-resistant colonies. The following day, an isolated colony was picked and transferred to 5 ml LB broth and incubated overnight at 37 °C in a shaker at 225 rpm.

2.5.6 Isolation of plasmid DNA

The QIAprep Spin Miniprep kit (QIAGEN) was used to extract plasmid DNA in accordance with the manufacturer's instructions. Briefly, the bacterial cells were pelleted and resuspended in 250 µl buffer P1 prior to the addition and thorough mixing

of 250 µl buffer P2. Then, 350 µl Buffer N3 was added and rapidly mixed by inverting the tube four to six times followed by 10 min of centrifugation at 13,000 rpm. The supernatant was then transferred to the supplied spin columns, centrifuged at 13,000 rpm for 1 min, and the flow-through discarded. After washing twice with Buffer PB, the plasmids were eluted in distilled water. The Nanodrop1000 (ND-1000 v3.2.1) was used to measure the absorbance and the software (Labtech) then calculated the plasmid DNA concentration. The plasmid purity and insert size were identified following restriction digestion and electrophoresis on a 1% agarose gel, stained with ethidium bromide and visualised using the ultraviolet transilluminator.

2.5.7 DNA sequence analysis

The isolated DNA was sequenced to confirm integration of the fragments of interest and the absence of mutations. Sequencing reactions were performed on purified DNA (100 ng/µl) using the Applied Biosystems 3730 DNA analyser at the Core Genomic Facility, University of Sheffield. Following confirmation by sequencing reaction, the isolated plasmids were stored at −70 °C with glycerol (Sigma-Aldrich) at a 1:1 ratio.

2.6 Cell culture

The 4D6 cell line (provided by Dr. Matthieu Giraud, Centre for Research in Transplantation and Immunology, Nantes, France) was originally derived from human thymic medullary epithelium from a child undergoing cardiac surgery. The cells were maintained in RPMI 1640 medium with L-glutamine (Lonza) supplemented with 10% foetal calf serum (Lonza) and 1% penicillin-streptomycin antibiotic.

2.7 Dual luciferase assay

A dual luciferase reporter assay (Promega) was used to test the ability of AIRE to activate the promoter of candidate genes. The vectors used to perform the assay included the pGL3-Basic vector that contains a modified coding region for firefly

(*Photinus pyralis*) luciferase and the pBIND vector that expresses the *Renilla reniformis* luciferase to normalise for differences in transfection efficiency. First, 4D6 cells were seeded in 12-well plates. The following day, half of the wells were co-transfected with AIRE-pBIND and the pGL3-Basic vector containing the promoters of candidate genes, a positive control (*INS2*) and a negative control (*FOXN1*). The other half was co-transfected with pBIND and the pGL3-basic vector containing the promoters of candidate genes, positive control (*INS2*) and negative control (*FOXN1*) to observe differences in expression with or without AIRE. After 48 h, the cells were lysed by adding 200 µl of 1X passive lysis buffer and incubated for 15 min at room temperature with mixing. Thereafter, the firefly luciferase signal was generated and measured by adding luciferase assay reagent II to produce luminescence. Then, Stop & Glo® reagent was added to the same samples to measure the Renilla luciferase signal.

2.8 Site-directed mutagenesis

Mutations were introduced in the AIRE gene to test the expression of AIRE-dependent genes (*SOX9* and *TH)* in the presence of wild-type and mutated AIRE. The QuickChange Lightning Site-directed Mutagenesis Kit (Agilent Technologies) was used according to the manufacturer's instructions to introduce the mutations responsible for APS-1 in Finns, Sardinians and Iranian Jews, i.e., c.769C>T (R257X), c.415C>T (R139X) and c.254A>G (Y85C), respectively. To design the primers (Table 2.9), the QuickChange Primer Design program available on the Agilent Technology website was used.

Table 2.9: Primers for mutations causing APS-1 in Finns, Sardinians and Iranian Jews.

Mutation	**Primer sequence**
c.769C>T (R257X)	Fwd 5' CTCCCTTGGCTCAAACCAGAGGCTTCGGG 3' Rev 5' CCCGAAGCCTCTGGTTTGAGCCAAGGGAG 3'
c.415C>T (R139X)	Fwd 5' GGCGCGGCAGCTCAAGCCTCTTCTGAG 3' Rev 5' CTCAGAAGAGGCTTGAGCTGCCGCGCC 3'
c.254A>G (Y85C)	Fwd 5' ATAGCGCTCCAGGTTGCAGTCCTTGAACAGCAC 3' Rev 5' GTGCTGTTCAAGGACTGCAACCTGGAGCGCTAT 3'

To introduce the mutations, a supercoiled double-stranded DNA plasmid with an insert of interest (pBIND AIRE) and primers including the desired mutations were mixed with reaction buffer, dNTP mix and P*fu* Ultra high-fidelity (HF) DNA polymerase DNA polymerase to be run in a thermal cycler. After an initial heating step at 95 °C for 2 min, the mixture was subjected to 17 cycles of 20 s at 95 °C, 10 s at 60 °C, 4 min at 68 °C, and an additional 5 min at 68 °C followed by cooling to 4 °C for permanent storage.

Following the reaction, 2 µl of Dpn I restriction enzyme was added to each sample and incubated at 37 °C for 5 min to digest the non-mutated supercoiled double-stranded plasmid DNA. Thereafter, transformation and isolation of plasmid DNA was done as described in Sections 2.5.5 and 2.5.6. Finally, the samples were sequenced to verify the successful introduction of the mutations.

2.9 Chromatin immunoprecipitation (ChIP) assay

4D6 cells (5×10^5 cells) were seeded in three 35-mm dishes one day before transfection. The cells in two dishes were transfected with 3.3 µg of the pCMV3-AIRE-

CMYC plasmid and the cells in the third dish were transfected with the Pcmv3-c-myc negative control vector using 9.9 µl of FuGENE® HD (Promega) at a ratio of 3:1 according to the manufacturer's protocol. The following day, when the number of cells reached around 4 × 10^6 cells per dish, the ChIP assay was done using the Pierce Magnetic ChIP Kit (Thermo Fisher) according to the manufacturer's protocol. The ChIP procedure involves the following four steps:

A- Crosslinking and cell pellet isolation

To each dish containing around 4 × 10^6 cells and 2 ml cell culture medium, 20 µl of 16% formaldehyde was added and mixed well by gently swirling the dish, and the dishes were incubated for 10 min at room temperature in a chemical fume hood. Thereafter, 200 µl of 10X Glycine Solution was added to each dish. The mixture was mixed well by gently swirling the dish and then incubated at room temperature for 5 min in the chemical fume hood. Thereafter, the mixture was removed by aspiration and then the cells were washed twice with 4.2 ml ice-cold PBS. After removing the PBS by aspiration, 30 µl of the Halt Cocktail was added to 3 ml of ice-cold PBS, 1 ml of this solution was added to each dish, and the cells were detached by scraping. The cells were transferred to 1.5-mL microcentrifuge tubes and centrifuged at 3000 × g for 5 min to remove the PBS.

B- Lysis and MNase digestion

The cell pellet was resuspended with 200 µL of Membrane Extraction Buffer containing protease/phosphatase inhibitors (2 µl Halt Cocktail in 200 µL Membrane Extraction Buffer) to break up the pellet. The tubes were vortexed and incubated on ice for 10 min. The tubes then were centrifuged at 9000 × g for 3 min, and the supernatant was removed. Thereafter, nuclei were resuspended in 200 µL MNase Digestion Buffer Working Solution (0.21 µL of 1M DTT in 210 µL of MNase Digestion Buffer). ChIP-

grade MNase (10 U/µl) was diluted by adding 0.5 µl to 4.5 µl MNase Digestion Buffer Working Solution, and 2 µl of the diluted MNase was added to each tube, vortexed and incubated in a 37 °C water bath for 15 min, mixing by inversion every 5 min. To stop the reaction, 20 µL of MNase Stop Solution was added, vortexed and incubated on ice for 5 min. The nuclei were recovered by centrifuging at 9000 × g for 5 min and removing the supernatant. The nuclei were resuspended in 500 µL of 1X IP Dilution Buffer containing protease/phosphatase inhibitors (100 µL of 5X IP Dilution/Wash Buffer and 5 µL of Thermo Scientific Halt Cocktail in 395 µL nuclease-free water). Sonication was done on ice with several pulses using MSE Soniprep 150 plus Ultrasonic Disintegrator (MSS150-CX4-1) to break the nuclear membrane, and the tubes were incubated for 20 s on ice between pulses.

C- Immunoprecipitation

After sonication, 10 µl of the supernatant containing the digested chromatin (input control) was transferred to a 1.5-ml tube and stored at −20 °C, and after thawing 90 µl of the supernatant was added to 410 µl of 1X IP Dilution Buffer for immunoprecipitation. Thereafter, 1.5 µl of primary antibody (c-Myc monoclonal antibody (9E10), ThermoFisher) was added to the samples. The IP reaction was incubated overnight at 4 °C with mixing. The following day, the tubes were vortexed and 20 µl of the magnetic beads was added to each IP sample and incubated for 2 h at 4 °C with mixing. Then, the beads were collected with a magnetic stand and the supernatant was discarded. Then, 1 mL of IP Wash Buffer 1 was added, vortexed briefly and incubated for 5 min with mixing, and the beads were collected with a magnetic stand. The wash step with IP Wash Buffer 1 was repeated three times, followed by a final wash step with IP Wash Buffer 2.

D- IP elution and DNA recovery

After washing with IP Wash Buffers 1 and 2, 150 µL of 1X IP Elution Buffer was added to the washed beads and incubated at 65 °C for 30 min with vigorous shaking. The magnetic beads were collected with the magnetic stand and the supernatant (containing the eluted protein-chromatin complex) was removed and mixed with NaCl and proteinase K in 1.5-mL centrifuge tubes. Thereafter, the IP samples were vortexed and incubated in a 65 °C heat block for 1.5 h. Then, 750 µl of DNA Binding Buffer was added to each eluted IP sample and transferred to a DNA Clean-Up Column inserted into a 2-mL collection tube, which was centrifuged at 10,000 × g for 1 min before discarding the flow-through. Thereafter, 750 µl of DNA Column Wash Buffer was added, the column was centrifuged at 10,000 × g for 1 min, the flow-through was discarded and the column was centrifuged again for 2 min. Finally, the DNA was purified by adding 50 µL of DNA Column Elution Solution directly into the centre of each column and centrifuging at 10,000 × g for 1 min.

Chapter Three

Manifestations of the brain in *Aire*$^{-/-}$ mice and the role of AIRE in the expression of genes encoding brain autoantigens associated with APS-1

3.1 Outline

In this chapter, the histological and molecular characteristics of the brain and thymus of *Aire*$^{-/-}$ mice were compared with those of wild-type mice. The accumulation of infiltrating lymphocytes in a specific organ can lead to tissue destruction and atrophy (Kluger et al., 2012). Therefore, lymphocytic infiltration was assessed in the brain sections of *Aire*$^{-/-}$ mice via H&E staining. In addition, brain weight was measured to test whether total brain volume is reduced in *Aire*$^{-/-}$ mice. Finally, the regulation of *Gad65, Gh, Ghrh, Th, Tph, Sox9 and Sox10* in the thymus and brain by AIRE was tested. These seven genes encode brain antigens (BAs) and autoantibodies to these antigens have been detected in APS-1 patients (Cocco et al., 2005; Serguei et al., 2009; Cocco et al., 2012).

3.2 Genotyping and AIRE expression

Aire$^{+/+}$, *Aire*$^{+/-}$ and *Aire*$^{-/-}$ mice were divided into four groups by age (5, 15, 30 and 38 weeks), each containing a total of 15 mice (five mice per genotype). The brain and thymus were collected from these mice for further experiments. Genotyping was performed using DNA extracted from ear clippings. A representative genotyping PCR result is shown in Figure 3.1.

The brain and thymus tissues of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice were stained with AIRE-specific antibody to detect its expression. As expected, Aire was expressed in the thymus and the brain in *Aire*$^{+/+}$ but the antibody does not detect the impaired Aire in *Aire*$^{-/-}$ mice as shown in Figures 3.2 and 3.3. AIRE was detected in the medulla of the thymus and cerebral cortex of the brain (Figures 3.2 and 3.3). The data in figure 1.2 indicates that the expression of Aire in whole brain at the RNA level is very low, however, our findings are in keeping with a previous study, which found that a large number of the granular neurons and Purkinje cells of the cerebellar cortex showed a

strong Aire immunoreactivity in *Aire*$^{+/+}$ mice with C57BL/6 background but not in Aire$^{-/-}$ that produced by Peltonen group (Ramsey et al., 2002b). It seems that the expression of Aire in the granular neurons and Purkinje cells of cerebral cortex particularly high but not in whole brain which may explain the low expression of Aire at RNA level since only a few specific cells in the brain express Aire.

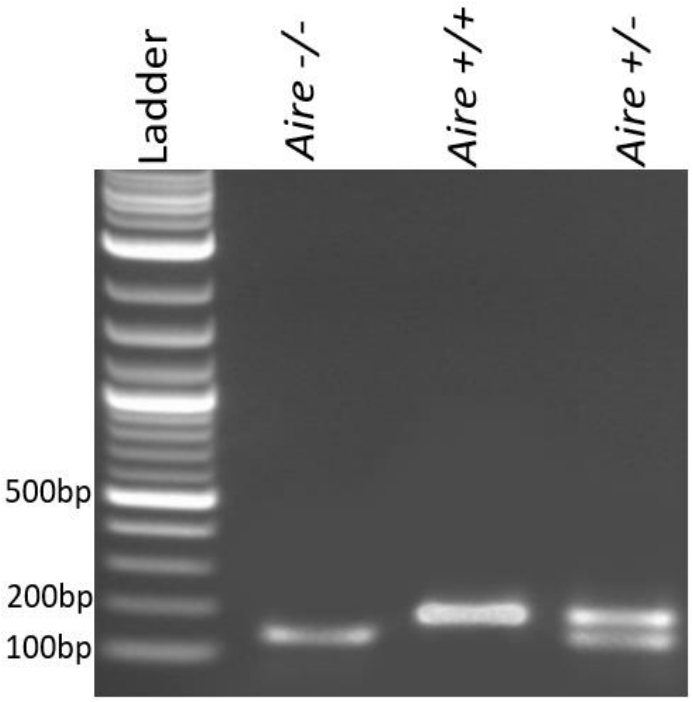

Figure 3.1: *Aire* genotyping in mice. DNA was extracted from the ear clippings of mice. *Aire* knockout (−/−) mice showed one band at 140 bp, *Aire* wild-type (+/+) mice showed one band at 195 bp and heterozygous (+/−) mice showed two bands at 195 bp and 140 bp.

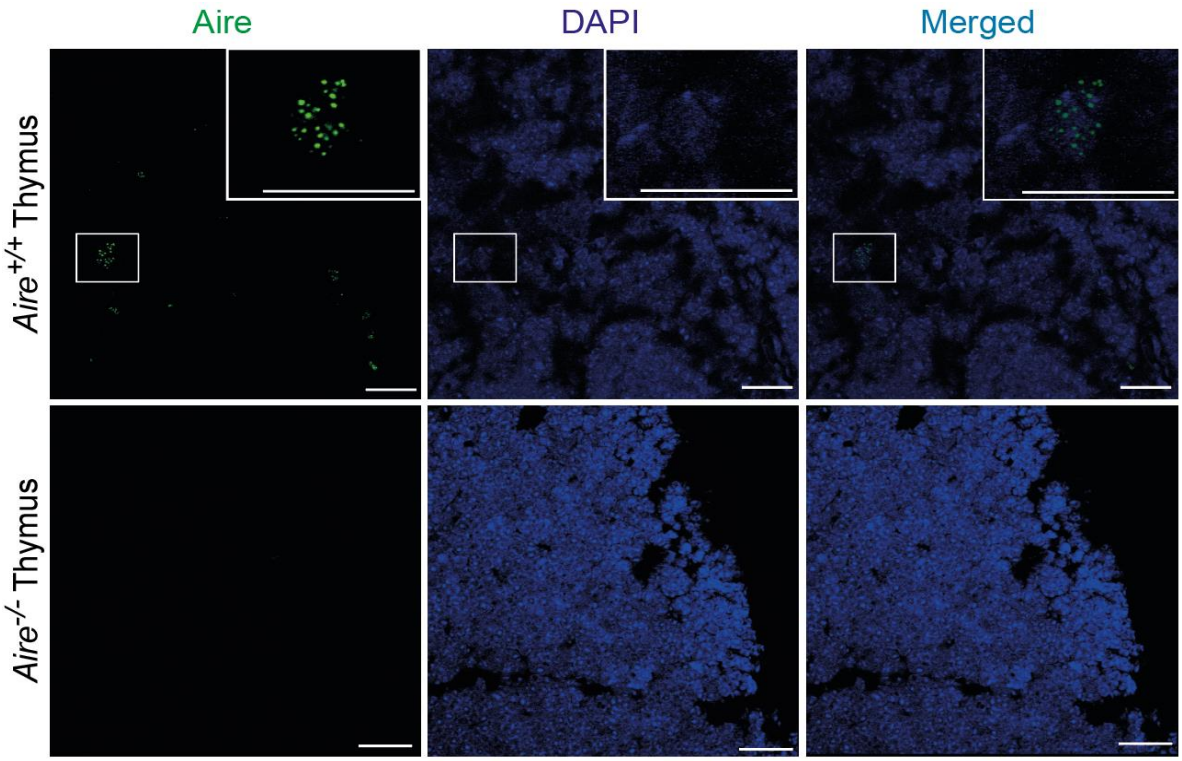

Figure 3.2: The expression of AIRE in the thymus. This picture shows the medulla of the thymus. The expression of AIRE in the thymus medulla in *Aire*$^{+/+}$ (upper panel) but the antibody doesn't recognise the impaired Aire in *Aire*$^{-/-}$ (lower panel). One cell is zoomed in and Aire is expressed in the nucleus of thymic cells in punctate pattern. Few cells positive for Aire were detected because the antibody might be degraded. The slides were incubated with polyclonal goat anti-AIRE D-17 antibody as a primary antibody linked to Alexa Fluor 488 (green) and the nuclei were counterstained with DAPI (blue). Scale bar 20 μm.

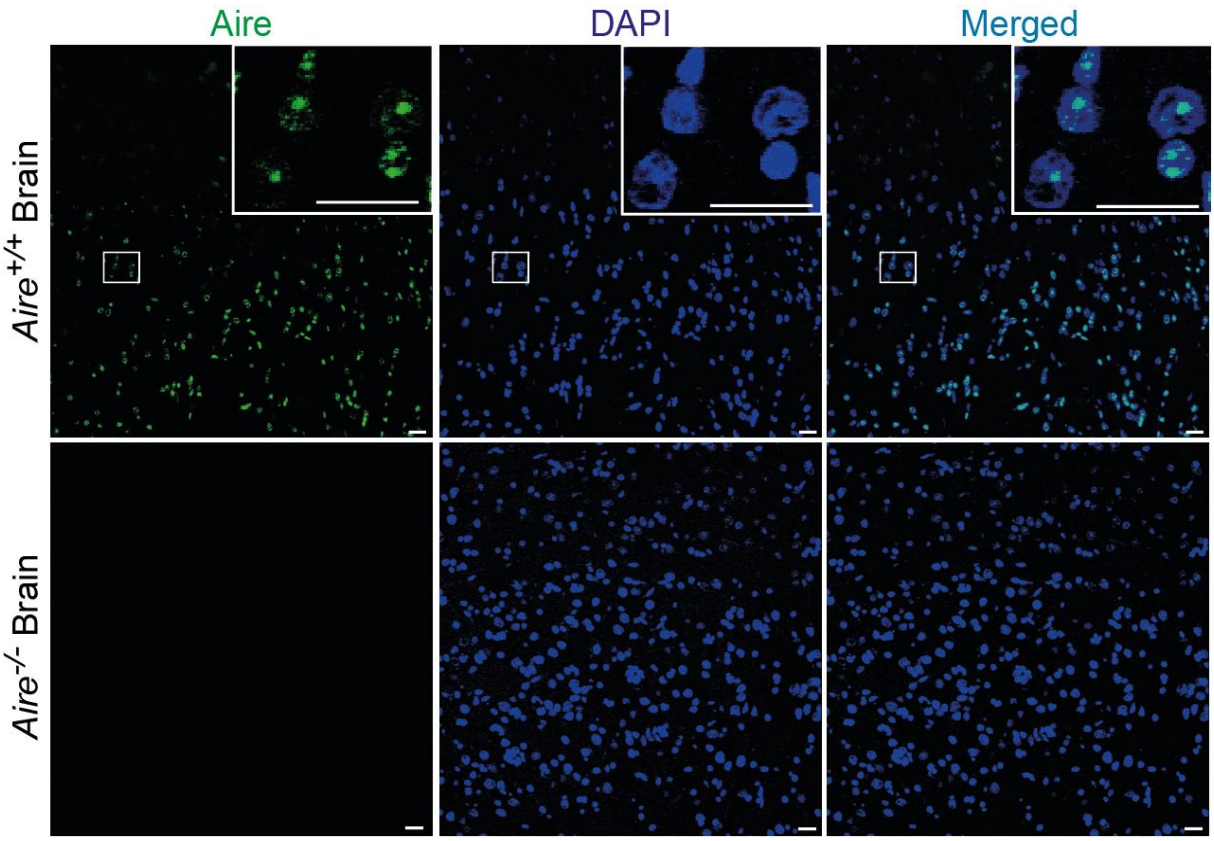

Figure 3.3: The expression of AIRE in the brain. The picture shows high expression of Aire in cells of cerebral cortex. No distinct punctate nuclear dot structure is formed as seen in thymus. The slides were incubated with polyclonal goat anti-AIRE D-17 antibody linked to Alexa Fluor 488 (green) and the nuclei were counterstained with DAPI (blue). Scale bar 20 μm.

3.3 Brain and thymus size

The weights of the brains and thymuses collected from *Aire*$^{+/+}$, *Aire*$^{+/-}$ and *Aire*$^{-/-}$ mice belonging to four age groups (5, 15, 30 and 38 weeks) were compared to assess the effects of the loss of AIRE. Initially, the thymus and brain weight of mice of different Aire genotype were measured using tissues from 36 mice. Each genotype group contained 12 mice and the number of males and females were equal in each age group. Organ weights were normalised to the body weight of each mouse. As seen in Figures 3.4 and 3.5, there was no significant difference in the weight of the thymus and brain among different genotypes. After that, to test the size of the thymus and brain at different stages of development, the weight of the thymus and brain of 60 mice were measured (Figures 3.6 and 3.7). 5 *Aire* wild-type (+/+), 5 heterozygous (+/−) and

5 knockout (−/−) mice were used at each time point (5, 15, 30 and 38 weeks), making a total of 60 mice. Each genotype group contained 20 mice and their sex was not considered. Figure 3.6 clearly shows a decrease in thymus size with age, reflecting a natural process known as age-related thymic involution (Gui et al., 2012). A precipitous drop in the weight of the thymus was observed from the age of 5 to 15 weeks at the same level across all genotypes. In addition, the thymus weight decreased slightly from the age of 15 to 38 weeks at the same level across all genotypes (Figure 3.6). These observations suggest that age-related thymic involution is not affected by the absence of *Aire*. On the other hand, no age-related change in brain weight was observed. Brain size increased slightly from the age of 5 to 15 weeks but no difference in brain weight was observed among 15-, 30- and 38-week-old mice across all genotypes (Figure 3.7). Thus, the size of the thymus and/or brain were not affected in C57BL/6 mice by the absence of *Aire*.

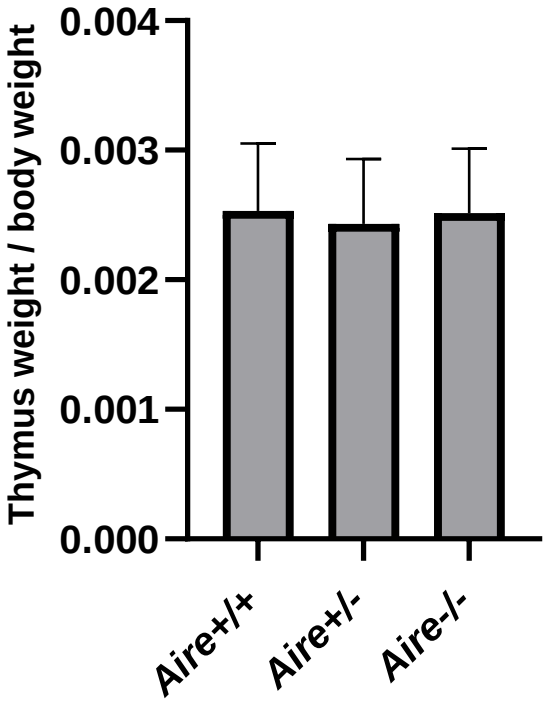

Figure 3.4: The thymus weight of mice of different Aire genotype. Thymuses were collected from 12 *Aire* wild-type (+/+), 12 heterozygous (+/−) and 12 *knockout* (−/−) mice and the weight was measured. Data were analysed by one-way ANOVA. Error bars represent 1SEM.

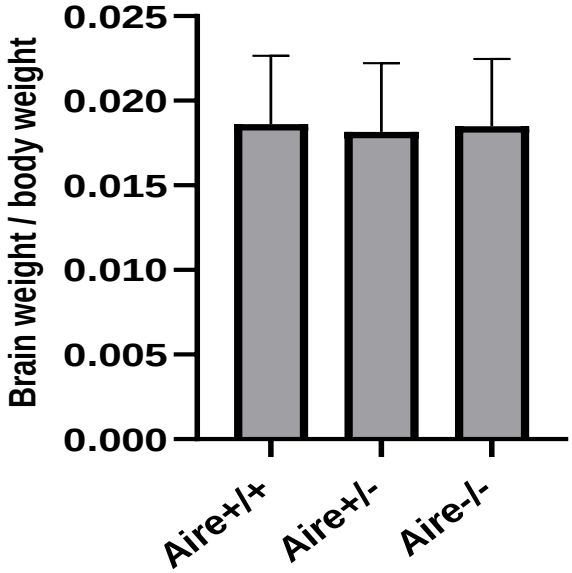

Figure 3.5: The brain weight of mice of different Aire genotype. Brains were collected from 12 *Aire* wild-type (+/+), 12 heterozygous (+/−) and 12 knockout (−/−) mice and the weight was measured. Data were analysed by one-way ANOVA. Error bars represent 1SEM.

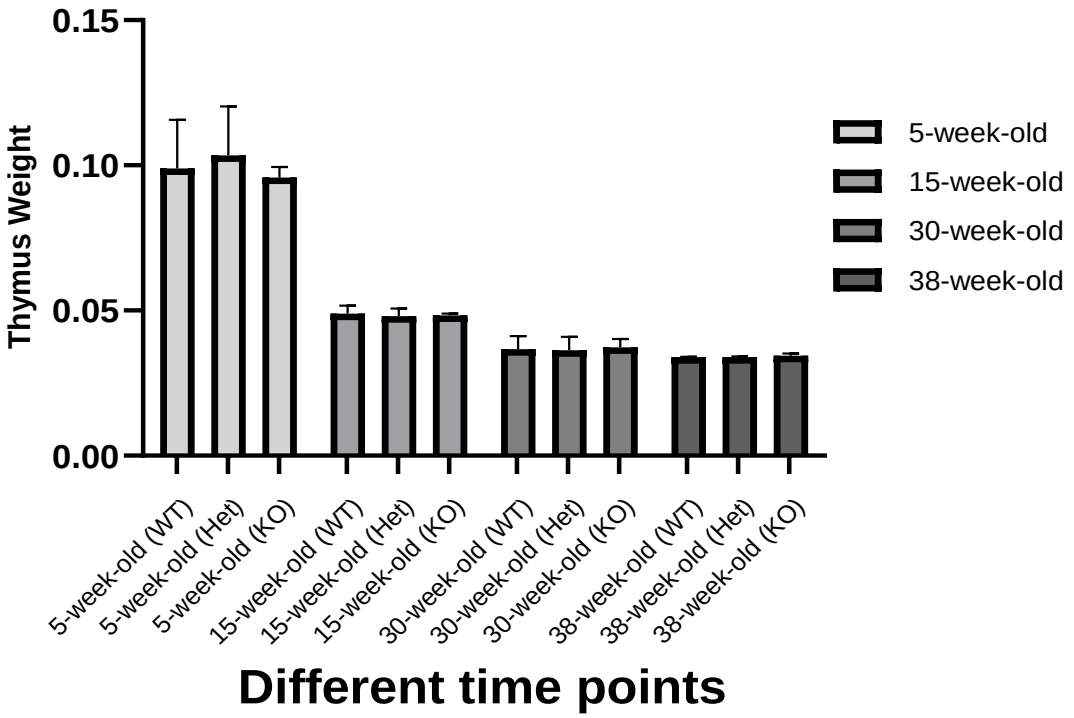

Figure 3.6: The thymus weight of mice of different Aire genotype at different time points. Thymus were collected from 60 mice, 20 *Aire* wild-type (+/+), 20 heterozygous (+/−) and 20 knockout (−/−) mice and the weight was measured. 5 *Aire* wild-type (+/+), 5 heterozygous (+/−) and 5 knockout (−/−) mice were used at each time point (5, 15, 30 and 38 weeks). Note the sharp decrease in thymus weight between the age of 5 and 15 weeks. The weight continues to decrease slightly from the age of 15 weeks onward. Data were analysed by one-way ANOVA. Error bars represent 1SEM.

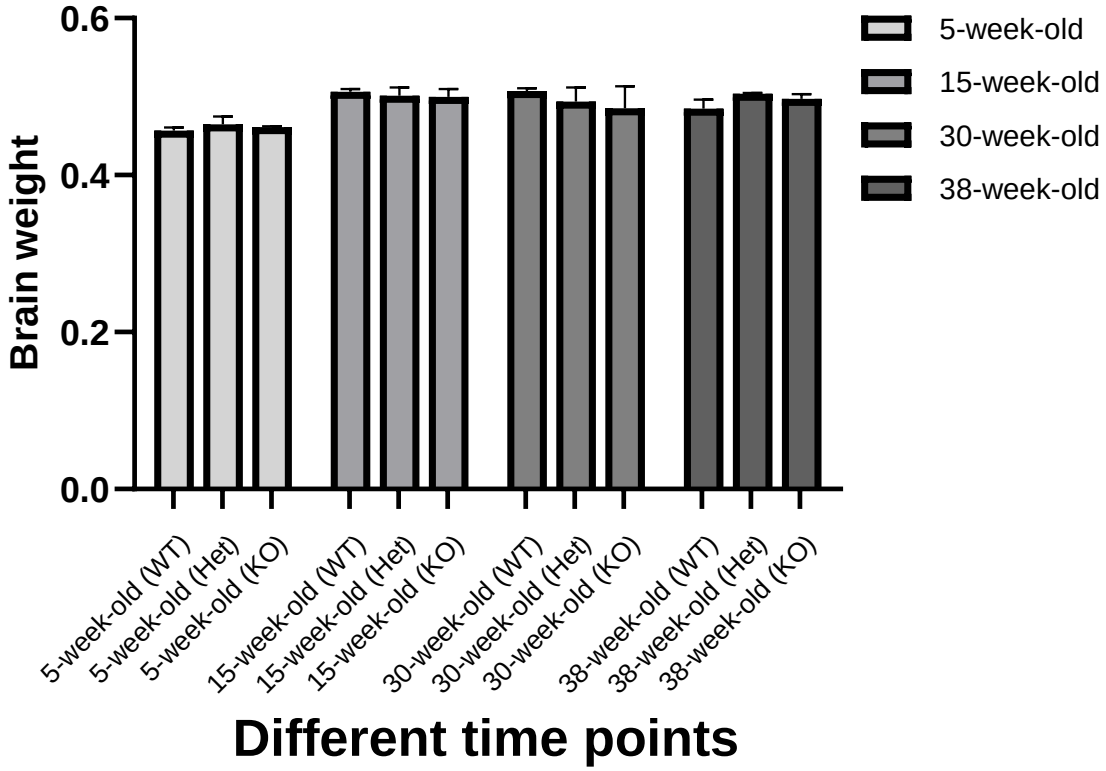

Figure 3.7: The brain weight of mice of different Aire genotype at different time points. Brains were collected from 60 mice, 20 *Aire* wild-type (+/+), 20 heterozygous (+/−) and 20 knockout (−/−) mice and the weight was measured. 5 *Aire* wild-type (+/+), 5 heterozygous (+/−) and 5 knockout (−/−) mice were used at each time point (5, 15, 30 and 38 weeks). A slight increase in the weight of the brain is seen from the age of 5 weeks 15 weeks. Data were analysed by one-way ANOVA. Error bars represent 1SEM.

3.4 Histological analysis of the thymus and brain in *Aire*-deficient mice

H&E staining was performed on the brain and thymus of 15- and 30-week-old mice to investigate immune-mediated destruction as reflected by lymphocytic infiltration in *Aire* knockout mice. Tissues from 8 mice were collected, 2 $Aire^{+/+}$ mice and 2 $Aire^{-/-}$ at each age group. No lymphocytic infiltration in the brain and thymus was observed in either $Aire^{+/+}$ or $Aire^{-/-}$ mice (Figures 3.8, 3.9, 3.10 and 3.11). This suggests that mice with the C57BL/6 genetic background do not mimic the neurological manifestations

associated with APS-1 in humans. This is not surprising since C57BL/6 mice develop a milder illness compared with mice of other genetic backgrounds (Jiang et al., 2005).

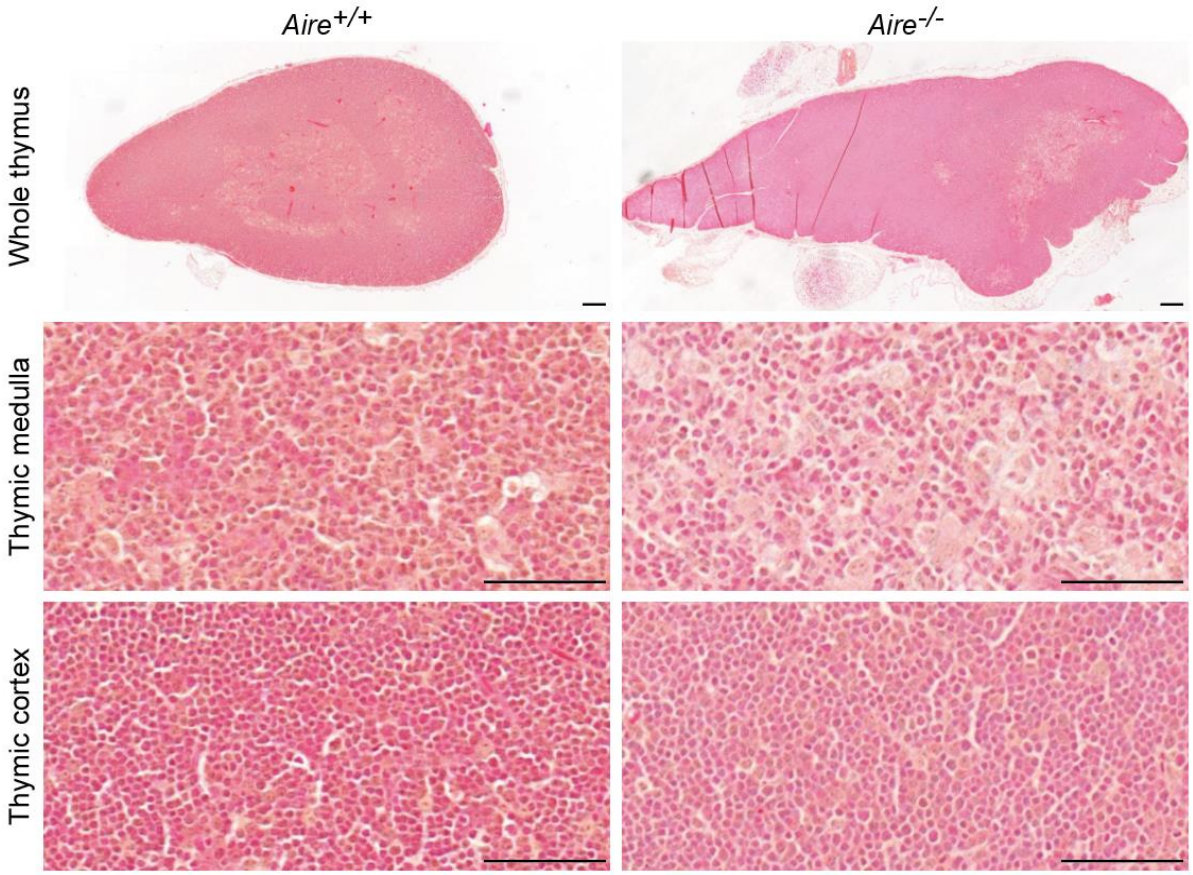

Figure 3.8: Representative thymus sections from 15-week-old *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. Thymus sections from *Aire*$^{+/+}$ (left lane) and *Aire*$^{-/-}$ (right lane) with H&E staining. Scale bar: 200 µm (whole thymus) and 50 µm (cortex and medulla).

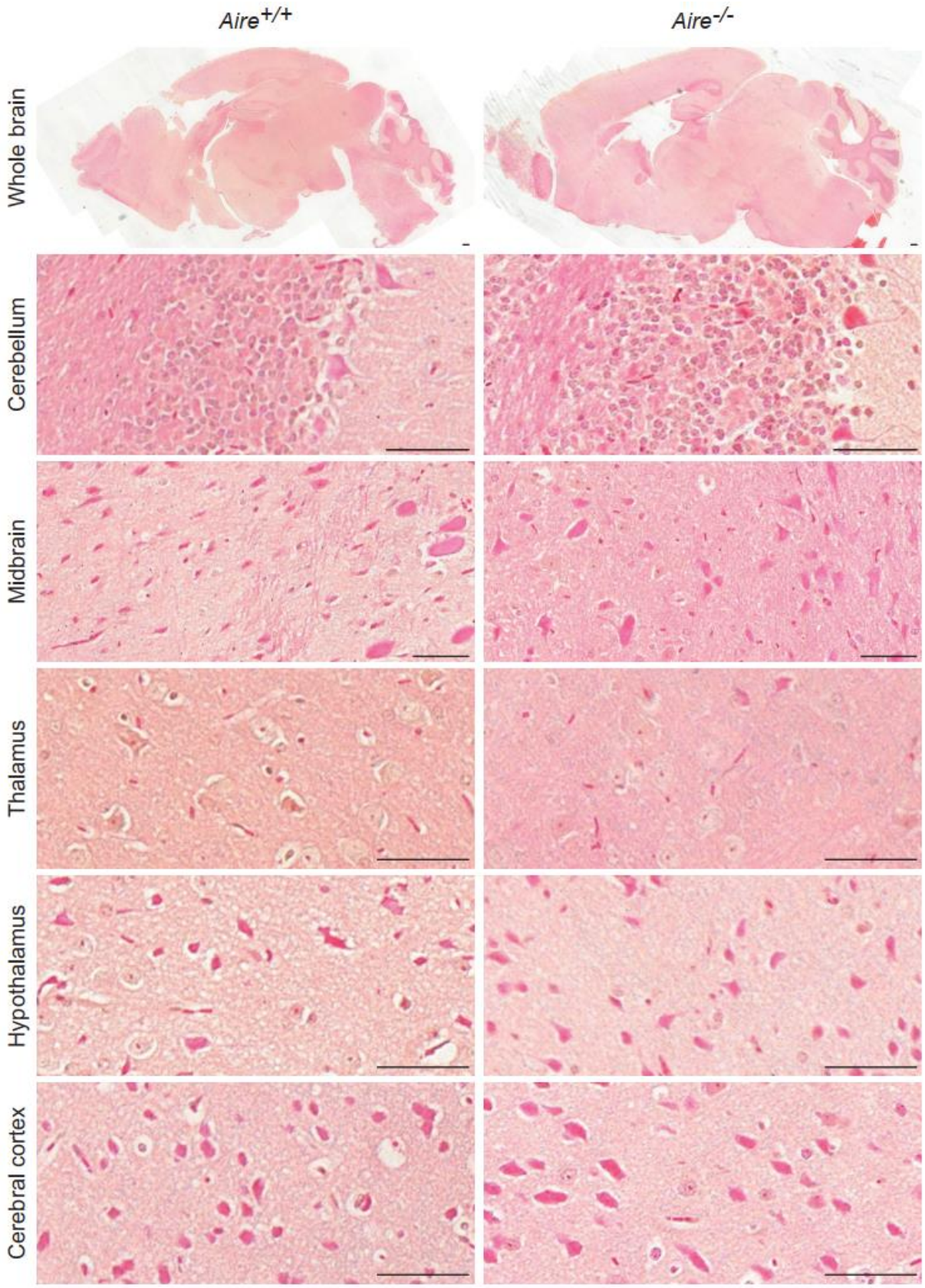

Figure 3.9: Representative brain sections from 15-week-old *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. Brain sections from *Aire*$^{+/+}$ (left lane) and *Aire*$^{-/-}$ (right lane) mice with H&E staining. Scale bar: 200 µm (whole brain) and 50 µm (cerebellum, midbrain, thalamus, hypothalamus and cerebral cortex).

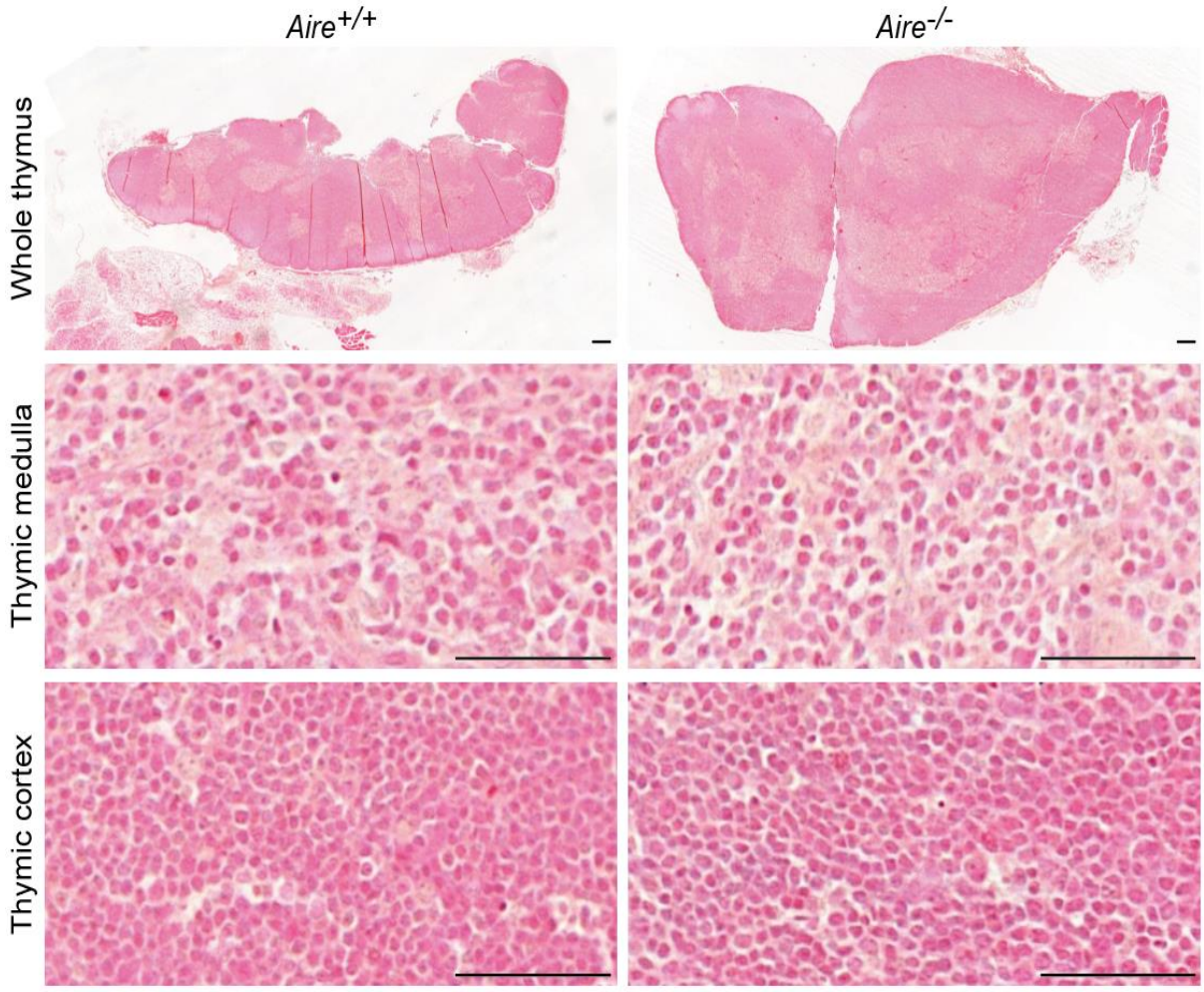

Figure 3.10: Representative thymus sections from 30-week-old *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. Thymus sections from *Aire*$^{+/+}$ (left lane) and *Aire*$^{-/-}$ (right lane) mice with H&E staining. Scale bar: 200 µm (whole thymus) and and 50 µm (cortex and medulla).

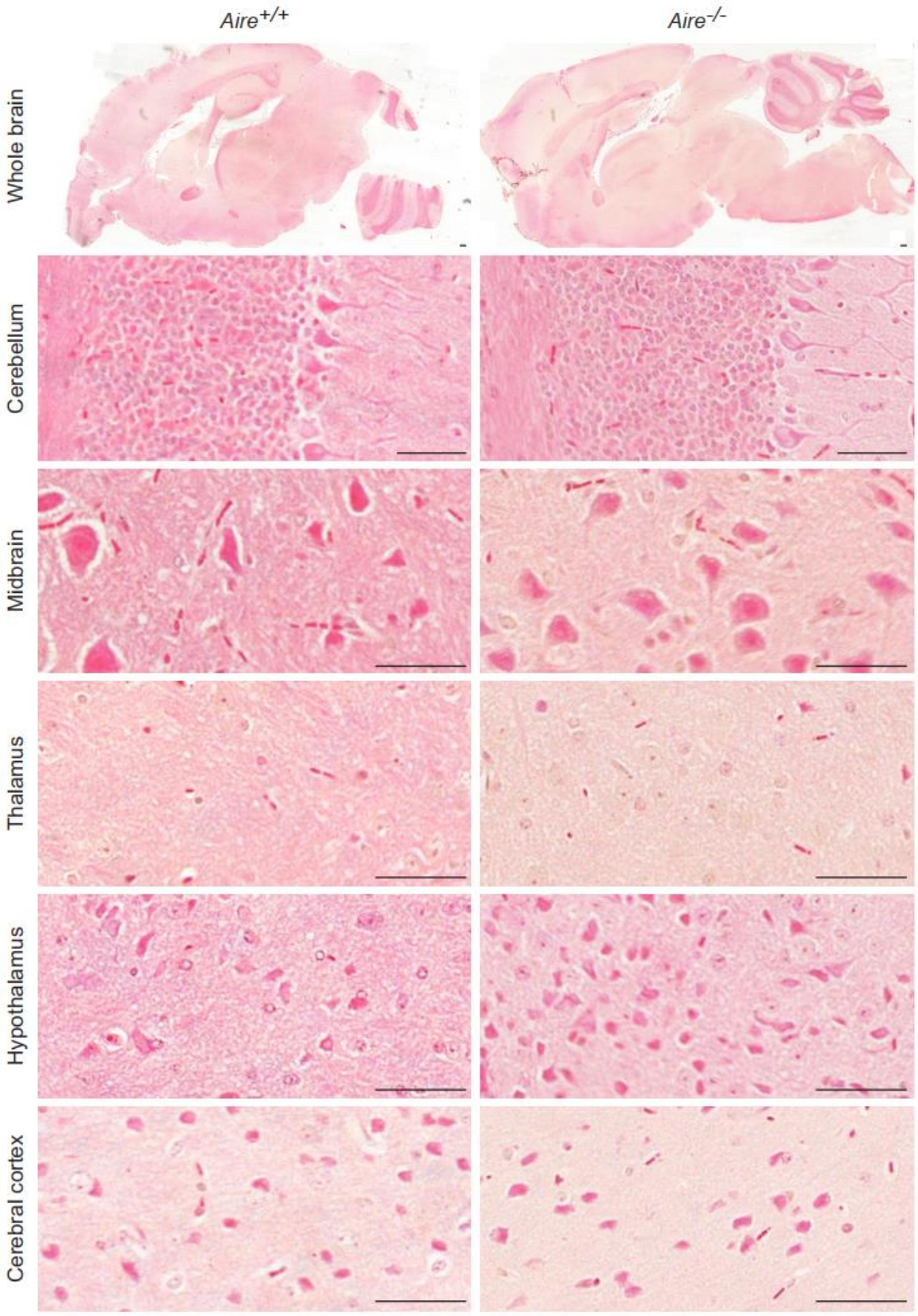

Figure 3.11: Representative brain sections from 30-week-old *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. Brain sections from *Aire*$^{+/+}$ (left lane) and *Aire*$^{-/-}$ (right lane) mice with H&E staining. Scale bar: 200 µm (whole brain) and 50 µm (cerebellum, midbrain, thalamus, hypothalamus and cerebral cortex).

3.5 Expression of genes encoding BAs

To test the ability of AIRE to regulate the genes encoding BAs (*Th*, *Tph, Gh, Gad65, Sox9, Sox10* and *Ghrh*), RNA was extracted from 5-week-old wild-type C57BL/6 mice and RT-PCR was performed to analyse the expression of these genes in the thymus and brain. All genes encoding BAs and AIRE were expressed in both thymus and brain (Figures 3.12 and 3.13).

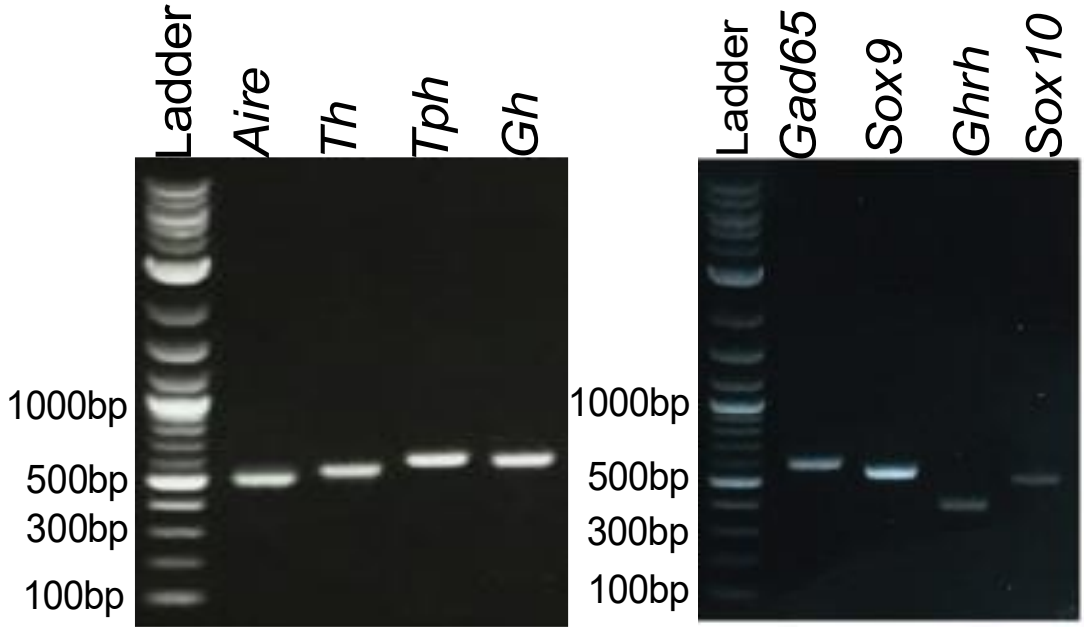

Figure 3.12: Expression of *Aire* and BA genes in the thymus. The expression of *Aire, Th, Tph, Gh, Gad65, Sox9, Ghrh* and *Sox10* in the thymus is shown. *Aire*, 509 bp, Th, 512 bp, *Tph*, 591 bp, *Gh*, 583 bp, *Gad65*, 562 bp, *Sox9*, 545 bp*, Ghrh*, 403 bp, *Sox10*, 522 bp. The PCR products were subjected to 35 cycles. 2 log DNA ladder was used to confirm that bands are of the expected size.

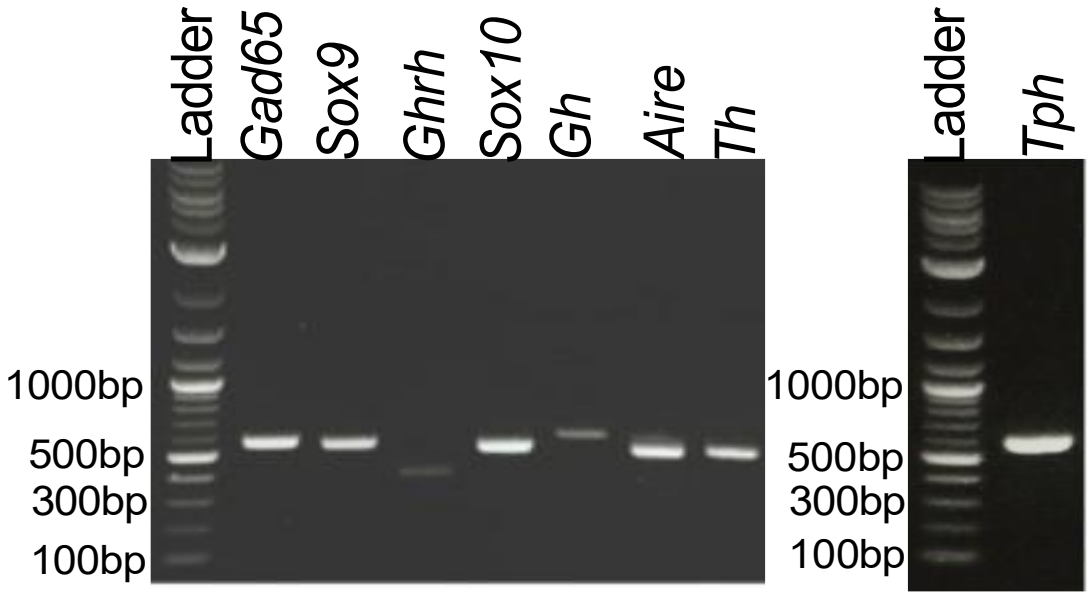

Figure 3.13: Expression of *Aire* and BA in the brain, Gad65 562 bp, Sox9, 545 bp, Ghrh, 403 bp, *Sox10*, 522 bp, *Gh*, 583 bp, *Aire*, 509 bp, *Th*, 512 bp, *Tph*, 591 bp. The expression of *Aire, Th, Tph, Gh, Gad65, Sox9, Ghrh* and *Sox10* in the brain is shown. The PCR products were subjected to 35 cycles. 2 log DNA ladder was used to confirm that bands are of the expected size.

3.6 Expression of *Aire* and the BA genes

RT-qPCR was performed to investigate whether the absence of *Aire* affected the expression of BA genes in mice. In addition, the expression level of *Aire* mRNA in the thymus and brain was also investigated at different stages of development. For this purpose, RNA was extracted from the thymus and brain of 40 mice belonging to different age groups (5, 15, 30 and 38 weeks). RNA was extracted from 10 mice at each age group, five $Aire^{+/+}$ and five $Aire^{-/-}$ mice, making a total of 40 mice and RT-qPCR experiments were repeated at least for three technical replicates. It can be clearly seen from figure 1.2 and figure 3.14 that the expression of Aire in whole brain at the RNA level is very low, however, the immunofluorescence results in figure 3.3 show that AIRE is highly expressed in specific cells of cerebral cortex. Our immunofluorescence results are in keeping with a previous study, which found that a large number of the granular neurons and Purkinje cells of the cerebellar cortex showed a strong Aire immunoreactivity in $Aire^{+/+}$ mice with C57BL/6 background but

not in Aire$^{-/-}$ that produced by Peltonen group (Ramsey et al., 2002b). It seems that the expression of Aire in the granular neurons and Purkinje cells of cerebral cortex particularly high but not in whole brain which may explain the low expression of Aire at RNA level since only a few specific cells in the brain express Aire (*See section 3.8*).

Expression of *Aire* mRNA decreased slightly with age in the thymus and brain; however, the difference was not statistically significant (Figure 3.14). The AIRE-dependent gene *Ins2*, which is expressed in the thymus and upregulated by AIRE, was used as a positive control, while the AIRE-independent gene *Foxn1* was used as a negative control (Figure 3.15). As expected, a marked increase in *Ins2* expression was observed in the thymus of *Aire*$^{+/+}$ mice compared with *Aire*$^{-/-}$ mice, while *Foxn1* expression was comparable in both groups.

RT-qPCR analysis of BA gene expression in the thymus showed that *Sox10* and *Th* expression was reduced significantly in *Aire*$^{-/-}$ mice compared with *Aire*$^{+/+}$ mice (Figure 3.16). *Sox10* expression was not altered significantly among the different age groups of *Aire*$^{+/+}$ mice, although a slight increase in *Sox10* expression in 30-week-old mice was observed compared with the other age groups (Figure 3.16). However, the expression of *Sox10* in *Aire*$^{-/-}$ mice decreased significantly from 5 to 15 weeks ($p = 0.028$) but increased significantly in 30-week-old mice compared with 15-week-old mice ($p = 0.029$). In 38-week-old *Aire*-deficient mice, the expression of *Sox10* decreased slightly compared with 30-week-old-mice ($p = 0.055$; Figure 3.16). However, the expression of *Sox10* showed a decline in *Aire*$^{-/-}$ mice compared with *Aire*$^{+/+}$ mice by 2.4, 9.2, 3.4 and 6.7 folds in 5-, 15-, 30- and 38-week-old mice, respectively (Figure 3.16). The expression of *Sox10* was significantly greater in *Aire*$^{+/+}$ mice compared with *Aire*$^{-/-}$ mice, with p-values of < 0.0001, 0.01, 0.0002 and < 0.0001 in 5-, 15-, 30- and 38-week-old mice, respectively (Figure 3.16). The RT-qPCR data

also showed that the expression of *Th* in *Aire*$^{+/+}$ mice fluctuated with age, but the difference was not significant (Figure 3.16). On the other hand, the expression of *Th* in *Aire*$^{-/-}$ mice decreased significantly in mice 15 weeks and older compared with 5-week-old mice, with p-values of 0.014, < 0.0001 and < 0.0001 in 15-, 30- and 38-week-old mice, respectively. In addition, the expression of *Th* decreased significantly in 38-week-old *Aire*$^{-/-}$ mice compared with 30-week-old mice (p = 0.023; Figure 3.16). Nevertheless, the expression of *Th* showed a decline in *Aire*$^{-/-}$ mice compared with *Aire*$^{+/+}$ mice by 2.2, 4.7, 13.6 and 4.6 folds in 5-, 15-, 30- and 38-week-old mice, respectively (Figure 3.16). *Aire*$^{+/+}$ mice show significantly greater expression of *Th* at all time periods tested, with p-values of 0.003, 0.004, < 0.0001 and < 0.0001 in 5-, 15-, 30- and 38-week-old mice, respectively (Figure 3.16). The observed variability in the expression of *Sox10* and *Th* in *Aire*$^{-/-}$ mice, especially the drop occurring between the ages of 5 and 15 weeks, suggests that the expression of *Sox10* and *Th* is regulated by another transcription factor besides AIRE (*see section 3.7*).

No significant difference was observed between *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice regarding the other BA genes in the thymus (Figure 3.17). The expression of *Gad65* in *Aire*$^{+/+}$ mice seemed to show a slight increase in the 30-week and 38-week groups compared with the younger mice, but the difference was not significant. The same was observed in *Aire*$^{-/-}$ mice but the difference between 15-week-old and 38-week-old mice was significant (p = 0.04; Figure 3.17). The expression of *Gh* increased in 30-week-old mice whether *Aire*$^{+/+}$ or *Aire*$^{-/-}$, but the increase was not significant. The increase in expression between 15-week-old mice and 38-week-old mice was significant in *Aire*$^{+/+}$ mice (p = 0.016) but not in *Aire*$^{-/-}$ mice (Figure 3.17). *Ghrh* expression did not show any significant difference in 38-week-old mice compared with the other age groups across genotypes (Figure 3.17). The expression of *Sox9*, whether in *Aire*$^{+/+}$ or *Aire*$^{-/-}$

mice, increased in 30-week-old mice and then decreased again in 38-week-old mice with no significant differences in either group (Figure 3.17). The expression of *Tph* increased from the age of 5 weeks to 30 weeks in both *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice but the difference was not significant (Figure 3.17).

In contrast with the thymus, no significant difference was observed in the expression of any BA gene between *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice in the brain, suggesting that these genes may be AIRE-independent in the brain. The expression of *Tph* in the brain of *Aire*$^{+/+}$ mice decreased significantly in 15-week-old mice compared with 5-week-old mice ($p = 0.005$) but then increased slightly until 38 weeks of age without significant differences (Figure 3.18). The same pattern was observed in *Aire*$^{-/-}$ mice, with a significant decrease in Tph expression between 5 weeks and 15 weeks ($p = 0.008$) and a slight, non-significant increase up to 38 weeks. The expression of *Gad65* was also decreased in 15-week-old mice compared with 5-week-old mice in Aire$^{+/+}$ mice but the difference was not significant; thereafter, the expression increased slightly. In *Aire*$^{-/-}$ mice, the expression of *Gad65* was stable until 15 weeks of age but then increased slightly in 30-week-old and 38-week-old mice (Figure 3.18). The expression of *Gh* decreased with age in both *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice but the difference was not significant in either group, even when comparing the youngest and oldest groups (Figure 3.18). The expression of *Ghrh* in *Aire*$^{+/+}$ mice decreased at 15 weeks compared with the younger group, but the difference was not significant. However, in *Aire*$^{-/-}$ mice, the expression of *Ghrh* increased in15-week-old mice compared with 5-week-old mice, although the difference was again not significant (Figure 3.18). The expression of *Sox10* fluctuated with age in both genotypes without any significant difference (Figure 3.18). *Sox9* showed significantly increased expression in 38-week-old mice compared with 30-week-old mice, with a p-value 0.0001 of in *Aire*$^{+/+}$ mice and 0.04 in *Aire*$^{-/-}$ mice

(Figure 3.18). Similarly, *Th* expression increased significantly across genotypes, with a p-value of 0.03 in *Aire*$^{+/+}$ mice and 0.01 in *Aire*$^{-/-}$ mice (Figure 3.18). Overall, comparing the expression of BA genes in the brains of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice at different stages of development did not provide any obvious evidence showing that AIRE has a regulatory effect in their expression. The data in figure 1.2 and 3.14 indicates that the expression of Aire in whole brain at the RNA level is very low and may not be sufficient to drive Aire-dependent expression but figure 3.3 and the results revealed by Ramsey et al. (2002b) indicates that Aire is expressed highly at a large number of the granular neurons and Purkinje cells of the cerebellar cortex in C57BL/6 mice. Therefore, Aire may not be expressed in the same cells as BA.

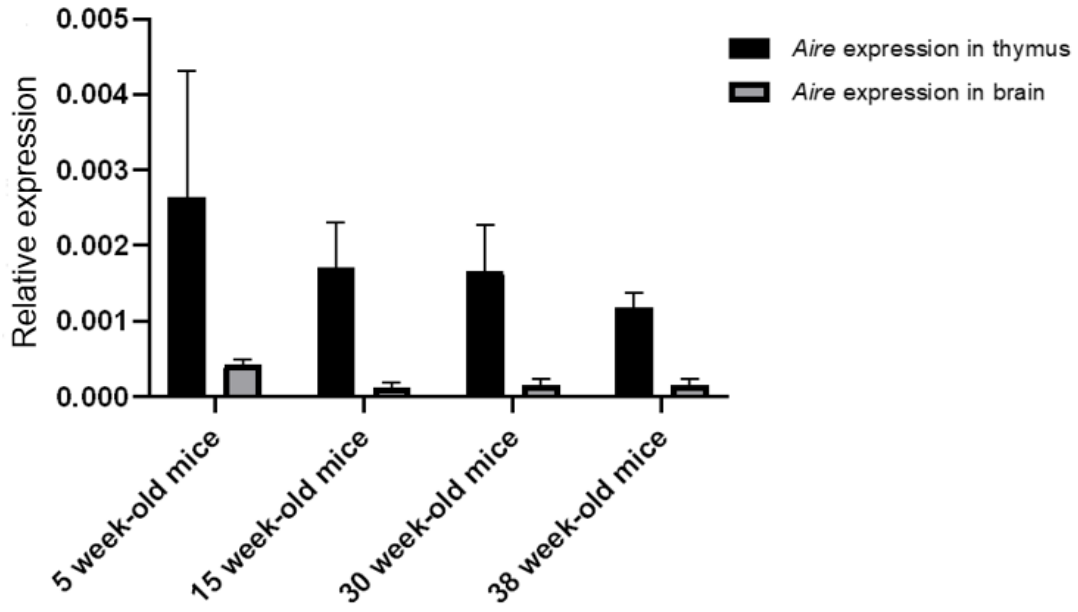

Figure 3.14: Expression of *Aire* in the thymus and brain. RT-qPCR analysis of *Aire* expression in the thymus and brain in different age groups. The black bars represent expression of *Aire* in the thymus and the grey bars represent the expression of *Aire* in the brain. The expression level of *Aire* mRNA was higher in the thymus than the brain in all age groups. Expression of *Aire* mRNA decreased slightly with mouse development. Results are normalised to the expression of the housekeeping gene *Actb* (β-actin). n = 3. Error bars represent 1SEM.

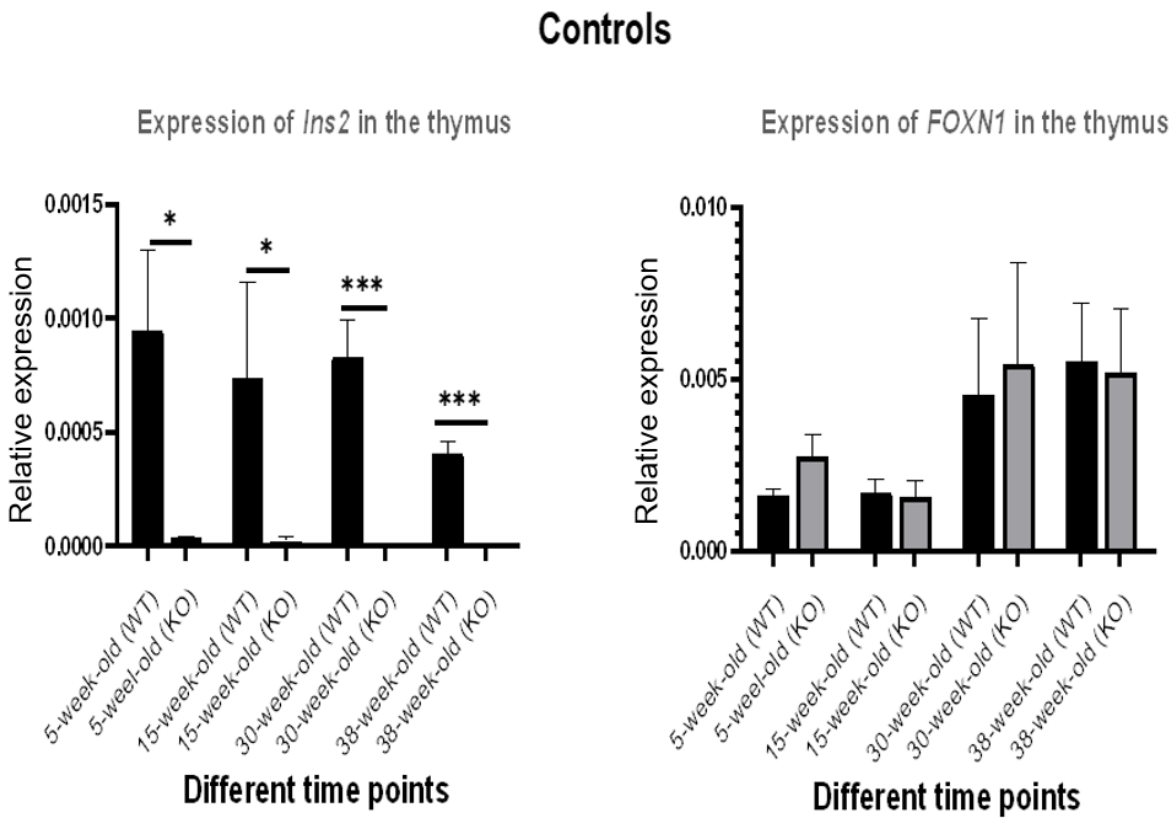

Figure 3.15: Expression of *Ins2* (AIRE-dependent) and *Foxn1* (AIRE-independent) in the thymus. RT-qPCR analysis of the expression of *Ins2* (AIRE-dependent) and *Foxn1* (AIRE-independent) in the thymus of mice of different age groups. *Ins2* was upregulated in *Aire*$^{+/+}$, while the expression of *Foxn1* remained the same in *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. Results are normalised to the expression of the housekeeping gene *Actb* (β-actin). n = 3. Data were analysed by t test. **** $P < 0.0001$, *** $P < 0.0005$, ** $P < 0.005$ and * $P < 0.05$. Error bars represent 1SEM.

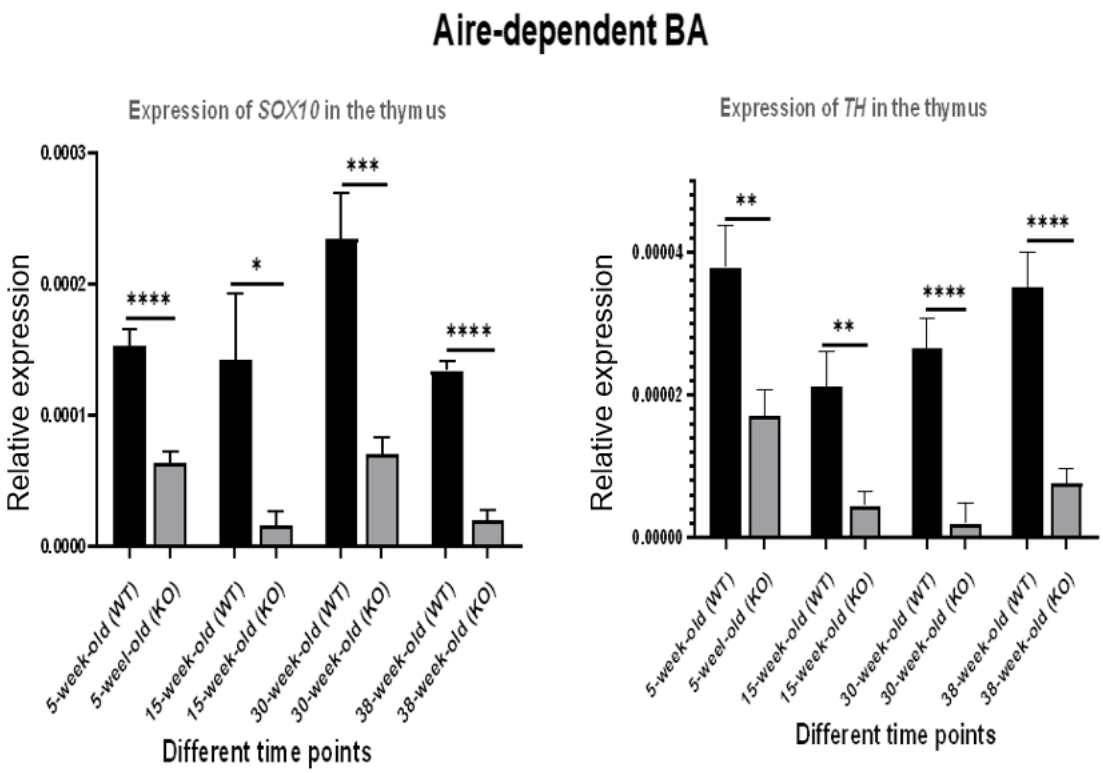

Figure 3.16: Expression of *Sox10* and *Th* in the thymus. RT-qPCR analysis of the expression of *Sox10* and *Th* in the thymus in mice of different age groups. Expression of *Sox10* and *Th* increased significantly in *Aire*$^{+/+}$ mice. Results are normalised to the expression of the housekeeping gene *Actb* (β-actin). n = 3. Data were analysed by t test. **** $P < 0.0001$, *** $P < 0.0005$, ** $P < 0.005$ and * $P < 0.05$. Error bars represent 1SEM.

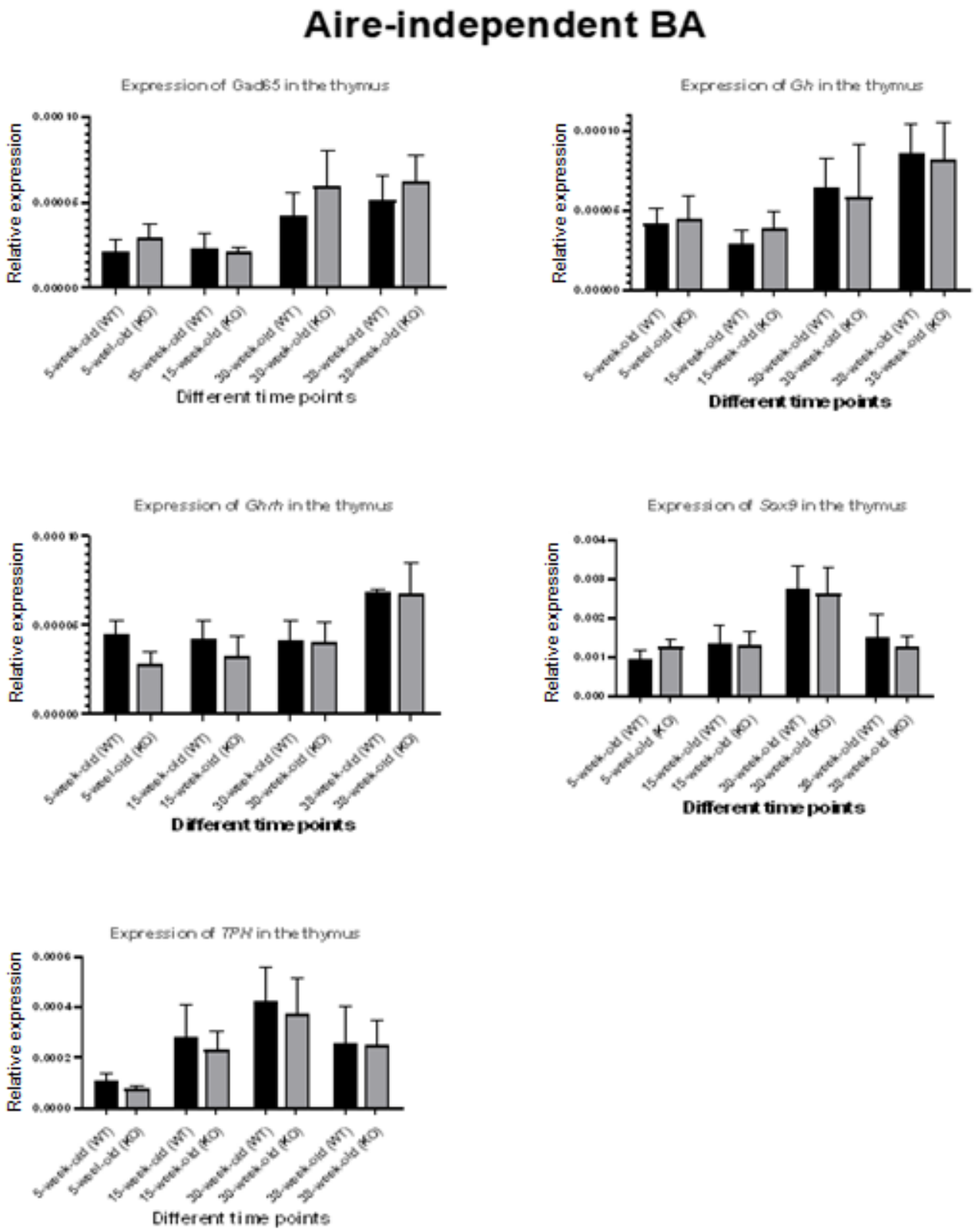

Figure 3.17: Expression of AIRE-independent BA genes in the thymus. RT-qPCR analysis of the expression of *Sox9, Gh, Ghrh, Gad65* and *Tph* in the thymus of mice of different age groups. There was no significant difference in the expression of these genes between *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. Results are normalised to the expression of the housekeeping gene *Actb* (β-actin). n = 3. Data were analysed by t test. Error bars represent 1SEM.

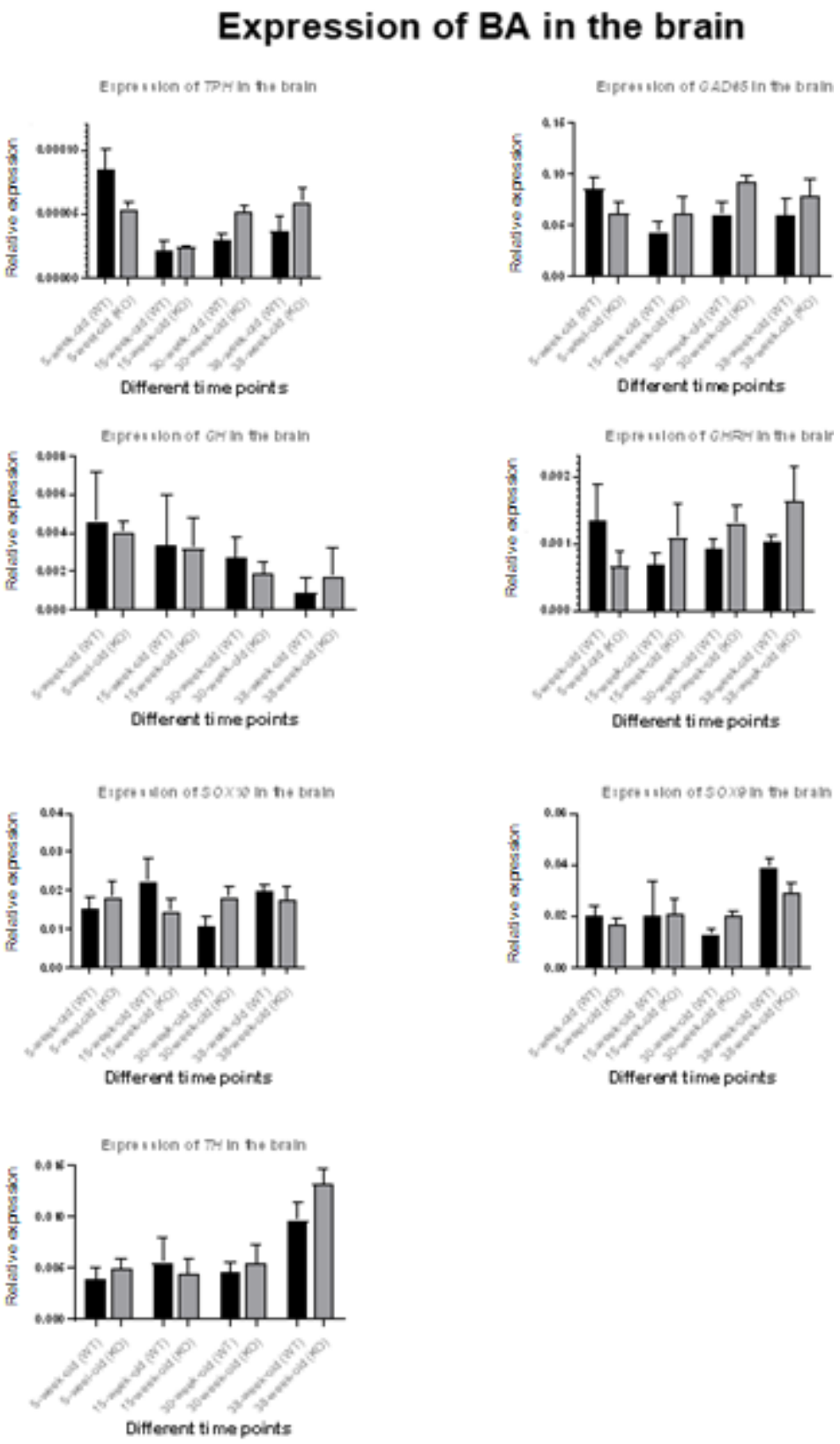

Figure 3.18: Expression of BAs in the brain of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. RT-qPCR analysis of the expression of genes encoding BAs in the brain of mice of different age groups. There was no significant difference in the expression of BAs in the brain between *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. Results are normalised to the expression of the housekeeping gene *Actb* (β-actin). n = 3. Error bars represent 1SEM.

3.7 Expression of SOX10 and TH at the protein level

To test whether the observed changes at the mRNA level were also reflected at the protein level, total protein was extracted from the brain and thymus tissues of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice at different stages of development (5, 15, 30 and 38 weeks old). Thereafter, antibodies against SOX10 and TH were used to detect the expression level of these two proteins in *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. In the thymus, the expression of TH and SOX10 was higher in *Aire*$^{+/+}$ mice comparing to *Aire*$^{-/-}$ mice at all stages of development as expected, but expression from 15 weeks onward increased significantly in *Aire*$^{+/+}$ mice (Figures 3.20 and 3.24). Additionally, to our knowledge, this is the first time that fluctuations in the expression of AIRE-dependent proteins during development have been detected in *Aire*$^{-/-}$ mice. Interestingly, at the protein level, the expression of TH in *Aire*$^{-/-}$ mice was higher in 5-week-old mice than in later age groups and the difference became more significant with age (Figure 3.20). In these mice, TH expression decreased compared with 5-week-old mice with p-values of 0.09, 0.04 and 0.01 in 15-, 30- and 38-week-old mice, respectively, showing that the gap increases with time (Figure 3.20). In addition, at the mRNA level, *Sox10* expression in *Aire*$^{-/-}$ mice decreased significantly between the age of 5 weeks and 15 weeks ($p = 0.028$) but increased significantly from 15 weeks to 30 weeks ($p = 0.029$; Figure 3.16). However, although the expression of SOX10 protein in *Aire*$^{-/-}$ mice decreased slightly from 5 weeks to 15 weeks and then increased again in 30-week-old mice as observed at the mRNA level, it was not detected at 38 weeks (Figure 3.24).

TH and SOX10 expression in the brain was not significantly different between *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice (Figures 3.22 and 3.26). Presumably because Aire may not be expressed in the same cells as BA since Aire detected at high levels in specific cells of the cerebral cortex and hippocampus in C57BL/6 mice (Ramsey et al., 2002b). In

addition, our results show that Aire is expressed in many cells of the cerebral cortex (Figure 3.3). However, Level of Aire expression in the brain is low as seen by RT-qPCR so may not be sufficient to be functional which needs further investigations to be confirmed. The experiments were repeated three times and relative band intensity was quantified using ImageJ.

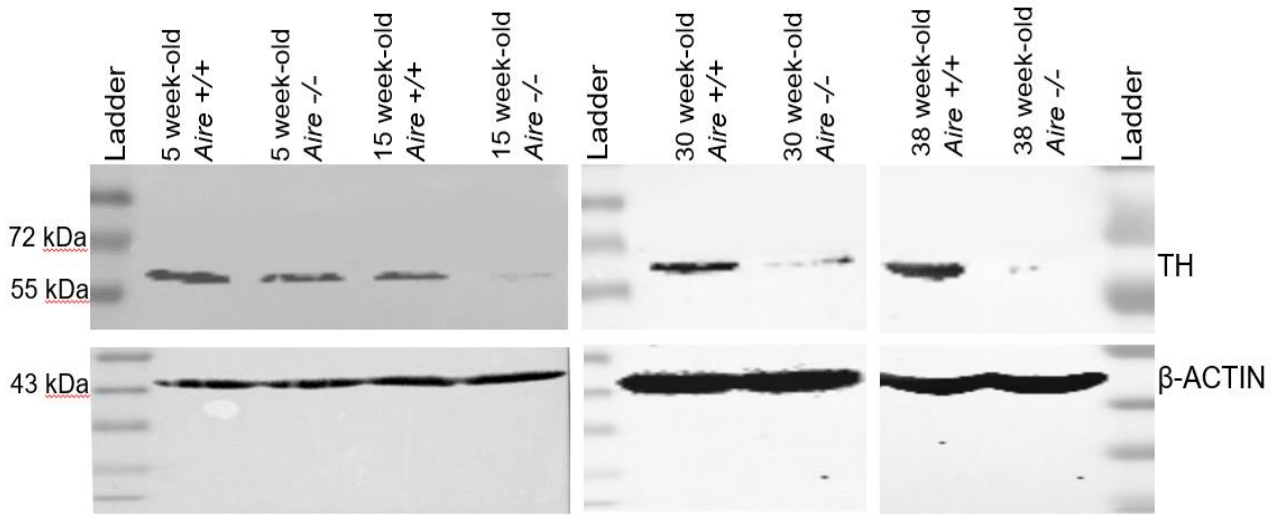

Figure 3.19: Expression of TH protein in the thymus of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. Western blotting was performed to examine the expression level of TH protein in the thymus of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice at different stages of development (5, 15, 30 and 38 weeks old). Expression of TH protein in *Aire*$^{+/+}$ mice was much higher than in *Aire*$^{-/-}$ mice. Data were normalised to the housekeeping protein β-actin.

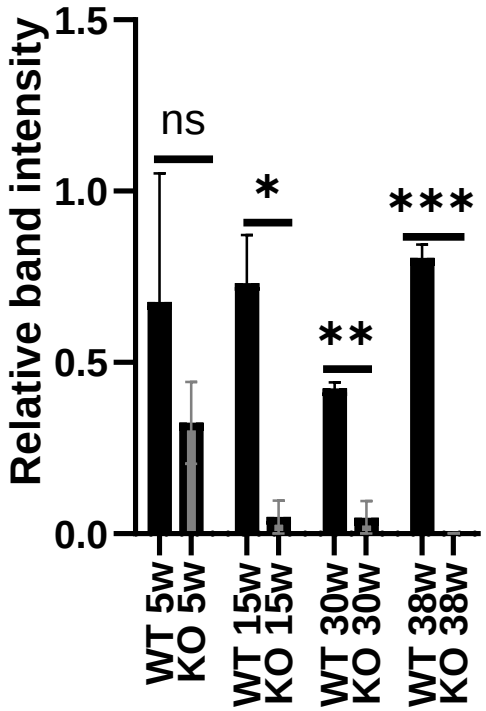

Figure 3.20: The relative band intensity of TH expression in the thymus. The relative band intensity of TH protein in the thymus of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice was quantified at different stages of development (5, 15, 30 and 38 weeks old) using ImageJ. The bar chart shows that the expression of TH protein in *Aire*$^{+/+}$ mice was higher than in *Aire*$^{-/-}$ mice at all ages and significantly higher from 15 weeks onward. Data were normalised to the housekeeping protein β-actin. Data were analysed by t test. **** $P < 0.0001$, *** $P < 0.0005$, ** $P < 0.005$ and * $P < 0.05$. Error bars represent 1SEM.

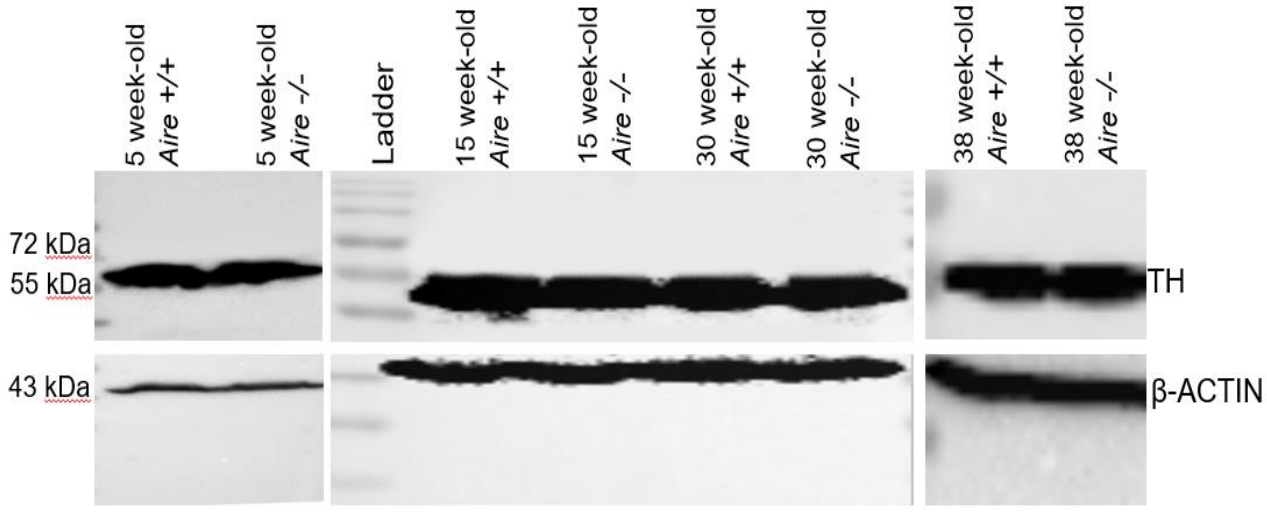

Figure 3.21: Expression of TH protein in the brain of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. Western blotting was performed to examine the expression level of TH protein in the brain of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice at different stages of development (5, 15, 30 and 38 weeks old). There was no significant difference in the expression of TH protein in the brain between *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. Data were normalised to the housekeeping protein β-actin.

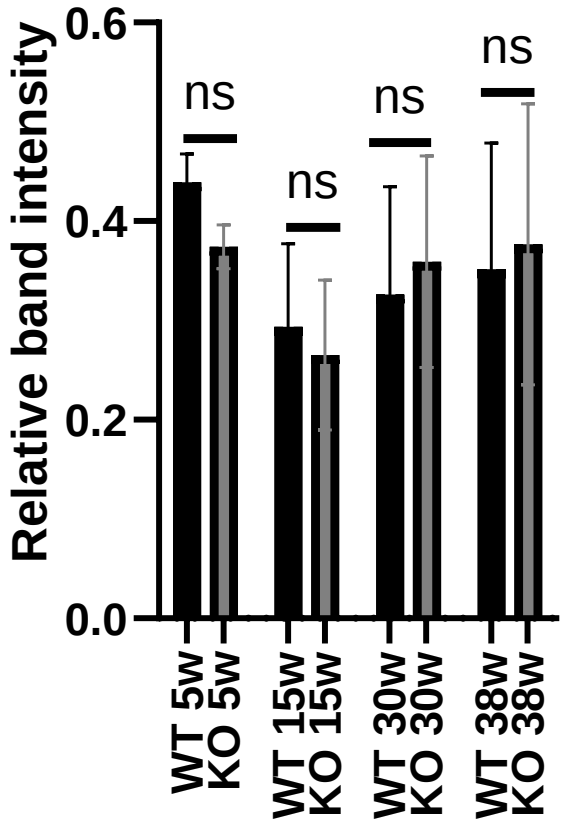

Figure 3.22: The relative band intensity of TH expression in the brain. The relative band intensity of TH protein in the brain of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice was quantified at different stages of development (5, 15, 30 and 38 weeks old) using ImageJ. The bar chart shows that the expression of TH protein in the brain of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice remained the same. Data were normalised to the housekeeping protein β-actin. Data were analysed by t test. Error bars represent 1SEM.

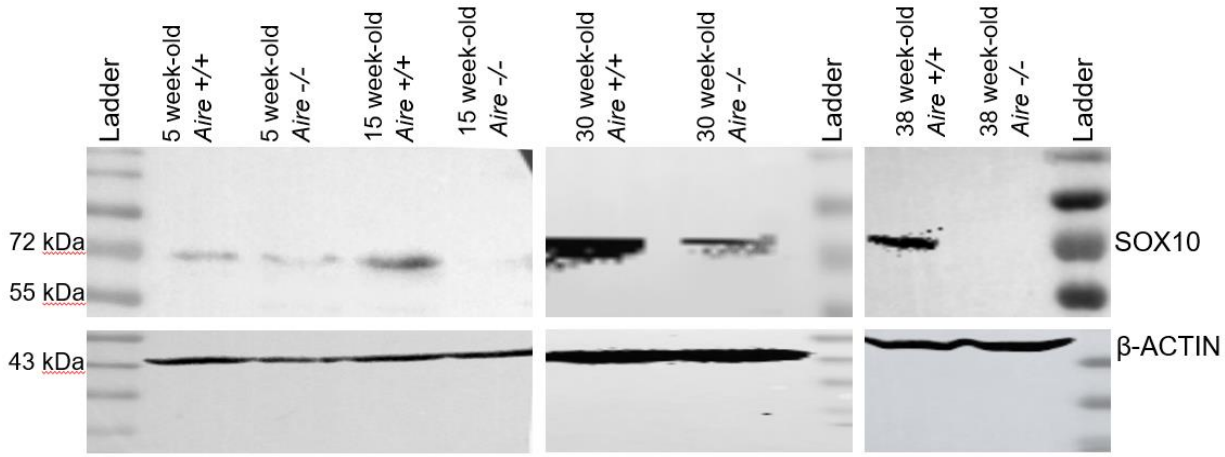

Figure 3.23: Expression of SOX10 protein in the thymus of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. Western blotting was performed to examine the expression level of SOX10 protein in the thymus of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice at different stages of development (5, 15 and 30 weeks old). Expression of SOX10 protein in *Aire*$^{+/+}$ mice was much higher than in *Aire*$^{-/-}$ mice. Data were normalised to the housekeeping protein β-actin.

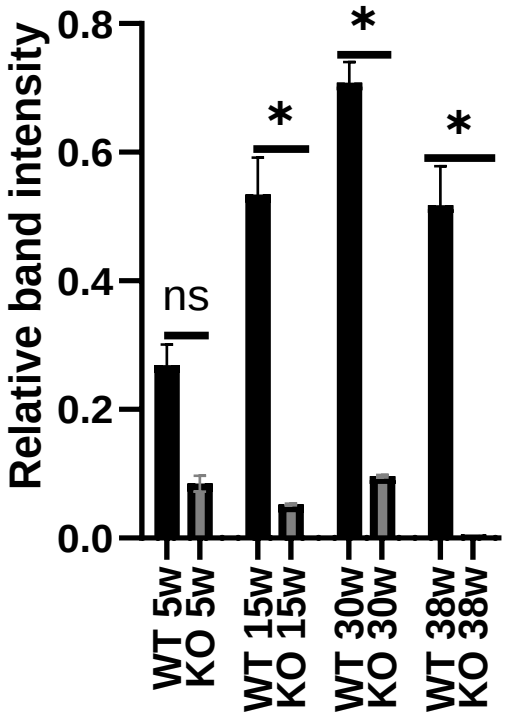

Figure 3.24: The relative band intensity of SOX10 expression in the thymus. The relative band intensity of SOX10 protein in the thymus of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice was quantified at different stages of development (5, 15, 30 and 38 weeks old) using ImageJ. The bar chart shows that the expression of SOX10 protein in *Aire*$^{+/+}$ mice was higher than in *Aire*$^{-/-}$ mice at all ages and significantly higher from 15 weeks onward. Data were normalised to the housekeeping protein β-actin. Data were analysed by t test. **** $P < 0.0001$, *** $P < 0.0005$, ** $P < 0.005$ and * $P < 0.05$. Error bars represent 1SEM.

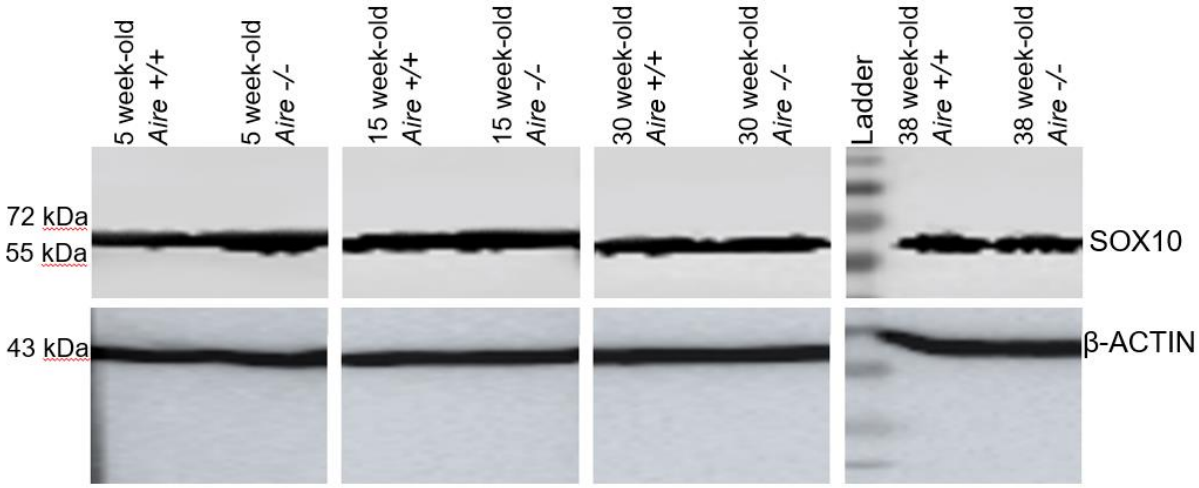

Figure 3.25: Expression of SOX10 protein in the brain of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. Western blotting was performed to examine the expression level of SOX10 protein in the brain of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice at different stages of development (5, 15 and 30 weeks old). There was no significant difference in the expression of SOX10 protein in the brain between *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. Data were normalised to the housekeeping protein β-actin. Data were analysed by t test. Error bars represent 1SEM.

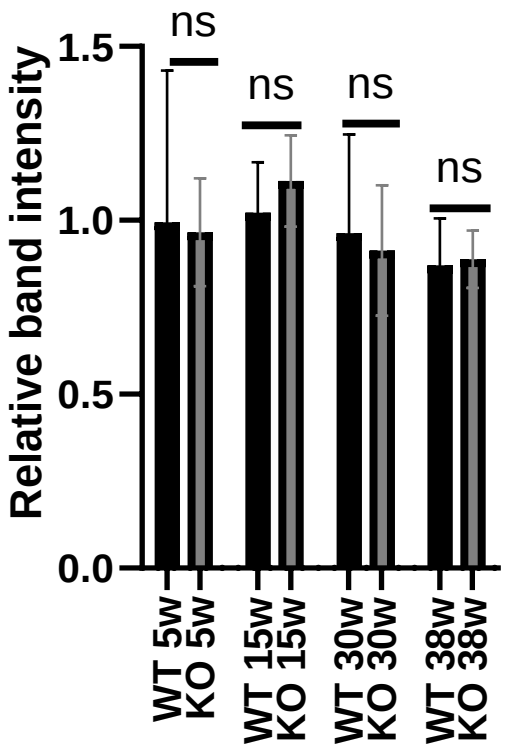

Figure 3.26: The relative band intensity of SOX10 expression in the brain. The relative band intensity of SOX10 protein in the brain of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice was quantified at different stages of development (5, 15, 30 and 38 weeks old) using ImageJ. The bar chart shows that the expression of SOX10 protein in the brain of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice remained the same. Data were normalised to the housekeeping protein β-actin. Data were analysed by t test. Error bars represent 1SEM.

3.8 Discussion

A recent study observed a reduction in total brain and cerebellum volumes in APS-1 patients and correlated this reduction to autoantibodies against GAD+ and TH+ neurons (Meloni et al., 2019). The authors observed and reported alterations in brain volume together with neuronal autoantibodies for the first time in 78.6% patients with APS-1 from Sardinia. Therefore, we hypothesised that *Aire* knockout C57BL/6 mice would mimic the neurological manifestations associated with APS-1 in humans, such as reduction in total brain volume and lymphocytic infiltration. To investigate this hypothesis, *Aire* knockout mice were generated and genotyped. The genotyping PCR was sufficient to define $Aire^{+/+}$, $Aire^{+/-}$ and $Aire^{-/-}$ mice but AIRE expression was verified by immunofluorescence microscopy for further validation. The results indicated that AIRE was expressed in the thymus and brain of $Aire^{+/+}$ but not $Aire^{-/-}$ mice. AIRE was shown to be nucleus-specific, forming distinct punctate nuclear dot structures expressed in a subset of thymic cells, verifying previous findings (Halonen et al., 2001, Blechschmidt et al., 1999, Heino et al., 1999). RT-qPCR results in figure 1.2 and figure 3.14 show that the expression of Aire in whole brain at the RNA level is very low, however, the immunofluorescence results in figure 3.3 show that AIRE is highly expressed in specific cells of cerebral cortex. Our immunofluorescence results are in keeping a previous study, which found that a large number of the granular neurons and Purkinje cells of the cerebellar cortex showed a strong Aire immunoreactivity in $Aire^{+/+}$ mice with C57BL/6 background but not in $Aire^{-/-}$ that produced by Peltonen group (Ramsey et al., 2002b). it appears that the expression of Aire only can be detected at high level in particular cells of the cerebral cortex and hippocampus (Ramsey et al., 2002b). As a result, the expression of *Aire* seems to be low as detected by RT-qPCR because some technical issues such as partly degradation of RNA and/or there is dilution effect since RNA is extracted from whole

brain, but *Aire* is only expressed at high level in specific cells of specific part of the brain (Figure 3.3; Ramsey et al., 2002b).

Brains and thymuses were collected from *Aire*$^{+/+}$, *Aire*$^{+/-}$ and *Aire*$^{-/-}$ mice of four age groups (5, 15, 30 and 38 weeks old) and their weights were compared to investigate whether loss of AIRE affects their size. No significant difference was observed in the weight of the thymus and brain among the different genotypes (Figures 3.4 and 3.5). However, a decrease in thymus size with age was observed, a phenomenon previously described as age-related thymic involution (Gui et al., 2012). Next, the thymus and brain were fixed and sectioned for H&E staining to observe the presence of lymphocytic infiltration, as this can be a sign of tissue damage. No apparent lymphocytic infiltration was observed in the thymus and brain of *Aire*$^{-/-}$ mice (Figures 3.8, 3.9, 3.10 and 3.11). It is not surprising that the thymus appeared to be unaffected by the absence of *Aire* as all *Aire* knockout mice displayed healthy thymic compartment morphologies (Hubert et al., 2009). However, there is more than one possible explanation for why changes appearing in the brain of APS-1 patients were not observed in *Aire* knockout mice with the C57BL/6 genetic background. One possibility may be the genetic background as it strongly influences the autoimmune phenotype in *Aire* knockout mice. For example, C57BL/6 mice have been shown to develop a milder illness than, for example, the NOD strain (discussed in detail in Section 1.3; Jiang et al., 2005). Moreover, a group recently generated *Aire*$^{-/-}$ rats that developed a severe autoimmune phenotype not observed in *Aire*$^{-/-}$ mice, including alopecia, nail dystrophy, vitiligo and impaired thymus development (Ossart et al., 2018). Several organs and tissues have been evaluated for autoimmune lesions in *Aire*$^{-/-}$ rats, including the thymus, liver, testis, thyroid, pancreas, skin, kidney, stomach, lung, salivary gland, ovary, spleen, lymph nodes, Peyer patches, duodenum, ileum,

colon and eye; however, the brain has not been evaluated (Ossart et al., 2018). Therefore, it is essential to observe the brain of *Aire*$^{-/-}$ rats as neurological alterations, such as alopecia, nail dystrophy and vitiligo, may be detected in *Aire*$^{-/-}$ rats but not in *Aire*$^{-/-}$ mice. Furthermore, the age of the mouse should be considered in the development of autoimmunity as it corresponds primarily to the duration of the disease. It may be too early for mice between 15 and 30 weeks of age to develop signs of autoimmunity in the brain since lesions or signs of T lymphocyte infiltration in some organs of *Aire*$^{-/-}$ mice or *Aire*$^{-/-}$ rats can appear at > 10 months (> 42 weeks; Anderson et al., 2002; Ossart et al., 2018). The prevalence of autoimmune components not only increases with age in APS-1 animal models but also in APS-1 patients (Perheentupa et al., 2006; Meloni et al., 2012). As a result, examining the brains of *Aire*-deficient rats or mice with different genetic backgrounds at > 1 year of age may be essential to find a suitable APS-1 model that mimics the neurological manifestations associated with APS-1 in humans for later use in gene therapy experiments.

Although structural features may not be affected in *Aire* knockout C57BL/6 mice 30 weeks or younger, gene expression in the thymus or brain could be altered. To evaluate this, the expression of seven BA genes, namely *Th, Tph, Gad65, Gh, Ghrh, Sox9* and *Sox10* was studied in these mice. We investigated whether the absence of AIRE affects the expression of BA genes in the thymus and brain by quantifying their mRNA levels in *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. The data showed that AIRE upregulated *Sox10* and *Th* but not the other genes in the thymus (Figures 3.16 and 3.17). TH is an enzyme responsible for producing dopamine from tyrosine by catalysing the conversion of L-tyrosine to L-3,4-dihydroxyphenylalanine (L-DOPA), which is a precursor of dopamine (Daubner et al., 2001; Smith-Anttila et al., 2017). Dopamine is

also converted into two other brain neurotransmitters, noradrenaline and adrenaline. Therefore, TH catalyse the rate-limiting step in the synthesis of the catecholamines dopamine, noradrenaline and adrenaline (Daubner et al., 2011; Smith-Anttila et al., 2017), and a lack of TH compromises the synthesis of these neurotransmitters (Daubner et al., 2011; Smith-Anttila et al., 2017). Impaired synthesis of these crucial neurotransmitter can lead to neurological manifestations that range from mild to severe. TH deficiency can lead to impaired movement, stiff muscles and involuntary eye movement, and brain magnetic resonance imaging (MRI) often reveals mild to moderate cerebellar atrophy (Furukawa and Kish, 1993; Hoffmann et al., 2003). Movement difficulties may increase with age, and in severe cases, patients become dependent on wheelchairs because of progressively uncontrollable movements (Hoffmann et al., 2003). Patients with TH deficiency share most clinical phenotypes with APS-1 patients, with neurological manifestations such as movement difficulties, stiff muscles and cerebral atrophy (Furukawa and Kish, 1993; Meloni et al., 2019), suggesting that autoantibodies targeting the TH+ neurons in APS-1 patients are responsible for these clinical manifestations.

SRY-box transcription factors 10 (*SOX10*) is a transcription factor that is expressed at all stages of development in embryonic and adult melanocytes and is required for their generation and homeostasis (Shakhova et al., 2015; Haseeb and Lefebvre, 2019). It has been reported that loss of SOX10 leads to severe decreases in melanocyte numbers and causes Waardenburg-Hirschsprung syndrome, which includes vitiligo; likewise, autoantibodies against SOX10 in APS-1 patients are associated with vitiligo (Hedstrand et al., 2001; Bondurand et al., 2007; Shakhova et al., 2015).

RT-qPCR analysis in this study showed no significant difference in the expression of BA genes between *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice in the brain (Figure 3.16). AIRE does not

seem to regulate the genes encoding for BAs in the brain as it is in the thymus. This may be because AIRE binds and forms complexes with several proteins known as AIRE partners, such as SIRT-1, TOP2α and DNA-PK to control the transcription of AIRE-dependent genes in mTEC (discussed in detail in Section 1.2.4; Liiv et al., 2008; Giraud et al., 2012; Abramson and Husebye, 2016). However, AIRE partners may either not be expressed in the same cells as AIRE or not expressed at all in the brain. In addition, Level of Aire expression in the brain is low as seen by RT-qPCR so may not be sufficient to be functional by this needs further investigations to be confirmed. Hence, AIRE may not function as a transcription regulator in the brain as it does in the thymus.

The non-involvement of AIRE in the expression of the genes encoding for BAs in the brain revealed for the first time by our results is extremely interesting and research to discover the role of AIRE in the brain is needed. AIRE function in the brain or periphery is, in general, yet to be discovered. However, in terms of autoimmunity, defective AIRE in the thymus and/or secondary lymphoid organs is sufficient to cause autoimmune manifestation in several organs (Zhao et al., 2018). Even though studies have been conducted to discover the function of AIRE in peripheral tolerance, its role remains unclear (Zhao et al., 2018). Gardner et al. (2008) suggested that stromal cells that express AIRE in the spleen and lymph nodes may induce tolerance through the deletion of autoreactive T cells. Microarray analysis of these stromal cells showed that these cells express several unique TSAs that are not expressed in mTEC. Remarkably, 163 genes were regulated by AIRE in the stromal cells of the spleen and lymph nodes while 1835 were regulated in mTEC; nevertheless, only seven genes overlapped between the stromal cells and mTEC (Gardner et al., 2008; Zhao et al., 2018). Therefore, autoreactive T cells against some BAs may undergo selection in the

secondary lymphoid organs (lymph nodes and spleen) instead of the thymus under the control of AIRE and more research about the function of AIRE in extra-thymic contexts and peripheral tolerance is needed. However, immune tolerance is not only enforced by negative selection but is also enforced by Foxp3+ regulatory T cells that induced through interaction with $mTEC^{Hi}$ with intermediate affinity (*see section 1.2.1*; Stritesky et al., 2012; Kisand and Peterson, 2015).

Western blotting was used to test the expression of SOX10 and TH at the protein level to see whether the changes at the mRNA level carried over to the protein level. The results showed that the expression of SOX10 and TH decreased in the thymus at all stages of development in the absence of AIRE, but the difference became statistically significant only from the age of 15 weeks onward (Figures 3.20 and 3.24). For example, the p-value was 0.19 and 0.14 between *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice in 5-week-old mice for TH and SOX10, respectively, but less than 0.05 in 15-, 30- and 38-week-old mice. The transcription of TH and SOX10 may pass through two stages, which can be seen at the protein level. The first stage may partially depend on AIRE, as seen in 5-week-old mice, and the second stage may be heavily AIRE-dependent, from the age of 15 weeks onward.

Chapter Four

The effect of AIRE and APS-1 mutations on the expression of BAs in 4D6 cells

4.1 Outline

Aire-deficient mice are essential to study the biochemical properties of AIRE and its biological effects at the cellular, tissue and systemic level (Anderson et al., 2002; Sato et al., 2004; Venanzi et al., 2008; Lovewell et al., 2018). However, using human cell lines to study AIRE function has some advantages since cell lines can be genetically and chemically manipulated and comprise the only available model of human origin (Lovewell et al., 2011; Anderson and Su, 2016). In addition, the number of $AIRE^+$ cells are limited to a single tissue in mice, i.e., less than 50,000 in a single murine thymus (Lovewell et al., 2011; Anderson and Su, 2016). Therefore, it is more appropriate to use cell line models to study the molecular mechanisms of AIRE. In this chapter, we investigated whether AIRE regulates the genes encoding BAs associated with APS-1 using RT-qPCR and the dual luciferase reporter assay system. In addition, we examined the effect of the most common mutations of *AIRE* on the expression of AIRE-dependent genes that encode BA.

4.2 Validating *AIRE* expression in vitro

4D6 cells, which were originally derived from the thymic medullary epithelium of a child undergoing cardiac surgery (a gift from Dr Matthieu Giraud, Centre for Research in Transplantation and Immunology, Nantes, France), were tested for *AIRE* expression using RT-qPCR and western blotting. Low levels of mRNA were detected using RT-qPCR, and no AIRE protein was detected by western blotting using anti-AIRE antibody (sc- 373703, SantaCruz Biotechnology). After the transfection of 4D6 cells with pCMV-AIRE plasmid (see Appendix Section 8.5), the expression of *AIRE* at the mRNA and protein levels was evaluated by qRT-PCR and western blotting.

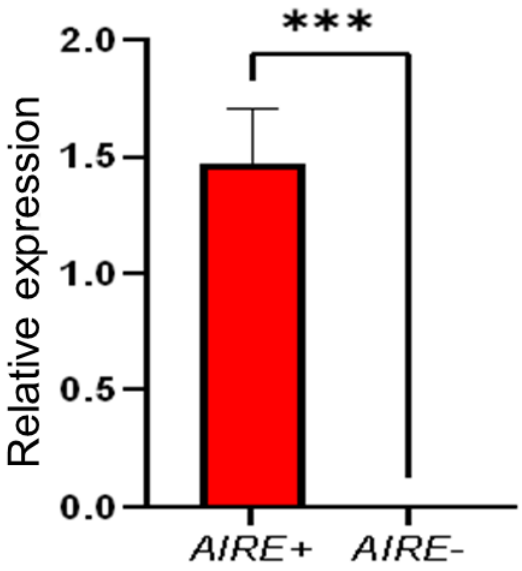

Figure 4.1: *AIRE* expression in transfected and untransfected 4D6 cells. 4D6 cells were transfected with pCMV-AIRE. Quantitative RT-PCR analysis was performed to assess the difference in *AIRE* expression between cells transfected or untransfected with *AIRE*. The red bar represents the expression of *AIRE* in cells transfected with pCMV-AIRE. Results are normalised to the expression of *ACTB* (β-actin). Data were analysed by t test. Error bars represent 1SEM.

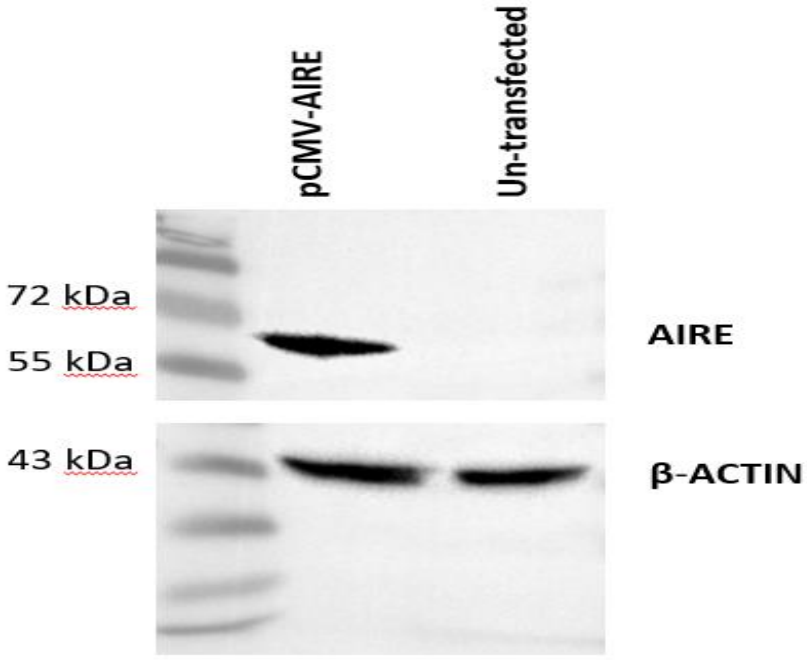

Figure 4.2: AIRE expression in transfected 4D6 cells. 4D6 cells were transfected with pCMV-AIRE or pCMV empty plasmid. Expression was evaluated by western blotting using anti-AIRE antibody (sc- 373703, SantaCruz Biotechnology) after 48h of transfection. The AIRE band is visible (AIRE 58 kDa) in transfected 4D6 cells as well as the loading control (β-actin 42 kDa).

4.3 Regulation by AIRE in vitro

To assess the ability of AIRE to regulate the genes encoding BAs, 4D6 cells were transfected with pCMV-AIRE or pCMV empty plasmid to compare the expression of BA genes in the presence and absence of AIRE. After the transfection, RNA was extracted and cDNA was synthesised for use in RT-qPCR. Among all the genes tested, the expression of *SOX9* and *TH* increased significantly in the cells transfected with pCMV-AIRE with p-values of 0.02 and 0.04, respectively (Figure 4.3). *SOX9* and *TH* expression levels increased in the cells transfected with pCMV-AIRE by 1.4 and 1.3 folds, respectively (Figure 4.3). *INS* (AIRE-dependent) was used as a positive control and was upregulated as expected ($p = 0.015$). INS expression levels increased in the cells transfected with pCMV-AIRE by 11 folds (Figure 4.5). *FOXN1* (AIRE-independent) was used as a negative control and as expected, its expression remained the same in cells with or without AIRE (Figure 4.5). The expression of *GAD65, GH, GHRH, SOX10* and *TPH* remained the same regardless of the presence or absence of AIRE (Figure 4.4).

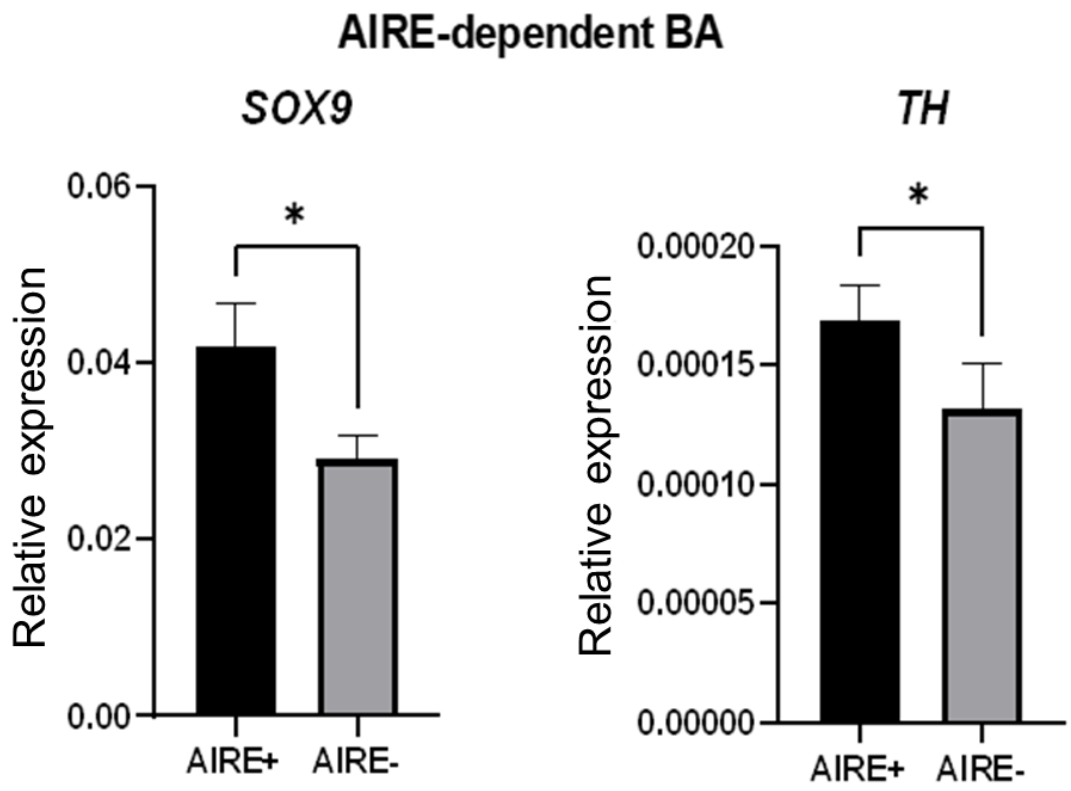

Figure 4.3: Expression of AIRE-dependent genes in AIRE-positive and AIRE-negative 4D6 cells. The expression of *SOX9* and *TH* were tested in AIRE-positive and AIRE-negative 4D6 cells. *SOX9* and *TH* were significantly upregulated in cells that express AIRE. The black bars represent expression in 4D6 cells transfected with pCMV-AIRE. The grey bars represent expression in 4D6 cells transfected with pCMV empty plasmid. Data were analysed by t test. **** $P < 0.0001$, *** $P < 0.0005$, ** $P < 0.005$ and * $P < 0.05$. Error bars represent 1SEM.

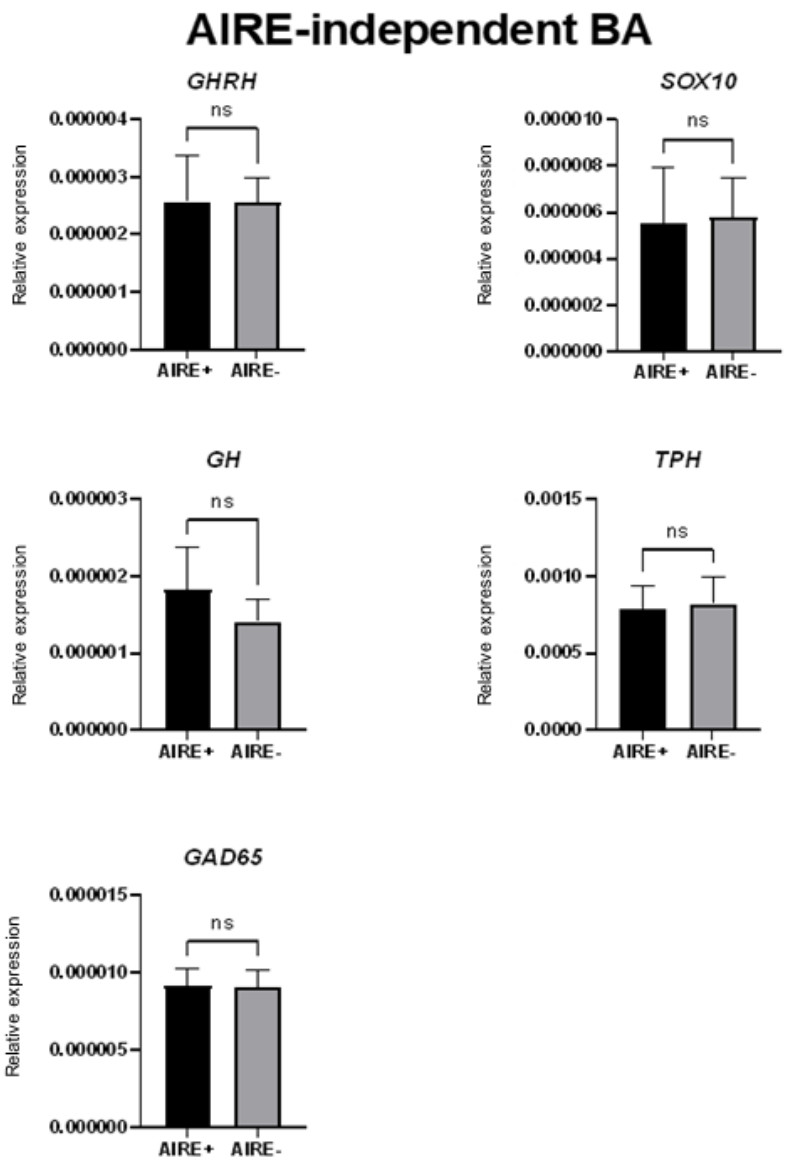

Figure 4.4: Expression of AIRE-independent genes in AIRE-positive and AIRE-negative 4D6 cells. The expression of *GAD65, GH, GHRH, SOX10* and *TPH* was tested in AIRE-positive and AIRE-negative 4D6 cells. Expression was the same with no significant difference regardless of the presence or absence of AIRE. The black bars represent expression in 4D6 cells transfected with pCMV-AIRE. The grey bars represent expression in 4D6 cells transfected with pCMV empty plasmid. Error bars represent 1SEM.

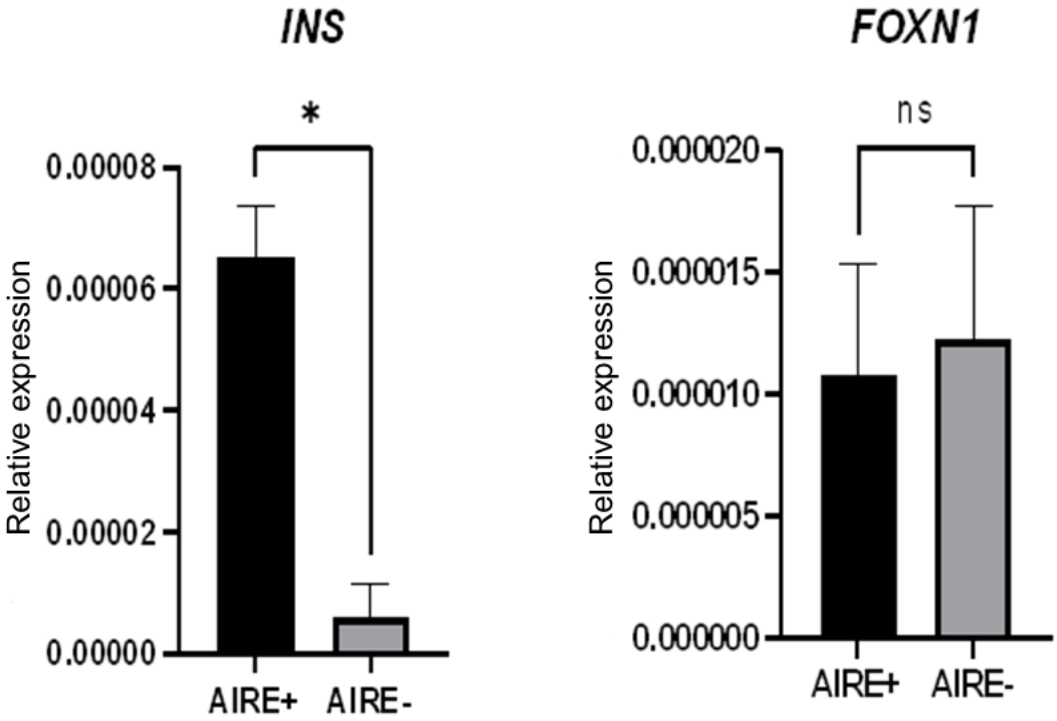

Figure 4.5: Expression of *INS* and *FOXN1* in AIRE-positive and AIRE-negative 4D6 cells. The expression of *INS* (AIRE-dependent) and *FOXN1* (AIRE-independent) was tested in AIRE-positive and AIRE-negative 4D6 cells. The expression of *INS* increased significantly in cells expressing AIRE. The expression of *FOXN1* showed no significant difference in the presence and absence of AIRE. The black bars represent expression in 4D6 cells transfected with pCMV-AIRE. The grey bars represent expression in 4D6 cells transfected with pCMV empty plasmid. Error bars represent 1SEM.

4.4 Activation of the promoters of genes encoding BA by AIRE

4.4.1 Identifying and cloning the promoter region

Dual luciferase assay was performed to test the ability of AIRE to activate the promoters of the genes encoding BAs. For this purpose, the promoters of the BA genes (*GAD65, GH, GHRH, SOX9, SOX10, TH and TPH*), positive control (*INS*) and negative control (*FOXN1*) were cloned into pGL3-Basic vectors. Promoters were identified using the EPD and the sequences of the promoters must include the TATA box or CAAT box (Table 4.1). Chromosomal positions were identified using UCSC Genome Browser on Human (GRCh38/hg38; https://genome-euro.ucsc.edu/cgi-

bin/hgGateway). The primers were designed using Primer3plus and DNA was extracted from 4D6 cells to amplify the promoter regions. After amplification, the samples were electrophoresed and amplified DNA fragments were extracted from the gel, purified and ligated into the Pgl3-Basic vector. Thereafter, competent cells were transformed to amplify the plasmids. Finally, the plasmids were isolated for later use in the dual luciferase reporter assay. The sequences of the promoter sequences are in the appendix (see Appendix Section 8.2).

Table 4.1. Location of the promoters of genes encoding BAs and controls:

Gene	Band	Promoter size	Strand	Chromosomal positions
SOX9	17q24.3	135 bp	+	Chr17: 72,120,898 - 72,121,032
TH	11p15.5	79 bp	-	Chr11: 2,171,806 - 2,171,884
TPH	11p15.1	149 bp	-	Chr11: 18,040,789 - 18,040,937
SOX10	22q13.1	528 bp	-	Chr22: 37,984,392 - 37,984,919
GAD65	10p12.1	141 bp	+	Chr10: 26,216,166 - 26,216,306
GH	17q23.3	75 bp	-	Chr17: 63,918,853 - 63,918,927
GHRH	20q11.23	301 bp	-	Chr20: 37,261,799 - 37,262,078
INS	11p15.5	123 bp	-	Chr11: 2,161,210 - 2,161,332
FOXN1	17q11.2	977 bp	+	Chr17: 28,505,234 - 28,506,210

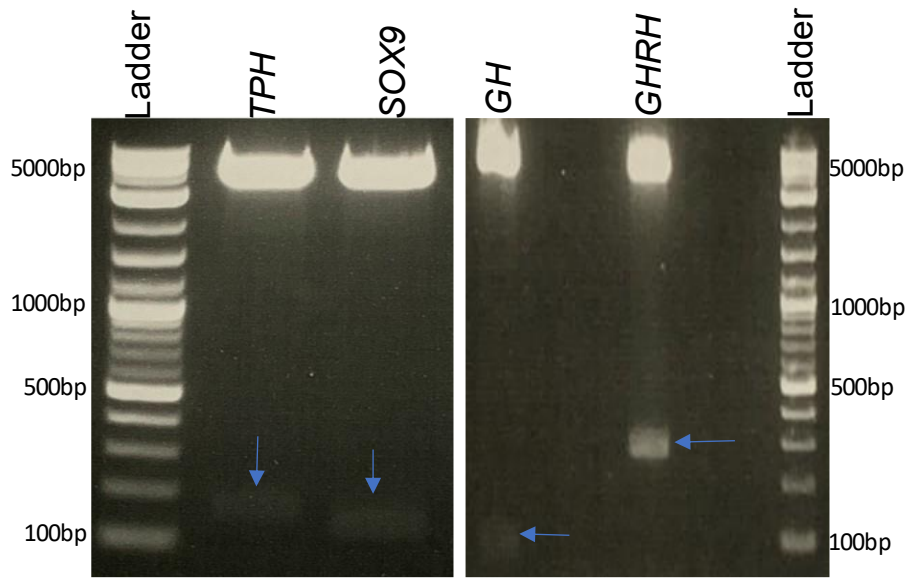

Figure 4.3.1: Digested PgL3-Basic vector containing the promoters of genes encoding BAs. Plasmids containing the promoters of *TPH, SOX9, GH and GHRH* were digested. The upper bands represent the digested PgL3 vector and the lower bands represent the promoters at the expected sizes.

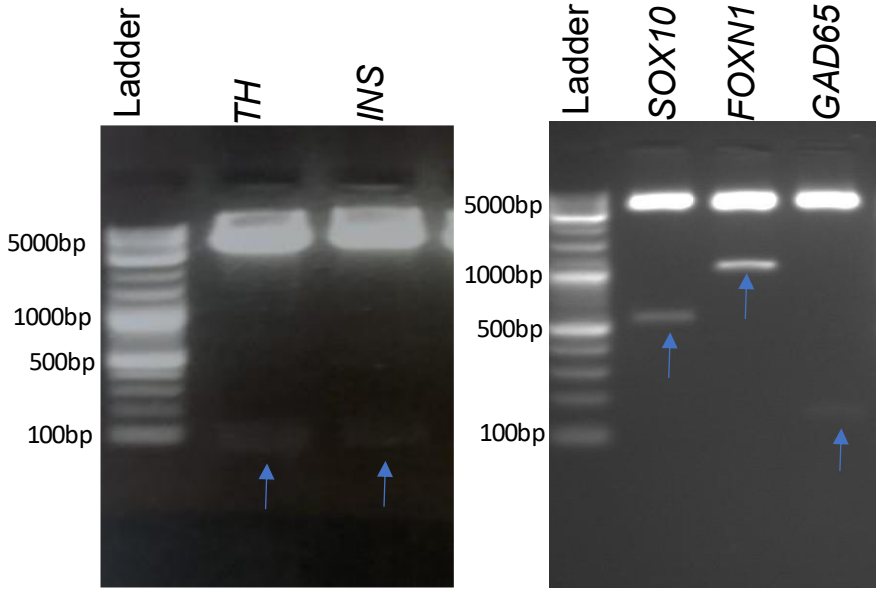

Figure 4.3.2: Digested PgL3-Basic vector containing the promoters of genes encoding BAs and the control. Plasmids containing the promoters of *TH, INS, SOX10, FOXN1* and *GAD65* were digested. The upper bands represent the digested PgL3 vector and the lower bands represent the promoters that marked with arrows.

4.4.2 Transfection and dual luciferase reporter assay

After successful cloning of all the desired promoters, transfection was performed to test the ability of AIRE to activate the promoters of genes encoding BAs by the

luciferase reporter assay. Cells were co-transfected with the pGL3-Basic vector and pBIND vector (See Appendix Section 8.5). The PgL3-Basic vector contained the promoters and firefly luciferase while the pBIND vector was used as a normalising transfection control since it contains a Renilla luciferase gene under control of a constitutive SV40 promoter and in addition was cloned with full-length *AIRE* to compare the activity of the promoters with or without AIRE. Protein lysates from cells transfected with pBind-*AIRE* or pBind empty plasmid were used to detect the expression of AIRE protein by western blotting using anti-AIRE antibody (sc- 373703, SantaCruz Biotechnology; Figure 4.3.5). Thereafter, the dual luciferase assay was performed to assess the ability of AIRE to activate the promoters of the genes encoding BA. The results showed that the expression of the vectors containing the promoters of *SOX9* and *TH* increased significantly in cells expressing AIRE compared with cells lacking AIRE ($p = 0.004$ and 0.0009, respectively; Figure 4.3.5). The expression of the vectors containing the promoters of *SOX9* and *TH* increased in cells expressing AIRE compared with cells lacking AIRE by 2.9 and 6.9 folds, respectively (Figure 4.3.5). However, the expression of the promoters belonging to the other genes encoding BA remained the same with or without AIRE (Figure 4.3.7). In addition, the promoter of *INS* (AIRE-dependent) was activated in the presence of AIRE as expected ($p < 0.0001$); however, there was no difference in the expression of the vector containing the promoter of *FOXN1* in the presence or absence of AIRE.

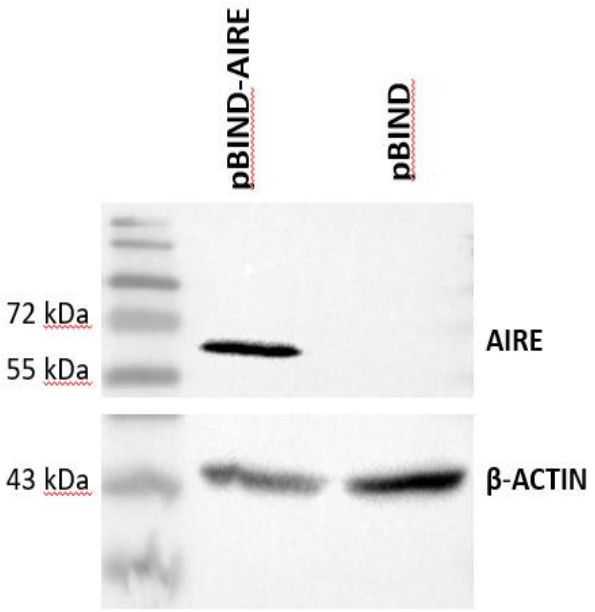

Figure 4.3.5: AIRE expression in transfected 4D6 cells. 4D6 cells were transfected with pBIND-AIRE or pBIND negative control. AIRE expression was evaluated via western blotting using anti-AIRE antibody. The AIRE band is visible (AIRE 58 kDa) in transfected 4D6 cells along with the loading control (β-actin 42 kDa).

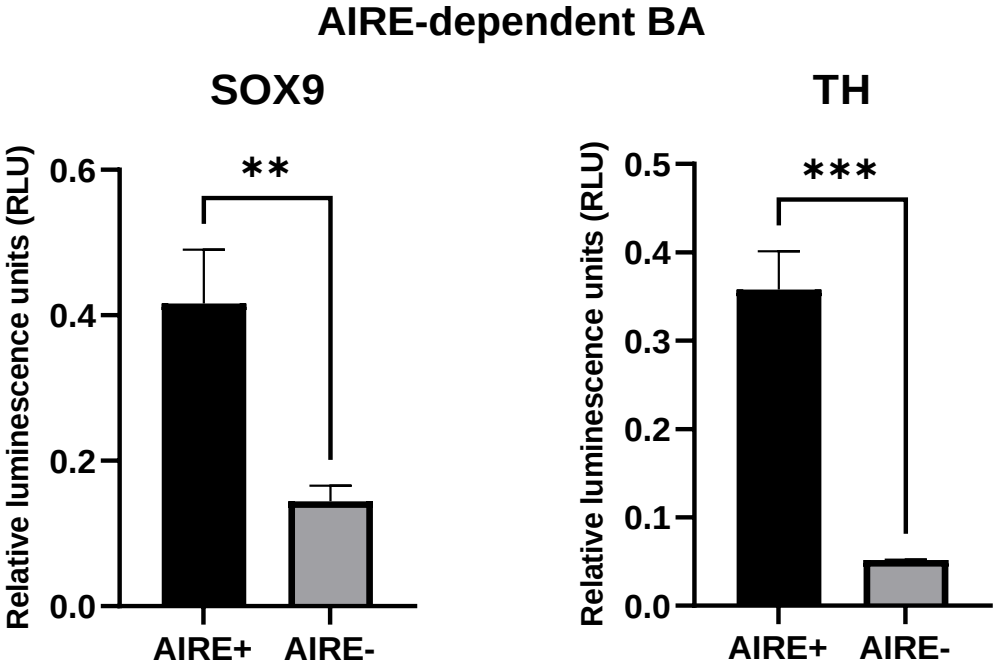

Figure 4.3.6: Dual luciferase assay in the presence and absence of AIRE to assess the promoters of AIRE-dependent BA genes relative to renilla activity. The promoters of *SOX9* and *TH* were activated in cells expressing AIRE. The black bars represent expression in 4D6 cells transfected with pBIND-AIRE and the grey bars represent expression in cells that do not express AIRE. Data were analysed by t test. **** $P < 0.0001$, *** $P < 0.0005$, ** $P < 0.005$ and * $P < 0.05$. Error bars represent 1SEM.

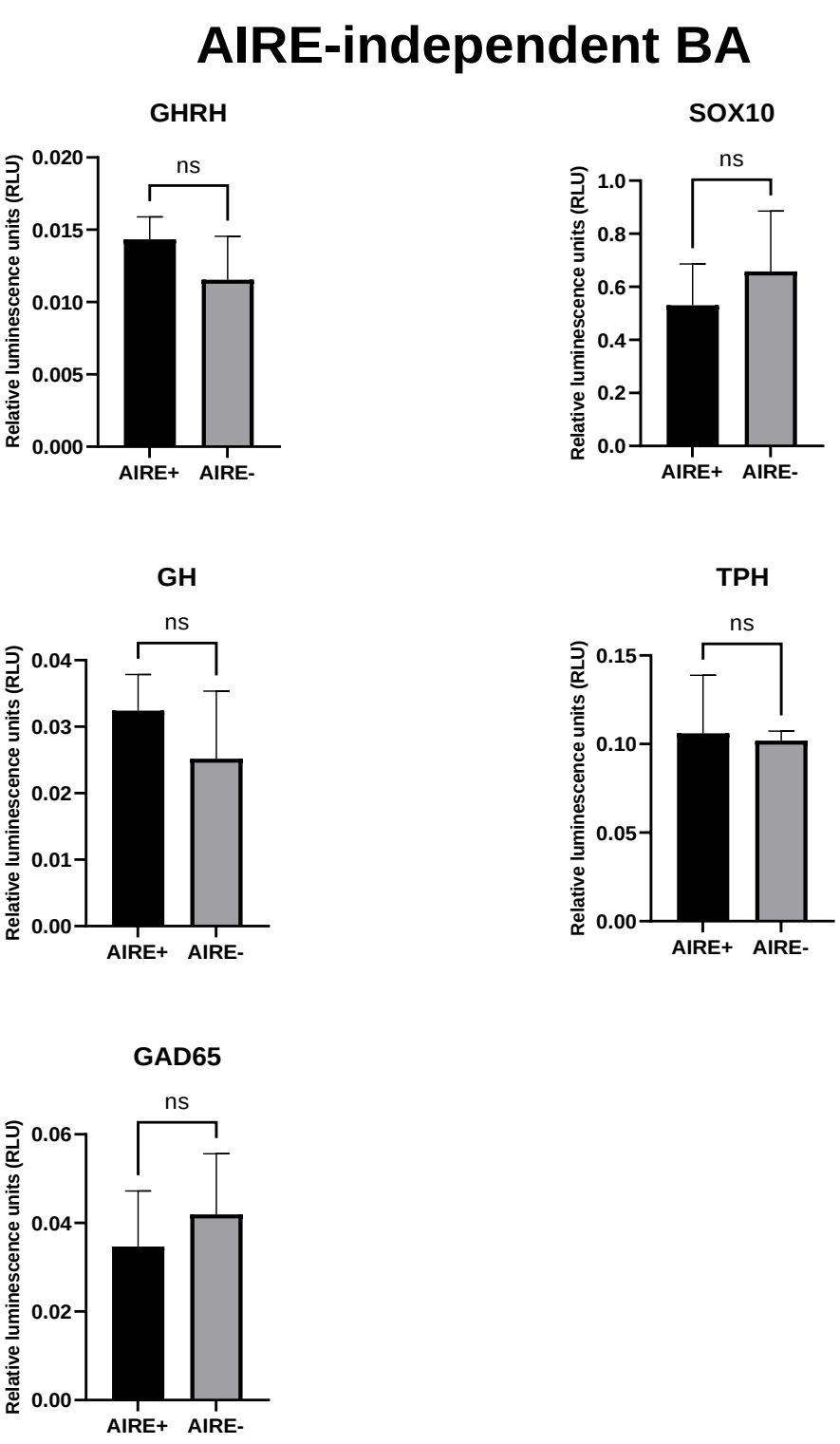

Figure 4.3.7: Dual luciferase assay in the presence and absence of AIRE to assess the promoters of AIRE-independent BA genes relative to renilla activity. No significant difference in the expression of vectors containing *GAD65, GH, GHRH, SOX10* and *TPH* was observed in the presence or absence of AIRE. The black bars represent expression in 4D6 cells transfected with pBIND-AIRE and the grey bars represent expression in cells that do not express AIRE. Data were analysed by t test. Error bars represent 1SEM.

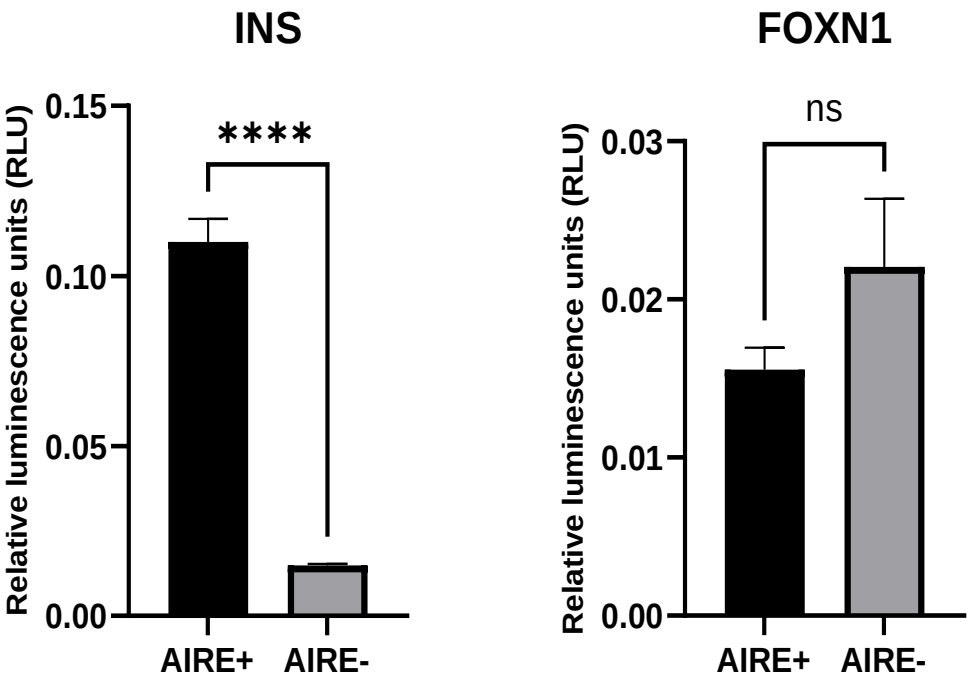

Figure 4.3.8: Dual luciferase assay in the presence and absence of AIRE to assess promoters of positive and negative control genes relative to renilla activity. The expression of the vector containing the promoter of *INS* was activated in the presence of AIRE, while *FOXN1* showed no significant difference in the presence or absence of AIRE. The black bars represent expression in 4D6 cells transfected with pBIND-AIRE and the grey bars represent expression in cells that do not express AIRE. Data were analysed by t test. **** $P < 0.0001$, *** $P < 0.0005$, ** $P < 0.005$ and * $P < 0.05$. Error bars represent 1SEM.

4.5 The effect of APS-1 mutations on AIRE transactivation potential

To test the effect of the most common AIRE mutations on the expression of BAs, three common APS-1 mutations, c.254A>G (Y85C), c.415C>T (R139X) and c.769C>T (R257X), were introduced into the *AIRE* gene using the QuickChange Lightning Site-directed Mutagenesis Kit (Agilent Technologies; see Chapter 2). These three mutants and c.682T>G (p.G228W), provided by the lab, were used to test the effect of APS-1 mutations on the activity of *SOX9* and *TH* promoters by dual luciferase assay. In addition, the promoters of *INS* and *FOXN1* were used as the positive and a negative control, respectively. Interestingly, the promoters of *SOX9* and *TH* were activated with no significant difference between the cells expressing wild-type AIRE and those

expressing the AIRE Y85C mutant (Figure 4.4.1). The Y85C mutation results in an amino acid change from tyrosine to cysteine in exon 2, which affects the CARD domain. The CARD domain is involved in AIRE homomultimerisation, and it seems that the Y85C mutation does not affect the transactivation potential of AIRE directly. Instead, the mutated polypeptide degrades rapidly intracellularly compared with the wild-type polypeptides (Ramsey et al., 2002a). This may explain why APS-1 patients harbouring the Y85C mutation (Iranian Jewish) show milder clinical phenotypes compared with patients having the R139X or R257X mutations, since the mutation does not affect the transactivation potential of AIRE directly but the generated polypeptide has a shorter half-life. However, AIRE mutations R139X, R257X and G228W seemed to affect the expression of AIRE-dependent genes. The stop codon introduced the expressed cDNA construct, which is not divided into separate exons may not cause nonsense- mediated mRNA destabilisation and loss of mRNA as expected in the endogenous gene. R257X mutation causes C to T substitution at nucleotide 769 situated in exon 6 which leads to a truncated 257 residue protein. R257X mutation cause deletion of part of the SNAD domain as well as both PHD domains (Halonen et al., 2002). R139X mutation causes C to T substitution at nucleotide 415 located in exon 3 which leads to a truncated 139 residue protein which cause a total absence of AIRE (De Martino et al., 2013). The activity of *SOX9*, *TH* and *INS* differed significantly between cells expressing wild-type AIRE and those expressing mutated AIRE variants R139X, R257X and G228W (Figures 4.4.2–4.4.4). The activity of *SOX9* and *TH* promoters was reduced in the cells expressing AIRE mutant R139X by 3.5 and 17.6 folds, respectively (Figure 4.4.2). The activity of *SOX9* and *TH* promoters was reduced significantly with AIRE mutant R139X (p-values of 0.002 and < 0.0001, respectively; Figure 4.4.2). Similarly, AIRE mutants R257X and

G228W adversely affected the activity of *SOX9* and *TH* promoters. The activity of *SOX9* and *TH* promoters was reduced in the cells expressing AIRE mutant R257X by 3.7 and 17.7 folds, respectively (Figure 4.4.3). The activity of *SOX9* and *TH* promoters decreased significantly with AIRE mutant R257X (p-values of 0.0039 and < 0.0001, respectively; Figure 4.4.3). The activity of *SOX9* and *TH* promoters was reduced in the cells expressing AIRE mutant G228W by 20.8 and 13.6 folds, respectively (Figure 4.4.4). The corresponding p-values with AIRE mutant G228W were 0.006 for *SOX9* and 0.0001 for *TH* (Figure 4.4.4). In addition, the *INS* promoter (AIRE-dependent) was not activated with AIRE variants R139X, R257X and G228W, similar to *SOX9* and *TH*, whereas no difference was observed in the expression of the vector containing the promoter of *FOXN1* in the presence of either wild-type or mutated AIRE.

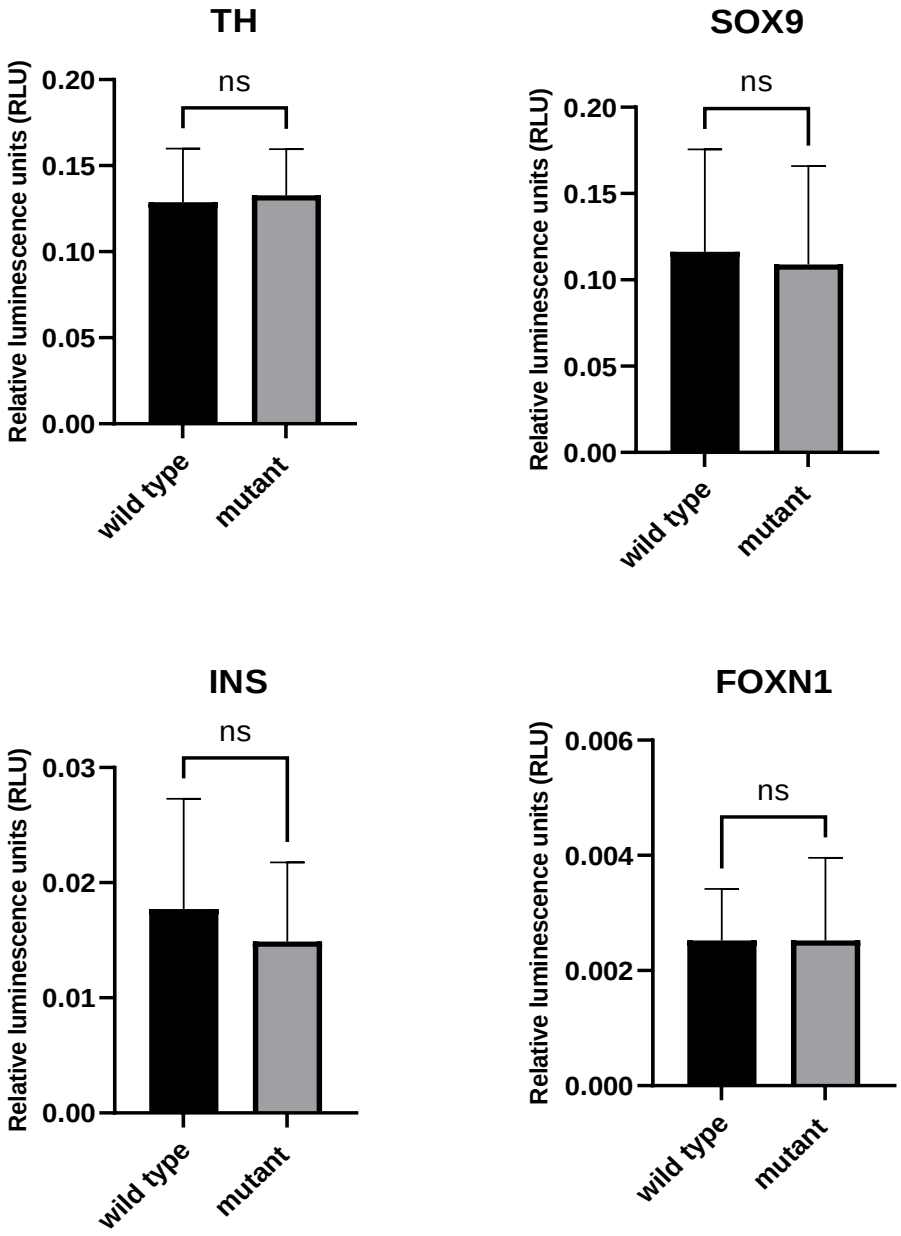

Figure 4.4.1: Dual luciferase assay of *SOX9* and *TH* with wild-type AIRE and c.254A>G (Y85) mutant AIRE relative to renilla activity. The promoters of *SOX9* and *TH* were activated in cells expressing wild-type AIRE and the c.254A>G AIRE mutant with no significant difference. The promoters of *INS* and *FOXN1* were activated in cells expressing wild-type AIRE and the c.254A>G AIRE mutant with no significant difference. The black bars represent expression in cells expressing wild-type AIRE while the grey bars represent expression in cells expressing mutant AIRE. Data were analysed by t test. Error bars represent 1SEM.

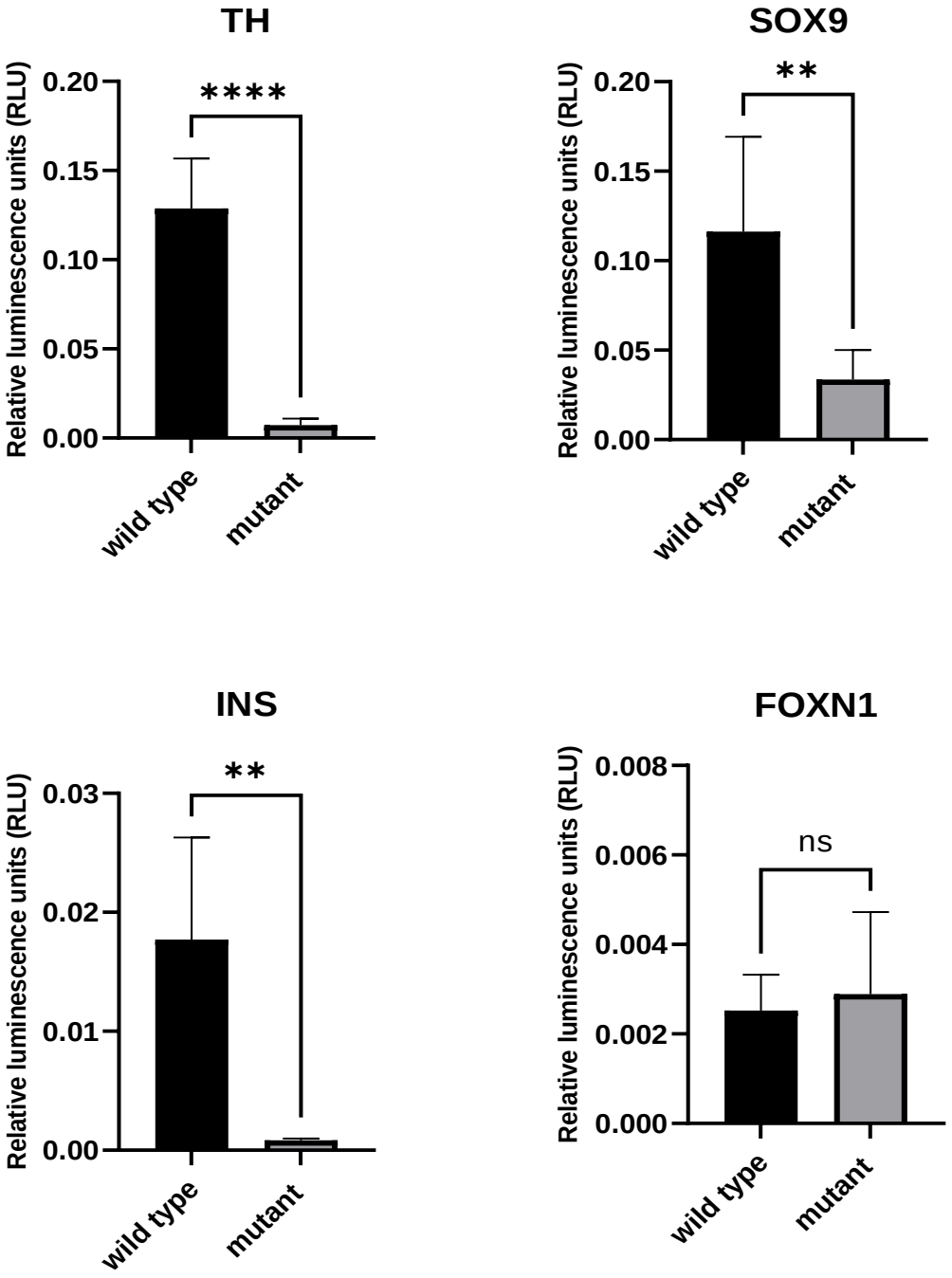

Figure 4.4.2: Dual luciferase assay of *SOX9* and *TH* with wild-type AIRE and c.415C>T mutant AIRE relative to renilla activity. The expression of *SOX9* and *TH* decreased significantly in cells expressing the mutant version of AIRE compared with the wild type. The expression of *INS* decreased significantly in cells expressing the mutant version of AIRE compared with the wild type, while *FOXN1* showed no changes. The black bars represent expression in cells expressing wild-type AIRE while the grey bars represent expression in cells expressing mutant AIRE. Data were analysed by t test. Error bars represent 1SEM.

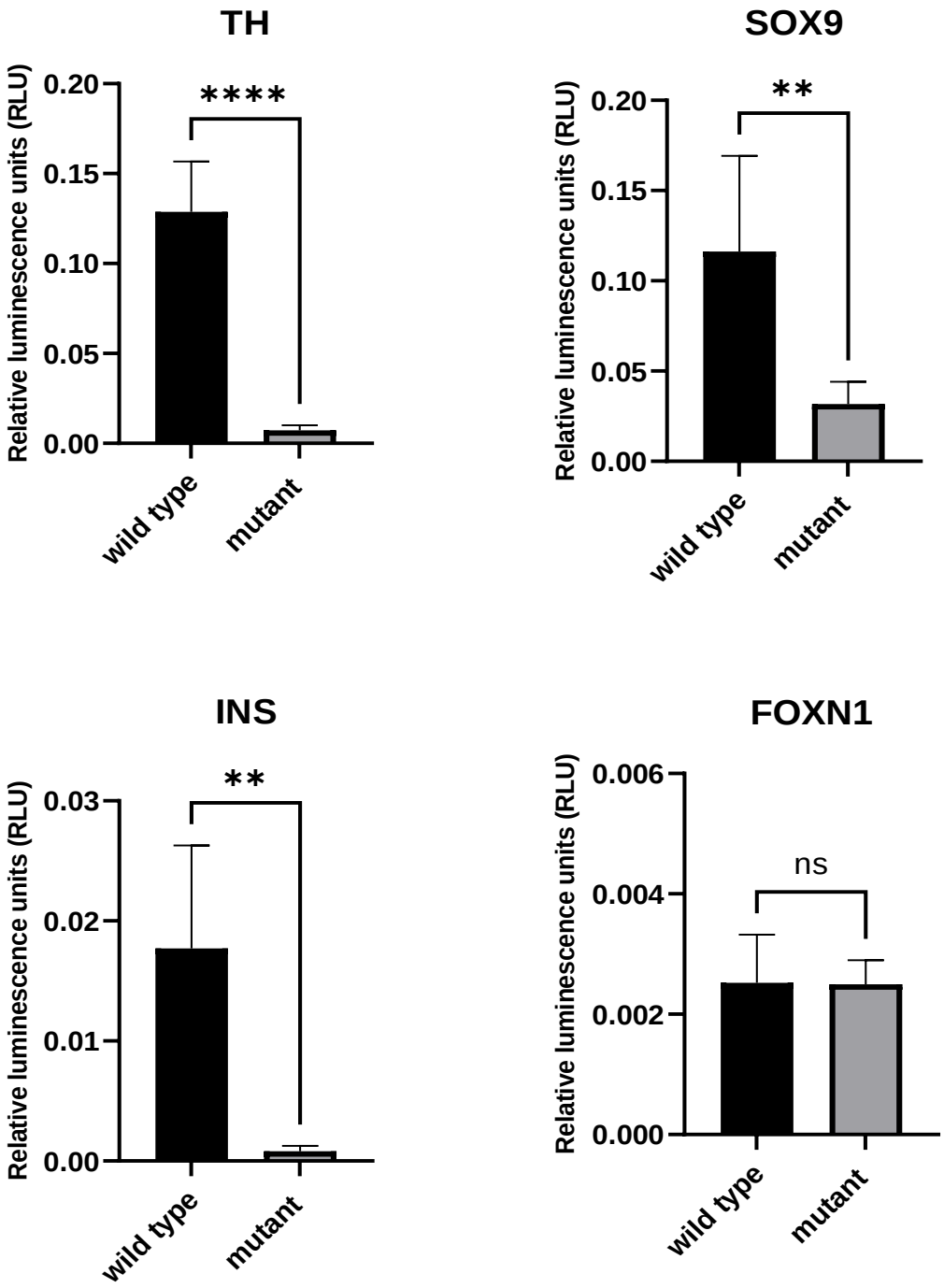

Figure 4.4.3: Dual luciferase assay of *SOX9* and *TH* with wild-type AIRE and c.769C>T mutant AIRE relative to renilla activity. The expression of *SOX9* and *TH* decreased significantly in cells expressing the mutant version of AIRE compared with the wild type. The expression of *INS* decreased significantly in cells expressing the mutant version of AIRE compared with the wild type, while *FOXN1* showed no changes. The black bars represent expression in cells expressing wild-type AIRE while the grey bars represent expression in cells expressing mutant AIRE. Data were analysed by t test. Error bars represent 1SEM.

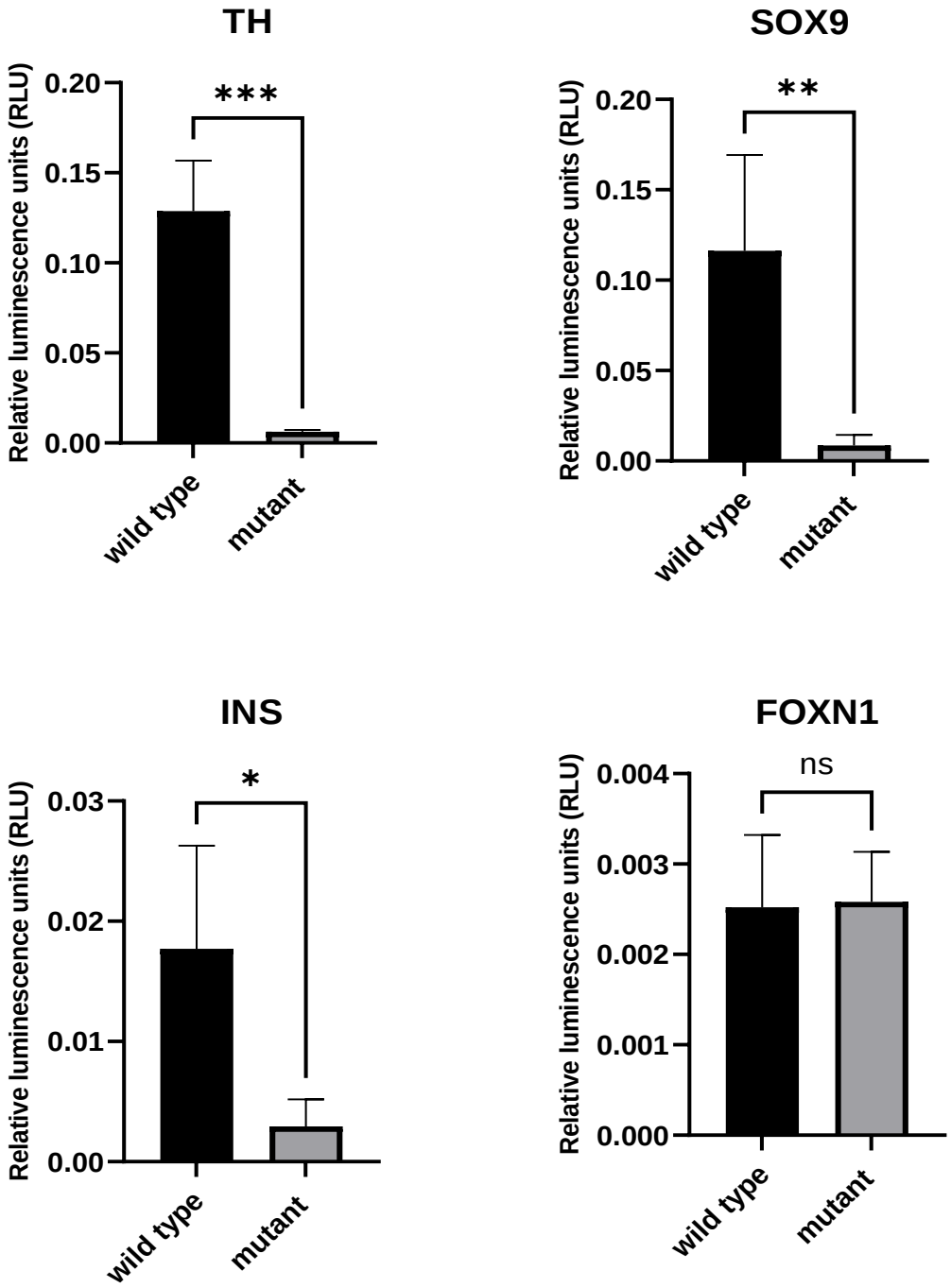

Figure 4.4.4: Dual luciferase assay of *SOX9* and *TH* with wild-type AIRE and c.682T>G mutant AIRE relative to renilla activity. The expression of *SOX9* and *TH* decreased significantly in cells expressing the mutant version of AIRE compared with the wild type. The expression of *INS* decreased significantly in cells expressing the mutant version of AIRE compared with the wild type, while *FOXN1* showed no changes. The black bars represent expression in cells expressing wild-type AIRE while the grey bars represent expression in cells expressing mutant AIRE. Data were analysed by t test. Error bars represent 1SEM.

4.6 Discussion

As mentioned at the beginning of this chapter, cell lines are more appropriate models to study the molecular mechanisms of AIRE (Lovewell et al., 2011). In this project, 4D6 cells were used to assess the effect of AIRE expression on BA gene expression and promoter activity via RT-qPCR and the dual luciferase reporter assay. RT-qPCR and western blotting were performed to test the expression of *AIRE*/AIRE in transfected and untransfected cells. Cells transfected with the pCMV-AIRE vector showed much higher amounts of AIRE mRNA, with more than a 1000-fold change compared with untransfected cells (Figure 4.1). At the protein level, AIRE was detected in transfected cells by western blotting but not in untransfected cells (Figure 4.2).

To assess the ability of AIRE to regulate the genes encoding BA, 4D6 cells transfected with pCMV-AIRE or with pCMV empty plasmid were used to compare the expression of genes with overexpression of AIRE or without AIRE. Analysis of the data obtained by RT-qPCR demonstrated that the expression of *SOX9* and *TH* increased significantly in cells transfected with *AIRE* compared with cells that do not express AIRE, with p-values of 0.02 and 0.04, respectively (Figure 4.3). The expression of other genes encoding BAs showed no significant difference with or without AIRE. Our data show that AIRE can upregulate *TH* in C57BL/6 mice (Chapter 3, Figure 3.11) and in human mTEC; however, *SOX9* can be upregulated in human cell line model but not in mice and *SOX10* is only upregulated in mice. *SOX9* and *SOX10* are among the SRY-box genes which are a family of transcription factors. SOX10 and SOX9 are not like other SOX proteins since the majority of SOX proteins function as transcriptional activators but SOX9 and SOX10 can function as transcriptional repressors as well (Shakhova et al., 2015). SOX9 and SOX10 are expressed in neural crest cells during

embryonic development (Lee et al., 2016). SOX10 is expressed at all stages of development in embryonic and adult melanocytes and is required for their generation and homeostasis (Shakhova et al., 2015; Haseeb and Lefebvre, 2019). The loss of SOX10 is reported to lead to a severe decrease of most neural crest derivatives and causes Waardenburg-Hirschsprung syndrome, which includes pigmentary abnormalities and aganglionic megacolon (blockage of the large intestine; Bondurand et al., 2007; Shakhova et al., 2015). However, loss of SOX9 affects the development of mesectodermal derivatives like cartilage, craniofacial bones and smooth muscle cells. In addition, heterozygous mutations in *SOX9/Sox9* lead to campomelic dysplasia in both humans and mice, characterised by skeletal malformation, cleft palate and XY sex reversal syndrome (Shakhova et al., 2015). Interestingly, SOX9 and SOX10 are not only brain autoantigens but also associated with skin manifestations such as vitiligo and alopecia areata in APS-1. In addition, it was suggested by Hedstrand et al. (2001) that SOX9 and SOX10 share epitopes since all APS-1 patients who showed reactivity against SOX9 in their study also displayed reactivity against SOX10 (Haseeb and Lefebvre, 2019).

To test the ability of AIRE to activate the promoters of the genes encoding BAs, the dual luciferase reporter assay was performed, for which the promoters of the genes encoding BAs, a positive control (*INS*) and a negative control (*FOXN*1) were cloned into pGL3-Basic vectors. The obtained data showed that the expression of the vectors containing the promoters of *SOX9* and *TH* increased significantly in cells expressing AIRE compared with cells lacking AIRE expression with p-values of 0.004 and 0.0009, respectively (Figure 4.3.6). However, no significant difference in the activity of the other tested promoters with or without AIRE was observed (Figure 4.3.7).

In addition, different APS-1 mutations were introduced to test the effect of the most common *AIRE* mutations on the expression of *SOX9* and *TH*. Three common APS-1 mutations c.254A>G (Y85C), c.415C>T (R139X) and c.769C>T (R257X) were introduced into *AIRE,* besides the *AIRE* mutation G228W. These mutations were selected because they cause APS-1 in Finns, Sardinians and Iranian Jews, populations in which APS-1 occurs is highly prevalent (1:14,000 in Sardinians, 1:25,000 in Finns and 1:6500 to 1:9000 in Iranian Jews). Moreover, the activity of *SOX9* and *TH* promoters was tested with the AIRE G228W mutant because it is the only mutation described to be inherited in a dominant negative mode, which prevents the wild-type variant from forming the complex with its partners that is needed for transactivation by binding to wild-type AIRE (Ilmarinen et al., 2005; Oftedal et al., 2015). The selected mutations R139X, and R257X are nonsense mutations that result in truncated versions of AIRE, while Y85C and G228W are missense mutations (Ilmarinen et al., 2005; De Martino et al., 2013; Oftedal et al., 2015). All tested AIRE mutations adversely affected the transactivation potential of AIRE except Y85C. The promoters of *SOX9* and *TH* were activated with no significant difference between cells expressing wild-type AIRE and those expressing the AIRE mutant Y85C (Figure 4.4.1). Our findings are in keeping with previous studies, which found that the AIRE Y85C mutant stimulated the tested promoters as effectively as wild-type AIRE (Björses et al., 2000; Ramsey et al., 2002a). However, pulse-chase experiments revealed that the polypeptide generated from the AIRE Y85C mutated sequence had a shorter half-life compared with wild-type polypeptides (Ramsey et al., 2002a). This observation suggests that the Y85C mutation does not affect the transactivation activity of AIRE directly; instead, the mutant protein is rapidly degraded. In addition, it should be noted that APS-1 patients with the Y85C mutation show milder clinical symptoms; for

example, Iranian Jews with the AIRE Y85C mutation do not develop chronic mucocutaneous candidiasis, which is the most common manifestation of APS-1 (Ramsey et al., 2002a; Arstila et al., 2013). Further research is required to verify that AIRE degradation is the main cause of the milder clinical symptoms associated with Iranian Jewish APS-1 patients, as even if the Y85C mutation leads to rapid degradation of most of the protein, the remaining protein may retain some biological activity. On the other hand, as expected, the nonsense mutations R139X and R257X, which produce totally non-functional truncated proteins, adversely affected the activity of AIRE-dependent genes. The activity of *SOX9* and *TH* promoters was significantly higher with wild-type AIRE than with truncated AIRE R139X, with p-values of 0.002 and < 0.0001, respectively (Figure 4.4.2). The activity of *SOX9* and *TH* promoters also decreased significantly with the AIRE mutant R257X, with p-values of 0.0039 and < 0.0001, respectively (Figure 4.4.3). Similarly, the AIRE mutant G228W adversely affected the activity of the *SOX9* promoter (p = 0.006) and *TH* promoter (p = 0.0001; Figure 4.4.4).

The results obtained from experiments in *Aire* mice and human mTEC (4D6 cell line) show that not all the genes encoding BAs are regulated by AIRE in the thymus. However, it cannot be confirmed that AIRE does not regulate these genes encoding BA in 4D6 cell line because only short promoters are used in this experiment so using longer promoters which might include enhancer upstream might be crucial. In addition, AIRE expression in this cell line is overexpressed in the absence of overexpression of any cofactor which might affect the regulation of AIRE. Moreover, AIRE may regulate these genes extra-thymically in the secondary lymphoid organs such as the spleen and lymph nodes. As a result, autoreactive T cells against these antigens may be selected and deleted in the secondary lymphoid organs (spleen and lymph nodes) as

suggested by Gardner et al. (2008); see Chapter 3. In addition, it is not surprising that AIRE upregulates *SOX10* in mice but not in humans whereas *SOX9* is only upregulated in humans. Previous work by our group showed that AIRE regulates different sets of genes in human cell line and mice (Lovewell et al., 2018). Microarray data of AIRE-dependent genes in humans obtained by our group from mTEC (TEC1A3 cell line) was compared with microarray data obtained previously by two other groups using mouse mTEC from *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice on the NOD, Balb/c and B6 genetic backgrounds (Guerau-de-Arellano et al., 2008; Venanzi et al., 2008; Lovewell et al., 2018). Comparison of the human dataset with the four datasets from mice with different genetic backgrounds revealed 2000 human probes that have at least one murine probe of the three genetic backgrounds belonging to a homologous gene. However, the comparison between the human cell line and mouse models revealed only 428 conserved human genes that had at least one homologous gene in at least three out of the four murine datasets (Lovewell et al., 2018). These findings indicate that AIRE-regulated genes vary between mice and human cell line (Lovewell et al., 2018). AIRE-dependent genes vary between mice and human cell line because overexpression of AIRE in these cells induce much lower number of TSA genes comparing to the number of genes induced in vivo (Abramson et al., 2010, Giraud et al., 2014, Besnard et al., 2021). The number of TSA genes induced by AIRE is much lower in human cell line than in vivo because 3D medullary microenvironment and certain signalling is necessary for the expression of AIRE and TSA genes and disruption of medullary microenvironment leads to the reduction of AIRE and TSA genes expression (*see section 1.2.4*; Rossi et al., 2007; Irla et al., 2008).

Chapter Five

Defining the role of AIRE as a DNA-binding element by ChIP-seq

5.1 Outline

As mentioned previously, contradicting ideas exist about the role of AIRE as a DNA-binding element. However, there is no strong evidence showing that AIRE binds to DNA even if it has been shown by gel shift assay that AIRE can bind to two different consensus binding sequence motifs, namely ATTGGTTA and TTATTA (Kumar et al., 2001; Purohit et al., 2005). In addition, Purohit et al. (2005) demonstrated that the PHD domains in AIRE are responsible for binding to the ATTGGTTA sequence motif, while the TTATTA motif was shown to be bound by the SAND domain. However, later studies have contradicted these findings by showing that AIRE is not a conventional transcription factor since it recognises its dependent genes through recognition of the repressive epigenetic signature directly through its PHD1 domain or indirectly by binding to other proteins (AIRE partners) instead of binding to DNA (*see section 1.2.4*; Org et al., 2008). In addition, it was found that AIRE binds weakly and non-specifically to DNA and no common motif has been identified for AIRE-dependent genes (Koh et al., 2008; Org et al., 2008). Moreover, The SAND domain in other proteins has been shown to be a DNA binding domain, however, the SNAD of AIRE lacks the motif responsible for DNA binding (KDWK; Waterfield et al., 2014). In this chapter, Chip-seq was performed to identify locations in the genome bound by AIRE using 4D6 cells to shed light on the molecular function of AIRE.

5.2 ChIP-seq assay

4D6 cells (5×10^5 cells) were seeded in three 35-mm dishes one day before transfection. The cells in two dishes were transfected with pCMV3-AIRE-c-myc, and the cells in the third dish were transfected with pCMV3-c-myc as a negative control vector using FuGENE® HD (Promega) at a ratio of 3:1 according to the manufacturer's protocol. The following day, when the number of cells reached around 4×10^6 cells per dish, the ChIP assay was performed using the Pierce Magnetic ChIP Kit

(Thermofisher) according to the manufacturer's protocol. The ChIP assay consists of four steps, namely crosslinking and cell pellet isolation, lysis and MNase digestion, immunoprecipitation and IP elution, and DNA recovery (see Chapter 2). After the DNA was purified, its concentration in the samples (positive and negative) was measured using a NanoDrop™ 2000 spectrophotometer. After elution, the DNA concentration of ChIP sample 1 was 3.68 ng/μl and that of ChIP sample 2 was 5.49 ng/μl as means of triplicate values, while the negative control sample (ChIP NC) had no DNA, as expected (Table 5.1).

Table 5.1. Amount of recovered DNA after ChIP.

ChIP sample	**DNA quantity (ng/μl)**	**Total DNA (ng)**
ChIP sample 1	3.68	110.4
ChIP sample 2	5.49	164.7
Negative control	-	-

5.3 Next-generation sequencing (NGS)

The immunoprecipitated (ChIP samples 1 and 2) and input (inputs 1 and 2) samples were sent to Macrogen (South Korea) for next-generation sequencing (NGS) and bioinformatic analysis. Before sequencing was performed, samples were prepared and the library was constructed. For library construction, DNA is extracted from samples and after quality control (QC), qualified samples proceed to library construction. After that, the DNA was randomly fragmented, followed by 5' and 3' adapter ligation. Then, the fragments containing the adapters were amplified by PCR and the DNA was purified following gel electrophoresis. For cluster generation, the DNA was loaded into a flow cell slide containing primers on the surface that can capture and bind to the adapters ligated to the fragments. When cluster generation is complete, the templates are ready for sequencing. Finally, sequencing was done by Illumina sequencing by synthesis (SBS) technology and sequencing data is converted

into raw data for the analysis. The total number of bases sequenced for ChIP samples 1 and 2 was 2,408,633,012 and 2,857,609,466 bp, respectively, and for inputs 1 and 2 was 3,652,885,998 and 3,678,919,606 bp, respectively (Table 5.2). However, the total reads, i.e, the sum of reads 1 and 2 were 15,951,212 and 18,924,566, respectively (Table 5.2). The ratio of bases that have Phred quality score of over 20 (Q20) of the raw data was over 97% and the ratio of bases that have Phred quality score of over 30 (Q30) was over 93% (Table 5.2). Phred quality score numerically expresses the accuracy of each nucleotide. Greater Q number indicates higher accuracy. For example, Q20 represents the chances of having base call error are 1 in 100 with a corresponding call accuracy of 99% and Q30 the chances of having base call error are 1 in 1000 with a corresponding call accuracy of 99.9%.

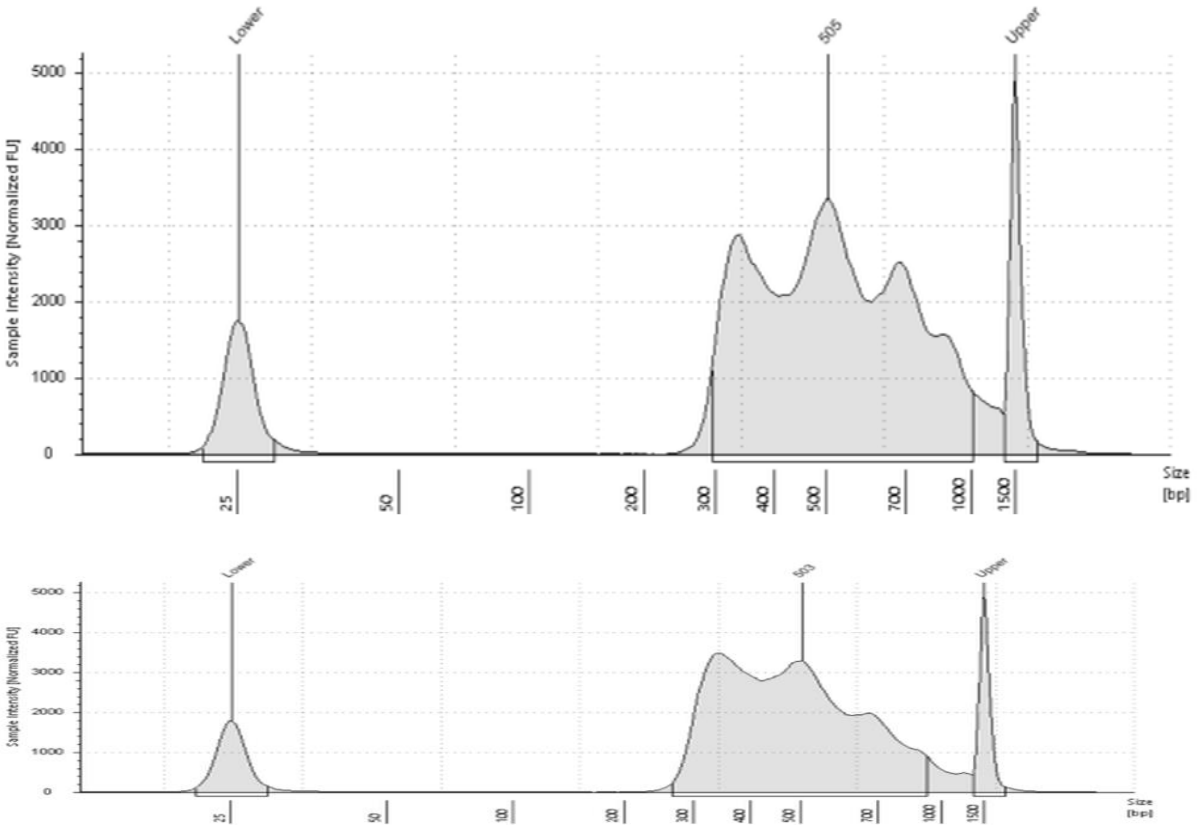

Figure 5.1: Distribution and library size of immunoprecipitated samples. The library sizes of ChIP sample 1 (upper panel) and ChIP sample 2 (lower panel) are shown to lie between the lower and upper markers (around 500 bp). It appears that that ChIP sample 2 has a lower size distribution.

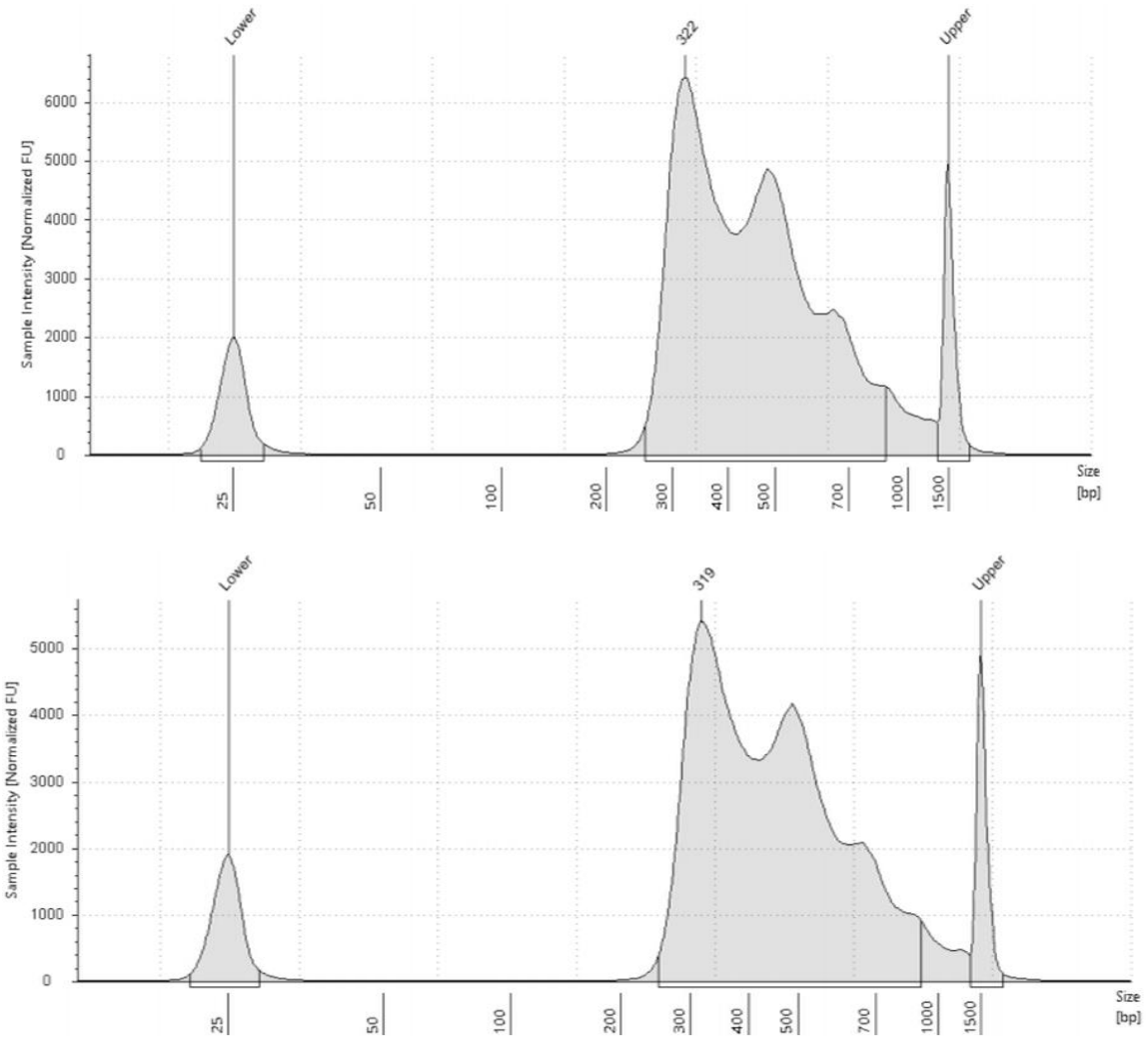

Figure 5.2: Distribution and library size of input samples. The library sizes of input 1 (upper panel) and input 2 (lower panel) are shown to lie between the lower and upper markers (around 300 bp).

Table 5.2: Raw data statistics.

Sample ID	Total reads bases (bp)	Total reads	GC (%)	AT (%)	Q20 (%)	Q30 (%)
ChIP sample 1	2,408,633,012	15,951,212	42.19	57.81	97.52	93.27
Input 1	3,652,885,998	24,191,298	41.97	58.03	97.67	93.74
ChIP sample 2	2,857,609,466	18,924,566	44.83	55.17	97.74	93.78
Input 2	3,678,919,606	24,363,706	41.76	58.24	97.77	93.96

5.4 Bioinformatic analysis of the ChIP-seq data

Prior to data analysis, QC of the raw sequencing data was done using the BBDUK software to remove low-quality reads while preserving those of high quality. The minimum length was set to 35 bp and the quality score to 25 (Q25). The number of reads was counted before and after QC (Table 5.3). The number of reads was

trimmed from 7,975,606 to 7,970,778 in ChIP sample 1 and from 9,462,283 to 9,457,802 in ChIP sample 2. The read quality was checked before and after the trimming step and the quality was improved after trimming.

Table 5.3: Number of reads before and after QC.

Sample name	Raw reads	Trimmed reads
ChIP sample 1	7,975,606	7,970,778
Input 1	12,095,649	11,068,595
ChIP sample 2	9,462,283	9,457,802
Input 2	12,181,853	11,202,674

The high-quality reads were aligned against the reference genome with Minimap2. MACS2 (https://github.com/macs3project/MACS, version 2.2.4) was used for peak calling. Model-based analysis of ChIP-seq (MACS) is a tool used for NGS ChIP-seq data analysis. MACS uses dynamic parameter to successfully capture the biases in the sequence of the genome, allowing for better and sensitive predictions of binding sites. The number of peaks obtained by MACS2 was 127 for ChIP sample 1 and 50 for ChIP sample 2 (Table 5.4). These data indicated that the number of AIRE-binding sites were low, suggesting that AIRE binds DNA poorly, as reported previously (Koh et al., 2008). To perform peak annotation, ChIPseeker was used. ChIPseeker supports the annotation of ChIP peaks and provides functions to visualise ChIP peak coverage over chromosomes and profiles of peaks binding to TSS regions. Most of the peaks were found in distal intergenic regions in the two samples. Most ChIP sample 1 peaks (124 out of 127) were annotated as AIRE binding to the distal intergenic DNA of 30 chromosomal regions, and the remaining three were in the introns of three different genes (Tables 5.4 and 5.5). On the other hand, all 50 peaks of ChIP sample 2 were

annotated as AIRE binding to the distal intergenic DNA of 19 chromosomal regions and none were annotated as intronic DNA (Tables 5.4 and 5.6). The peaks were found very far from TSS since the nearest peak were 70 kb away from TSS but in most cases the distance to TSS is more than 100 kb. Fifteen chromosomal regions were common to both samples, indicating regions where AIRE could bind preferentially (Figure 5.3 and Table 5.7). As only four regions were obtained in ChIP sample 2 and 15 regions in ChIP sample 1, AIRE binding to these regions could be uncertain. Interestingly, our data showed that AIRE did not bind to any promoter or enhancer of AIRE-dependent genes but bound the distal intergenic sites of some genes, most of which were functionless pseudogenes (Table 5.7). However, interestingly, all 15 regions covered by AIRE were in the pericentromeric region (the chromosomal region around the centromere) of different chromosomes (Figure 5.4). Not only were the 15 regions covered in both samples found in pericentromeric region but also most of the regions exclusively covered in sample 1 or sample 2. In addition, it was found that 19 peaks overlap between the two samples (Table 5.8). The chromosomal positions of the peaks were identified using UCSC Genome Browser on Human (GRCh38/hg38; https://genome-euro.ucsc.edu/cgi-bin/hgGateway). The overlapping peaks were found in the centromeric region of different chromosomes (Figure 5.4). These data suggest that AIRE does not bind to the promoter or enhancer of AIRE-dependent genes but may instead form a complex with cohesins to ensure robust cohesion of sister chromatids or be involved in the process of chromosome bi-orientation during mitosis when kinetochores attach to microtubules from opposite spindle poles by binding to specific DNA sequences in the pericentromeric region, by binding either indirectly to pericentromeric cohesions or directly to specific DNA sequences in the

pericentromeric region (see Discussion). Additionally, the peaks called by MACS2 were further annotated using HOMER to verify the presence of a common motif in the peaks or any transcription factor binding motif. HOMER analyses regulatory elements in genomic applications using a novel motif discovery algorithm. Homer uses two sets of DNA sequences (immunoprecipitated and input samples) and identify the specifically enriched regulatory elements in the immunoprecipitated sample in relative to the input sample. The previously reported motifs ATTGGTTA (G box) and TTATTA (T box) were not detected among the AIRE-enriched sequences; however, previous research found that AIRE binds to motifs with multiple boxes such as GG and TGG boxes, and motifs with multiple boxes have higher binding affinity than motifs with a single box (Ruan et al., 2007). The HOMER de novo motif results showed that AIRE bound mostly to TCTGCAAGTGGA, which has a TGG box, in both samples (ChIP1 and ChIP2), with p-values of 1e-195 and 1e-95 compared with the input samples, respectively. In addition, the HOMER de novo motif results revealed two more motifs that AIRE might bind, namely GTGTGTGTTCAA, with p-values of 1e-137 in ChIP sample 1 and 1e-90 in ChIP sample 2, and CGCTTTGAGGAC, with p-values of 1e-101 in ChIP sample 1 and 1e-43 in ChIP sample 2, compared with the input samples (Table 5.10). These three motifs do not resemble known binding sites for any known transcription factors. HOMER de novo motif results show that AIRE bound to the three motifs in both samples (ChIP1 and ChIP2), with very low P-values comparing to the input samples, which indicates strong evidence that the results is very significant and unlikely to be random (Table 5.10).

Table 5.4: Number of peaks called by MACS2.

Sample name	Peak number
ChIP sample 1	127
ChIP sample 2	50

Table 5.5: Number of annotation events in ChIP sample 1.

Annotation event	Number of events	Region covered
Distal intergenic	124	30
Intron	3	3

Table 5.6: Number of annotation events in ChIP sample 2.

Annotation event	Number of events	Regions covered
Distal intergenic	50	19

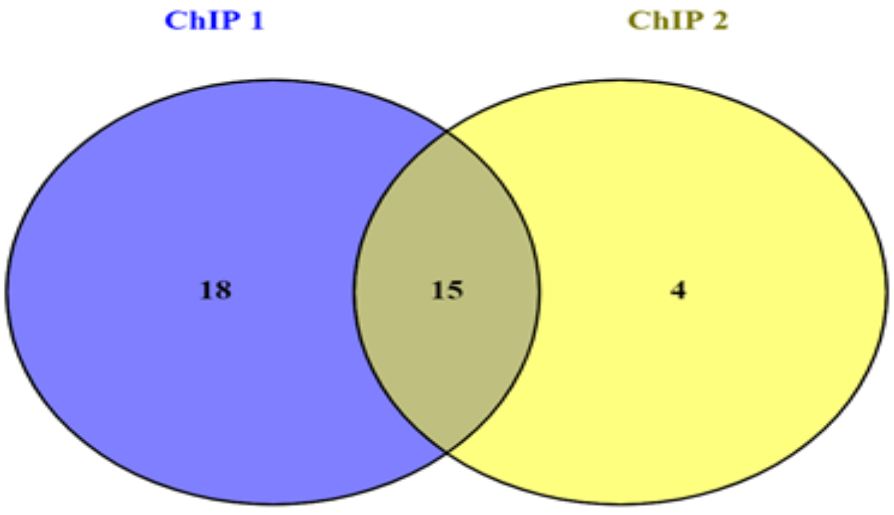

Figure 5.3: Chromosomal regions enriched by AIRE. ChIP sample 1 covered 33 chromosomal regions while ChIP sample 2 covered 19 chromosomal regions. Fifteen regions were common in both.

Table 5.7: List of annotated genes or pseudogenes bound by AIRE and common to the two sets of ChIP-seq experiments performed.

15 common elements in ChIP samples 1 and 2	18 elements exclusive to ChIP sample 1	4 elements exclusive to ChIP sample 2
ACTR3BP2 ALG10 BMS1P14 CWH43 EMB EMBP1 HAVCR1P1 HCN1 LINC00662 LOC441666 LOC644669 LOC646813 MIR4522 ROCK1 ZNF716	ASNSP1 C12orf71 CHEK2P2 CYTOR DCUN1D4 FAM230C FRG1CP KCNU1 LINC00680 LINC02167 LSP1P5 MIR663AHG MTRNR2L9 OR11H12 OR4C46 POTEA XPR1 ZXDA	ANKRD26P1 ANKRD30BP2 LOC101927050 PROS1

Table 5.8: Locations of the peaks on hg38 genome and the length of each peak of the two samples. The overlapping peaks between the two samples were highlighted with the same colour in each chromosome.

Chromosome number	**Locations of the peaks of sample 1 on hg38 (length of peak)**	**Locations of the peaks of sample 2 on hg38 (length of peak)**
Chromosome 1	chr1:121775143-121775562 (420) chr1: 122524659-122525030 (372) chr1: 122912494-122912854 (361) chr1: 124196867-124197207 (341) chr1: 124204941-124205280 (340) chr1: 124280116-124280456 (341) chr1: 124475766-124476183 (418) chr1: 124603572-124603983 (412) chr1: 125179629-125180555 (927) chr1: 143214388-143214794 (407) chr1: 143217269-143217836 (568) chr1: 143263616-143264146 (531) chr1: 180701616-180702062 (447)	chr1: 122524644-122524966 (323) chr1: 122603037-122603366 (330) chr1: 122734866-122735224 (359) chr1: 123679094-123679442 (349) chr1: 123960073-123960541 (469) chr1: 124603545-124604005 (461) chr1: 125179764-125180210 (447)
Chromosome 2	chr2: 87419009-87419577 (569) chr2: 92292247-92292644 (398) chr2: 92295997-92297009 (1013) chr2: 92339371-92339709 (339) chr2: 92367794-92368434 (641) chr2: 92378175-92378777 (603)	chr2: 90387870-90388387 (518) chr2: 92296018-92296493 (476) chr2: 92367856-92368465 (610) chr2: 92482495-92482936 (442) chr2: 92911032-92911447 (416) chr2: 93249096-93249441 (346)

Chromosome number	Locations of the peaks of sample 1 on hg38 (length of peak)	Locations of the peaks of sample 2 on hg38 (length of peak)
	chr2: 92387398-92387809 (412) chr2: 92468202-92468984 (783) chr2: 92482483-92482948 (466) chr2: 92489967-92490526 (560) chr2: 92561996-92562389 (394) chr2: 92603261-92603640 (380) chr2: 92605776-92606461 (686) chr2: 92617319-92617742 (424) chr2: 92780289-92780635 (347) chr2: 92911025-92911453 (429) chr2: 92934623-92935011 (389) chr2: 93346399-93346776 (378) chr2: 93605461-93605837 (377) chr2: 93639145-93639482 (338) chr2: 93851836-93852308 (473) chr2: 94066500-94067020 (521) chr2: 94077968-94078494 (527)	chr2: 93564759-93565156 (398) chr2: 93851796-93852314 (519)
Chromosome 3	-	chr3: 93323708-93324053 (346)
Chromosome 4	chr4: 49099383-49099732 (350) chr4: 49102338-49102889 (552) chr4: 49110364-49110738 (375) chr4: 49149157-49149792 (636) chr4: 49154192-49154566 (375) chr4: 49633398-49634378 (981) chr4: 49635183-49635592 (410) chr4: 51547629-51548130 (502)	-
Chromosome 5	chr5: 47472886-47473263 (378) chr5: 47486199-47486607 (409) chr5: 47862735-47863105 (371) chr5: 48485028-48485389 (362) chr5: 48861676-48862029 (354) chr5: 49045610-49046093 (484) chr5: 49243967-49244378 (412) chr5: 49310873-49311258 (386) chr5: 49409482-49409987 (506) chr5: 49536464-49536820 (357) chr5: 49601596-49602113 (518) chr5: 49659729-49660192 (464)	chr5: 47359607-47359937 (331) chr5: 48485026-48485400 (375) chr5: 48931046-48931443 (398) chr5: 49409606-49409976 (371)
Chromosome 6	chr6: 58755331-58755669 (339) chr6: 59807802-59808140 (339)	-
Chromosome 7	chr7: 58437277-58437681 (405) chr7: 58699486-58699943 (458) chr7: 58722295-58722645 (351) chr7: 58946468-58946806 (339) chr7: 59473525-59473885 (361) chr7: 59535633-59536107 (475) chr7: 59734571-59734931 (361) chr7: 59883305-59883694 (390) chr7: 59908418-59908794 (377) chr7: 60253425-60253815 (391)	chr7: 58345397-58345824 (428) chr7: 58437277-58437689 (413) chr7: 58556769-58557141 (373) chr7: 58722260-58722688 (429) chr7: 59205044-59205402 (359) chr7: 59686989-59687356 (368) chr7: 59751642-59751984 (343) chr7: 59811517-59811899 (383) chr7: 60298271-60298588 (318) chr7: 60338151-60338660 (510)
Chromosome 8	chr8: 36489093-36489570 (478) chr8: 44523370-44523743 (374) chr8: 44930258-44930713 (456) chr8: 45328091-45328543 (453)	-
Chromosome 10	chr10: 41883161-41884046 (886)	chr10: 41883400-41883794 (395)

Chromosome number	Locations of the peaks of sample 1 on hg38 (length of peak)	Locations of the peaks of sample 2 on hg38 (length of peak)
Chromosome 11	chr11: 51489558-51490155 (598) chr11: 51569621-51570062 (442) chr11: 51643588-51643957 (370) chr11: 51965679-51966169 (491) chr11: 52645026-52645448 (423) chr11: 52802659-52803017 (359) chr11: 54049404-54049802 (399)	chr11: 51243479-51243822 (344)
Chromosome 12	chr12: 27151729-27152177 (449) chr12: 34970375-34970765 (391) chr12: 35930666-35931050 (385) chr12: 36013924-36014260 (337) chr12: 36038696-36039164 (469)	chr12: 36168680-36169064 (385)
Chromosome 13	chr13: 16415019-16415511 (493)	-
Chromosome 14	chr14: 17107521-17108031 (511) chr14: 17250835-17251184 (350) chr14: 17301826-17302299 (474)	-
Chromosome 15	chr15: 19114753-19115333 (581) chr15: 19221991-19222530 (540)	-
Chromosome 16	chr16: 36722458-36722808 (351) chr16: 37268325-37268672 (348)	-
Chromosome 17	chr17: 25353066-25353446 (381) chr17: 26603804-26604266 (463) chr17: 26619485-26620149 (665)	chr17: 25353074-25353412 (339) chr17: 26619550-26620107 (558)
Chromosome 18	chr18: 16706165-16706527 (363) chr18: 16822491-16822854 (364) chr18: 16993494-16993847 (354) chr18: 17301919-17302462 (544) chr18: 18115194-18115703 (510) chr18: 18225311-18225755 (445) chr18: 18479724-18480139 (416) chr18: 18737664-18738079 (416) chr18: 18769937-18770277 (341) chr18: 18827541-18827932 (392) chr18: 19413148-19413604 (457) chr18: 19942991-19943357 (367) chr18: 20450271-20450742 (472)	chr18: 16382474-16382837 (364) chr18: 17301907-17302340 (434) chr18: 18071442-18071910 (469) chr18: 18225334-18225762 (429) chr18: 18827549-18827930 (382) chr18: 19097751-19098102 (352)
Chromosome 19	chr19: 24916345-24916703 (359) chr19: 24930062-24930493 (432) chr19: 25050202-25050620 (419) chr19: 25140298-25140672 (375) chr19: 25259401-25259755 (355) chr19: 25873934-25874278 (345) chr19: 26426681-26427076 (396) chr19: 26812867-26813250 (384) chr19: 26866243-26866846 (604) chr19: 26923164-26923627 (464) chr19: 27110903-27111455 (553)	chr19: 24930085-24930409 (325) chr19: 26084533-26084891 (359) chr19: 26150412-26150907 (496) chr19: 26385412-26385751 (340) chr19: 26601471-26601802 (332) chr19: 26734187-26734505 (319)
Chromosome 20	chr20: 26604973-26605471 (499) chr20: 26902002-26902517 (516) chr20: 27074802-27075160 (359) chr20: 27407612-27408077 (466) chr20: 28237720-28238203 (484)	**-**
Chromosome 21	**-**	chr21: 12277734-12278207 (474)
Chromosome x	chrX: 59121056-59121512 (457) chrX: 60567552-60567888 (337)	

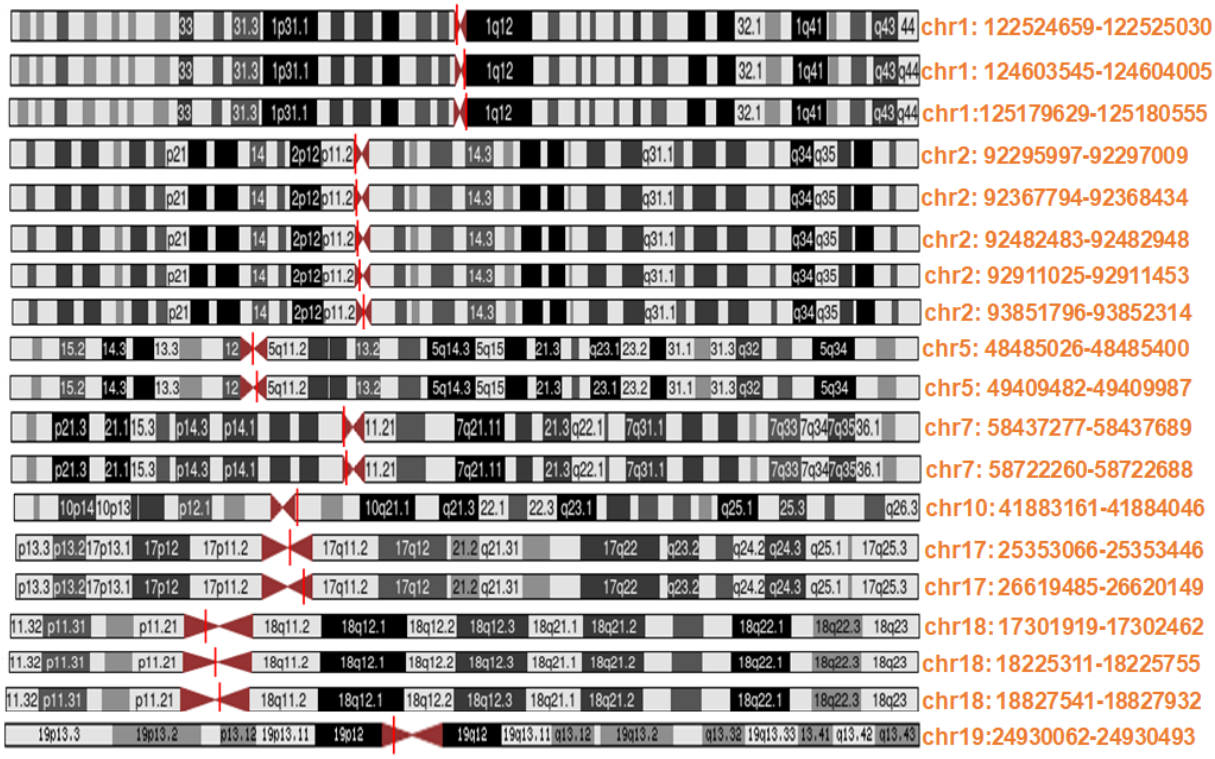

Figure 5.4: Chromosomal locations of 19 peaks overlapped between ChIP samples 1 and 2. The locations of overlapping peaks in the centromeric region marked with red line. The chromosomal locations of the peaks are found with orange colour in the right of the picture. The chromosomal locations of peaks identified using UCSC Genome Browser on Human (GRCh38/hg38).

Table 5.9: Chromosomes numbers of pericentromeric regions enriched by AIRE.

Chromosomes covered in both samples	Chromosomes covered exclusively in ChIP 1	Chromosomes covered exclusively in ChIP 2
Chromosome 1	Chromosome 6	Chromosome 3
Chromosome 2	Chromosome 8	Chromosome 21
Chromosome 4	Chromosome 13	
Chromosome 5	Chromosome 14	
Chromosome 10	Chromosome 15	
Chromosome 11	Chromosome 16	
Chromosome 12	Chromosome 20	
Chromosome 17	Chromosome X	
Chromosome 18		
Chromosome 19		

Table 5.10: Common HOMER de novo motifs in ChIP samples 1 and 2. These motifs do not resemble known binding sites for any known transcription factors.

Common motifs in both samples	p-value (ChIP 1)	p-value (ChIP 2)
TCTGCAAGTGGA	1e-195	1e-95
GTGTGTGTTCAA	1e-137	1e-90
CGCTTTGAGGAC	1e-101	1e-43

5.5 Discussion

In this ChIP-seq analysis, the number of peaks obtained by MACS2 was 127 for ChIP sample 1 and 50 for ChIP sample 2 (Table 5.3). The ChIP-seq data showed that AIRE did not bind to any promoter or enhancer of AIRE-dependent genes. These results are in keeping with previous studies that found that AIRE does not bind to the promoters or near the TSSs of its dependent genes (Koh et al., 2008; Org et al., 2008) and contradict the previous reports by Kumar et al. (2001) and Purohit et al. (2005) showing AIRE binding on ATTGGTTA or TTATTA motifs. However, HOMER de novo motif results showed that AIRE bound mostly to TCTGCAAGTGGA, which has a TGG box, in both samples (ChIP1 and ChIP2), with p-values of 1e-195 and 1e-95, respectively, compared with the input samples. Previous reports state that the SAND and PHD domains of AIRE bind strongly to guanine-containing motifs such as GG and TGG oligonucleotides (Purohit et al., 2005). Interestingly, HOMER de novo motif results show that AIRE bound to the three motifs in both samples (ChIP1 and ChIP2), with very low P-values comparing to the input samples, which suggests that the results is very significant and unlikely to be false positive (Table 5.10).

Interestingly our data showed that AIRE enriched the intergenic region in the pericentromeric region of many chromosomes, including 1, 2, 4, 5, 10, 11, 12, 17, 18 and 19, in both ChIP samples. The pericentromeric region is the region around the centromere in the chromosome that is associated with robust sister chromatid cohesions by the cohesin complex. Cohesins have been found to be bound along chromosome arms but enriched preferentially in intergenic regions surrounding the centromere in budding yeast (Blat and Klenckner, 1999; Tanaka et al., 2013). Cohesins in the pericentromeric region form a ring-shaped complex with some

proteins such as RAD21, SMC1 and SMC3 to ensure cohesion of sister chromatids and facilitate the attachment of sister kinetochores to microtubules from opposite spindle poles, a process known as chromosome bi-orientation or sister kinetochore bi-orientation, which occurs in mitosis (Tanaka et al., 2013). Interestingly, RAD21 and SMC3 have been found to be among the putative AIRE-associated proteins (Abramson et al., 2010), suggesting that AIRE binds to pericentromeric regions directly by binding the motifs revealed by de novo HOMER motif analysis (Table 5.10) or indirectly by binding to other proteins associated with the cohesin complex. The main role of the centromeric region, which includes the centromere and pericentromere, is to attach the chromosome to the spindle microtubules for chromosome segregation at the beginning of mitosis in prometaphase. For chromosome segregation to be performed, more than 100 proteins called kinetochores form a complex at the centromeric region (Cheeseman, 2014). Kinetochore proteins can be divided into three groups by location and function: the inner kinetochores which bind to the DNA of the centromeric region; the outer kinetochores, which bind to the microtubules; and the regulatory proteins that control the activities of the kinetochores (Cheeseman, 2014). Although $mTEC^{Hi}$ had been thought to be non-proliferative (postmitotic) and undergo apoptosis after the expression of AIRE, it was shown that a fraction of $mTEC^{Hi}$ differentiate to post-AIRE mTEC (Wang et al., 2012). The expression of AIRE, most AIRE-dependent genes, MHCII, CD80 molecules are lost in post-AIRE mTEC (Wang et al., 2012). Metzger et al. (2013) has been proposed that post-AIRE mTEC can convert to $mTEC^{lo}$ and then differentiated into $mTEC^{Hi}$ again. Therefore, AIRE may be involved in the cell division of $mTEC^{Hi}$. In addition, we cannot exclude its role in mitosis in extra-thymic AIRE-expressing cells. AIRE interact to TOP2α via its PHD1 domain

which is known to mediate chromatid separation and the relief of supercoiled DNA during mitosis and depletion of TOP2α from precondensed chromosomes reverses chromosome condensation and their separating is prevented, which indicate that TOP2α is necessary for maintenance of chromosome structure during mitosis (Bansal et al., 2017; Nielsen et al., 2020). This might suggest that AIRE has a completely separate function from that of a gene activator by recruiting TOP2α to sites of chromosome adhesion. There is no evidence that there are homologues of AIRE in organisms that do not have thymuses since this would suggest that its role as a thymic regulator of TSA genes expression is a recent co-option. The exact role of AIRE in the centromeric region and cohesions, whether in mTEC or extra-thymic AIRE-expressing cells, is yet to be discovered. It may be necessary to co-localise AIRE with cohesins and its partners that involved in mitosis on chromosomes to confirm its involvement in this region. In addition, validating whether AIRE can bind to the motifs identified in Table 5.10 using other DNA-binding assays such as the electrophoretic mobility shift assay (EMSA) may further our understanding of whether AIRE binds to these sequences directly or indirectly by binding to the cohesin complex or other partners.

Chapter Six

General discussion and Future plan

6.1 General discussion

APS-1 is a monogenic autosomal recessive disease that occurs due to mutations in *AIRE*. AIRE is a transcriptional regulator; it is expressed mainly in mTEC and plays an essential role in TSA expression, which is part of the process by which to eliminate autoreactive T cells (negative selection). Patients with APS-1 commonly show high levels of autoantibodies against proteins of the brain such as TPH, GH, GHRH, GAD65, SOX9, SOX10 and TH. According to recent studies, some APS-1 patients show neurological manifestations such as stiff person syndrome, cerebellar ataxia and memory loss. Besides learning and memory deficiencies, a study has also observed a reduction in total brain and cerebellum volumes in APS-1 and correlated this reduction to the presence of autoantibodies against GAD+ and TH+ neurons (Meloni et al., 2019). In this study, the role of AIRE in the pathogenesis of neurological disorders associated with APS-1 was investigated.

We investigated the brain of C57BL/6 *Aire*$^{-/-}$ mice but changes reported in the brains of APS-1 patients were not observed in *Aire* knockout mice. There are several possible reasons to explain this observation in the context of the mouse C57BL/6 genetic background. First, the genetic background strongly influences autoimmune phenotypes in *Aire* knockout mice. For example, mice with the C57BL/6 genetic background have been shown to develop a milder illness than, for example, the NOD strain. In addition, a group recently bred *Aire*$^{-/-}$ rats that developed a severe autoimmune phenotype that was not observed in *Aire*$^{-/-}$ mice, including alopecia, nail dystrophy, vitiligo and impaired thymus development (Ossart et al., 2018). Furthermore, the age of the mouse is relevant in the development of autoimmunity as it corresponds primarily to the duration of the disease. Mice between 15 and 30 weeks of age may be too young to develop signs of autoimmunity in the brain because the

lesions or T lymphocyte infiltration in some organs of *Aire*$^{-/-}$ mice or *Aire*$^{-/-}$ rats can appear at > 10 months (> 42 weeks) of age (Anderson et al., 2002; Ossart et al., 2018). As a result, examining the brains of *Aire*-deficient rats or mice with different genetic backgrounds at > 1 year of age may be essential to find out if they are able to mimic the neurological manifestations associated with APS-1 in humans.

Ex vivo and in vitro experiments in this study revealed that AIRE upregulates the expression of *Th/TH* in the thymus of C57BL/6 mice and in human mTEC (4D6 cell line). However, *SOX9* was only upregulated in 4D6 cells in the presence of AIRE, and *Sox10* was only upregulated in mice. It is not surprising that AIRE upregulates *Sox10* in mice but not in human cell line and *SOX9* is only upregulated in human cell line, as previous work by our group showed that AIRE regulates different sets of genes in immortal model cell line for human mTEC and mice (Lovewell et al., 2018). AIRE-dependent genes vary between mice and human cell line because overexpression of AIRE in these cells induce much lower number of TSA genes comparing to the number of genes induced in vivo (Abramson et al., 2010, Giraud et al., 2014, Besnard et al., 2021). The number of TSA genes induced by AIRE is much lower in human cell line than in vivo because 3D medullary microenvironment and certain signalling is necessary for the expression of AIRE and TSA genes and disruption of medullary microenvironment leads to the reduction of AIRE and TSA genes expression (*see section 1.2.4*; Rossi et al., 2007; Irla et al., 2008).

Different APS-1 mutations were introduced to test the effect of the most common AIRE mutations on the expression of *SOX9* and *TH*. Three common APS-1 mutations, c.254A>G (Y85C), c.415C>T (R139X) and c.769C>T (R257X), were introduced into AIRE apart from the AIRE mutation G228W. The selected mutations R139X and R257X are nonsense mutations that result in truncated versions of AIRE, while Y85C

and G228W are missense mutations (Ilmarinen et al., 2005; De Martino et al., 2013; Oftedal et al., 2015). All tested AIRE mutations adversely affected the transactivation potential of AIRE except for Y85C. The promoters of *SOX9* and *TH* were activated with no significant difference between cells express wild-type AIRE and those expressing the AIRE mutant Y85C. Our findings are in keeping with previous studies reporting that the AIRE Y85C mutant stimulated the tested promoters as effectively as wild-type AIRE (Björses et al., 2000; Ramsey et al., 2002a). However, pulse-chase experiments revealed that the polypeptide generated from the AIRE Y85C mutation has a shorter half-life compared with the wild-type polypeptides (Ramsey et al., 2002a), suggesting that the Y85C mutation does not affect the transactivation activity of AIRE directly but leads to rapid degradation of the mutant protein. In addition, it should be noted that APS-1 patients with the Y85C mutation show milder clinical symptoms than others. For example, Iranian Jews with the AIRE Y85C mutation do not develop chronic mucocutaneous candidiasis, which is the most common component of APS-1 (Ramsey et al., 2002a; Arstila et al., 2013). Further research is required to verify that AIRE degradation is the main cause of the milder clinical symptoms associated with Iranian Jewish APS-1 patients, as even if the Y85C mutation leads to rapid degradation of most of the protein, the remaining protein may retain some biological activity.

Our ChIP-seq data conflicted with previous reports by Kumar et al. (2001) and Purohit et al. (2005) showing AIRE binding on ATTGGTTA or TTATTA motifs. However, HOMER de novo motif results showed that AIRE bound to three motifs, mostly the TCTGCAAGTGGA motif. Most interestingly, our data showed that AIRE enriched the intergenic region in the pericentromeric region of many chromosomes where cohesins are enriched. Cohesins in the pericentromeric region form a ring-shaped complex with

some proteins such as RAD21, SMC1 and SMC3 to ensure the cohesion of sister chromatids and facilitate the attachment of sister kinetochores to microtubules from opposite spindle poles, a process known as chromosome bi-orientation or sister kinetochore bi-orientation, which occurs in mitosis (Tanaka et al., 2013). Interestingly, RAD21 and SMC3 were found to be among the putative AIRE-associated proteins (Abramson et al., 2010). In addition, AIRE interact to TOP2α via its PHD1 domain which is known to mediate chromatid separation and the relief of supercoiled DNA during mitosis and depletion of TOP2α from precondensed chromosomes reverses chromosome condensation and their separating is prevented, which indicate that TOP2α is necessary for maintenance of chromosome structure during mitosis (Bansal et al., 2017; Nielsen et al., 2020). This might suggest that AIRE has another function beside gene activator by recruiting TOP2α to sites of chromosome adhesion. There is no evidence that there are homologues of AIRE in organisms that do not have thymuses since this would suggest that its role as a thymic regulator of TSA genes expression is a recent co-option. Our data suggest that AIRE binds to pericentromeric regions directly by binding the motifs revealed by the de novo HOMER motif analysis (Table 5.10) or indirectly by binding to AIRE partners associated with the cohesin complex or involved in mitosis such as TOP2α. Co-localisation of AIRE with cohesins and TOP2α on chromosomes may be essential to confirm the involvement of AIRE in this region. In addition, validating the binding of AIRE to the motifs identified in Table 5.10 by other DNA binding assays such as electrophoretic mobility shift assay (EMSA) may shed light on whether AIRE binds to these sequences directly or indirectly.

6.2 Future plans

A group recently bred *Aire*$^{-/-}$ rats that developed a severe autoimmune phenotype that was not observed in *Aire*$^{-/-}$ mice, including alopecia, nail dystrophy, vitiligo and

impaired thymus development (Ossart et al., 2018). Furthermore, the age of the mouse is relevant in the development of autoimmunity as it corresponds primarily to the duration of the disease. Mice between 15 and 30 weeks of age may be too young to develop signs of autoimmunity in the brain because the lesions or T lymphocyte infiltration in some organs of *Aire*$^{-/-}$ mice or *Aire*$^{-/-}$ rats can appear at > 10 months (> 42 weeks) of age (Anderson et al., 2002; Ossart et al., 2018). Therefore, we will examine the brains of *Aire*-deficient rats or mice with different genetic backgrounds at > 1 year of age to find a suitable APS-1 model that mimics the neurological manifestations associated with APS-1 in humans.

HOMER de novo motif results showed that AIRE bound to three motifs and these three motifs do not resemble known binding sites for any known transcription factors (Table 5.10). We will validate whether AIRE can bind to the motifs identified in Table 5.10 using other DNA-binding assays such as EMSA to further our understanding of whether AIRE binds to these sequences directly or indirectly by binding to the cohesin complex. In addition, our data showed that AIRE enriched the intergenic region in the pericentromeric region of many chromosomes in both ChIP samples. Therefore, AIRE may be involved in the cohesion complex to ensure robust cohesion of sister chromatids or may be involved in the process of chromosome bi-orientation during mitosis in mTEC^{Hi} and/or in extra-thymic AIRE-expressing cells. As a result, we will co-localise AIRE with cohesins and TOP2α on chromosomes to validate that AIRE binds indirectly to the pericentromic region by binding to cohesions such as RAD21 and SMC3.

Bibliography

7- References

AALTONEN, J., BJORSES, P., PERHEENTUPA, J., HORELLIKUITUNEN, N., PALOTIE, A., PELTONEN, L., LEE, Y. S., FRANCIS, F., HENNIG, S., THIEL, C., LEHRACH, H. & YASPO, M. L. 1997. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nature Genetics,* 17**,** 399-403.

ABRAMSON, J., GIRAUD, M., BENOIST, C. & MATHIS, D. 2010. Aire's partners in the molecular control of immunological tolerance. *Cell,* 140**,** 123-135.

ABRAMSON, J. & GOLDFARB, Y. 2016. AIRE: From promiscuous molecular partnerships to promiscuous gene expression. *European Journal of Immunology,* 46**,** 22-33.

ABRAMSON, J. & HUSEBYE, E. S. 2016. Autoimmune regulator and self-tolerance – molecular and clinical aspects. *Immunological Reviews,* 271**,** 127-140.

ADAMSON, K. A., PEARCE, S. H. S., LAMB, J. R., SECKL, J. R. & HOWIE, S. E. M. 2004. A comparative study of mRNA and protein expression of the autoimmune regulator gene (Aire) in embryonic and adult murine tissues. *Journal of Pathology,* 202**,** 180-187.

AHONEN, P., MYLLÄRNIEMI, S., SIPILÄ, I. & PERHEENTUPA, J. 1990. Clinical Variation of Autoimmune Polyendocrinopathy–Candidiasis–Ectodermal Dystrophy (APECED) in a Series of 68 Patients. *New England Journal of Medicine,* 322**,** 1829-1836.

ALIMOHAMMADI, M., BJORKLUND, P., HALLGREN, A., PONTYNEN, N., SZINNAI, G., SHIKAMA, N., KELLER, M. P., EKWALL, O., KINKEL, S. A., HUSEBYE, E. S., GUSTAFSSON, J., RORSMAN, F., PELTONEN, L., BETTERLE, C., PERHEENTUPA, J., AKERSTROM, G., WESTIN, G., SCOTT, H. S., HOLLANDER, G. A. & KAMPE, O. 2008. Autoimmune polyendocrine syndrome type 1 and NALP5, parathyroid autoantigen. *New England Journal of Medicine,* 358**,** 1018-1028.

ALMAGHRABI, S., AZZOUZ, M. & TAZI AHNINI, R. 2020. AAV9-mediated AIRE gene delivery clears circulating antibodies and tissue T-cell infiltration in a mouse model of autoimmune polyglandular syndrome type-1. *Clinical & translational immunology,* 9**,** e1166-e1166.

ANDERSON, M. S. & SU, M. A. 2016. AIRE expands: new roles in immune tolerance and beyond. *Nat Rev Immunol,* 16**,** 247-58.

ANDERSON, M. S., VENANZI, E. S., KLEIN, L., CHEN, Z., BERZINS, S. P., TURLEY, S. J., VON BOEHMER, H., BRONSON, R., DIERICH, A., BENOIST, C. & MATHIS, D. 2002. Projection of an Immunological Self Shadow Within the Thymus by the Aire Protein. *Science,* 298**,** 1395.

ARSTILA, T. P. & JARVA, H. 2013. Human APECED; a Sick Thymus Syndrome? *Frontiers in immunology,* 4**,** 313-313.

BANSAL, K., YOSHIDA, H., BENOIST, C. & MATHIS, D. 2017. The transcriptional regulator Aire binds to and activates super-enhancers. *Nature Immunology,* 18**,** 263-273.

BENSING, S., FETISSOV, S. O., MULDER, J., PERHEENTUPA, J., GUSTAFSSON, J., HUSEBYE, E. S., OSCARSON, M., EKWALL, O., CROCK, P. A., HÖKFELT, T., HULTING, A.-L. & KÄMPE, O. 2007. Pituitary autoantibodies in autoimmune polyendocrine syndrome type 1. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 949-954.

BESNARD, M., PADONOU, F., PROVIN, N., GIRAUD, M., GUILLONNEAU, C. 2021. AIRE Deficiency, From Preclinical Models to Human APECED Disease. PMID: 33729987; PMCID: PMC7875492.

BETTERLE, C., GREGGIO, N. A. & VOLPATO, M. 1998. Autoimmune Polyglandular Syndrome Type 1. *The Journal of Clinical Endocrinology & Metabolism,* 83**,** 1049-1055.

BJÖRK, E., VELLOSO, L. A., KÄMPE, O. & KARLSSON, F. A. 1994. GAD Autoantibodies in IDDM, Stiff-Man Syndrome, and Autoimmune Polyendocrine Syndrome Type I Recognize Different Epitopes. *Diabetes,* 43**,** 161.

BJÖRSES, P., AALTONEN, J., VIKMAN, A., PERHEENTUPA, J., BEN-ZION, G., CHIUMELLO, G., DAHL, N., HEIDEMAN, P., HOORWEG-NIJMAN, J. J., MATHIVON, L., MULLIS, P. E., POHL, M., RITZEN, M., ROMEO, G., SHAPIRO, M. S., SMITH, C. S., SOLYOM, J., ZLOTOGORA, J. & PELTONEN, L. 1996. Genetic homogeneity of autoimmune polyglandular disease type I. *American journal of human genetics,* 59**,** 879-886.

BJÖRSES, P., HALONEN, M., PALVIMO, J. J., KOLMER, M., AALTONEN, J., ELLONEN, P., PERHEENTUPA, J., ULMANEN, I. & PELTONEN, L. 2000. Mutations in the AIRE Gene: Effects on Subcellular Location and Transactivation Function of the Autoimmune Polyendocrinopathy-Candidiasis–Ectodermal Dystrophy Protein. *The American Journal of Human Genetics,* 66**,** 378-392.

BLAT, Y. & KLECKNER, N. 1999. Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. *Cell,* 98**,** 249-59.

BLECHSCHMIDT, K., SCHWEIGER, M., WERTZ, K., POULSON, R., CHRISTENSEN, H. M., ROSENTHAL, A., LEHRACH, H. & YASPO, M. L. 1999. The mouse Aire gene: comparative genomic sequencing, gene organization, and expression. *Genome Res,* 9**,** 158-66.

BONDURAND, N., DASTOT-LE MOAL, F., STANCHINA, L., COLLOT, N., BARAL, V., MARLIN, S., ATTIE-BITACH, T., GIURGEA, I., SKOPINSKI, L., REARDON, W., TOUTAIN, A., SARDA, P., ECHAIEB, A., LACKMY-PORT-LIS, M., TOURAINE, R., AMIEL, J., GOOSSENS, M. & PINGAULT, V. 2007. Deletions at the SOX10 gene locus cause Waardenburg syndrome types 2 and 4. *American journal of human genetics,* 81**,** 1169-1185.

CARPINO, A., BUGANZA, R., MATARAZZO, P., TULI, G., PINON, M., CALVO, P. L., MONTIN, D., LICCIARDI, F. & DE SANCTIS, L. 2021. Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy in Two Siblings: Same Mutations but Very Different Phenotypes. *Genes,* 12**,** 169.

CETANI, F., BARBESINO, G., BORSARI, S., PARDI, E., CIANFEROTTI, L., PINCHERA, A. & MARCOCCI, C. 2001. A Novel Mutation of the Autoimmune Regulator Gene in an Italian Kindred with Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy, Acting in a Dominant Fashion and Strongly Cosegregating with Hypothyroid Autoimmune Thyroiditis. *The Journal of Clinical Endocrinology & Metabolism,* 86**,** 4747-4752.

CHEESEMAN, I. M. 2014. The kinetochore. *Cold Spring Harbor perspectives in biology,* 6**,** a015826-a015826.

CHI, X., LI, Y. & QIU, X. 2020. V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. *Immunology,* 160**,** 233-247.

CHIGNOLA, F., GAETANI, M., REBANE, A., ORG, T., MOLLICA, L., ZUCCHELLI, C., SPITALERI, A., MANNELLA, V., PETERSON, P. & MUSCO, G. 2009. The solution structure of the first PHD finger of autoimmune regulator in complex with non-modified histone H3 tail reveals the antagonistic role of H3R2 methylation. *Nucleic Acids Research,* 37**,** 2951-2961.

CIHAKOVA, D., TREBUSAK, K., HEINO, M., FADEYEV, V., TIULPAKOV, A., BATTELINO, T., TAR, A., HALÁSZ, Z., BLÜMEL, P., TAWFIK, S., KROHN, K., LEBL, J., PETERSON, P. & GROUP, M.-A. S. 2001. Novel AIRE mutations and P450 cytochrome autoantibodies in Central and Eastern European patients with APECED. *Human Mutation,* 18**,** 225-232.

COCCO, C., FERRI, G.-L., BOI, F., PINNA, G., MARIOTTI, S., MELONI, A. & POSSENTI, R. 2005. Median Eminence Dopaminergic Nerve Terminals: A Novel Target in Autoimmune Polyendocrine Syndrome? *The Journal of Clinical Endocrinology & Metabolism,* 90**,** 4108-4111.

COCCO, C., MELONI, A., MARIOTTI, S., COSSU, E., D'AMATO, F., ZULIAN, S., TONGIORGI, E. & FERRI, G. L. 2012. Novel neuronal and endocrine autoantibody targets in autoimmune polyendocrine syndrome type 1. *Autoimmunity,* 45**,** 485-494.

DANAN-GOTTHOLD M, GUYON C, GIRAUD M, LEVANON EY, ABRAMSON J. 2016. Extensive RNA Editing and Splicing Increase Immune Self-Representation Diversity in Medullary Thymic Epithelial Cells. Genome Biology, 24;17(1):219.

DAUBNER, S. C., LE, T. & WANG, S. 2011. Tyrosine hydroxylase and regulation of dopamine synthesis. *Archives of biochemistry and biophysics,* 508**,** 1-12.

DE MARTINO, L., CAPALBO, D., IMPRODA, N., D' ELIA, F., DI MASE, R., D'ASSANTE, R., D' ACUNZO, I., PIGNATA, C. & SALERNO, M. 2013. APECED: A Paradigm of Complex Interactions between Genetic Background and Susceptibility Factors. *Frontiers in Immunology,* 4.

DEPNER, M., FUCHS, S., RAABE, J., FREDE, N., GLOCKER, C., DOFFINGER, R., GKRANIA-KLOTSAS, E., KUMARARATNE, D., ATKINSON, T. P., SCHROEDER, H. W., NIEHUES, T., DUCKERS, G., STRAY-PEDERSEN, A., BAUMANN, U., SCHMIDT, R., FRANCO, J. L., ORREGO, J., BEN-SHOSHAN, M., MCCUSKER, C., JACOB, C. M. A., CARNEIRO-SAMPAIO, M., DEVLIN, L. A., EDGAR, J. D. M., HENDERSON, P., RUSSELL, R. K., SKYTTE, A. B., SENEVIRATNE, S. L., WANDERS, J., STAUSS, H., MEYTS, I., MOENS, L., JESENAK, M., KOBBE, R., BORTE, S., BORTE, M., WRIGHT, D. A., HAGIN, D., TORGERSON, T. R. & GRIMBACHER, B. 2016. The Extended Clinical Phenotype of 26 Patients with Chronic Mucocutaneous Candidiasis due to Gain-of-Function Mutations in STAT1. *Journal of Clinical Immunology,* 36**,** 73-84.

DERBINSKI, J., GÄBLER, J., BRORS, B., TIERLING, S., JONNAKUTY, S., HERGENHAHN, M., PELTONEN, L., WALTER, J. & KYEWSKI, B. 2005. Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. *The Journal of Experimental Medicine,* 202**,** 33.

DERBINSKI, J., PINTO, S., RÖSCH, S., HEXEL, K., KYEWSKI, B. 2008. Promiscuous gene expression patterns in single medullary thymic epithelial cells argue for a stochastic mechanism PNAS 105:657–662

DESANTI, G. E., COWAN, J. E., BAIK, S., PARNELL, S. M., WHITE, A. J., PENNINGER, J. M., LANE, P. J., JENKINSON, E. J., JENKINSON, W. E. & ANDERSON, G. 2012. Developmentally regulated availability of RANKL and

CD40 ligand reveals distinct mechanisms of fetal and adult cross-talk in the thymus medulla. *J Immunol,* 189**,** 5519-26.

ELDERSHAW, S. A., SANSOM, D. M. & NARENDRAN, P. 2011. Expression and function of the autoimmune regulator (Aire) gene in non-thymic tissue. *Clinical and Experimental Immunology,* 163**,** 296-308.

FERLETTA, M., UHRBOM, L., OLOFSSON, T., PONTEN, F. & WESTERMARK, B. 2007. Sox10 has a broad expression pattern in gliomas and enhances platelet-derived growth factor-B--induced gliomagenesis. *Mol Cancer Res,* 5**,** 891-7.

FETISSOV, S. O., BENSING, S., MULDER, J., LE MAITRE, E., HULTING, A. L., HARKANY, T., EKWALL, O., SKOLDBERG, F., HUSEBYE, E. S., PERHEENTUPA, J., RORSMAN, F., KAEMPE, O. & HOKFELT, T. 2009. Autoantibodies in Autoimmune Polyglandular Syndrome Type I Patients React with Major Brain Neurotransmitter Systems. *Journal of Comparative Neurology,* 513**,** 1-20.

FIERABRACCI, A. 2016. Type 1 Diabetes in Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy Syndrome (APECED): A "Rare" Manifestation in a "Rare" Disease. *International journal of molecular sciences,* 17**,** 1106.

FURUKAWA, Y. & KISH, S. 1993. Tyrosine Hydroxylase Deficiency. *In:* ADAM, M. P., ARDINGER, H. H., PAGON, R. A., WALLACE, S. E., BEAN, L. J. H., MIRZAA, G. & AMEMIYA, A. (eds.) *GeneReviews(®).* Seattle (WA): University of Washington, Seattle Copyright © 1993-2021, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.

GARDNER, J. M., DEVOSS, J. J., FRIEDMAN, R. S., WONG, D. J., TAN, Y. X., ZHOU, X., JOHANNES, K. P., SU, M. A., CHANG, H. Y., KRUMMEL, M. F. & ANDERSON, M. S. 2008. Deletional tolerance mediated by extrathymic Aire-expressing cells. *Science (New York, N.Y.),* 321**,** 843-847.

GARELLI, S., DALLA COSTA, M., SABBADIN, C., BAROLLO, S., RUBIN, B., SCARPA, R., MASIERO, S., FIERABRACCI, A., BIZZARRI, C., CRINÒ, A., CAPPA, M., VALENZISE, M., MELONI, A., DE BELLIS, A. M., GIORDANO, C., PRESOTTO, F., PERNIOLA, R., CAPALBO, D., SALERNO, M. C., STIGLIANO, A., RADETTI, G., CAMOZZI, V., GREGGIO, N. A., BOGAZZI, F., CHIODINI, I., PAGOTTO, U., BLACK, S. K., CHEN, S., REES SMITH, B., FURMANIAK, J., WEBER, G., PIGLIARU, F., DE SANCTIS, L., SCARONI, C. & BETTERLE, C. 2021. Autoimmune polyendocrine syndrome type 1: an Italian survey on 158 patients. *J Endocrinol Invest,* 44**,** 2493-2510.

GIRAUD, M., JMARI, N., DU, L., CARALLIS, F., NIELAND, T. J., PEREZ-CAMPO, F. M., BENSAUDE, O., ROOT, D. E., HACOHEN, N., MATHIS, D. & BENOIST, C. 2014. An RNAi screen for Aire cofactors reveals a role for Hnrnpl in polymerase release and Aire-activated ectopic transcription. *Proc Natl Acad Sci U S A,* 111**,** 1491-6.

GIRAUD, M., YOSHIDA, H., ABRAMSON, J., RAHL, P. B., YOUNG, R. A., MATHIS, D. & BENOIST, C. 2012. Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 535-540.

GOTTER, J., BRORS, B., HERGENHAHN, M. & KYEWSKI, B. 2004. Medullary Epithelial Cells of the Human Thymus Express a Highly Diverse Selection of

Tissue-specific Genes Colocalized in Chromosomal Clusters. *The Journal of Experimental Medicine,* 199**,** 155.

GRAY, D., ABRAMSON, J., BENOIST, C. & MATHIS, D. 2007. Proliferative arrest and rapid turnover of thymic epithelial cells expressing Aire. *Journal of Experimental Medicine,* 204**,** 2521-2528.

GUERAU-DE-ARELLANO, M., MATHIS, D. & BENOIST, C. 2008. Transcriptional impact of Aire varies with cell type. *Proceedings of the National Academy of Sciences,* 105**,** 14011.

GUI, J., MUSTACHIO, L. M., SU, D.-M. & CRAIG, R. W. 2012. Thymus Size and Age-related Thymic Involution: Early Programming, Sexual Dimorphism, Progenitors and Stroma. *Aging and disease,* 3**,** 280-290.

GUO, C. J., LEUNG, P. S. C., ZHANG, W. C., MA, X. & GERSHWIN, M. E. 2018. The immunobiology and clinical features of type 1 autoimmune polyglandular syndrome (APS-1). *Autoimmunity Reviews,* 17**,** 78-85.

HALONEN, M., ESKELIN, P., MYHRE, A.-G., PERHEENTUPA, J., HUSEBYE, E. S., KÄMPE, O., RORSMAN, F., PELTONEN, L., ULMANEN, I. & PARTANEN, J. 2002. AIRE Mutations and Human Leukocyte Antigen Genotypes as Determinants of the Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy Phenotype. *The Journal of Clinical Endocrinology & Metabolism,* 87**,** 2568-2574.

HALONEN, M., KANGAS, H., RUPPELL, T., ILMARINEN, T., OLLILA, J., KOLMER, M., VIHINEN, M., PALVIMO, J., SAARELA, J., ULMANEN, I. & ESKELIN, P. 2004. APECED-causing mutations in AIRE reveal the functional domains of the protein. *Human Mutation,* 23**,** 245-257.

HALONEN, M., PELTO-HUIKKO, M., ESKELIN, P., PELTONEN, L., ULMANEN, I. & KOLMER, M. 2001. Subcellular location and expression pattern of autoimmune regulator (Aire), the mouse orthologue for human gene defective in autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED). *Journal of Histochemistry & Cytochemistry,* 49**,** 197-208.

HASEEB, A. & LEFEBVRE, V. 2019. The SOXE transcription factors-SOX8, SOX9 and SOX10-share a bi-partite transactivation mechanism. *Nucleic acids research,* 47**,** 6917-6931.

HEDSTRAND, H., EKWALL, O., OLSSON, M. J., LANDGREN, E., KEMP, E. H., WEETMAN, A. P., PERHEENTUPA, J., HUSEBYE, E., GUSTAFSSON, J., BETTERLE, C., KÄMPE, O. & RORSMAN, F. 2001. The Transcription Factors SOX9 and SOX10 Are Vitiligo Autoantigens in Autoimmune Polyendocrine Syndrome Type I. *Journal of Biological Chemistry,* 276**,** 35390-35395.

HEERY, D. M., KALKHOVEN, E., HOARE, S. & PARKER, M. G. 1997. A signature motif in transcriptional co-activators mediates binding to nuclear receptor. *Nature,* 387**,** 733-736.

HEINO, M., PETERSON, P., KUDOH, J., NAGAMINE, K., LAGERSTEDT, A., OVOD, V., RANKI, A., RANTALA, I., NIEMINEN, M., TUUKKANEN, J., SCOTT, H. S., ANTONARAKIS, S. E., SHIMIZU, N. & KROHN, K. 1999. Autoimmune regulator is expressed in the cells regulating immune tolerance in thymus medulla. *Biochemical and Biophysical Research Communications,* 257**,** 821-825.

HIKOSAKA, Y., NITTA, T., OHIGASHI, I., YANO, K., ISHIMARU, N., HAYASHI, Y., MATSUMOTO, M., MATSUO, K., PENNINGER, J. M., TAKAYANAGI, H., YOKOTA, Y., YAMADA, H., YOSHIKAI, Y., INOUE, J., AKIYAMA, T. & TAKAHAMA, Y. 2008. The cytokine RANKL produced by positively selected

thymocytes fosters medullary thymic epithelial cells that express autoimmune regulator. *Immunity,* 29**,** 438-50.

HOFFMANN, G. F., ASSMANN, B., BRÄUTIGAM, C., DIONISI-VICI, C., HÄUSSLER, M., DE KLERK, J. B. C., NAUMANN, M., STEENBERGEN-SPANJERS, G. C. H., STRASSBURG, H.-M. & WEVERS, R. A. 2003. Tyrosine hydroxylase deficiency causes progressive encephalopathy and dopa-nonresponsive dystonia. *Annals of Neurology,* 54**,** S56-S65.

HUBERT, F. X., KINKEL, S. A., CREWTHER, P. E., CANNON, P. Z. F., WEBSTER, K. E., LINK, M., UIBO, R., O'BRYAN, M. K., MEAGER, A., FOREHAN, S. P., SMYTH, G. K., MITTAZ, L., ANTONARAKIS, S. E., PETERSON, P., HEATH, W. R. & SCOTT, H. S. 2009. Aire-Deficient C57BL/6 Mice Mimicking the Common Human 13-Base Pair Deletion Mutation Present with Only a Mild Autoimmune Phenotype. *Journal of Immunology,* 182**,** 3902-3918.

ILMARINEN, T., ESKELIN, P., HALONEN, M., RÜPPELL, T., KILPIKARI, R., TORRES, G. D., KANGAS, H. & ULMANEN, I. 2005. Functional analysis of SAND mutations in AIRE supports dominant inheritance of the G228W mutation. *Hum Mutat,* 26**,** 322-31.

IRLA, M., HUGUES, S., GILL, J., NITTA, T., HIKOSAKA, Y., WILLIAMS, I. R., HUBERT, F. X., SCOTT, H. S., TAKAHAMA, Y., HOLLÄNDER, G. A. & REITH, W. 2008. Autoantigen-specific interactions with CD4+ thymocytes control mature medullary thymic epithelial cell cellularity. *Immunity,* 29**,** 451-63.

ISHII, T., SUZUKI, Y., ANDO, N., MATSUO, N. & OGATA, T. 2000. Novel mutations of the autoimmune regulator gene in two siblings with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy. *J Clin Endocrinol Metab,* 85**,** 2922-6.

JIANG, W. Y., ANDERSON, M. S., BRONSON, R., MATHIS, D. & BENOIST, C. 2005. Modifier loci condition autoimmunity provoked by Aire deficiency. *Journal of Experimental Medicine,* 202**,** 805-815.

JIN, L., UNDERHILL, PA., DOCTOR, V., DAVIS, RW., SHEN, P., CAVALLI-SFORZA, LL., OEFNER, PJ. 1999. Distribution of Haplotypes from a Chromosome 21 Region Distinguishes Multiple Prehistoric Human Migrations. Proc Natl Acad Sci U S A. PMID: 10097117; PMCID: PMC22374

KAHALY, G. & FROMMER, L. 2018. Polyglandular autoimmune syndromes. *Journal of Endocrinological Investigation,* 41**,** 91-98.

KEKALAINEN, E., TUOVINEN, H., JOENSUU, J., GYLLING, M., FRANSSILA, R., PONTYNEN, N., TALVENSAARI, K., PERHEENTUPA, J., MIETTINEN, A. & ARSTILA, T. P. 2007. A defect of regulatory T cells in patients with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy. *J Immunol,* 178**,** 1208-15.

KIM, J., LO, L., DORMAND, E. & ANDERSON, D. J. 2003. SOX10 Maintains Multipotency and Inhibits Neuronal Differentiation of Neural Crest Stem Cells. *Neuron,* 38**,** 17-31.

KISAND, K. & PETERSON, P. 2015. Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy. *Journal of Clinical Immunology,* 35**,** 463-478.

KLUGER, N., RANKI, A. & KROHN, K. 2012. APECED: is this a model for failure of T cell and B cell tolerance? *Frontiers in Immunology,* 3.

KO, H.-J., KINKEL, S.A., HUBERT, F.-X., NASA, Z., CHAN, J., SIATSKAS, C., HIRUBALAN, P., TOH, B.-H., SCOTT, H.S. AND ALDERUCCIO, F. 2010. Transplantation of Autoimmune Regulator-Encoding Bone Marrow Cells Delays

the Onset of Experimental Autoimmune Encephalomyelitis. Eur. J. Immunol., 40: 3499-3509.

KOH, A. S., KUO, A. J., PARK, S. Y., CHEUNG, P., ABRAMSON, J., BUA, D., CARNEY, D., SHOELSON, S. E., GOZANI, O., KINGSTON, R. E., BENOIST, C. & MATHIS, D. 2008. Aire employs a histone-binding module to mediate immunological tolerance, linking chromatin regulation with organ-specific autoimmunity. *Proceedings of the National Academy of Sciences,* 105**,** 15878-15883.

KOH, A. S., MILLER, E. L., BUENROSTRO, J. D., MOSKOWITZ, D. M., WANG, J., GREENLEAF, W. J., CHANG, H. Y. & CRABTREE, G. R. 2018. Rapid chromatin repression by Aire provides precise control of immune tolerance. *Nature Immunology,* 19**,** 162-172.

KONDO, K., OHIGASHI, I. & TAKAHAMA, Y. 2019. Thymus machinery for T-cell selection. *Int Immunol,* 31**,** 119-125.

KUMAR, P. G., LALORAYA, M., WANG, C.-Y., RUAN, Q., DAVOODI-SEMIROMI, A., KAO, K. & SHE, J. 2001. The Autoimmune Regulator (AIRE) Is a DNA-binding Protein*. *The Journal of Biological Chemistry,* 276**,** 41357 - 41364.

KURODA, N., MITANI, T., TAKEDA, N., ISHIMARU, N., ARAKAKI, R., HAYASHI, Y., BANDO, Y., IZUMI, K., TAKAHASHI, T., NOMURA, T., SAKAGUCHI, S., UENO, T., TAKAHAMA, Y., UCHIDA, D., SUN, S. J., KAJIURA, F., MOURI, Y., HAN, H. W., MATSUSHIMA, A., YAMADA, G. & MATSUMOTO, M. 2005. Development of autoimmunity against transcriptionally unrepressed target antigen in the thymus of aire-deficient mice. *Journal of Immunology,* 174**,** 1862-1870.

KYEWSKI, B. & KLEIN, L. 2006. A central role for central tolerance. *Annu Rev Immunol,* 24**,** 571-606.

LEE, E. M., YUAN, T., BALLIM, R. D., NGUYEN, K., KELSH, R. N., MEDEIROS, D. M. & MCCAULEY, D. W. 2016. Functional constraints on SoxE proteins in neural crest development: The importance of differential expression for evolution of protein activity. *Developmental Biology,* 418**,** 166-178.

LIIV, I., REBANE, A., ORG, T., SAARE, M., MASLOVSKAJA, J., KISAND, K., JURONEN, E., VALMU, L., BOTTOMLEY, M. J., KALKKINEN, N. & PETERSON, P. 2008. DNA-PK contributes to the phosphorylation of AIRE: importance in transcriptional activity. *Biochimica et biophysica acta,* 1783**,** 74-83.

LISTON, A. 2006. There and back again: Autoimmune Polyendocrinopathy Syndrome Type I and the Aire knockout mouse. *Drug Discovery Today: Disease Models,* 3**,** 33-40.

LOVEWELL, T. & TAZI-AHNINI, R. 2011. Models to explore the molecular function and regulation of AIRE. *Egyptian Journal of Medical Human Genetics,* 12**,** 107-115.

LOVEWELL, T. R. J., MCDONAGH, A. J. G., MESSENGER, A. G., AZZOUZ, M. & TAZI-AHNINI, R. 2018. Meta-Analysis of Autoimmune Regulator-Regulated Genes in Human and Murine Models: A Novel Human Model Provides Insights on the Role of Autoimmune Regulator in Regulating STAT1 and STAT1-Regulated Genes. *Frontiers in Immunology,* 9**,** 1380.

MATHIS, D., BENOIST C. 2009. AIRE. Annual Review of Immunology. 27: 287-312. PMID 19302042

MELONI, A., CORDA, G., SABA, L., FERRI, G. L., MARIOTTI, S. & COCCO, C. 2019. Reduction of Total Brain and Cerebellum Volumes Associated With Neuronal

Autoantibodies in Patients With APECED. *J Clin Endocrinol Metab,* 104**,** 150-162.

MELONI, A., WILLCOX, N., MEAGER, A., ATZENI, M., WOLFF, A. S. B., HUSEBYE, E. S., FURCAS, M., ROSATELLI, M. C., CAO, A. & CONGIA, M. 2012. Autoimmune Polyendocrine Syndrome Type 1: An Extensive Longitudinal Study in Sardinian Patients. *The Journal of Clinical Endocrinology & Metabolism,* 97**,** 1114-1124.

MEREDITH, M., ZEMMOUR, D., MATHIS, D. & BENOIST, C. 2015. Aire controls gene expression in the thymic epithelium with ordered stochasticity. *Nat Immunol,* 16**,** 942-9.

METZGER, T. C., KHAN, I. S., GARDNER, J. M., MOUCHESS, M. L., JOHANNES, K. P., KRAWISZ, A. K., SKRZYPCZYNSKA, K. M. & ANDERSON, M. S. 2013. Lineage tracing and cell ablation identify a post-Aire-expressing thymic epithelial cell population. *Cell Rep,* 5**,** 166-79.

MEYER, S., WOODWARD, M., HERTEL, C., VLAICU, P., HAQUE, Y., KÄRNER, J., MACAGNO, A., ONUOHA, S. C., FISHMAN, D., PETERSON, H., METSKÜLA, K., UIBO, R., JÄNTTI, K., HOKYNAR, K., WOLFF, A. S. B., MELONI, A., KLUGER, N., HUSEBYE, E. S., PODKRAJSEK, K. T., BATTELINO, T., BRATANIC, N., PEET, A., KROHN, K., RANKI, A., PETERSON, P., KISAND, K. & HAYDAY, A. 2016. AIRE-Deficient Patients Harbor Unique High-Affinity Disease-Ameliorating Autoantibodies. *Cell,* 166**,** 582-595.

MYHRE, A. G., HALONEN, M., ESKELIN, P., EKWALL, O., HEDSTRAND, H., RORSMAN, F., KAMPE, O. & HUSEBYE, E. S. 2001. Autoimmune polyendocrine syndrome type 1 (APS I) in Norway. *Clin Endocrinol (Oxf),* 54**,** 211-7.

NAGAMINE, K., PETERSON, P., SCOTT, H. S., KUDOH, J., MINOSHIMA, S., HEINO, M., KROHN, K. J., LALIOTI, M. D., MULLIS, P. E., ANTONARAKIS, S. E., KAWASAKI, K., ASAKAWA, S., ITO, F. & SHIMIZU, N. 1997. Positional cloning of the APECED gene. *Nat Genet,* 17**,** 393-8.

NIELSEN, C. F., ZHANG, T., BARISIC, M., KALITSIS, P. & HUDSON, D. F. 2020. Topoisomerase IIα is Essential for Maintenance of Mitotic Chromosome Structure. *Proc. Natl Acad. Sci. USA* **117**, 12131–12142.

OFTEDAL, B. E., HELLESEN, A., ERICHSEN, M. M., BRATLAND, E., VARDI, A., PERHEENTUPA, J., KEMP, E. H., FISKERSTRAND, T., VIKEN, M. K., WEETMAN, A. P., FLEISHMAN, S. J., BANKA, S., NEWMAN, W. G., SEWELL, W. A., SOZAEVA, L. S., ZAYATS, T., HAUGARVOLL, K., ORLOVA, E. M., HAAVIK, J., JOHANSSON, S., KNAPPSKOG, P. M., LØVÅS, K., WOLFF, A. S., ABRAMSON, J. & HUSEBYE, E. S. 2015. Dominant Mutations in the Autoimmune Regulator AIRE Are Associated with Common Organ-Specific Autoimmune Diseases. *Immunity,* 42**,** 1185-96.

ORG, T., CHIGNOLA, F., HETENYI, C., GAETANI, M., REBANE, A., LIIV, I., MARAN, U., MOLLICA, L., BOTTOMLEY, M. J., MUSCO, G. & PETERSON, P. 2008. The Autoimmune Regulator PHD Finger Binds to Non-Methylated Histone H3K4 to Activate Gene Expression. *Embo Reports,* 9**,** 370-376.

ORG, T., REBANE, A., KISAND, K., LAAN, M., HALJASORG, U., ANDRESON, R. & PETERSON, P. 2009. AIRE activated tissue specific genes have histone modifications associated with inactive chromatin. *Hum Mol Genet,* 18**,** 4699-710.

OSSART, J., MOREAU, A., AUTRUSSEAU, E., MÉNORET, S., MARTIN, J. C., BESNARD, M., OUISSE, L.-H., TESSON, L., FLIPPE, L., KISAND, K.,

PETERSON, P., HUBERT, F.-X., ANEGON, I., JOSIEN, R. & GUILLONNEAU, C. 2018. Breakdown of Immune Tolerance in AIRE-Deficient Rats Induces a Severe Autoimmune Polyendocrinopathy–Candidiasis–Ectodermal Dystrophy–like Autoimmune Disease. *The Journal of Immunology,* 201**,** 874.

OVEN, I., BRDICKOVA, N., KOHOUTEK, J., VAUPOTIC, T., NARAT, M. & PETERLIN, B. M. 2007. AIRE recruits P-TEFb for transcriptional elongation of target genes in medullary thymic epithelial cells. *Molecular and Cellular Biology,* 27**,** 8815-8823.

PALMER, J. C., BAIG, S., KEHOE, P. G. & LOVE, S. 2009. Endothelin-converting enzyme-2 is increased in Alzheimer's disease and up-regulated by Abeta. *The American journal of pathology,* 175**,** 262-270.

PARK, Y. S., GAUNA, A. E. & CHA, S. 2015. Mouse Models of Primary Sjogren's Syndrome. *Curr Pharm Des,* 21**,** 2350-64.

PASSOS, G. A., MENDES-DA-CRUZ, D. A. & OLIVEIRA, E. H. 2015. The Thymic Orchestration Involving Aire, miRNAs, and Cell-Cell Interactions during the Induction of Central Tolerance. *Front Immunol,* 6**,** 352.

PASSOS, G. A., SPECK-HERNANDEZ, C. A., ASSIS, A. F. & MENDES-DA-CRUZ, D. A. 2018. Update on Aire and thymic negative selection. *Immunology,* 153**,** 10-20.

PEARCE, SH., CHEETHAM, T., IMRIE, H., VAIDYA, B., BARNES, ND., BILOUS, RW., CARR, D., MEERAN, K., SHAW, NJ., SMITH, CS., TOFT, AD., WILLIAMS, G., KENDALL-TAYLOR, P. 1998. A Common and Recurrent 13-bp Deletion in the Autoimmune Regulator Gene in British Kindreds with Autoimmune Polyendocrinopathy Type 1. PMID: 9837820; PMCID: PMC1377639.

PEREIRA, L. E., BOSTIK, P. & ANSARI, A. A. 2005. The development of mouse APECED models provides new insight into the role of AIRE in immune regulation. *Clinical & developmental immunology,* 12**,** 211-216.

PERHEENTUPA, J. 2006. Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy. *The Journal of Clinical Endocrinology & Metabolism,* 91**,** 2843-2850.

PERNIOLA, R., FIERABRACCI, A. & FALORNI, A. 2021. Autoimmune Addison's Disease as Part of the Autoimmune Polyglandular Syndrome Type 1: Historical Overview and Current Evidence. *Frontiers in immunology,* 12**,** 606860-606860.

PETERSON, P., ORG, T. & REBANE, A. 2008. Transcriptional regulation by AIRE: molecular mechanisms of central tolerance. *Nature reviews. Immunology,* 8**,** 948-957.

PETERSON, P. & PELTONEN, L. 2005. Autoimmune polyendocrinopathy syndrome type 1 (APS1) and AIRE gene: New views on molecular basis of autoimmunity. *Journal of Autoimmunity,* 25**,** 49-55.

PINTO S, MICHEL C, SCHMIDT-GLENEWINKEL H, HARDER N, ROHR K, WILD S, BRORS B, KYEWSKI B. 2013. Overlapping Gene Coexpression Patterns in Human Medullary Thymic Epithelial Cells Generate Self-Antigen Diversity. Proc Natl Acad Sci U S A. 10;110(37):E3497-505.

PITKÄNEN, J., REBANE, A., ROWELL, J., MURUMÄGI, A., STRÖBEL, P., MÖLL, K., SAARE, M., HEIKKILÄ, J., DOUCAS, V., MARX, A., PETERSON, P. 2005. Cooperative Activation of Transcription by Autoimmune Regulator AIRE and CBP. Biochem Biophys Res Commun 333:944–953.

PÖNTYNEN, N., MIETTINEN, A., PETTERI ARSTILA, T., KÄMPE, O., ALIMOHAMMADI, M., VAARALA, O., PELTONEN, L. & ULMANEN, I. 2006.

Aire Deficient Mice do not Develop the Same Profile of Tissue-Specific Autoantibodies as APECED Patients. *Journal of Autoimmunity,* 27**,** 96-104.

PROUST-LEMOINE, E., CARDOT-BAUTERS, C., SAUGIER-VEBER, P., LEFEBVRE, H., LALAU, J. D., DESAILLOUD, R., REZNIK, Y., LEFRANC, D., PRIN, L. & WERNEAU, J. D. 2007. Clinical, immunological and genetic evaluation of APECED syndrome in north-west France: concerning 18 observations. *Diabetes & Metabolism,* 33**,** S11-S11.

PROUST-LEMOINE, E., SAUGIER-VEBER, P. & WÉMEAU, J.-L. 2012. Polyglandular Autoimmune Syndrome Type I. *La Presse Médicale,* 41**,** e651-e662.

PUROHIT, S., KUMAR, P. G., LALORAYA, M. & SHE, J.-X. 2005. Mapping DNA-binding domains of the autoimmune regulator protein. *Biochemical and Biophysical Research Communications,* 327**,** 939-944.

RAMSEY, C., BUKRINSKY, A. & PELTONEN, L. 2002a. Systematic mutagenesis of the functional domains of AIRE reveals their role in intracellular targeting. *Human Molecular Genetics,* 11**,** 3299-3308.

RAMSEY, C., WINQVIST, O., PUHAKKA, L., HALONEN, M., MORO, A., KAMPE, O., ESKELIN, P., PELTO-HUIKKO, M. & PELTONEN, L. 2002b. Aire deficient mice develop multiple features of APECED phenotype and show altered immune response. *Hum Mol Genet,* 11**,** 397-409.

RODRIGUIZ, R. M., GADNIDZE, K., RAGNAUTH, A., DORR, N., YANAGISAWA, M., WETSEL, W. C. & DEVI, L. A. 2008. Animals lacking endothelin-converting enzyme-2 are deficient in learning and memory. *Genes Brain Behav,* 7**,** 418-26.

ROSATELLI, M. C., MELONI, A., DEVOTO, M., CAO, A., SCOTT, H. S., PETERSON, P., HEINO, M., KROHN, K. J., NAGAMINE, K., KUDOH, J., SHIMIZU, N. & ANTONARAKIS, S. E. 1998. A common mutation in Sardinian autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy patients. *Hum Genet,* 103**,** 428-34.

ROSSI, S. W., KIM, M. Y., LEIBBRANDT, A., PARNELL, S. M., JENKINSON, W. E., GLANVILLE, S. H., MCCONNELL, F. M., SCOTT, H. S., PENNINGER, J. M., JENKINSON, E. J., LANE, P. J. & ANDERSON, G. 2007. RANK signals from CD4(+)3(-) inducer cells regulate development of Aire-expressing epithelial cells in the thymic medulla. *J Exp Med,* 204**,** 1267-72.

RUAN, Q.-G., TUNG, K., EISENMAN, D., SETIADY, Y., ECKENRODE, S., YI, B., PUROHIT, S., ZHENG, W.-P., ZHANG, Y., PELTONEN, L. & SHE, J.-X. 2007. The Autoimmune Regulator Directly Controls the Expression of Genes Critical for Thymic Epithelial Function. *The Journal of Immunology,* 178**,** 7173.

SANSOM, S. N., SHIKAMA-DORN, N., ZHANYBEKOVA, S., NUSSPAUMER, G., MACAULAY, I. C., DEADMAN, M. E., HEGER, A., PONTING, C. P. & HOLLÄNDER, G. A. 2014. Population and single-cell genomics reveal the Aire dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. *Genome Res,* 24**,** 1918-31.

SATO, K., SATO, U., TATEISHI, S., KUBO, K., HORIKAWA, R., MIMURA, T., YAMAMOTO, K. & KANDA, H. 2004. Aire downregulates multiple molecules that have contradicting immune-enhancing and immune-suppressive functions. *Biochem Biophys Res Commun,* 318**,** 935-40.

SCOTT, C. E., WYNN, S. L., SESAY, A., CRUZ, C., CHEUNG, M., GAVIRO, M.-V. G., BOOTH, S., GAO, B., CHEAH, K. S. E., LOVELL-BADGE, R. & BRISCOE, J. 2010. SOX9 induces and maintains neural stem cells. *Nature Neuroscience,* 13**,** 1181.

SERREZE, D. V. & LEITER, E. H. 1994. Genetic and pathogenic basis of autoimmune diabetes in NOD mice. *Current Opinion in Immunology,* 6**,** 900-906.

SHAKHOVA, O., CHENG, P., MISHRA, P. J., ZINGG, D., SCHAEFER, S. M., DEBBACHE, J., HÄUSEL, J., MATTER, C., GUO, T., DAVIS, S., MELTZER, P., MIHIC-PROBST, D., MOCH, H., WEGNER, M., MERLINO, G., LEVESQUE, M. P., DUMMER, R., SANTORO, R., CINELLI, P. & SOMMER, L. 2015. Antagonistic cross-regulation between Sox9 and Sox10 controls an anti-tumorigenic program in melanoma. *PLoS genetics,* 11**,** e1004877-e1004877.

SMITH-ANTTILA, C. J. A., BENSING, S., ALIMOHAMMADI, M., DALIN, F., OSCARSON, M., ZHANG, M. D., PERHEENTUPA, J., HUSEBYE, E. S., GUSTAFSSON, J., BJORKLUND, P., FRANSSON, A., NORDMARK, G., RONNBLOM, L., MELONI, A., SCOTT, R. J., HOKFELT, T., CROCK, P. A. & KAMPE, O. 2017. Identification of endothelin-converting enzyme-2 as an autoantigen in autoimmune polyendocrine syndrome type 1. *Autoimmunity,* 50**,** 223-231.

SPARKS, A. E., CHEN, C., BRESLIN, M. B. & LAN, M. S. 2016. Functional Domains of Autoimmune Regulator (AIRE) Modulate INS-VNTR Transcription in Human Thymic Epithelial Cells. *Journal of Biological Chemistry,* 291**,** 11313-11322.

ST-PIERRE C, TROFIMOV A, BROCHU S, LEMIEUX S, PERREAULT C. 2015. Differential Features of AIRE-Induced and AIRE-Independent Promiscuous Gene Expression in Thymic Epithelial Cells. Journal of Immunology 498-506.

STRITESKY, G. L., JAMESON, S. C. & HOGQUIST, K. A. 2012. Selection of self-reactive T cells in the thymus. *Annual review of immunology,* 30**,** 95-114.

SU, M. A., GIANG, K., ŽUMER, K., JIANG, H., OVEN, I., RINN, J. L., DEVOSS, J. J., JOHANNES, K. P. A., LU, W., GARDNER, J., CHANG, A., BUBULYA, P., CHANG, H. Y., PETERLIN, B. M. & ANDERSON, M. S. 2008. Mechanisms of an autoimmunity syndrome in mice caused by a dominant mutation in Aire. *The Journal of Clinical Investigation,* 118**,** 1712-1726.

TAKABA, H., MORISHITA, Y., TOMOFUJI, Y., DANKS, L., NITTA, T., KOMATSU, N., KODAMA, T. & TAKAYANAGI, H. 2015. Fezf2 Orchestrates a Thymic Program of Self-Antigen Expression for Immune Tolerance. *Cell,* 163**,** 975-87.

TANAKA, T. U., CLAYTON, L. & NATSUME, T. 2013. Three wise centromere functions: see no error, hear no break, speak no delay. *EMBO reports,* 14**,** 1073-1083.

UIBO, R., AAVIK, E., PETERSON, P., PERHEENTUPA, J., ARANKO, S., PELKONEN, R. & KROHN, K. J. 1994. Autoantibodies to cytochrome P450 enzymes P450scc, P450c17, and P450c21 in autoimmune polyglandular disease types I and II and in isolated Addison's disease. *The Journal of Clinical Endocrinology & Metabolism,* 78**,** 323-328.

VENANZI, E. S., MELAMED, R., MATHIS, D. & BENOIST, C. 2008. The variable immunological self: Genetic variation and nongenetic noise in Aire-regulated transcription. *Proceedings of the National Academy of Sciences,* 105**,** 15860.

VILLASEÑOR, J., BESSE, W., BENOIST, C., MATHIS, D. 2008. Ectopic Expression of Peripheral-Tissue Antigens in The Thymic Epithelium: Probabilistic, Monoallelic, Misinitiated. Proc Natl Acad Sci U S A 105:15854–15859

WANG, X. P., LAAN, M., BICHELE, R., KISAND, K., SCOTT, H. S. & PETERSON, P. 2012. Post-Aire maturation of thymic medullary epithelial cells involves selective expression of keratinocyte-specific autoantigens. *Frontiers in Immunology,* 3.

WATERFIELD, M., KHAN, I. S., CORTEZ, J. T., FAN, U., METZGER, T., GREER, A., FASANO, K., MARTINEZ-LLORDELLA, M., POLLACK, J. L., ERLE, D. J., SU, M. & ANDERSON, M. S. 2014. The transcriptional regulator Aire coopts the repressive ATF7ip-MBD1 complex for the induction of immunotolerance. *Nature Immunology,* 15**,** 258-265.

WOLFF, A. S., OFTEDAL, B. E., KISAND, K., ERSVAER, E., LIMA, K. & HUSEBYE, E. S. 2010. Flow cytometry study of blood cell subtypes reflects autoimmune and inflammatory processes in autoimmune polyendocrine syndrome type I. *Scand J Immunol,* 71**,** 459-67.

WOLFF, A. S. B., ERICHSEN, M. M., MEAGER, A., MAGITTA, N. W. F., MYHRE, A. G., BOLLERSLEV, J., FOUGNER, K. J., LIMA, K., KNAPPSKOG, P. M. & HUSEBYE, E. S. 2007. Autoimmune Polyendocrine Syndrome Type 1 in Norway: Phenotypic Variation, Autoantibodies, and Novel Mutations in the Autoimmune Regulator Gene. *The Journal of Clinical Endocrinology & Metabolism,* 92**,** 595-603.

YANG, S. Y., BANSAL, K., LOPES, J., BENOIST, C. & MATHIS, D. 2013. Aire's plant homeodomain(PHD)-2 is critical for induction of immunological tolerance. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 1833-1838.

ZHAO, B. J., CHANG, L., FU, H. Y., SUN, G. Y. & YANG, W. 2018. The Role of Autoimmune Regulator (AIRE) in Peripheral Tolerance. *Journal of Immunology Research*.

ZLOTOGORA, J. & SHAPIRO, M. S. 1992. Polyglandular autoimmune syndrome type I among Iranian Jews. *Journal of medical genetics,* 29**,** 824-826.

Chapter Eight

Appendices

8.1 Validation of RT-qPCR samples by gel electrophoresis

After quantifying the expression of the candidate genes by RT-qPCR, the samples were run on a gel electrophoresis to confirm that the signal detected belong to the desired genes at the expected size.

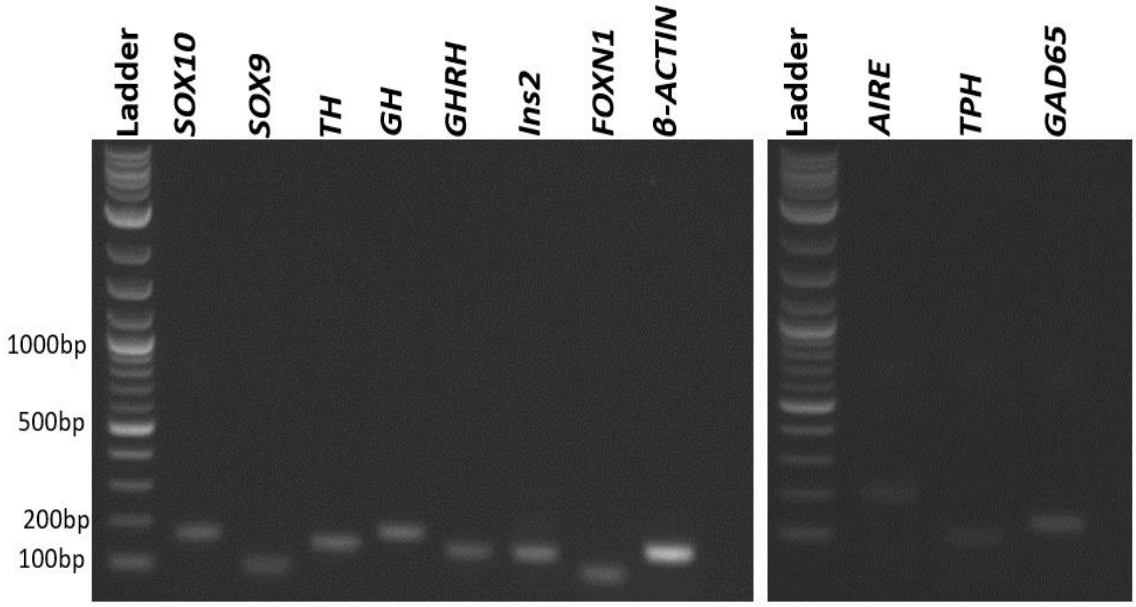

Figure 7.1: Electrophoresis gel picture for qPCR mice thymus samples. To confirm that the signal we detected by the RT-qPCR is belonging to our desired genes, we ran the sample on electrphoresis gel to verify that our amplified genes in the correct size. As the picture show the genes are at the expected size indicated in table 2.3. *Aire*, 199 bp, Th, 158 bp, *Tph*, 102 bp, *Gh*, 191 bp, *Gad65*, 135 bp, *Sox9*, 101 bp*, Ghrh*, 148 bp, *Sox10*, 182 bp, *Ins* 147bp, *Foxn1* 95bp, *B-actin* 154bp.

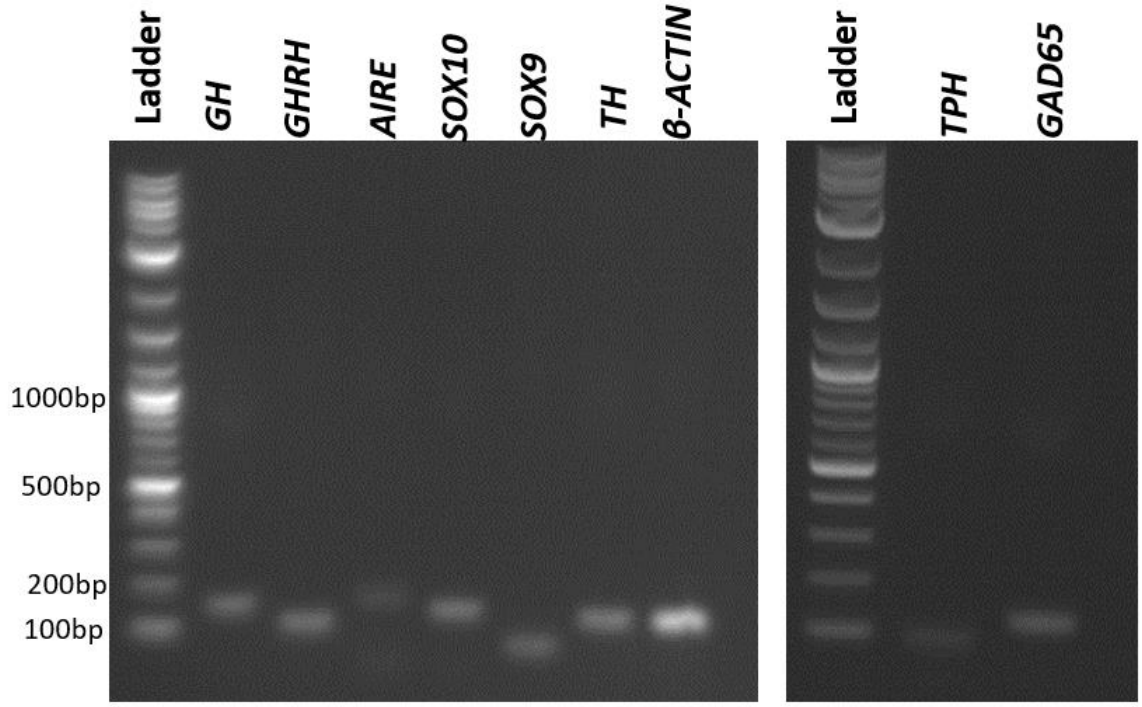

Figure 7.3: Electrophoresis gel picture for qPCR of mice brain samples . To confirm that the signal we detected by the RT-qPCR is belonging to our desired genes, we ran the sample on electrphoresis gel to verify that our amplified genes in the correct size. As the picture show the genes are at the expected size indicated in table 2.3. *Aire*, 199 bp, Th, 158 bp, *Tph*, 102 bp, *Gh*, 191 bp, *Gad65*, 135 bp, *Sox9*, 101 bp*, Ghrh*, 148 bp, *Sox10*, 182 bp, *Ins* 147bp, *Foxn1* 95bp, *B-actin* 154bp.

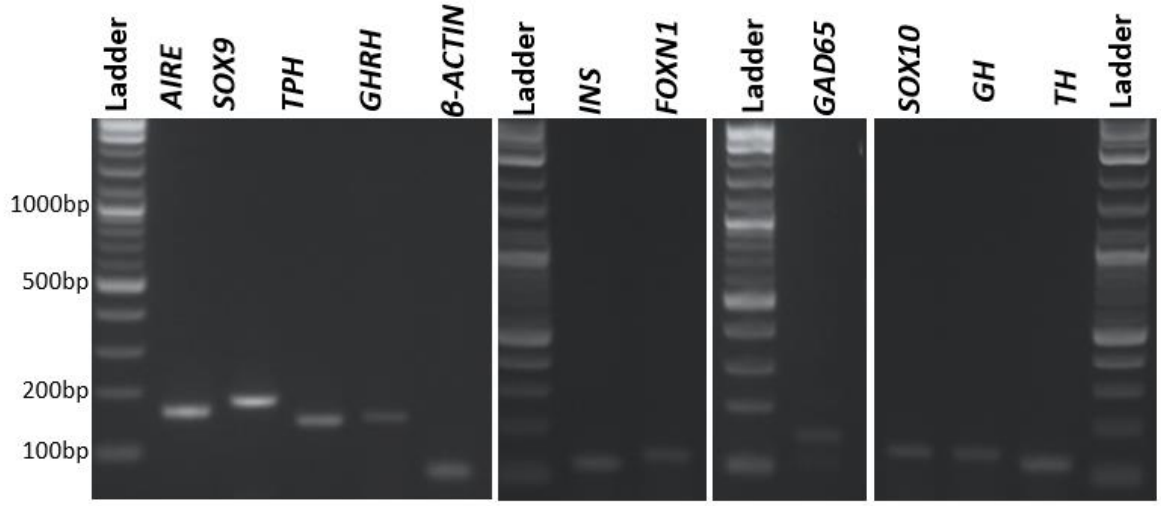

Figure 7.3: Electrophoresis gel picture for 4D6 cells samples. To confirm that the signal we detected by the RT-qPCR is belonging to our desired genes, we ran the sample on electrphoresis gel to verify that our amplified genes in the correct size. As the picture show the genes are at the expected size indicated in table 2.4. *Aire*, 165 bp, Th, 130 bp, *Tph*, 148 bp, *Gh*, 152 bp, *Gad65*, 146 bp, *Sox9*, 183 bp*, Ghrh*, 154 bp, *Sox10*, 159 bp, *Ins* 134bp, *Foxn1* 146bp, *B-actin* 71bp.

8.2 In silico sequences of promoters cloned in pGL3

> pGL3 *SOX9*

GGTACCCCTCCCTCCCCAAAATCGGGTCCAATCAGCTGCCTGCCAACCCTGGGACTGCTGTG
CTGTGATTGGCGGGTGGCTCTAAGGTGAGGCGGAGTATTTATTAAAGAGACCCTGGGCTGGG
AGTTGGAGAGCCGAAAGAAGCTTGGCATTCCGGTACTGTTGGTAAAGCCACCATGGAAGACG
CCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAAGATGGAACCGCTGGAGAG
CAACTGCATAAGGCTATGAAGAGATACGCCCTGGTTCCTGGAACAATTGCTTTTACAGATGC
ACATATCGAGGTGGACATCACTTACGCTGAGTACTTCGAAATGTCCGTTCGGTTGGCAGAAG
CTATGAAACGATATGGGCTGAATACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTCTT
CAATTCTTTATGCCGGTGTTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGA
CATTTATAATGAACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACCGTGGTGTTCG
TTTCCAAAAAGGGGTTGCAAAAAATTTTGAACGTGCAAAAAAAGCTCCCAATCATCCAAAAA
ATTATTATCATGGATTCTAAAACGGATTACCAGGGATTTCAGTCGATGTACACGTTCGTCAC
ATCTCATCTACCTCCCGGTTTTAATGAATACGATTTTGTGCCAGAGTCCTTCGATAGGGACA
AGACAATTGCACTGATCATGAACTCCTCTGGATCTACTGGTCTGCCTAAAGGTGTCGCTCTG
CCTCATAGAACTGCCTGCGTGAGATTCTCGCATGCCAGAGATCCTATTTTTGGCAATCAAAT
CATTCCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTA
CACTCGGATATTTGATATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTG
TTTCTGAGGAGCCTTCAGGATTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCTATTCTC
CTTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTTACACGAAATTGCTT
CTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCCAAGAGGTTCCATCTGCCA
GGTATCAGGCAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTGATTACACCCGAGGG
GGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAAGCGAAGGTTGTGGATC
TGGATACCGGGAAAACGCTGGGCGTTAATCAAAGAGGCGAACTGTGTGTGAGAGGTCCTATG
ATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATTGACAAGGATGGATG
GCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAACACTTCTTCATCGTTGACCGCC
TGAAGTCTCTGATTAAGTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCTTG
CTCCAACACCCCAACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGACGATGACGCCGGTGA
ACTTCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGATCGTGG
ATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAGTTGTGTTTGTGGAC
GAAGTACCGAAAGGTCTTACCGGAAAACTCGACGCAAGAAAAATCAGAGAGATCCTCATAAA
GGCCAAGAAGGGCGGAAAGATCGCCGTGTAATTCTAGAGTCGGGGCGGCCGGCCGCTTCGAG
CAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAA
TGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAA
ACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGG
TTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAATCGATAAGGATCCGTCGACCGAT
GCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCG
CCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTTC
CGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTC
ACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGA
GCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAG
GCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGA
CAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCG
ACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCA
ATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGC
ACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAAC
CCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAG
GTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGA
CAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCT

TGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTAC
GCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGT
GGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAG
ATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTC
TGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCAT
CCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGC
CCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAA
CCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGT
CTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTT
GTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTC
CGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCT
CCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATG
GCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGA
GTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGT
CAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGT
TCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCAC
TCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAA
CAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATA
CTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT
ATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGC
CACCTGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTG
ACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGC
CACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTA
GTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCA
TCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACT
CTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGA
TTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAAT
TTTAACAAAATATTAACGTTTACAATTTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGA
AGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCCCAAGCTACCATGATAAGTAAGTA
ATATTAAGGTACGGGAGGTACTTGGAGCGGCCGCAATAAAATATCTTTATTTTCATTACATC
TGTGTGTTGGTTTTTTGTGTGAATCGATAGTACTAACATACGCTCTCCATCAAAACAAAACG
AAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCT
CTATCGATA

> pGL3 *GAD65*

GGTACCAGCTGGAGCCAGACCTCAAACAAAACCCCAATCGATGCACACAGAAAACTCCTCTG
GGCCACGCTTCCCGCCTCGCCGAGGTCTCCCCAGTCTGCCCCTCGCTGACGCTGGCGCGCAG
CGGCTGTGGCAGCACCCGGGACAAAGCTTGGCATTCCGGTACTGTTGGTAAAGCCACCATGG
AAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAAGATGGAACCGCT
GGAGAGCAACTGCATAAGGCTATGAAGAGATACGCCCTGGTTCCTGGAACAATTGCTTTTAC
AGATGCACATATCGAGGTGGACATCACTTACGCTGAGTACTTCGAAATGTCCGTTCGGTTGG
CAGAAGCTATGAAACGATATGGGCTGAATACAAATCACAGAATCGTCGTATGCAGTGAAAAC
TCTCTTCAATTCTTTATGCCGGTGTTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGC
GAACGACATTTATAATGAACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACCGTGG
TGTTCGTTTCCAAAAAGGGGTTGCAAAAAATTTTGAACGTGCAAAAAAAGCTCCCAATCATC
CAAAAAATTATTATCATGGATTCTAAAACGGATTACCAGGGATTTCAGTCGATGTACACGTT
CGTCACATCTCATCTACCTCCCGGTTTTAATGAATACGATTTTGTGCCAGAGTCCTTCGATA
GGGACAAGACAATTGCACTGATCATGAACTCCTCTGGATCTACTGGTCTGCCTAAAGGTGTC
GCTCTGCCTCATAGAACTGCCTGCGTGAGATTCTCGCATGCCAGAGATCCTATTTTTGGCAA
TCAAATCATTCCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGT

TTACTACACTCGGATATTTGATATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAA
GAGCTGTTTCTGAGGAGCCTTCAGGATTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCT
ATTCTCCTTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTTACACGAAA
TTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCCAAGAGGTTCCAT
CTGCCAGGTATCAGGCAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTGATTACACC
CGAGGGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAAGCGAAGGTTG
TGGATCTGGATACCGGGAAAACGCTGGGCGTTAATCAAAGAGGCGAACTGTGTGTGAGAGGT
CCTATGATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATTGACAAGGA
TGGATGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAACACTTCTTCATCGTTG
ACCGCCTGAAGTCTCTGATTAAGTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCC
ATCTTGCTCCAACACCCCAACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGACGATGACGC
CGGTGAACTTCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGA
TCGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAGTTGTGTTT
GTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGCAAGAAAAATCAGAGAGATCCT
CATAAAGGCCAAGAAGGGCGGAAAGATCGCCGTGTAATTCTAGAGTCGGGGCGGCCGGCCGC
TTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGA
AAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTG
CAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGT
GGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAATCGATAAGGATCCGTCG
ACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTA
TCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCG
CTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTAT
CAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC
ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTT
CCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAA
ACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT
GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCT
TTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCT
GTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAG
TCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAG
AGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
GAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGT
AGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCA
GATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACG
CTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTC
ACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAAC
TTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCA
TCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGC
AATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCA
TCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGC
AACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATT
CAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGG
TTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATG
GTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGAC
TGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC
CGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGA
AAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTA
ACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAG
CAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATA
CTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGG

ATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAA
AAGTGCCACCTGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGC
AGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTT
TCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCC
GATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGT
GGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAG
TGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTAT
AAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAAC
GCGAATTTTAACAAAATATTAACGTTTACAATTTCCCATTCGCCATTCAGGCTGCGCAACTG
TTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCCCAAGCTACCATGATAAG
TAAGTAATATTAAGGTACGGGAGGTACTTGGAGCGGCCGCAATAAAATATCTTTATTTTCAT
TACATCTGTGTGTTGGTTTTTTGTGTGAATCGATAGTACTAACATACGCTCTCCATCAAAAC
AAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAAC
ATTTCTCTATCGATA

> pGL3 *TPH*

GGTACCAGATTCTTGCAAAGCTTATCTGCATATCCAGGTTTTATAAAAGGTTAAGAAACTGA
AGCTCCTAGAAGTAAATGAGCCCTTAGTTATCAAATATTGTCTGCATTGTGCAACTTCTTTT
TTAGCCCGTAGTCTTTGTTTTCATACTTTAGGCTAGCGGCATTCCGGTACTGTTGGTAAAGC
CACCATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAAGATG
GAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCCCTGGTTCCTGGAACAATT
GCTTTTACAGATGCACATATCGAGGTGGACATCACTTACGCTGAGTACTTCGAAATGTCCGT
TCGGTTGGCAGAAGCTATGAAACGATATGGGCTGAATACAAATCACAGAATCGTCGTATGCA
GTGAAAACTCTCTTCAATTCTTTATGCCGGTGTTGGGCGCGTTATTTATCGGAGTTGCAGTT
GCGCCCGCGAACGACATTTATAATGAACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCC
TACCGTGGTGTTCGTTTCCAAAAAGGGGTTGCAAAAAATTTTGAACGTGCAAAAAAAGCTCC
CAATCATCCAAAAAATTATTATCATGGATTCTAAAACGGATTACCAGGGATTTCAGTCGATG
TACACGTTCGTCACATCTCATCTACCTCCCGGTTTTAATGAATACGATTTTGTGCCAGAGTC
CTTCGATAGGGACAAGACAATTGCACTGATCATGAACTCCTCTGGATCTACTGGTCTGCCTA
AAGGTGTCGCTCTGCCTCATAGAACTGCCTGCGTGAGATTCTCGCATGCCAGAGATCCTATT
TTTGGCAATCAAATCATTCCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATCACGGTTT
TGGAATGTTTACTACACTCGGATATTTGATATGTGGATTTCGAGTCGTCTTAATGTATAGAT
TTGAAGAAGAGCTGTTTCTGAGGAGCCTTCAGGATTACAAGATTCAAAGTGCGCTGCTGGTG
CCAACCCTATTCTCCTTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTT
ACACGAAATTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCCAAGA
GGTTCCATCTGCCAGGTATCAGGCAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTG
ATTACACCCGAGGGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAAGC
GAAGGTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTTAATCAAAGAGGCGAACTGTGTG
TGAGAGGTCCTATGATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATT
GACAAGGATGGATGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAACACTTCTT
CATCGTTGACCGCCTGAAGTCTCTGATTAAGTACAAAGGCTATCAGGTGGCTCCCGCTGAAT
TGGAATCCATCTTGCTCCAACACCCCAACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGAC
GATGACGCCGGTGAACTTCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACGATGACGGA
AAAAGAGATCGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAG
TTGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGCAAGAAAAATCAGA
GAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGTGTAATTCTAGAGTCGGGGCGG
CCGGCCGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAA
TGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATT
ATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGG
GGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAATCGATAAGG

ATCCGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGG
CATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGC
CGGCAGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCG
AGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG
GAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTG
GCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAG
GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGC
GCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGC
GTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAA
GCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATC
GTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGG
ATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGG
CTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAA
GAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGC
AAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGG
GTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAA
GGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATAT
GAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTG
TCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGG
GCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGAT
TTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATC
CGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATA
GTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATG
GCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAA
AAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTAT
CACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTT
TCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTG
CTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCA
TCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGT
TCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTC
TGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAAT
GTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTC
ATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATT
TCCCCGAAAAGTGCCACCTGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGG
TTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTC
CCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTT
AGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTT
CACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTC
TTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTT
TGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAA
AATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCCCATTCGCCATTCAGGCTG
CGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCCCAAGCTACC
ATGATAAGTAAGTAATATTAAGGTACGGGAGGTACTTGGAGCGGCCGCAATAAAATATCTTT
ATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGATAGTACTAACATACGCTCTCC
ATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGT
GCCAGAACATTTCTCTATCGATA

> pGL3 *GHRH*

GGTACCTGAGCCACCGCGGTTCTCCAATCACTATTATAGCAGTATATATTCTCTATATCCTC
TTGGAATAATGTTACACCTTTGTACTATGTCCACTGTGCCAAAGATAAAAGGAGACTTTACC
AGGAGTCTAAGTCTGCAAGGGGCCAAACCTCTTTCACCAACAGGGTTTGTCAGTGTGATATG
ATGCTAAAAACAGTCCTTTGGTTGACTTGTGGGTAATTGATTCTCTGACGCTGACAACGCTT
AGGAAAATGAAGAGATAAATGATGGGAACGCCAGGCGGCTGCCAGAGCAAACACCCAGCAAG
CTTGGCATTCCGGTACTGTTGGTAAAGCCACCATGGAAGACGCCAAAAACATAAAGAAAGGC
CCGGCGCCATTCTATCCGCTGGAAGATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAA
GAGATACGCCCTGGTTCCTGGAACAATTGCTTTTACAGATGCACATATCGAGGTGGACATCA
CTTACGCTGAGTACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGGGCTG
AATACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTTATGCCGGTGTT
GGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACATTTATAATGAACGTGAAT
TGCTCAACAGTATGGGCATTTCGCAGCCTACCGTGGTGTTCGTTTCCAAAAAGGGGTTGCAA
AAAATTTTGAACGTGCAAAAAAAGCTCCCAATCATCCAAAAAATTATTATCATGGATTCTAA
AACGGATTACCAGGGATTTCAGTCGATGTACACGTTCGTCACATCTCATCTACCTCCCGGTT
TTAATGAATACGATTTTGTGCCAGAGTCCTTCGATAGGGACAAGACAATTGCACTGATCATG
AACTCCTCTGGATCTACTGGTCTGCCTAAAGGTGTCGCTCTGCCTCATAGAACTGCCTGCGT
GAGATTCTCGCATGCCAGAGATCCTATTTTTGGCAATCAAATCATTCCGGATACTGCGATTT
TAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACACTCGGATATTTGATATGT
GGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTGTTTCTGAGGAGCCTTCAGGA
TTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCTATTCTCCTTCTTCGCCAAAAGCACTC
TGATTGACAAATACGATTTATCTAATTTACACGAAATTGCTTCTGGTGGCGCTCCCCTCTCT
AAGGAAGTCGGGGAAGCGGTTGCCAAGAGGTTCCATCTGCCAGGTATCAGGCAAGGATATGG
GCTCACTGAGACTACATCAGCTATTCTGATTACACCCGAGGGGGATGATAAACCGGGCGCGG
TCGGTAAAGTTGTTCCATTTTTTGAAGCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTG
GGCGTTAATCAAAGAGGCGAACTGTGTGTGAGAGGTCCTATGATTATGTCCGGTTATGTAAA
CAATCCGGAAGCGACCAACGCCTTGATTGACAAGGATGGATGGCTACATTCTGGAGACATAG
CTTACTGGGACGAAGACGAACACTTCTTCATCGTTGACCGCCTGAAGTCTCTGATTAAGTAC
AAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCTTGCTCCAACACCCCAACATCTT
CGACGCAGGTGTCGCAGGTCTTCCCGACGATGACGCCGGTGAACTTCCCGCCGCCGTTGTTG
TTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGATCGTGGATTACGTCGCCAGTCAAGTA
ACAACCGCGAAAAAGTTGCGCGGAGGAGTTGTGTTTGTGGACGAAGTACCGAAAGGTCTTAC
CGGAAAACTCGACGCAAGAAAAATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGA
TCGCCGTGTAATTCTAGAGTCGGGGCGGCCGGCCGCTTCGAGCAGACATGATAAGATACATT
GATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTG
TGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATT
GCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAAC
CTCTACAAATGTGGTAAAATCGATAAGGATCCGTCGACCGATGCCCTTGAGAGCCTTCAACC
CAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTC
TTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTTCCGCTTCCTCGCTCACTGACT
CGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGG
TTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGC
ATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAG
GCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATA
CCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATC
TCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC
GACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATC
GCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAG
AGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCT
CTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCAC
CGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTC

AAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAA
GGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATG
AAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAA
TCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCC
GTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACC
GCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCG
AGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAA
GCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCAT
CGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGC
GAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTT
GTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCT
TACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCT
GAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCG
CCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTC
AAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTT
CAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCA
AAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTA
TTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAA
ATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGCGCCCTGTAGC
GGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGC
CCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCC
GTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGAC
CCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTT
TCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAA
CACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTAT
TGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTT
TACAATTTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCT
CTTCGCTATTACGCCAGCCCAAGCTACCATGATAAGTAAGTAATATTAAGGTACGGGAGGTA
CTTGGAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGT
GAATCGATAGTACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAA
ATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGATA

> pGL3 *SOX10*

GGTACCTCACGTCTCTCCCTAGTCCCCTGCCAAGTGCCCATCCTGCCACCTGAGGCAGACGT
TAGGAGTCCTGGTTGGGCCCAATTTTCTTCCATCCAACAGGTTCTGACCCTCCAGCCTACAT
CTCTTCCTTCCTCGGCCTAACCTCCCCTCCACCTCCCAGACTAAGCCCAAGATAAGCCTCAG
CAAGATGGGCCTGGGCAGTAGGGCAGGGAGCTGGGTGGGGGGGCCTGGGCTCCGGGTGGAGG
AGCCCTGAGTGTTGGGGATGAAGGAAAGCCTGGAGGCTCCACCTTCTGCCCTTCCCTCCCTC
CTCTCCCCTCCCCCCTCTCCCCCTCCTTCCTTCCCCAGCCCTCCGCAGCGGCTCAGGCTCAG
TCGCTCAGTCAGTCTCGGGCTGTCCGGCCAGGGTGGTTGGTGGTAAGGATTCAGGCTCCGTC
CTAACGAGGCCGTGGCCTGAGGCTCAGGGCCCCCCGCCCCTCCCTCCCAGCCCACCAGCGTC
ACCTCCCAGCCCCGAGCTGGACCGCACACCTTGGGACAAAGCTTGGCATTCCGGTACTGTTG
GTAAAGCCACCATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTG
GAAGATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCCCTGGTTCCTGG
AACAATTGCTTTTACAGATGCACATATCGAGGTGGACATCACTTACGCTGAGTACTTCGAAA
TGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGGGCTGAATACAAATCACAGAATCGTC
GTATGCAGTGAAAACTCTCTTCAATTCTTTATGCCGGTGTTGGGCGCGTTATTTATCGGAGT
TGCAGTTGCGCCCGCGAACGACATTTATAATGAACGTGAATTGCTCAACAGTATGGGCATTT
CGCAGCCTACCGTGGTGTTCGTTTCCAAAAAGGGGTTGCAAAAAATTTTGAACGTGCAAAAA
AAGCTCCCAATCATCCAAAAAATTATTATCATGGATTCTAAAACGGATTACCAGGGATTTCA

GTCGATGTACACGTTCGTCACATCTCATCTACCTCCCGGTTTTAATGAATACGATTTTGTGC
CAGAGTCCTTCGATAGGGACAAGACAATTGCACTGATCATGAACTCCTCTGGATCTACTGGT
CTGCCTAAAGGTGTCGCTCTGCCTCATAGAACTGCCTGCGTGAGATTCTCGCATGCCAGAGA
TCCTATTTTTGGCAATCAAATCATTCCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATC
ACGGTTTTGGAATGTTTACTACACTCGGATATTTGATATGTGGATTTCGAGTCGTCTTAATG
TATAGATTTGAAGAAGAGCTGTTTCTGAGGAGCCTTCAGGATTACAAGATTCAAAGTGCGCT
GCTGGTGCCAACCCTATTCTCCTTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTAT
CTAATTTACACGAAATTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTT
GCCAAGAGGTTCCATCTGCCAGGTATCAGGCAAGGATATGGGCTCACTGAGACTACATCAGC
TATTCTGATTACACCCGAGGGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTTT
TTGAAGCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTTAATCAAAGAGGCGAA
CTGTGTGTGAGAGGTCCTATGATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGC
CTTGATTGACAAGGATGGATGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAAC
ACTTCTTCATCGTTGACCGCCTGAAGTCTCTGATTAAGTACAAAGGCTATCAGGTGGCTCCC
GCTGAATTGGAATCCATCTTGCTCCAACACCCCAACATCTTCGACGCAGGTGTCGCAGGTCT
TCCCGACGATGACGCCGGTGAACTTCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACGA
TGACGGAAAAAGAGATCGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTGCGC
GGAGGAGTTGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGCAAGAAA
AATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGTGTAATTCTAGAGTC
GGGGCGGCCGGCCGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACA
ACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGT
AACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGG
TTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAATC
GATAAGGATCCGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGG
CGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGA
CAGGTGCCGGCAGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGC
TGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT
AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGC
GTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAA
GTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCC
CTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTC
GGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTC
GCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGT
AACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGG
TAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTA
ACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTC
GGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTT
TGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTT
CTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTA
TCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAG
TATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAG
CGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATA
CGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGC
TCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAA
CTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCA
GTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTT
TGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGT
TGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCA
GTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAG
ATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGAC
CGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAA

GTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAG
ATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCA
GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACA
CGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTA
TTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGC
GCACATTTCCCCGAAAAGTGCCACCTGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGT
GTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGC
TTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGC
TCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGT
GATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTC
CACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCT
ATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATT
TAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCCCATTCGCCATT
CAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCCCA
AGCTACCATGATAAGTAAGTAATATTAAGGTACGGGAGGTACTTGGAGCGGCCGCAATAAAA
TATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGATAGTACTAACATAC
GCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAG
TGCAGGTGCCAGAACATTTCTCTATCGATA

> pGL3 *GH*

GGTACCCCATGCATAAATGTACACAGAAACAGGTGGGGTCAACAGTGGGAGAGAAGGGGCCA
GGGTATAAAAAGGGCCCACAAGCTTGGCATTCCGGTACTGTTGGTAAAGCCACCATGGAAGA
CGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAAGATGGAACCGCTGGAG
AGCAACTGCATAAGGCTATGAAGAGATACGCCCTGGTTCCTGGAACAATTGCTTTTACAGAT
GCACATATCGAGGTGGACATCACTTACGCTGAGTACTTCGAAATGTCCGTTCGGTTGGCAGA
AGCTATGAAACGATATGGGCTGAATACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTC
TTCAATTCTTTATGCCGGTGTTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAAC
GACATTTATAATGAACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACCGTGGTGTT
CGTTTCCAAAAAGGGGTTGCAAAAAATTTTGAACGTGCAAAAAAAGCTCCCAATCATCCAAA
AAATTATTATCATGGATTCTAAAACGGATTACCAGGGATTTCAGTCGATGTACACGTTCGTC
ACATCTCATCTACCTCCCGGTTTTAATGAATACGATTTTGTGCCAGAGTCCTTCGATAGGGA
CAAGACAATTGCACTGATCATGAACTCCTCTGGATCTACTGGTCTGCCTAAAGGTGTCGCTC
TGCCTCATAGAACTGCCTGCGTGAGATTCTCGCATGCCAGAGATCCTATTTTTGGCAATCAA
ATCATTCCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTAC
TACACTCGGATATTTGATATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGC
TGTTTCTGAGGAGCCTTCAGGATTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCTATTC
TCCTTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTTACACGAAATTGC
TTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCCAAGAGGTTCCATCTGC
CAGGTATCAGGCAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTGATTACACCCGAG
GGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAAGCGAAGGTTGTGGA
TCTGGATACCGGGAAAACGCTGGGCGTTAATCAAAGAGGCGAACTGTGTGTGAGAGGTCCTA
TGATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATTGACAAGGATGGA
TGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAACACTTCTTCATCGTTGACCG
CCTGAAGTCTCTGATTAAGTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCT
TGCTCCAACACCCCAACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGACGATGACGCCGGT
GAACTTCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGATCGT
GGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAGTTGTGTTTGTGG
ACGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGCAAGAAAAATCAGAGAGATCCTCATA
AAGGCCAAGAAGGGCGGAAAGATCGCCGTGTAATTCTAGAGTCGGGGCGGCCGGCCGCTTCG
AGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAA

AATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAAT
AAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGA
GGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAATCGATAAGGATCCGTCGACCG
ATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGT
CGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCT
TCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGC
TCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGT
GAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCAT
AGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCC
GACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTC
CGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCT
CAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGT
GCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA
ACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCG
AGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG
GACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCT
CTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATT
ACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCA
GTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCT
AGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGG
TCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTC
ATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTG
GCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATA
AACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCA
GTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACG
TTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGC
TCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAG
CTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTA
TGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGT
GAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGC
GTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAAC
GTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCC
ACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAA
AACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCA
TACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATAC
ATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGT
GCCACCTGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCG
TGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTC
GCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATT
TAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGC
CATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGA
CTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGG
GATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGA
ATTTTAACAAAATATTAACGTTTACAATTTCCCATTCGCCATTCAGGCTGCGCAACTGTTGG
GAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCCCAAGCTACCATGATAAGTAAG
TAATATTAAGGTACGGGAGGTACTTGGAGCGGCCGCAATAAAATATCTTTATTTTCATTACA
TCTGTGTGTTGGTTTTTTGTGTGAATCGATAGTACTAACATACGCTCTCCATCAAAACAAAA
CGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTT
CTCTATCGATA

> pGL3 *TH*

GGTACCAGGTGGGGGACCCAGAGGGGGCTTTGACGTCAGCTCAGCTTATAAGAGGCTGCTGG
GCCAGGGCTGTGGAGACGGAGCCAAGCTTGGCATTCCGGTACTGTTGGTAAAGCCACCATGG
AAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAAGATGGAACCGCT
GGAGAGCAACTGCATAAGGCTATGAAGAGATACGCCCTGGTTCCTGGAACAATTGCTTTTAC
AGATGCACATATCGAGGTGGACATCACTTACGCTGAGTACTTCGAAATGTCCGTTCGGTTGG
CAGAAGCTATGAAACGATATGGGCTGAATACAAATCACAGAATCGTCGTATGCAGTGAAAAC
TCTCTTCAATTCTTTATGCCGGTGTTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGC
GAACGACATTTATAATGAACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACCGTGG
TGTTCGTTTCCAAAAAGGGGTTGCAAAAAATTTTGAACGTGCAAAAAAAGCTCCCAATCATC
CAAAAAATTATTATCATGGATTCTAAAACGGATTACCAGGGATTTCAGTCGATGTACACGTT
CGTCACATCTCATCTACCTCCCGGTTTTAATGAATACGATTTTGTGCCAGAGTCCTTCGATA
GGGACAAGACAATTGCACTGATCATGAACTCCTCTGGATCTACTGGTCTGCCTAAAGGTGTC
GCTCTGCCTCATAGAACTGCCTGCGTGAGATTCTCGCATGCCAGAGATCCTATTTTTGGCAA
TCAAATCATTCCGGATACTGCGATTTTAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGT
TTACTACACTCGGATATTTGATATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAA
GAGCTGTTTCTGAGGAGCCTTCAGGATTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCT
ATTCTCCTTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTTACACGAAA
TTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCCAAGAGGTTCCAT
CTGCCAGGTATCAGGCAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTGATTACACC
CGAGGGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAAGCGAAGGTTG
TGGATCTGGATACCGGGAAAACGCTGGGCGTTAATCAAAGAGGCGAACTGTGTGTGAGAGGT
CCTATGATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATTGACAAGGA
TGGATGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAACACTTCTTCATCGTTG
ACCGCCTGAAGTCTCTGATTAAGTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCC
ATCTTGCTCCAACACCCCAACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGACGATGACGC
CGGTGAACTTCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGA
TCGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAGTTGTGTTT
GTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGCAAGAAAAATCAGAGAGATCCT
CATAAAGGCCAAGAAGGGCGGAAAGATCGCCGTGTAATTCTAGAGTCGGGGCGGCCGGCCGC
TTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGA
AAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTG
CAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGT
GGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAATCGATAAGGATCCGTCG
ACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTA
TCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCG
CTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTAT
CAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC
ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTT
CCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAA
ACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT
GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCT
TTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCT
GTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAG
TCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAG
AGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
GAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGT
AGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCA
GATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACG

CTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTC
ACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAAC
TTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCA
TCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGC
AATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCA
TCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGC
AACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATT
CAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGG
TTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATG
GTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGAC
TGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC
CGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGA
AAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTA
ACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAG
CAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATA
CTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGG
ATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAA
AAGTGCCACCTGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGC
AGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTT
TCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCC
GATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGT
GGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAG
TGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTAT
AAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAAC
GCGAATTTTAACAAAATATTAACGTTTACAATTTCCCATTCGCCATTCAGGCTGCGCAACTG
TTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCCCAAGCTACCATGATAAG
TAAGTAATATTAAGGTACGGGAGGTACTTGGAGCGGCCGCAATAAAATATCTTTATTTTCAT
TACATCTGTGTGTTGGTTTTTTGTGTGAATCGATAGTACTAACATACGCTCTCCATCAAAAC
AAAACGAAACAAAACAAACTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAAC
ATTTCTCTATCGATA

> pGL3 *INS*

GGTACCCCTCAGCCCCCAGCCATCTGCCGACCCCCCCACCCCAGGCCCTAATGGGCCAGGCG
GCAGGGGTTGAGAGGTAGGGGAGATGGGCTCTGAGACTATAAAGCCAGCGGGGGCCCAGCAG
CCCTCAAGCTTGGCATTCCGGTACTGTTGGTAAAGCCACCATGGAAGACGCCAAAAACATAA
AGAAAGGCCCGGCGCCATTCTATCCGCTGGAAGATGGAACCGCTGGAGAGCAACTGCATAAG
GCTATGAAGAGATACGCCCTGGTTCCTGGAACAATTGCTTTTACAGATGCACATATCGAGGT
GGACATCACTTACGCTGAGTACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGAT
ATGGGCTGAATACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTTATG
CCGGTGTTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACATTTATAATGA
ACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACCGTGGTGTTCGTTTCCAAAAAGG
GGTTGCAAAAAATTTTGAACGTGCAAAAAAGCTCCCAATCATCCAAAAAATTATTATCATG
GATTCTAAAACGGATTACCAGGGATTTCAGTCGATGTACACGTTCGTCACATCTCATCTACC
TCCCGGTTTTAATGAATACGATTTTGTGCCAGAGTCCTTCGATAGGGACAAGACAATTGCAC
TGATCATGAACTCCTCTGGATCTACTGGTCTGCCTAAAGGTGTCGCTCTGCCTCATAGAACT
GCCTGCGTGAGATTCTCGCATGCCAGAGATCCTATTTTTGGCAATCAAATCATTCCGGATAC
TGCGATTTTAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACACTCGGATATT

TGATATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTGTTTCTGAGGAGC
CTTCAGGATTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCTATTCTCCTTCTTCGCCAA
AAGCACTCTGATTGACAAATACGATTTATCTAATTTACACGAAATTGCTTCTGGTGGCGCTC
CCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCCAAGAGGTTCCATCTGCCAGGTATCAGGCAA
GGATATGGGCTCACTGAGACTACATCAGCTATTCTGATTACACCCGAGGGGGATGATAAACC
GGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAAGCGAAGGTTGTGGATCTGGATACCGGGA
AAACGCTGGGCGTTAATCAAAGAGGCGAACTGTGTGTGAGAGGTCCTATGATTATGTCCGGT
TATGTAAACAATCCGGAAGCGACCAACGCCTTGATTGACAAGGATGGATGGCTACATTCTGG
AGACATAGCTTACTGGGACGAAGACGAACACTTCTTCATCGTTGACCGCCTGAAGTCTCTGA
TTAAGTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCTTGCTCCAACACCCC
AACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGACGATGACGCCGGTGAACTTCCCGCCGC
CGTTGTTGTTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGATCGTGGATTACGTCGCCA
GTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAGTTGTGTTTGTGGACGAAGTACCGAAA
GGTCTTACCGGAAAACTCGACGCAAGAAAAATCAGAGAGATCCTCATAAAGGCCAAGAAGGG
CGGAAAGATCGCCGTGTAATTCTAGAGTCGGGGCGGCCGGCCGCTTCGAGCAGACATGATAA
GATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGT
GAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAA
CAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCA
AGTAAAACCTCTACAAATGTGGTAAAATCGATAAGGATCCGTCGACCGATGCCCTTGAGAGC
CTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGA
CTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTTCCGCTTCCTCGCT
CACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG
TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAG
CAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCC
TGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTT
ACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTG
TAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCG
TTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACAC
GACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGG
TGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTA
TCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAA
CAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAA
AGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACT
CACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAAT
TAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCA
ATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCT
GACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCA
ATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGG
AAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTT
GCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCT
ACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACG
ATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTC
CGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCAT
AATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAA
GTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATA
ATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGA
AAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAA
CTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAA
ATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTT
CAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTAT

TTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGCGC
CCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTT
GCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGG
CTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGC
ACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAG
ACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAAC
TGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTT
CGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATA
TTAACGTTTACAATTTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGT
GCGGGCCTCTTCGCTATTACGCCAGCCCAAGCTACCATGATAAGTAAGTAATATTAAGGTAC
GGGAGGTACTTGGAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTT
TTTTGTGTGAATCGATAGTACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAA
CTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGATA

> pGL3 *FOXN1*

ACGCGTAGCAGAGAGAAGGAGGTGAGGACAGGGTCTGGGGGAACAGCAGAGAAAGCTGTCAG
GGAGCAGGGACAGGCTCCAGGCCTGAGTCCAGCTGGGACCAGGGAAGCCAATTCCTAGGGCT
TCACCCTCAACATGAACTCTAAGGGGCCAGAGACCAGGGCAACAGCCACTCAGGTTTCTGTT
TCTATCCTCCAAATAATGAAGAGTAGCCTGGGATGAAGACACAGAACCAGGGAGCCCCCATC
TCCACCCCCACAGACACCCACATGCTCACACAGGACTAGGTTGCTGACCCACAGCAGAGACG
TGTGCACAGACATGAGCATGTCAACACACGGACGACACACGGGCCTGGGTCCGGGGACACAG
AGACACAGACACAGGGGAAGACATGCAGACACACGCCCACCCAGGAGACAGATGTACAGACC
CAGGGAGAGATGCAGACAGACTCGCATGTATCTGCCCCTGAGGGAGAGGGAGATGGGAGTCA
CCAGCACATGCTCCAGTGGGTGCACACGCCCCCACCCACAAGGAGACACCCTCGAGGTGCTG
CGGGAGCTGTGGAGAGACAGACAGATGGCCGGCCCCATGGACAGACAGTTCCCAGCAAATCT
GTGCTAATGCTTGGCCCTTGAAGGGCCCAACCTGTCCATGCCGTGCCTCAGAGAACCCTGGG
CTGCCACCCAGCGCTGTTCCCCAGAACCTCATGAGTGGGCTCCATGAGCATACATGTGGCGG
CCAGGCTCTCCTGACACACGTGCATGCTCCAACAGGCCTGTGCCCCACCTGTGGGTGGGGGG
GCGACATCTCCCTTCTGGAGTGCAGGCGGGGGCCACGGGACAACCCCCTCCTCTACAAGCCA
CAGGCTGTGCGCACTGGCAGGGACAGGGGAGGGGGCCTGGGCTAGCTCAGGAGGAGGGAGGG
GTTGGGGGGCCAGGGTGCGGCTGGGGCGGGCACGGGAGCCGGTACCTGTCAAAAAGCTTGGC
ATTCCGGTACTGTTGGTAAAGCCACCATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCG
CCATTCTATCCGCTGGAAGATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATA
CGCCCTGGTTCCTGGAACAATTGCTTTTACAGATGCACATATCGAGGTGGACATCACTTACG
CTGAGTACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGGGCTGAATACA
AATCACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTTATGCCGGTGTTGGGCGC
GTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACATTTATAATGAACGTGAATTGCTCA
ACAGTATGGGCATTTCGCAGCCTACCGTGGTGTTCGTTTCCAAAAAGGGGTTGCAAAAAATT
TTGAACGTGCAAAAAAAGCTCCCAATCATCCAAAAAATTATTATCATGGATTCTAAAACGGA
TTACCAGGGATTTCAGTCGATGTACACGTTCGTCACATCTCATCTACCTCCCGGTTTTAATG
AATACGATTTTGTGCCAGAGTCCTTCGATAGGGACAAGACAATTGCACTGATCATGAACTCC
TCTGGATCTACTGGTCTGCCTAAAGGTGTCGCTCTGCCTCATAGAACTGCCTGCGTGAGATT
CTCGCATGCCAGAGATCCTATTTTTGGCAATCAAATCATTCCGGATACTGCGATTTTAAGTG
TTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACACTCGGATATTTGATATGTGGATTT
CGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTGTTTCTGAGGAGCCTTCAGGATTACAA
GATTCAAAGTGCGCTGCTGGTGCCAACCCTATTCTCCTTCTTCGCCAAAAGCACTCTGATTG
ACAAATACGATTTATCTAATTTACACGAAATTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAA
GTCGGGGAAGCGGTTGCCAAGAGGTTCCATCTGCCAGGTATCAGGCAAGGATATGGGCTCAC
TGAGACTACATCAGCTATTCTGATTACACCCGAGGGGGATGATAAACCGGGCGCGGTCGGTA
AAGTTGTTCCATTTTTTGAAGCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTT

AATCAAAGAGGCGAACTGTGTGTGAGAGGTCCTATGATTATGTCCGGTTATGTAAACAATCC
GGAAGCGACCAACGCCTTGATTGACAAGGATGGATGGCTACATTCTGGAGACATAGCTTACT
GGGACGAAGACGAACACTTCTTCATCGTTGACCGCCTGAAGTCTCTGATTAAGTACAAAGGC
TATCAGGTGGCTCCCGCTGAATTGGAATCCATCTTGCTCCAACACCCCAACATCTTCGACGC
AGGTGTCGCAGGTCTTCCCGACGATGACGCCGGTGAACTTCCCGCCGCCGTTGTTGTTTTGG
AGCACGGAAAGACGATGACGGAAAAAGAGATCGTGGATTACGTCGCCAGTCAAGTAACAACC
GCGAAAAAGTTGCGCGGAGGAGTTGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAA
ACTCGACGCAAGAAAAATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGATCGCCG
TGTAATTCTAGAGTCGGGGCGGCCGGCCGCTTCGAGCAGACATGATAAGATACATTGATGAG
TTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGC
TATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTC
ATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTAC
AAATGTGGTAAAATCGATAAGGATCCGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCA
GCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATC
ATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGC
GCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCC
ACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAA
CCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACA
AAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC
CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTT
CGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGC
TGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACT
GGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCT
TGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTG
AAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGG
TAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAG
ATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATT
TTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTT
TAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTG
AGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTG
TAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGA
CCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCA
GAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGA
GTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGT
GTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTA
CATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGA
AGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGT
CATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAAT
AGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACAT
AGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGAT
CTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCAT
CTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAG
GGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAG
CATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAAC
AAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGCGCCCTGTAGCGGCGCA
TTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGC
GCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAG
CTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAA
AAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCC
TTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCA

ACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTA
AAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAAT
TTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGC
TATTACGCCAGCCCAAGCTACCATGATAAGTAAGTAATATTAAGGTACGGGAGGTACTTGGA
GCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCG
ATAGTACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGC
TGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGATA

Red: Restriction enzymes

Yellow: Insert

Green: Forward primer for sequencing

Turquoise: Reverse primer for sequencing

8.4 In silico sequences of wild type and mutated AIRE in pBIND

>pBIND AIRE

TCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATT
GGCCATTGCATACGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATA
TGACCGCCATGTTGGCATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT
AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCT
GACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCA
ATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGT
ACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCG
CCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTA
TTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCG
GTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC
ACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTGCGATCGCCCGCCCCGTTGACGC
AAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGT
CAGATCACTAGAAGCTTTATTGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGTGC
TTCTGACACAACAGTCTCGAACTTAAGCTGCAGTGACTCTCTTAAGGTAGCCTTGCAGAAGT
TGGTCGTGAGGCACTGGGCAGGTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATA
GAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTA
CTGACATCCACTTTGCCTTTCTCTCCACAGGTGTCCACTCCCAGTTCAATTACAGCTCTTAA
GGCTAGAGTACTTAATACGACTCACTATAGGCTAGCCAGCTTGAAGCAAGCCTCCTGAAAGA
TGAAGCTACTGTCTTCTATCGAACAAGCATGCGATATTTGCCGACTTAAAAAGCTCAAGTGC
TCCAAAGAAAAACCGAAGTGCGCCAAGTGTCTGAAGAACAACTGGGAGTGTCGCTACTCTCC
CAAAACCAAAAGGTCTCCGCTGACTAGGGCACATCTGACAGAAGTGGAATCAAGGCTAGAAA
GACTGGAACAGCTATTTCTACTGATTTTTCCTCGAGAAGACCTTGACATGATTTTGAAAATG
GATTCTTTACAGGATATAAAAGCATTGTTAACAGGATTATTTGTACAAGATAATGTGAATAA
AGATGCCGTCACAGATAGATTGGCTTCAGTGGAGACTGATATGCCTCTAACATTGAGACAGC
ATAGAATAAGTGCGACATCATCATCGGAAGAGAGTAGTAACAAAGGTCAAAGACAGTTGACT
GTATCGCCGGAATTCCCGGGGATCCGTCGACTTGACGCGTCTATGGCGACGGACGCGGCGCT
ACGCCGGCTTCTGAGGCTGCACCGCACGGAGATCGCGGTGGCCGTGGACAGCGCCTTCCCAC
TGCTGCACGCGCTGGCTGACCACGACGTGGTCCCCGAGGACAAGTTTCAGGAGACGCTTCAT
CTGAAGGAAAAGGAGGGCTGCCCCCAGGCCTTCCACGCCCTCCTGTCCTGGCTGCTGACCCA
GGACTCCACAGCCATCCTGGACTTCTGGAGGGTGCTGTTCAAGGACTACAACCTGGAGCGCT

ATGGCCGGCTGCAGCCCATCCTGGACAGCTTCCCCAAAGATGTGGACCTCAGCCAGCCCCGG
AAGGGGAGGAAGCCCCCGGCCGTCCCCAAGGCTTTGGTACCGCCACCCAGACTCCCCACCAA
GAGGAAGGCCTCAGAAGAGGCTCGAGCTGCCGCGCCAGCAGCCCTGACTCCAAGGGGCACCG
CCAGCCCAGGCTCTCAACTGAAGGCCAAGCCCCCCAAGAAGCCGGAGAGCAGCGCAGAGCAG
CAGCGCCTTCCACTCGGGAACGGGATTCAGACCATGTCAGCTTCAGTCCAGAGAGCTGTGGC
CATGTCCTCCGGGGACGTCCCGGGAGCCCGAGGGGCCGTGGAGGGGATCCTCATCCAGCAGG
TGTTTGAGTCAGGCGGCTCCAAGAAGTGCATCCAGGTTGGCGGGGAGTTCTACACTCCCAGC
AAGTTCGAAGACTCCGGCAGTGGGAAGAACAAGGCCCGCAGCAGCAGTGGCCCGAAGCCTCT
GGTTCGAGCCAAGGGAGCCCAGGGCGCTGCCCCCGGTGGAGGTGAGGCTAGGCTGGGCCAGC
AGGGCAGCGTTCCCGCCCCTCTGGCCCTCCCCAGTGACCCCCAGCTCCACCAGAAGAATGAG
GACGAGTGTGCCGTGTGTCGGGACGGCGGGGAGCTCATCTGCTGTGACGGCTGCCCTCGGGC
CTTCCACCTGGCCTGCCTGTCCCCTCCGCTCCGGGAGATCCCCAGTGGGACCTGGAGGTGCT
CCAGCTGCCTGCAGGCAACAGTCCAGGAGGTGCAGCCCCGGGCAGAGGAGCCCCGGCCCCAG
GAGCCACCCGTGGAGACCCCGCTCCCCCCGGGGCTTAGGTCGGCGGGAGAGGAGGTAAGAGG
TCCACCTGGGGAACCCCTAGCCGGCATGGACACGACTCTTGTCTACAAGCACCTGCCGGCTC
CGCCTTCTGCAGCCCCGCTGCCAGGGCTGGACTCCTCGGCCCTGCACCCCCTACTGTGTGTG
GGTCCTGAGGGTCAGCAGAACCTGGCTCCTGGTGCGCGTTGCGGGGTGTGCGGAGATGGTAC
GGACGTGCTGCGGTGTACTCACTGCGCCGCTGCCTTCCACTGGCGCTGCCACTTCCCAGCCG
GCACCTCCCGGCCCGGGACGGGCCTGCGCTGCAGATCCTGCTCAGGAGACGTGACCCCAGCC
CCTGTGGAGGGGGTGCTGGCCCCCAGCCCCGCCCGCCTGGCCCCTGGGCCTGCCAAGGATGA
CACTGCCAGTCACGAGCCCGCTCTGCACAGGGATGACCTGGAGTCCCTTCTGAGCGAGCACA
CCTTCGATGGCATCCTGCAGTGGGCCATCCAGAGCATGGCCCGTCCGGCGGCCCCCTTCCCC
TCCGGGGGTGGAGGCTCTGAGCAGAAACTCATCTCAGAAGAGGATCTGTAAGCGGCCGCAGG
TACCTGAATAACTAAGGCCGCTTCCCTTTAGTGAGGGTTAATGCTTCGAGCAGACATGATAA
GATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGT
GAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAA
CAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGATGTGGGAGGTTTTTTAAAGCA
AGTAAAACCTCTACAAATGTGGTAAAATCCGATAAGGATCGATCCGGGCTGGCGTAATAGCG
AAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGC
CCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTT
GCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCACGCCACGTTCGCCGG
CTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGC
ACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAG
ACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAAC
TGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTT
CGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATA
TTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACAC
CGCATACGCGGATCTTCCGTACCTTCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAG
TTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAA
TTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCA
TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACT
CCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGC
CGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAG
GCTTTTGCAAAAAGCTTGATTCTTCTGACACAACAGTCTCGAACTTAAGGCTAGAATTCTGG
CTACAGGTAAGCGCCCCTAAAATCCCTTTGGGCACAATGTGTCCTGAGGGGAGAGGCAGCGA
CCTGTAGATGGGACGGGGGCACTAACCCTCAGGTTTGGGGCTTCTGAATGTGAGTATCGCCA
TGTAAGCCCAGTATTTGGCCAATCTCAGAAAGCTCCTGGTCCCTGGAGGGATGGAGAGAGAA

AAACAAACAGCTCCTGGAGCAGGGAGAGTGCTGGCCTCTTGCTCTCCGGCTCCCTCTGTTGC
CCTCTGGTTTCTCCCCAGGCTCCCGGACGTCCTCTAGCCACCATGACTTCGAAAGTTTATGA
TCCAGAACAAAGGAAACGGATGATAACTGGTCCGCAGTGGTGGGCCAGATGTAAACAAATGA
ATGTTCTTGATTCATTTATTAATTATTATGATTCAGAAAAACATGCAGAAAATGCTGTTATT
TTTTTACATGGTAACGCGGCCTCTTCTTATTTATGGCGACATGTTGTGCCACATATTGAGCC
AGTAGCGCGGTGTATTATACCAGACCTTATTGGTATGGGCAAATCAGGCAAATCTGGTAATG
GTTCTTATAGGTTACTTGATCATTACAAATATCTTACTGCATGGTTTGAACTTCTTAATTTA
CCAAAGAAGATCATTTTTGTCGGCCATGATTGGGGTGCTTGTTTGGCATTTCATTATAGCTA
TGAGCATCAAGATAAGATCAAAGCAATAGTTCACGCTGAAAGTGTAGTAGATGTGATTGAAT
CATGGGATGAATGGCCTGATATTGAAGAAGATATTGCGTTGATCAAATCTGAAGAAGGAGAA
AAAATGGTTTTGGAGAATAACTTCTTCGTGGAAACCATGTTGCCATCAAAAATCATGAGAAA
GTTAGAACCAGAAGAATTTGCAGCATATCTTGAACCATTCAAAGAGAAAGGTGAAGTTCGTC
GTCCAACATTATCATGGCCTCGTGAAATCCCGTTAGTAAAAGGTGGTAAACCTGACGTTGTA
CAAATTGTTAGGAATTATAATGCTTATCTACGTGCAAGTGATGATTTACCAAAAATGTTTAT
TGAATCGGACCCAGGATTCTTTTCCAATGCTATTGTTGAAGGTGCCAAGAAGTTTCCTAATA
CTGAATTTGTCAAAGTAAAAGGTCTTCATTTTTCGCAAGAAGATGCACCTGATGAAATGGGA
AAATATATCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAACAATAATTCTAGCCCTGAAT
AAGTGATAATAAGCGGATGAATGGCAGAAATTCGTCGAAGCGCAATAAAATATCTTTATTTT
CATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGATAGCGATAAGGATCGGAAGATCCGCG
TATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCG
CCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGC
TGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGA
GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCT
TAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTA
AATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATT
GAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCA
TTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTT
TTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTA
TTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGA
CTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAAT
TATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATC
GGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGA
TCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTG
TAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGG
CAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCT
TCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCA
TTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGT
CAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA
TTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTT
AATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGT
GAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC
TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTT
GTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAG
ATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGC
ACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT
CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGA

ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCT
ACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGG
TAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTAT
CTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTC
AGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT
GCTGGCCTTTTGCTCACATGGCTCGACAGATCT

>pBIND AIRE-Y85

TCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATT
GGCCATTGCATACGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATA
TGACCGCCATGTTGGCATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT
AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCT
GACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCA
ATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGT
ACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCG
CCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTA
TTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCG
GTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC
ACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTGCGATCGCCCGCCCCGTTGACGC
AAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGT
CAGATCACTAGAAGCTTTATTGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGTGC
TTCTGACACAACAGTCTCGAACTTAAGCTGCAGTGACTCTCTTAAGGTAGCCTTGCAGAAGT
TGGTCGTGAGGCACTGGGCAGGTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATA
GAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTA
CTGACATCCACTTTGCCTTTCTCTCCACAGGTGTCCACTCCCAGTTCAATTACAGCTCTTAA
GGCTAGAGTACTTAATACGACTCACTATAGGCTAGCCAGCTTGAAGCAAGCCTCCTGAAAGA
TGAAGCTACTGTCTTCTATCGAACAAGCATGCGATATTTGCCGACTTAAAAAGCTCAAGTGC
TCCAAAGAAAAACCGAAGTGCGCCAAGTGTCTGAAGAACAACTGGGAGTGTCGCTACTCTCC
CAAAACCAAAAGGTCTCCGCTGACTAGGGCACATCTGACAGAAGTGGAATCAAGGCTAGAAA
GACTGGAACAGCTATTTCTACTGATTTTTCCTCGAGAAGACCTTGACATGATTTTGAAAATG
GATTCTTTACAGGATATAAAAGCATTGTTAACAGGATTATTTGTACAAGATAATGTGAATAA
AGATGCCGTCACAGATAGATTGGCTTCAGTGGAGACTGATATGCCTCTAACATTGAGACAGC
ATAGAATAAGTGCGACATCATCATCGGAAGAGAGTAGTAACAAAGGTCAAAGACAGTTGACT
GTATCGCCGGAATTCCCGGGGATCCGTCGACTTGACGCGTCTATGGCGACGGACGCGGCGCT
ACGCCGGCTTCTGAGGCTGCACCGCACGGAGATCGCGGTGGCCGTGGACAGCGCCTTCCCAC
TGCTGCACGCGCTGGCTGACCACGACGTGGTCCCCGAGGACAAGTTTCAGGAGACGCTTCAT
CTGAAGGAAAAGGAGGGCTGCCCCCAGGCCTTCCACGCCCTCCTGTCCTGGCTGCTGACCCA
GGACTCCACAGCCATCCTGGACTTCTGGAGGGTGCTGTTCAAGGACTGCAACCTGGAGCGCT
ATGGCCGGCTGCAGCCCATCCTGGACAGCTTCCCCAAAGATGTGGACCTCAGCCAGCCCCGG
AAGGGGAGGAAGCCCCCGGCCGTCCCCAAGGCTTTGGTACCGCCACCCAGACTCCCCACCAA
GAGGAAGGCCTCAGAAGAGGCTCGAGCTGCCGCGCCAGCAGCCCTGACTCCAAGGGGCACCG
CCAGCCCAGGCTCTCAACTGAAGGCCAAGCCCCCCAAGAAGCCGGAGAGCAGCGCAGAGCAG
CAGCGCCTTCCACTCGGGAACGGGATTCAGACCATGTCAGCTTCAGTCCAGAGAGCTGTGGC
CATGTCCTCCGGGGACGTCCCGGGAGCCCGAGGGGCCGTGGAGGGGATCCTCATCCAGCAGG
TGTTTGAGTCAGGCGGCTCCAAGAAGTGCATCCAGGTTGGCGGGGAGTTCTACACTCCCAGC
AAGTTCGAAGACTCCGGCAGTGGGAAGAACAAGGCCCGCAGCAGCAGTGGCCCGAAGCCTCT
GGTTCGAGCCAAGGGAGCCCAGGGCGCTGCCCCCGGTGGAGGTGAGGCTAGGCTGGGCCAGC

AGGGCAGCGTTCCCGCCCCTCTGGCCCTCCCCAGTGACCCCCAGCTCCACCAGAAGAATGAG
GACGAGTGTGCCGTGTGTCGGGACGGCGGGGAGCTCATCTGCTGTGACGGCTGCCCTCGGGC
CTTCCACCTGGCCTGCCTGTCCCCTCCGCTCCGGGAGATCCCCAGTGGGACCTGGAGGTGCT
CCAGCTGCCTGCAGGCAACAGTCCAGGAGGTGCAGCCCCGGGCAGAGGAGCCCCGGCCCCAG
GAGCCACCCGTGGAGACCCCGCTCCCCCCGGGGCTTAGGTCGGCGGGAGAGGAGGTAAGAGG
TCCACCTGGGGAACCCCTAGCCGGCATGGACACGACTCTTGTCTACAAGCACCTGCCGGCTC
CGCCTTCTGCAGCCCCGCTGCCAGGGCTGGACTCCTCGGCCCTGCACCCCCTACTGTGTGTG
GGTCCTGAGGGTCAGCAGAACCTGGCTCCTGGTGCGCGTTGCGGGGTGTGCGGAGATGGTAC
GGACGTGCTGCGGTGTACTCACTGCGCCGCTGCCTTCCACTGGCGCTGCCACTTCCCAGCCG
GCACCTCCCGGCCCGGGACGGGCCTGCGCTGCAGATCCTGCTCAGGAGACGTGACCCCAGCC
CCTGTGGAGGGGGTGCTGGCCCCCAGCCCCGCCCGCCTGGCCCCTGGGCCTGCCAAGGATGA
CACTGCCAGTCACGAGCCCGCTCTGCACAGGGATGACCTGGAGTCCCTTCTGAGCGAGCACA
CCTTCGATGGCATCCTGCAGTGGGCCATCCAGAGCATGGCCCGTCCGGCGGCCCCCTTCCCC
TCCGGGGGGTGGAGGCTCTGAGCAGAAACTCATCTCAGAAGAGGATCTGTAAGCGGCCGCAGG
TACCTGAATAACTAAGGCCGCTTCCCTTTAGTGAGGGTTAATGCTTCGAGCAGACATGATAA
GATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGT
GAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAA
CAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGATGTGGGAGGTTTTTTAAAGCA
AGTAAAACCTCTACAAATGTGGTAAAATCCGATAAGGATCGATCCGGGCTGGCGTAATAGCG
AAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGC
CCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTT
GCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCACGCCACGTTCGCCGG
CTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGC
ACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAG
ACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAAC
TGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTT
CGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATA
TTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACAC
CGCATACGCGGATCTTCCGTACCTTCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAG
TTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAA
TTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCA
TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACT
CCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGC
CGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAG
GCTTTTGCAAAAAGCTTGATTCTTCTGACACAACAGTCTCGAACTTAAGGCTAGAATTCTGG
CTACAGGTAAGCGCCCCTAAAATCCCTTTGGGCACAATGTGTCCTGAGGGGAGAGGCAGCGA
CCTGTAGATGGGACGGGGGCACTAACCCTCAGGTTTGGGGCTTCTGAATGTGAGTATCGCCA
TGTAAGCCCAGTATTTGGCCAATCTCAGAAAGCTCCTGGTCCCTGGAGGGATGGAGAGAGAA
AAACAAACAGCTCCTGGAGCAGGGAGAGTGCTGGCCTCTTGCTCTCCGGCTCCCTCTGTTGC
CCTCTGGTTTCTCCCCAGGCTCCCGGACGTCCTCTAGCCACCATGACTTCGAAAGTTTATGA
TCCAGAACAAAGGAAACGGATGATAACTGGTCCGCAGTGGTGGGCCAGATGTAAACAAATGA
ATGTTCTTGATTCATTTATTAATTATTATGATTCAGAAAAACATGCAGAAAATGCTGTTATT
TTTTTACATGGTAACGCGGCCTCTTCTTATTTATGGCGACATGTTGTGCCACATATTGAGCC
AGTAGCGCGGTGTATTATACCAGACCTTATTGGTATGGGCAAATCAGGCAAATCTGGTAATG
GTTCTTATAGGTTACTTGATCATTACAAATATCTTACTGCATGGTTTGAACTTCTTAATTTA
CCAAAGAAGATCATTTTTGTCGGCCATGATTGGGGTGCTTGTTTGGCATTTCATTATAGCTA
TGAGCATCAAGATAAGATCAAAGCAATAGTTCACGCTGAAAGTGTAGTAGATGTGATTGAAT

CATGGGATGAATGGCCTGATATTGAAGAAGATATTGCGTTGATCAAATCTGAAGAAGGAGAA
AAAATGGTTTTGGAGAATAACTTCTTCGTGGAAACCATGTTGCCATCAAAAATCATGAGAAA
GTTAGAACCAGAAGAATTTGCAGCATATCTTGAACCATTCAAAGAGAAAGGTGAAGTTCGTC
GTCCAACATTATCATGGCCTCGTGAAATCCCGTTAGTAAAAGGTGGTAAACCTGACGTTGTA
CAAATTGTTAGGAATTATAATGCTTATCTACGTGCAAGTGATGATTTACCAAAAATGTTTAT
TGAATCGGACCCAGGATTCTTTTCCAATGCTATTGTTGAAGGTGCCAAGAAGTTTCCTAATA
CTGAATTTGTCAAAGTAAAAGGTCTTCATTTTTCGCAAGAAGATGCACCTGATGAAATGGGA
AAATATATCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAACAATAATTCTAGCCCTGAAT
AAGTGATAATAAGCGGATGAATGGCAGAAATTCGTCGAAGCGCAATAAAATATCTTTATTTT
CATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGATAGCGATAAGGATCGGAAGATCCGCG
TATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCG
CCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGC
TGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGA
GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCT
TAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTA
AATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATT
GAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCA
TTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTT
TTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTA
TTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGA
CTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAAT
TATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATC
GGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGA
TCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTG
TAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGG
CAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCT
TCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCA
TTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGT
CAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA
TTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTT
AATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGT
GAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC
TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTT
GTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAG
ATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGC
ACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT
CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGA
ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCT
ACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGG
TAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTAT
CTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTC
AGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT
GCTGGCCTTTTGCTCACATGGCTCGACAGATCT

>pBIND AIRE-R139X

TCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATT
GGCCATTGCATACGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATA
TGACCGCCATGTTGGCATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT
AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCT
GACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCA
ATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGT
ACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCG
CCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTA
TTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCG
GTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC
ACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTGCGATCGCCCGCCCCGTTGACGC
AAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGT
CAGATCACTAGAAGCTTTATTGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGTGC
TTCTGACACAACAGTCTCGAACTTAAGCTGCAGTGACTCTCTTAAGGTAGCCTTGCAGAAGT
TGGTCGTGAGGCACTGGGCAGGTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATA
GAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTA
CTGACATCCACTTTGCCTTTCTCTCCACAGGTGTCCACTCCCAGTTCAATTACAGCTCTTAA
GGCTAGAGTACTTAATACGACTCACTATAGGCTAGCCAGCTTGAAGCAAGCCTCCTGAAAGA
TGAAGCTACTGTCTTCTATCGAACAAGCATGCGATATTTGCCGACTTAAAAAGCTCAAGTGC
TCCAAAGAAAAACCGAAGTGCGCCAAGTGTCTGAAGAACAACTGGGAGTGTCGCTACTCTCC
CAAAACCAAAAGGTCTCCGCTGACTAGGGCACATCTGACAGAAGTGGAATCAAGGCTAGAAA
GACTGGAACAGCTATTTCTACTGATTTTTCCTCGAGAAGACCTTGACATGATTTTGAAAATG
GATTCTTTACAGGATATAAAAGCATTGTTAACAGGATTATTTGTACAAGATAATGTGAATAA
AGATGCCGTCACAGATAGATTGGCTTCAGTGGAGACTGATATGCCTCTAACATTGAGACAGC
ATAGAATAAGTGCGACATCATCATCGGAAGAGAGTAGTAACAAAGGTCAAAGACAGTTGACT
GTATCGCCGGAATTCCCGGGGATCCGTCGACTTGACGCGTCTATGGCGACGGACGCGGCGCT
ACGCCGGCTTCTGAGGCTGCACCGCACGGAGATCGCGGTGGCCGTGGACAGCGCCTTCCCAC
TGCTGCACGCGCTGGCTGACCACGACGTGGTCCCCGAGGACAAGTTTCAGGAGACGCTTCAT
CTGAAGGAAAAGGAGGGCTGCCCCCAGGCCTTCCACGCCCTCCTGTCCTGGCTGCTGACCCA
GGACTCCACAGCCATCCTGGACTTCTGGAGGGTGCTGTTCAAGGACTACAACCTGGAGCGCT
ATGGCCGGCTGCAGCCCATCCTGGACAGCTTCCCCAAAGATGTGGACCTCAGCCAGCCCCGG
AAGGGGAGGAAGCCCCCGGCCGTCCCCAAGGCTTTGGTACCGCCACCCAGACTCCCCACCAA
GAGGAAGGCCTCAGAAGAGGCTTGAGCTGCCGCGCCAGCAGCCCTGACTCCAAGGGGCACCG
CCAGCCCAGGCTCTCAACTGAAGGCCAAGCCCCCCAAGAAGCCGGAGAGCAGCGCAGAGCAG
CAGCGCCTTCCACTCGGGAACGGGATTCAGACCATGTCAGCTTCAGTCCAGAGAGCTGTGGC
CATGTCCTCCGGGGACGTCCCGGGAGCCCGAGGGGCCGTGGAGGGGATCCTCATCCAGCAGG
TGTTTGAGTCAGGCGGCTCCAAGAAGTGCATCCAGGTTGGCGGGGAGTTCTACACTCCCAGC
AAGTTCGAAGACTCCGGCAGTGGGAAGAACAAGGCCCGCAGCAGCAGTGGCCCGAAGCCTCT
GGTTCGAGCCAAGGGAGCCCAGGGCGCTGCCCCCGGTGGAGGTGAGGCTAGGCTGGGCCAGC
AGGGCAGCGTTCCCGCCCCTCTGGCCCTCCCCAGTGACCCCCAGCTCCACCAGAAGAATGAG
GACGAGTGTGCCGTGTGTCGGGACGGCGGGGAGCTCATCTGCTGTGACGGCTGCCCTCGGGC
CTTCCACCTGGCCTGCCTGTCCCCTCCGCTCCGGGAGATCCCCAGTGGGACCTGGAGGTGCT
CCAGCTGCCTGCAGGCAACAGTCCAGGAGGTGCAGCCCCGGGCAGAGGAGCCCCGGCCCCAG
GAGCCACCCGTGGAGACCCCGCTCCCCCCGGGGCTTAGGTCGGCGGGAGAGGAGGTAAGAGG
TCCACCTGGGGAACCCCTAGCCGGCATGGACACGACTCTTGTCTACAAGCACCTGCCGGCTC
CGCCTTCTGCAGCCCCGCTGCCAGGGCTGGACTCCTCGGCCCTGCACCCCCTACTGTGTGTG

GGTCCTGAGGGTCAGCAGAACCTGGCTCCTGGTGCGCGTTGCGGGGTGTGCGGAGATGGTAC
GGACGTGCTGCGGTGTACTCACTGCGCCGCTGCCTTCCACTGGCGCTGCCACTTCCCAGCCG
GCACCTCCCGGCCCGGGACGGGCCTGCGCTGCAGATCCTGCTCAGGAGACGTGACCCCAGCC
CCTGTGGAGGGGGTGCTGGCCCCCAGCCCCGCCCGCCTGGCCCCTGGGCCTGCCAAGGATGA
CACTGCCAGTCACGAGCCCGCTCTGCACAGGGATGACCTGGAGTCCCTTCTGAGCGAGCACA
CCTTCGATGGCATCCTGCAGTGGGCCATCCAGAGCATGGCCCGTCCGGCGGCCCCCTTCCCC
TCCGGGGGGTGGAGGCTCTGAGCAGAAACTCATCTCAGAAGAGGATCTGTAAGCGGCCGCAGG
TACCTGAATAACTAAGGCCGCTTCCCTTTAGTGAGGGTTAATGCTTCGAGCAGACATGATAA
GATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGT
GAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAA
CAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGATGTGGGAGGTTTTTTAAAGCA
AGTAAAACCTCTACAAATGTGGTAAAATCCGATAAGGATCGATCCGGGCTGGCGTAATAGCG
AAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGC
CCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTT
GCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCACGCCACGTTCGCCGG
CTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGC
ACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAG
ACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAAC
TGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTT
CGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATA
TTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACAC
CGCATACGCGGATCTTCCGTACCTTCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAG
TTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAA
TTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCA
TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACT
CCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGC
CGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAG
GCTTTTGCAAAAAGCTTGATTCTTCTGACACAACAGTCTCGAACTTAAGGCTAGAATTCTGG
CTACAGGTAAGCGCCCCTAAAATCCCTTTGGGCACAATGTGTCCTGAGGGGAGAGGCAGCGA
CCTGTAGATGGGACGGGGGCACTAACCCTCAGGTTTGGGGCTTCTGAATGTGAGTATCGCCA
TGTAAGCCCAGTATTTGGCCAATCTCAGAAAGCTCCTGGTCCCTGGAGGGATGGAGAGAGAA
AAACAAACAGCTCCTGGAGCAGGGAGAGTGCTGGCCTCTTGCTCTCCGGCTCCCTCTGTTGC
CCTCTGGTTTCTCCCCAGGCTCCCGGACGTCCTCTAGCCACCATGACTTCGAAAGTTTATGA
TCCAGAACAAAGGAAACGGATGATAACTGGTCCGCAGTGGTGGGCCAGATGTAAACAAATGA
ATGTTCTTGATTCATTTATTAATTATTATGATTCAGAAAAACATGCAGAAAATGCTGTTATT
TTTTTACATGGTAACGCGGCCTCTTCTTATTTATGGCGACATGTTGTGCCACATATTGAGCC
AGTAGCGCGGTGTATTATACCAGACCTTATTGGTATGGGCAAATCAGGCAAATCTGGTAATG
GTTCTTATAGGTTACTTGATCATTACAAATATCTTACTGCATGGTTTGAACTTCTTAATTTA
CCAAAGAAGATCATTTTTGTCGGCCATGATTGGGGTGCTTGTTTGGCATTTCATTATAGCTA
TGAGCATCAAGATAAGATCAAAGCAATAGTTCACGCTGAAAGTGTAGTAGATGTGATTGAAT
CATGGGATGAATGGCCTGATATTGAAGAAGATATTGCGTTGATCAAATCTGAAGAAGGAGAA
AAAATGGTTTTGGAGAATAACTTCTTCGTGGAAACCATGTTGCCATCAAAAATCATGAGAAA
GTTAGAACCAGAAGAATTTGCAGCATATCTTGAACCATTCAAAGAGAAAGGTGAAGTTCGTC
GTCCAACATTATCATGGCCTCGTGAAATCCCGTTAGTAAAAGGTGGTAAACCTGACGTTGTA
CAAATTGTTAGGAATTATAATGCTTATCTACGTGCAAGTGATGATTTACCAAAAATGTTTAT
TGAATCGGACCCAGGATTCTTTTCCAATGCTATTGTTGAAGGTGCCAAGAAGTTTCCTAATA
CTGAATTTGTCAAAGTAAAAGGTCTTCATTTTTCGCAAGAAGATGCACCTGATGAAATGGGA

AAATATATCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAACAATAATTCTAGCCCTGAAT
AAGTGATAATAAGCGGATGAATGGCAGAAATTCGTCGAAGCGCAATAAAATATCTTTATTTT
CATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGATAGCGATAAGGATCGGAAGATCCGCG
TATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCG
CCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGC
TGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGA
GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCT
TAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTA
AATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATT
GAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCA
TTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTT
TTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTA
TTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGA
CTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAAT
TATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATC
GGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGA
TCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTG
TAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGG
CAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCT
TCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCA
TTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGT
CAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA
TTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTT
AATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGT
GAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC
TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTT
GTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAG
ATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGC
ACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT
CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGA
ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCT
ACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGG
TAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTAT
CTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTC
AGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT
GCTGGCCTTTTGCTCACATGGCTCGACAGATCT

>pBIND AIRE- R257X

TCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATT
GGCCATTGCATACGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATA
TGACCGCCATGTTGGCATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT
AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCT
GACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCA
ATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGT
ACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCG
CCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTA

TTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCG
GTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC
ACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTGCGATCGCCCGCCCCGTTGACGC
AAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGT
CAGATCACTAGAAGCTTTATTGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGTGC
TTCTGACACAACAGTCTCGAACTTAAGCTGCAGTGACTCTCTTAAGGTAGCCTTGCAGAAGT
TGGTCGTGAGGCACTGGGCAGGTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATA
GAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTA
CTGACATCCACTTTGCCTTTCTCTCCACAGGTGTCCACTCCCAGTTCAATTACAGCTCTTAA
GGCTAGAGTACTTAATACGACTCACTATAGGCTAGCCAGCTTGAAGCAAGCCTCCTGAAAGA
TGAAGCTACTGTCTTCTATCGAACAAGCATGCGATATTTGCCGACTTAAAAAGCTCAAGTGC
TCCAAAGAAAAACCGAAGTGCGCCAAGTGTCTGAAGAACAACTGGGAGTGTCGCTACTCTCC
CAAAACCAAAAGGTCTCCGCTGACTAGGGCACATCTGACAGAAGTGGAATCAAGGCTAGAAA
GACTGGAACAGCTATTTCTACTGATTTTTCCTCGAGAAGACCTTGACATGATTTTGAAAATG
GATTCTTTACAGGATATAAAAGCATTGTTAACAGGATTATTTGTACAAGATAATGTGAATAA
AGATGCCGTCACAGATAGATTGGCTTCAGTGGAGACTGATATGCCTCTAACATTGAGACAGC
ATAGAATAAGTGCGACATCATCATCGGAAGAGAGTAGTAACAAAGGTCAAAGACAGTTGACT
GTATCGCCGGAATTCCCGGGGATCCGTCGACTTGACGCGTCTATGGCGACGGACGCGGCGCT
ACGCCGGCTTCTGAGGCTGCACCGCACGGAGATCGCGGTGGCCGTGGACAGCGCCTTCCCAC
TGCTGCACGCGCTGGCTGACCACGACGTGGTCCCCGAGGACAAGTTTCAGGAGACGCTTCAT
CTGAAGGAAAAGGAGGGCTGCCCCCAGGCCTTCCACGCCCTCCTGTCCTGGCTGCTGACCCA
GGACTCCACAGCCATCCTGGACTTCTGGAGGGTGCTGTTCAAGGACTACAACCTGGAGCGCT
ATGGCCGGCTGCAGCCCATCCTGGACAGCTTCCCCAAAGATGTGGACCTCAGCCAGCCCCGG
AAGGGGAGGAAGCCCCCGGCCGTCCCCAAGGCTTTGGTACCGCCACCCAGACTCCCCACCAA
GAGGAAGGCCTCAGAAGAGGCTCGAGCTGCCGCGCCAGCAGCCCTGACTCCAAGGGGCACCG
CCAGCCCAGGCTCTCAACTGAAGGCCAAGCCCCCCAAGAAGCCGGAGAGCAGCGCAGAGCAG
CAGCGCCTTCCACTCGGGAACGGGATTCAGACCATGTCAGCTTCAGTCCAGAGAGCTGTGGC
CATGTCCTCCGGGGACGTCCCGGGAGCCCGAGGGGCCGTGGAGGGGATCCTCATCCAGCAGG
TGTTTGAGTCAGGCGGCTCCAAGAAGTGCATCCAGGTTGGCGGGGAGTTCTACACTCCCAGC
AAGTTCGAAGACTCCGGCAGTGGGAAGAACAAGGCCCGCAGCAGCAGTGGCCCGAAGCCTCT
GGTTTGAGCCAAGGGAGCCCAGGGCGCTGCCCCCGGTGGAGGTGAGGCTAGGCTGGGCCAGC
AGGGCAGCGTTCCCGCCCCTCTGGCCCTCCCCAGTGACCCCCAGCTCCACCAGAAGAATGAG
GACGAGTGTGCCGTGTGTCGGGACGGCGGGGAGCTCATCTGCTGTGACGGCTGCCCTCGGGC
CTTCCACCTGGCCTGCCTGTCCCCTCCGCTCCGGGAGATCCCCAGTGGGACCTGGAGGTGCT
CCAGCTGCCTGCAGGCAACAGTCCAGGAGGTGCAGCCCCGGGCAGAGGAGCCCCGGCCCCAG
GAGCCACCCGTGGAGACCCCGCTCCCCCCGGGGCTTAGGTCGGCGGGAGAGGAGGTAAGAGG
TCCACCTGGGGAACCCCTAGCCGGCATGGACACGACTCTTGTCTACAAGCACCTGCCGGCTC
CGCCTTCTGCAGCCCCGCTGCCAGGGCTGGACTCCTCGGCCCTGCACCCCCTACTGTGTGTG
GGTCCTGAGGGTCAGCAGAACCTGGCTCCTGGTGCGCGTTGCGGGGTGTGCGGAGATGGTAC
GGACGTGCTGCGGTGTACTCACTGCGCCGCTGCCTTCCACTGGCGCTGCCACTTCCCAGCCG
GCACCTCCCGGCCCGGGACGGGCCTGCGCTGCAGATCCTGCTCAGGAGACGTGACCCCAGCC
CCTGTGGAGGGGGTGCTGGCCCCCAGCCCCGCCCGCCTGGCCCCTGGGCCTGCCAAGGATGA
CACTGCCAGTCACGAGCCCGCTCTGCACAGGGATGACCTGGAGTCCCTTCTGAGCGAGCACA
CCTTCGATGGCATCCTGCAGTGGGCCATCCAGAGCATGGCCCGTCCGGCGGCCCCCTTCCCC
TCCGGGGGGTGGAGGCTCTGAGCAGAAACTCATCTCAGAAGAGGATCTGTAAGCGGCCGCAGG
TACCTGAATAACTAAGGCCGCTTCCCTTTAGTGAGGGTTAATGCTTCGAGCAGACATGATAA
GATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGT

GAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAA
CAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGATGTGGGAGGTTTTTTAAAGCA
AGTAAAACCTCTACAAATGTGGTAAAATCCGATAAGGATCGATCCGGGCTGGCGTAATAGCG
AAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGC
CCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTT
GCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCACGCCACGTTCGCCGG
CTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGC
ACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAG
ACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAAC
TGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTT
CGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATA
TTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACAC
CGCATACGCGGATCTTCCGTACCTTCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAG
TTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAA
TTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCA
TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACT
CCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGC
CGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAG
GCTTTTGCAAAAAGCTTGATTCTTCTGACACAACAGTCTCGAACTTAAGGCTAGAATTCTGG
CTACAGGTAAGCGCCCCTAAAATCCCTTTGGGCACAATGTGTCCTGAGGGGAGAGGCAGCGA
CCTGTAGATGGGACGGGGGCACTAACCCTCAGGTTTGGGGCTTCTGAATGTGAGTATCGCCA
TGTAAGCCCAGTATTTGGCCAATCTCAGAAAGCTCCTGGTCCCTGGAGGGATGGAGAGAGAA
AAACAAACAGCTCCTGGAGCAGGGAGAGTGCTGGCCTCTTGCTCTCCGGCTCCCTCTGTTGC
CCTCTGGTTTCTCCCCAGGCTCCCGGACGTCCTCTAGCCACCATGACTTCGAAAGTTTATGA
TCCAGAACAAAGGAAACGGATGATAACTGGTCCGCAGTGGTGGGCCAGATGTAAACAAATGA
ATGTTCTTGATTCATTTATTAATTATTATGATTCAGAAAAACATGCAGAAAATGCTGTTATT
TTTTTACATGGTAACGCGGCCTCTTCTTATTTATGGCGACATGTTGTGCCACATATTGAGCC
AGTAGCGCGGTGTATTATACCAGACCTTATTGGTATGGGCAAATCAGGCAAATCTGGTAATG
GTTCTTATAGGTTACTTGATCATTACAAATATCTTACTGCATGGTTTGAACTTCTTAATTTA
CCAAAGAAGATCATTTTTGTCGGCCATGATTGGGGTGCTTGTTTGGCATTTCATTATAGCTA
TGAGCATCAAGATAAGATCAAAGCAATAGTTCACGCTGAAAGTGTAGTAGATGTGATTGAAT
CATGGGATGAATGGCCTGATATTGAAGAAGATATTGCGTTGATCAAATCTGAAGAAGGAGAA
AAAATGGTTTTGGAGAATAACTTCTTCGTGGAAACCATGTTGCCATCAAAAATCATGAGAAA
GTTAGAACCAGAAGAATTTGCAGCATATCTTGAACCATTCAAAGAGAAAGGTGAAGTTCGTC
GTCCAACATTATCATGGCCTCGTGAAATCCCGTTAGTAAAAGGTGGTAAACCTGACGTTGTA
CAAATTGTTAGGAATTATAATGCTTATCTACGTGCAAGTGATGATTTACCAAAAATGTTTAT
TGAATCGGACCCAGGATTCTTTTCCAATGCTATTGTTGAAGGTGCCAAGAAGTTTCCTAATA
CTGAATTTGTCAAAGTAAAAGGTCTTCATTTTTCGCAAGAAGATGCACCTGATGAAATGGGA
AAATATATCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAACAATAATTCTAGCCCTGAAT
AAGTGATAATAAGCGGATGAATGGCAGAAATTCGTCGAAGCGCAATAAAATATCTTTATTTT
CATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGATAGCGATAAGGATCGGAAGATCCGCG
TATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCG
CCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGC
TGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGA
GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCT
TAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTA
AATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATT

GAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCA
TTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTT
TTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTA
TTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGA
CTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAAT
TATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATC
GGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGA
TCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTG
TAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGG
CAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCT
TCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCA
TTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGT
CAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA
TTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTT
AATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGT
GAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC
TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTT
GTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAG
ATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGC
ACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT
CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGA
ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCT
ACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGG
TAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTAT
CTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTC
AGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT
GCTGGCCTTTTGCTCACATGGCTCGACAGATCT

>pBIND AIRE-G228W

TCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATT
GGCCATTGCATACGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATA
TGACCGCCATGTTGGCATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT
AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCT
GACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCA
ATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGT
ACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCG
CCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTA
TTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCG
GTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC
ACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTGCGATCGCCCGCCCCGTTGACGC
AAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGT
CAGATCACTAGAAGCTTTATTGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGTGC
TTCTGACACAACAGTCTCGAACTTAAGCTGCAGTGACTCTCTTAAGGTAGCCTTGCAGAAGT
TGGTCGTGAGGCACTGGGCAGGTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATA
GAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTA
CTGACATCCACTTTGCCTTTCTCTCCACAGGTGTCCACTCCCAGTTCAATTACAGCTCTTAA

GGCTAGAGTACTTAATACGACTCACTATAGGCTAGCCAGCTTGAAGCAAGCCTCCTGAAAGA
TGAAGCTACTGTCTTCTATCGAACAAGCATGCGATATTTGCCGACTTAAAAAGCTCAAGTGC
TCCAAAGAAAAACCGAAGTGCGCCAAGTGTCTGAAGAACAACTGGGAGTGTCGCTACTCTCC
CAAAACCAAAAGGTCTCCGCTGACTAGGGCACATCTGACAGAAGTGGAATCAAGGCTAGAAA
GACTGGAACAGCTATTTCTACTGATTTTTCCTCGAGAAGACCTTGACATGATTTTGAAAATG
GATTCTTTACAGGATATAAAAGCATTGTTAACAGGATTATTTGTACAAGATAATGTGAATAA
AGATGCCGTCACAGATAGATTGGCTTCAGTGGAGACTGATATGCCTCTAACATTGAGACAGC
ATAGAATAAGTGCGACATCATCATCGGAAGAGAGTAGTAACAAAGGTCAAAGACAGTTGACT
GTATCGCCGGAATTCCCGGGGATCCGTCGACTTGACGCGTCTATGGCGACGGACGCGGCGCT
ACGCCGGCTTCTGAGGCTGCACCGCACGGAGATCGCGGTGGCCGTGGACAGCGCCTTCCCAC
TGCTGCACGCGCTGGCTGACCACGACGTGGTCCCCGAGGACAAGTTTCAGGAGACGCTTCAT
CTGAAGGAAAAGGAGGGCTGCCCCCAGGCCTTCCACGCCCTCCTGTCCTGGCTGCTGACCCA
GGACTCCACAGCCATCCTGGACTTCTGGAGGGTGCTGTTCAAGGACTACAACCTGGAGCGCT
ATGGCCGGCTGCAGCCCATCCTGGACAGCTTCCCCAAAGATGTGGACCTCAGCCAGCCCCGG
AAGGGGAGGAAGCCCCCGGCCGTCCCCAAGGCTTTGGTACCGCCACCCAGACTCCCCACCAA
GAGGAAGGCCTCAGAAGAGGCTCGAGCTGCCGCGCCAGCAGCCCTGACTCCAAGGGGCACCG
CCAGCCCAGGCTCTCAACTGAAGGCCAAGCCCCCCAAGAAGCCGGAGAGCAGCGCAGAGCAG
CAGCGCCTTCCACTCGGGAACGGGATTCAGACCATGTCAGCTTCAGTCCAGAGAGCTGTGGC
CATGTCCTCCGGGGACGTCCCGGGAGCCCGAGGGGCCGTGGAGGGGATCCTCATCCAGCAGG
TGTTTGAGTCAGGCGGCTCCAAGAAGTGCATCCAGGTTGGCTGGGAGTTCTACACTCCCAGC
AAGTTCGAAGACTCCGGCAGTGGGAAGAACAAGGCCCGCAGCAGCAGTGGCCCGAAGCCTCT
GGTTCGAGCCAAGGGAGCCCAGGGCGCTGCCCCCGGTGGAGGTGAGGCTAGGCTGGGCCAGC
AGGGCAGCGTTCCCGCCCCTCTGGCCCTCCCCAGTGACCCCCAGCTCCACCAGAAGAATGAG
GACGAGTGTGCCGTGTGTCGGGACGGCGGGGAGCTCATCTGCTGTGACGGCTGCCCTCGGGC
CTTCCACCTGGCCTGCCTGTCCCCTCCGCTCCGGGAGATCCCCAGTGGGACCTGGAGGTGCT
CCAGCTGCCTGCAGGCAACAGTCCAGGAGGTGCAGCCCCGGGCAGAGGAGCCCCGGCCCCAG
GAGCCACCCGTGGAGACCCCGCTCCCCCCGGGGCTTAGGTCGGCGGGAGAGGAGGTAAGAGG
TCCACCTGGGGAACCCCTAGCCGGCATGGACACGACTCTTGTCTACAAGCACCTGCCGGCTC
CGCCTTCTGCAGCCCCGCTGCCAGGGCTGGACTCCTCGGCCCTGCACCCCCTACTGTGTGTG
GGTCCTGAGGGTCAGCAGAACCTGGCTCCTGGTGCGCGTTGCGGGGTGTGCGGAGATGGTAC
GGACGTGCTGCGGTGTACTCACTGCGCCGCTGCCTTCCACTGGCGCTGCCACTTCCCAGCCG
GCACCTCCCGGCCCGGGACGGGCCTGCGCTGCAGATCCTGCTCAGGAGACGTGACCCCAGCC
CCTGTGGAGGGGGTGCTGGCCCCCAGCCCCGCCCGCCTGGCCCCTGGGCCTGCCAAGGATGA
CACTGCCAGTCACGAGCCCGCTCTGCACAGGGATGACCTGGAGTCCCTTCTGAGCGAGCACA
CCTTCGATGGCATCCTGCAGTGGGCCATCCAGAGCATGGCCCGTCCGGCGGCCCCCTTCCCC
TCCGGGGGTGGAGGCTCTGAGCAGAAACTCATCTCAGAAGAGGATCTGTAAGCGGCCGCAGG
TACCTGAATAACTAAGGCCGCTTCCCTTTAGTGAGGGTTAATGCTTCGAGCAGACATGATAA
GATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGT
GAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAA
CAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGATGTGGGAGGTTTTTTAAAGCA
AGTAAAACCTCTACAAATGTGGTAAAATCCGATAAGGATCGATCCGGGCTGGCGTAATAGCG
AAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGC
CCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTT
GCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCACGCCACGTTCGCCGG
CTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGC
ACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAG
ACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAAC

TGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTT
CGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATA
TTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACAC
CGCATACGCGGATCTTCCGTACCTTCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAG
TTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAA
TTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCA
TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACT
CCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGC
CGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAG
GCTTTTGCAAAAAGCTTGATTCTTCTGACACAACAGTCTCGAACTTAAGGCTAGAATTCTGG
CTACAGGTAAGCGCCCCTAAAATCCCTTTGGGCACAATGTGTCCTGAGGGGAGAGGCAGCGA
CCTGTAGATGGGACGGGGGCACTAACCCTCAGGTTTGGGGCTTCTGAATGTGAGTATCGCCA
TGTAAGCCCAGTATTTGGCCAATCTCAGAAAGCTCCTGGTCCCTGGAGGGATGGAGAGAGAA
AAACAAACAGCTCCTGGAGCAGGGAGAGTGCTGGCCTCTTGCTCTCCGGCTCCCTCTGTTGC
CCTCTGGTTTCTCCCCAGGCTCCCGGACGTCCTCTAGCCACCATGACTTCGAAAGTTTATGA
TCCAGAACAAAGGAAACGGATGATAACTGGTCCGCAGTGGTGGGCCAGATGTAAACAAATGA
ATGTTCTTGATTCATTTATTAATTATTATGATTCAGAAAAACATGCAGAAAATGCTGTTATT
TTTTTACATGGTAACGCGGCCTCTTCTTATTTATGGCGACATGTTGTGCCACATATTGAGCC
AGTAGCGCGGTGTATTATACCAGACCTTATTGGTATGGGCAAATCAGGCAAATCTGGTAATG
GTTCTTATAGGTTACTTGATCATTACAAATATCTTACTGCATGGTTTGAACTTCTTAATTTA
CCAAAGAAGATCATTTTTGTCGGCCATGATTGGGGTGCTTGTTTGGCATTTCATTATAGCTA
TGAGCATCAAGATAAGATCAAAGCAATAGTTCACGCTGAAAGTGTAGTAGATGTGATTGAAT
CATGGGATGAATGGCCTGATATTGAAGAAGATATTGCGTTGATCAAATCTGAAGAAGGAGAA
AAAATGGTTTTGGAGAATAACTTCTTCGTGGAAACCATGTTGCCATCAAAAATCATGAGAAA
GTTAGAACCAGAAGAATTTGCAGCATATCTTGAACCATTCAAAGAGAAAGGTGAAGTTCGTC
GTCCAACATTATCATGGCCTCGTGAAATCCCGTTAGTAAAAGGTGGTAAACCTGACGTTGTA
CAAATTGTTAGGAATTATAATGCTTATCTACGTGCAAGTGATGATTTACCAAAAATGTTTAT
TGAATCGGACCCAGGATTCTTTTCCAATGCTATTGTTGAAGGTGCCAAGAAGTTTCCTAATA
CTGAATTTGTCAAAGTAAAAGGTCTTCATTTTTCGCAAGAAGATGCACCTGATGAAATGGGA
AAATATATCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAACAATAATTCTAGCCCTGAAT
AAGTGATAATAAGCGGATGAATGGCAGAAATTCGTCGAAGCGCAATAAAATATCTTTATTTT
CATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGATAGCGATAAGGATCGGAAGATCCGCG
TATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCG
CCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGC
TGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGA
GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCT
TAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTA
AATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATT
GAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCA
TTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA
GTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTT
TTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTA
TTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGA
CTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAAT
TATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATC
GGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGA
TCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTG

TAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGG
CAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCT
TCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCA
TTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGT
CAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA
TTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTT
AATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGT
GAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC
TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTT
GTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAG
ATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGC
ACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT
CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGA
ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCT
ACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGG
TAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTAT
CTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTC
AGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT
GCTGGCCTTTTGCTCACATGGCTCGACAGATCT

Yellow：Restriction enzyme

Turquoise : Mutation

Red: Insert

Green: T7 EEV primer for sequencing

8.5 Maps of plasmids

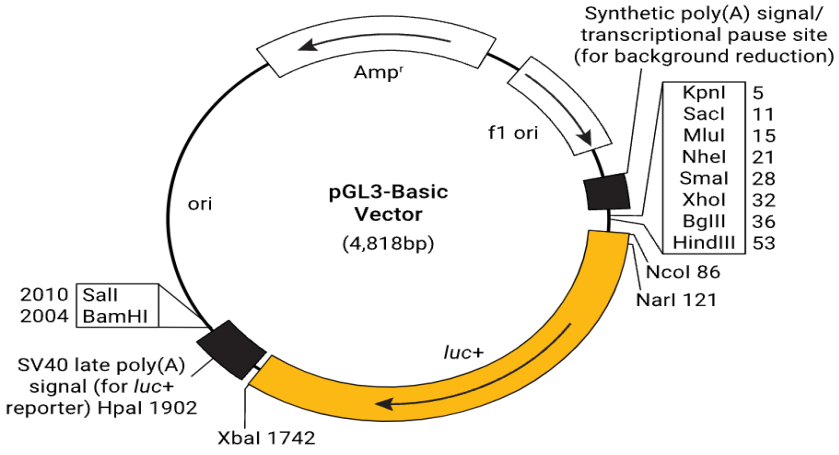

Figure 8.1: pGL3 Luciferase Reporter Vectors: pGL3 basic vector is a Luciferase Reporter Vector that was used to test the activity of the promoters of candidate genes with or without AIRE.

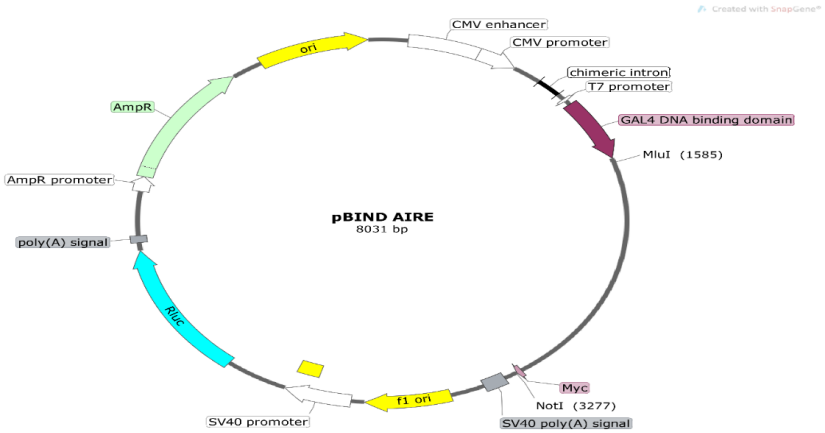

Figure 8.2: pBIND AIRE plasmid. It is expressed under the control of the CMV promoter and to the upstream Gal4 DBD. The region with turquoise colour is Renilla luciferase coding region and the origin of replication is shown in yellow. The plasmid is resistance to ampicillin. This plasmid is provided by the lab and was used in dual luciferase assay.

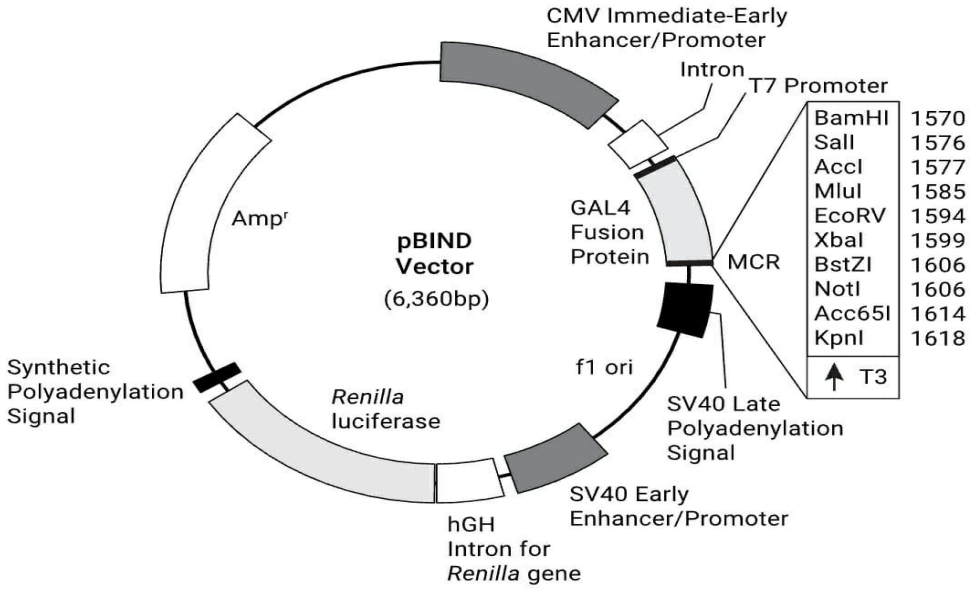

Figure 8.3: pBIND plasmid. It is expressed under the control of the CMV promoter and to the upstream Gal4 DBD. The region down the vector with light grey colour is Renilla luciferase coding region. The plasmid is resistance to ampicillin.

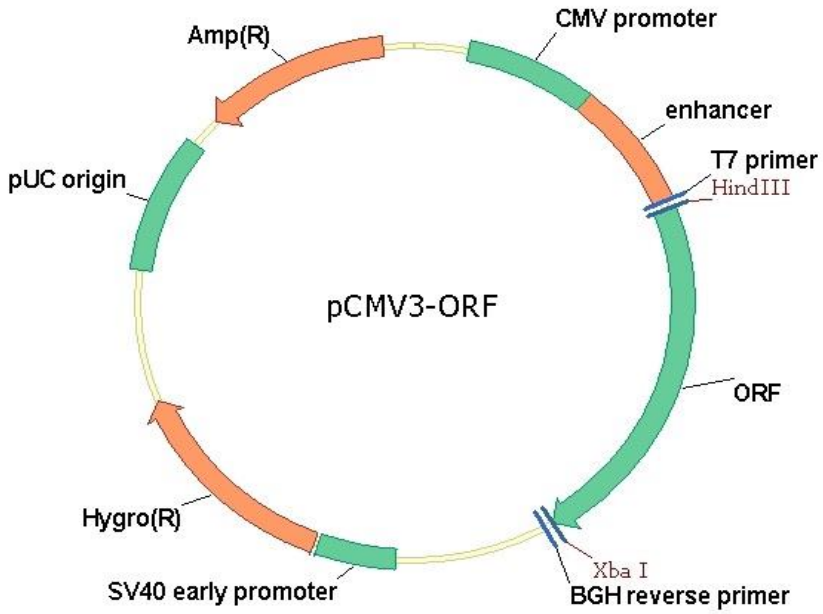

Figure 8.4: AIRE C-Myc: Full length AIRE including the c-cmyc tag was inserted in the multiple cloning site at the HindIII and Xba I restriction sites and and is under the control of the CMV promoter. The Plasmid is resistance to ampicillin. This was used for ChIP-seq and RT-qPCR.

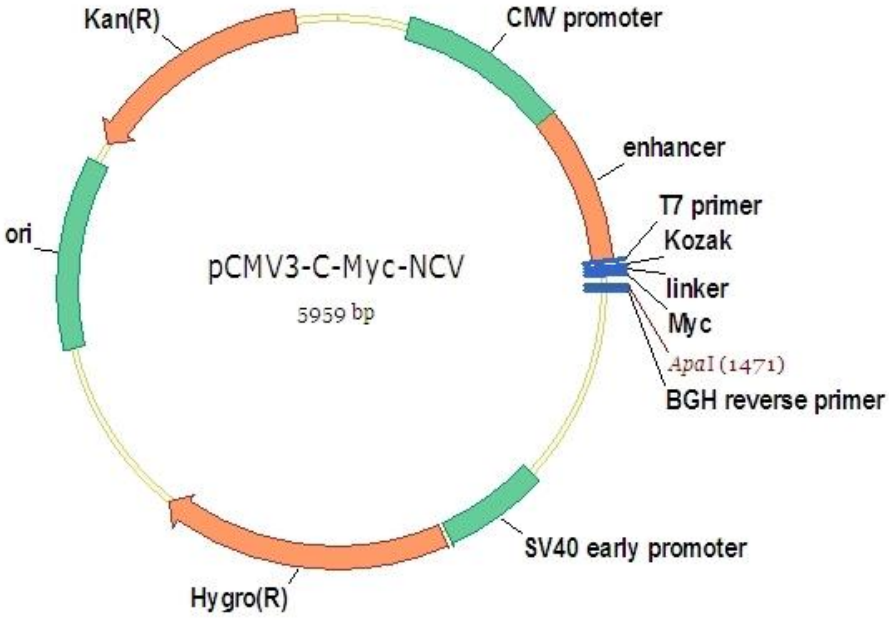

Figure 8.5: Pcmv3-c-myc negative control vector: Pcmv3-C-Myc negative control was used as a negative control for ChIP-seq and RT-qPCR.