Norwegian University of Life Sciences

Master's Thesis 2020 60 ECTS

Faculty of Chemistry, Biotechnology, and Food Science (KBM)

ChIP-seq analysis of AIRE, FEZF2, and DEAF1 transcription factors in human thymus tissue

Dina Ruud Aronsen

Biotechnology – molecular biology

ChIP-seq analysis of AIRE, FEZF2 and DEAF1 transcription factors in human thymus tissue

Dina Ruud Aronsen

Supervisors:

Prof. Dr. Benedicte A. Lie (Main supervisor)
Dr. Teodora Ribarska
Assoc. Prof. Siv Kjølsrud Bøhn

Master thesis

Department of Medical Genetics, Oslo University Hospital
Faculty of Chemistry, Biotechnology and Food science
Norwegian University of Life Sciences

June 2020

©Dina Ruud Aronsen

2020

ChIP-seq analysis of AIRE, FEZF2 and DEAF1 transcription factors in human thymus tissue

https://nmbu.brage.unit.no

Acknowledgements

The work of this thesis was carried out in collaboration with the Department of Medical Genetics, at Oslo University Hospital (OUS), as part of a Master´s Degree in Biotechnology at the Norwegian University of Life Sciences (NMBU), the main affiliation being the Faculty of Chemistry, Biotechnology and Food Science (KBM), from August 2019 to June 2020.

Foremost, I would like to thank my main supervisor Benedicte A. Lie for the opportunity to be a part of her thymus group and conduct such interesting research. I am grateful for her guidance, insightful advice and the enthusiasm she has shown for my work. I am thankful for the outstanding work of my co-supervisor Teodora Ribarska, Postdoc. She guided and encouraged me through both the writing process and laboratory work. Her invaluable advice and knowledge have greatly contributed to my understanding throughout this process. Additionally, her laboratory expertise and guidance have helped me improve my laboratory skills and expanded my expertise by introducing me to several laboratory techniques. Additionally, I would like to thank my supervisor at NMBU, Siv Kjølsrud Bøhn, for valuable feedback and guidance.

I would like to thank all past and present members of the Immgen-group for welcoming and including me and for providing advice along the way. I would like to extend an extra thank to the thymus-group for advice and for helping with laboratory work.

Lastly, I would like to thank my loved ones, who have supported me throughout the entire process. I will be grateful forever for your love and encouragement.

Oslo, June 2020

Dina Ruud Aronsen

Abstract

Autoimmune diseases develop as a consequence of incorrect establishment of self-tolerance. A key element in this process is the presentation of tissue-restricted antigens (TRA) by medullary thymic epithelial cells (mTECs) and antigen presenting cells (APCs) to the developing T-cells in the thymus. TRA expression is mostly driven by the transcription factors Autoimmune regulator (AIRE) and Forebrain Embryonic Zinc Finger-Like Protein 2 (FEZF2) in mTECs. Deformed Epidermal Autoregulatory Factor 1 Homolog (DEAF1), present in all thymic APCs, is suggested to also control TRA expression as it does in the lymph node. The purpose of this study was to establish and optimise chromatin immunoprecipitation (ChIP) method for use in human freshly frozen thymus tissues and apply it to determine the genomic binding sites of FEZF2, AIRE and DEAF1 in the infant human thymus in order to better understand the molecular mechanism of T-cell self-tolerance establishment.

Aiming to reach the highest possible sensitivity and specificity of the enrichment, we optimised the ChIP procedure using homogenised frozen human thymic tissue (*n*=3) by testing the parameters: fixation (single or double crosslinking), fragmentation by sonication, antibody type and amount, and immunoprecipitation conditions.

The optimal ChIP conditions were found to be: 1) Fixation for 5 min with 1 % formaldehyde and 2.5 mM disuccinimidyl glutarate with 20-30 mg tissue powder ; 2) Sonication to 300-500 bp average DNA fragment size; 3) IP with 5 µg anti-AIRE (GeneTex), 5 µg anti-DEAF1 (LSBio), and 1 µg FEZF2 (Abcam) per 100 µl chromatin (out of 2 mL chromatin isolated from 20-30 mg tissue powder) using 30 µl protein G magnetic beads (Dynabeads) per IP reaction. These condition provided the highest specific recovery of target regions for the three transcription factors, measured by qPCR. The optimal conditions were used in a pilot ChIP-seq experiment to locate binding sites of the transcription factors. Analysis of AIRE, FEZF2, and DEAF1 ChIP-seq data from human thymus and qPCR of ChIP using TEC and APC enriched cell sample isolated from half of a thymus showed low enrichment of target genes, which may be caused by scarcity of cells expressing AIRE, FEZF2, and DEAF1 in bulk thymus tissue or enriched for mTEC and APC thymus cell suspension. To assess the function of AIRE, FEZF2, and DEAF1 in mTECs and thymic APCs, further exploration using CUT&Tag or ChIP-seq will be needed using mTEC and APC enriched cell solution.

Table of Contents

Acknowledgements iv
Abstract v
Abbreviations ix
List of Figures xi
1 Introduction 1
1.1 The Immune System 1
1.1.1 Cells of the Immune System 1
1.1.2 The Innate Immune System 5
1.1.3 Human Leukocyte Antigen 5
1.1.4 The Adaptive Immune System 6
1.2 T-cell Development 8
1.2.1 Positive Selection 10
1.2.2 Negative Selection 11
1.2.3 Peripheral tolerance 13
1.2.4 Antigen-Presenting Cells in the Thymus 13
1.3 AIRE, FEZF2, and DEAF1 Functions in APCs 14
1.3.1 AIRE 14
1.3.2 FEZF2 15
1.3.3 DEAF1 17
1.4 Autoimmune Diseases 18
2 Aims 20
3 Materials and Methods 21
3.1 Thymus collection 21
3.2 Frozen thymus tissue preparation 21
3.3 Fresh thymus tissue dissociation 21
3.4 Chromatin Immunoprecipitation followed by Sequencing 23
3.4.1 Chromatin isolation 25
3.4.2 Chromatin Immunoprecipitation 27
3.4.3 Quantitative PCR (qPCR) 28
3.4.4 Library preparation and sequencing 31

3.5 ChIP DNA analysis 32
3.6 CUT&Tag 32
4 Results 36
4.1 AIRE, FEZF2, and DEAF1 expression in thymus 36
4.2 Chromatin Immunoprecipitation Optimisation 37
4.2.1 Crosslinking and Sonication Optimisation 37
4.2.2 Chromatin Immunoprecipitation optimisation using α-CTCF 39
4.2.3 Chromatin Immunoprecipitation optimisation with α-AIRE 41
4.2.4 Chromatin Immunoprecipitation optimisation with α-FEZF2 44
4.2.5 Chromatin immunoprecipitation optimisation with α-DEAF1 46
4.2.6 Test of Background Signal Produced During IP 47
4.3 AIRE, FEZF2, DEAF1 ChIP-seq in different Human Thymus Tissue samples 49
4.4 Chromatin Immunoprecipitation with TECs and APCs 59
4.5 CUT&Tag on Thymic cells 62
5 Discussion 63
5.1 ChIP optimisation 63
5.1.1 Antibody Selection 64
5.1.2 Target genes for AIRE, FEZF2 and DEAF1 used in qPCR 65
5.1.3 Transcription Factor Abundance 66
5.1.4 Biological Specificities of Transcription Factor Binding 67
5.1.5 ChIP-seq sequencing results 68
5.2 CUT&Tag 69
6 Conclusion 71
6.1 Future studies 71
References 73
Appendix I 79
Appendix II 85

Abbreviations

α	Anti
AID	Autoimmune diseases
AIRE	Autoimmune Repressor Protein
APC	Antigen Presenting Cell
APECED	Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy
BCR	B-cell Receptor
bp	Base pairs
BSA	Bovine Serum Albumin
C	Constant gene
°C	Degrees
Cat. No.	Catalogue number
CD	Cluster of differentiation
ChIP-seq	Chromatin Immunoprecipitation sequencing
ConA-beads	Concanavalin A-coated beads
Chr	Chromosome
CTCF	CCCTC-binding factor
cTEC	cortical Thymic Epithelial Cell
CUT&Tag	Cleavage under target and tagmentation
CZ	Santa Cruz Biotechnologies
D	Diversity gene
DEAF1	Deformed Epidermal Autoregulatory Factor Protein 1
DN	Double negative
DNA	Deoxyribonucleic Acid
DP	Double positive
DSG	Disuccinimidyl Glutarate
EDTA	Ethylenediaminetetraacetic Acid
ETP	Early Thymic Progenitors
Et. al.	And others (Et Alia)
eTAC	extrathymic AIRE-expressing cells
FA	Formaldehyde
FEZF2	Forebrain Embryonic Zinc Finger-Like Protein 2
FBS	Fetal Bovine Serum
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GTX	GeneTex
HLA	Human Leukocyte Antigen
HSC	Hematopoietic Stem Cells
IG	Immunoglobulin
IL	Interleukin
IP	immunoprecipitation
J	Joining gene
LNSC	Lymph Node Stromal Cells
MACS	Model-based Analysis of ChIP-Seq
mfold	Minimum fold-enrichment
MHC	Major Histocompatibility Complex

mTEC	medullary Thymic Epithelial Cell
N.D.	Not detected
NK	Natural Killer
NSC	Norwegian sequencing centre
REK	Regional Ethical Committee
RNA	Ribonucleic Acid
RT	Room Temperature
OUS	Oslo University Hospital
pA-Tn5	Tn5 transposase-protein A fusion protein
PAMP	Pathogen-Associated Molecular Pattern
PBS	Phosphate-Buffered Saline
PIC	Proteinase Inhibitor Cocktail
PRR	Pattern Recognition Receptors
qPCR	Quantitative Polymerase Chain Reaction
SDS	Sodium Dodecyl Sulphate
SP	Single positive
TCR	T-cell Receptor
TEC	Thymic Epithelial Cell
TF	Thermo Fisher
TRA	Tissue-Restricted Antigen
V	Variable gene
Yo	Years old

List of Figures

Figure 1.1 Schematic representation of hematopoiesis. 2
Figure 1.2. Cells of the immune system and their function. 4
Figure 1.3. Schematic drawing of an antibody and the genomic regions that make up immunoglobulins. 7
Figure 1.4. T-cell development in the thymus. 10
Figure 1.5. The T-cell repertoire is determined by positive and negative selection. 12
Figure 1.6. The ratio of TRAs regulated by Aire and Fezf2 in mice. 16
Figure 3.1. Schematic illustration of ChIP protocol steps. 23
Figure 4.1. DEAF1 (blue), fezf2 (orange), and AIRE (green) expression level assessed by RNA-sequencing of different APCs. 36
Figure 4.2. Fragment size of chromatin after sonication in three 5 minute intervals. 38
Figure 4.3. ChIP-qPCR with optimisation of crosslinking reagents with the α-CTCF antibody. 39
Figure 4.4. Optimisation of crosslinking reagents with α-AIRE (GTX) antibody by ChIP-qPCR. 40
Figure 4.5. ChIP-qPCR with optimisation of AIRE antibody and fixation method for AIRE. 42
Figure 4.6. ChIP-qPCR with optimisation of the amount of α-AIRE GTX antibody per IP. 43
Figure 4.7. ChIP-qPCR with optimisation for α-FEZF2 to test crosslinking reagents and amount of antibody per IP. 44
Figure 4.8. ChIP-qPCR with optimisation of α-DEAF1 antibody and the amount of antibody per IP. 46
Figure 4.9. AIRE, FEZF2 and DEAF1 enrichment of potential target and non-target genes. 47
Figure 4.10. Chip-qPCR with three biological samples. 49
Figure 4.11. Whole genome distribution of ChIP-seq peaks. 51
Figure 4.12. AIRE-, FEZF2- and DEAF1 ChIP-seq peaks accumulate in centromeric regions. 52
Figure 4.13. AIRE peak “MACS_peak_74” is a true peak with a high enrichment relative to IgG. 55
Figure 4.14. FEZF2 peak “MACS_peak_2”, is a true peak with a enrichment higher than IgG. 56
Figure 4.15. Aligned peak against human genome (hg38) viewing enriched area . 57
Figure 4.16. Tapestation image of APC chromatin sample after 8 minutes sonication. 58
Figure 4.17. ChIP-qPCR with APC enriched cell suspension. 60

1 Introduction

1.1 The Immune System

Immunity is resistance to disease, especially infectious diseases. The immune system is a collection of organs, tissue and cells that work together to defend the body against disease-causing microorganisms called pathogens, such as bacteria, viruses, parasites and other harmful microbes and proteins. The main organs of the immune system are skin, various mucous membranes, blood, lymphatic system, thymus, and bone marrow. The skin and the mucous membranes are considered the first line of defence and provide a physical protective barrier. The mucous membranes produce chemical barriers, including enzymes and acids, that dissolve and break down pathogens. The second line of defence is the innate immune system, providing immediately available mechanisms to combat pathogens and prevent spread. The third line of defence is the adaptive immune system, where specialised cells are designed to remember previously encountered pathogens and distinguish them from cells of the body ("the self"). The four main protective functions of the immune system; recognition, elimination, regulation, and memory are mediated by immune cells and specific molecules (Param, 2009). Misfunction of immune system mechanisms can result in autoimmune disease, inflammatory disease or cancer.

1.1.1 Cells of the Immune System

The cells of the immune system are mostly white blood cells, called leukocytes. There are two main types of leukocytes, phagocytes and lymphocytes. Immune cells derive from hematopoietic stem cells in the bone marrow and are continually generated by the body in a developmental process called hematopoiesis. These cells differentiate through distinct processes in the lymphatic organs.

Hematopoietic stem cells differentiate into lymphoid and myeloid progenitor cells, which develop further into different immune cells (Figure 1.1.). Myeloid progenitors give rise to granulocytes, which further turn develop into neutrophil, basophil and eosinophil, and monocytes. The latter further turn into immature dendritic cells, macrophages and mast cells. Immature dendritic cells, neutrophils and macrophages all function as phagocytic cells that engulf and disintegrate pathogens and present protein fragments (peptides) from pathogen by the major

histocompatibility complex II (MHC II) molecules on their surface. Immature dendritic mature after engulfing pathogenic microorganisms. (Param, 2009)

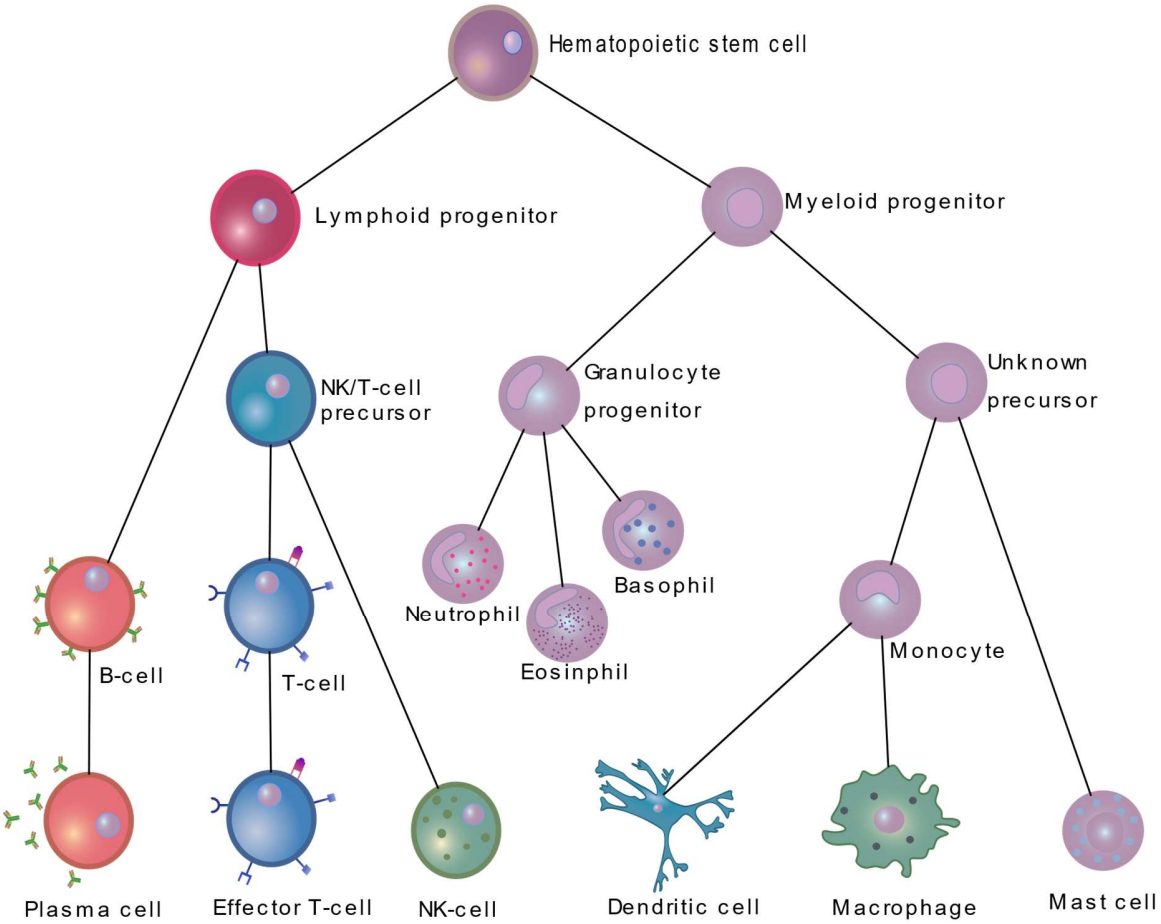

Figure 1.1. Schematic representation of hematopoiesis. Pluripotent hematopoietic cells differentiate into lymphoid and myeloid progenitor cells. Lymphoid progenitors (red) differentiate into either NK/T-cell precursors (αβ/γδT-cell) that further differentiate into NK-cells or effector T-cells, or B-cells that further differentiate into plasma cells. The myeloid progenitors differentiate into various types of cells: granulocytes, or through an unknown precursor, monocytes and mast cells. Granulocytes differentiate to neutrophil, eosinophil, and basophil. Monocytes give rise to the primary phagocytic cells, macrophages and dendritic cells. Figure based on (Param, 2009).

The cells of the immune system have an array of different functions that all are important in protecting the body against harm from pathogens (Figure 1.2.). Dendritic cells digest the pathogen and present parts of it on the surface of the cell and activate T-lymphocytes. Macrophages and neutrophils also engulf and digest pathogens, and together with mast cells, white blood cells, and other effector immune cells release cytokines and trigger inflammation and recruitment of other immune cells to the site of infection (Janeway, Travers, & Walport, 2001). During inflammation, blood vessels expand, allowing fluid, immune cells, and protein to flow into the tissue to combat the infection. Cytokines, including Interleukins (IL) and chemokines, are small proteins released by various cells of the body, that regulate and mediate immunity, inflammation and hematopoiesis by binding to receptors of either the cell that secreted them (autocrine signalling), neighbouring cells (paracrine signalling), or entering the circulation and affecting distant cells (endocrine signalling). Cytokines act either pro-inflammatory or anti-inflammatory depending on the cytokine and the phagocytic cells recruited (J.-M. Zhang & An, 2007).

Lymphoid progenitor cells give rise to natural killer cells (NK-cells), T-lymphocytes, and B-lymphocytes. NK-cells eliminate virally-infected and tumour cells lacking MHC class I molecules with self-peptides on their surface, by releasing cytotoxic molecules that lyse the target cell and induce apoptosis. Thereby they hinder the spread of infection to neighbouring cells. Description of B- and T-cells maturation and their functions follows in later chapters [1.1.4 and 1.2, respectively]. Myeloid progenitor-derived cells and NK-cells are part of the innate immune system but also aid the adaptive immune system. B- and T-lymphocytes are the primary cells of the adaptive immune system.

Type of cell	Function
Basophil	White blood cell that release histamine and several cytokines. Involved in parasite defence and allergy.
Eosinphil	White blood cell that releases histamine and kills antibody-coated intestinal parasites like helminth worms.
Neutrophil	White blood cell and effector cells stimulate inflammation and engulf and kill extracellular pathogens. They die at the site of infection, creating pus.
Mast cell	Release histamine when damaged and resident in connective tissue.
Macrophage	Engulf and digest microorganisms and release cytokines to recruit neutrophils and other leukocytes. Activate T-cells and are long-lived compared to neutrophils.
Dendritic cell	Cellular messengers that carry degraded pathogen out of infected site to a lymphoid organ that further activate adaptive immune responses. Also present peptides of the body to developing T-lymphocytes.
NK-cell	Prevent the spread of infection by attacking and lysing virus-infected or cancerous host cells. They release cytokines that hinder viral replication in infected cells.
T-cell	Kill virus-infected cells or cancerous host cells. Secrete cytokines to help other immune cells become fully activated effector cells.
B-cell	Differentiates to form memory and plasma cells that present and secrete immunoglobulin, respectively. Secreted immunoglobulin bind to pathogens and their toxic products.

Figure 1.2. Cells of the immune system and their function. Illustration based on (Sadava, Hillis, Heller, & Berenbaum, 2014)

1.1.2 The Innate Immune System

The innate immune system consists of the proteins of the complement system , phagocytic cells and NK-cells that recognise and eliminate pathogens. This is a nonspecific response in which the cells identify any foreign or nonself substance as a target and activate the innate immune response.

Innate immunity recognises pathogens based on surface markers. Phagocytic cells express invariant receptor molecules called pattern recognition receptors (PRRs) that recognise pathogen-associated molecular patterns (PAMPs) from many microorganisms, e.g. double-stranded RNA from viruses (Mogensen, 2009). The innate immune system is, therefore, able to distinguish between foreign (pathogen) and self (cells of the body). When a phagocytic cell's PRR recognise PAMPs, the phagocyte will engulf the pathogen and digest, or lyse the foreign cell or protein. In addition, phagocytes present peptides from the pathogen through their MHC (also called HLA in humans) molecules to T-cells of the adaptive immune system.

1.1.3 Human Leukocyte Antigen

Human leukocyte antigen (HLA) molecules are highly polymorphic glycoproteins encoded by genes in the HLA complex that consist of more than 200 genes and are known to compose the most polymorphic genetic system in humans (Choo, 2007) with more than 15,000 alleles (Dendrou, Petersen, Rossjohn, & Fugger, 2018). Its biological function is to bind peptides inside the cell and transport them to the cell surface for presentation. Each HLA molecule can bind a wide variety of peptides (Murata et al., 2007).

There are two types of HLA molecules, HLA class I and HLA class II. Class I molecules are expressed on the surface of most nucleated cells and present self -peptides and antigens from intracellular pathogens like a virus. HLA class II are only expressed on B-lymphocytes, antigen-presenting cells (APC), and activated T-lymphocytes and present antigens from extracellular pathogens like bacteria and foreign proteins (Janeway et al., 2001). Antigens are peptides or other molecules bound to an HLA-molecule. The HLA-molecule present the antigen to T-cells that recognize it, to activate the adaptive immune system.

APCs are a large group of cells with an HLA class II molecule on their surface that can activate cellular immune response by displaying peptides of a protein antigen to be recognised by a T-cell

receptor (TCR). For an antigen to be recognised by T-cell receptors, pathogen-derived proteins must be degraded into peptides and bound to an HLA molecule. The αβ-TCR recognise peptide fragment of 8-10 amino acids presented by HLA-I molecules, and 13-25 amino acids presented by HLA-II molecules (Param, 2009). This process is called antigen processing.

1.1.4 The Adaptive Immune System

The mechanism unique to adaptive immunity is pathogenic memory. Lymphocytes collectively have the ability to recognise a vast array of antigens through the development of a highly diverse population with specific antigen receptors, B-cell receptors (BCR) and TCRs. B-cells and T-cells represent, respectively, humoral immunity (in the body's fluids), that control extracellular pathogens, and cellular immunity, that control intracellular pathogens. BCRs of mature B-cells interact with pathogens and their toxic products in the extracellular spaces in the body. TCRs recognise antigens bound on HLA molecules on the cell surface. B- and T-cells use similar receptors to identify these pathogenic proteins.

BCRs are membrane-bound globulin proteins on the surface of B-cells that bind to antigens. The BCR is made up by a signal transduction protein (CD79) and an immunoglobulin (IG) protein. Each IG protein is a polypeptide consisting of a heavy- and a light-chain, bound by a disulphide bond. The chains consists of a constant (invariable) region that determines the effector class of the antibody and a variable region that make up the antigen-binding region of the IG. The variable region binds the antigen, and the constant region interacts with effector molecules and cells of the immune system (Figure 1.3.). The binding site has around 10^{10} possible arrangements achieved through re-combination (Janeway et al., 2001). The process of recombination occurs in the B-cell while in the bone marrow. The IG is encoded by different genes situated on chromosomes 2, 14 and 22. These genes encode variable gene segments (V), diversity gene segments (D), joining gene segments (J), and a constant gene segment (C). During the development of a B-cell, the V, D, and J segments are randomly rearranged by DNA recombinase complex. Single gene segments are brought together to form the gene sequence for the IG. In each individual B-cell, only one rearranged IG gene becomes functional, which makes the B-cell specific for one epitope. (Janeway et al., 2001)

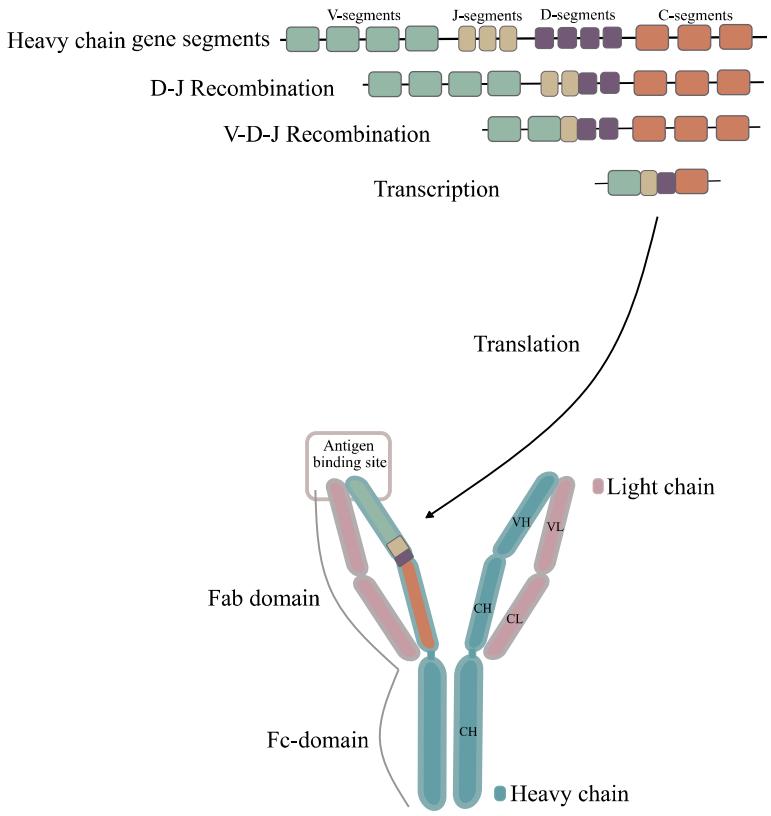

Figure 1.3. Schematic drawing of an antibody and the genomic regions that make up immunoglobulins. An antibody is made up by V, J, D and C gene segments through recombination. The antibody has variable regions and constant regions. The variable region at the top makes up the antigen binding site.

B-cell activation depends on helper T-cells that recognise the same antigen as the B-cell. Upon activation by an antigen, B-cells mature further in the periphery (outside of the central lymphoid organs) where they develop into effector B-cells, antibody-secreting plasma cells, or memory B-cells. Memory B-cells are APC together with macrophages and dendritic cells. They have HLA II molecules and present antigens to and activate T-cells.

Activated B cells can differentiate into plasma cells that produce antibodies specific to the antigen recognized by the mother B cell and are released into the circulation to neutralize cells presenting such antigens. The different effector mechanisms of antibodies are neutralisation, opsonisation, activation of the complement system. These mechanisms contribute to immunity in three main ways: neutralization of pathogens by binding to the virus or intracellular bacteria

prevents the pathogens from binding to and entering target cells; opsonization, by which the antibody coats the surface of a pathogen and enhances its phagocytosis by an immune cell; complement system activation whereby antibody-bound pathogen complex triggers the complement cascade to create pores in the membranes and lyse the infected cells or recruits phagocytic cells (Murphy, Travers, & Waltport, 2008). Antibody-coated cells are recognized by Fragment constant receptors (Fc receptors) on immune cells that in turn, destroy the infected cell by phagocytosis or cytotoxicity (DeFranco, Locksley, & Robertson, 2007).

1.2 T-cell Development

T-cells develop from hematopoietic stem cells that migrate to the thymus for further maturation; thus, they are called thymus-dependent lymphocytes, or T-cells. In the thymus, T-cells start to express TCRs. These receptors are generated through random recombination and recognise peptides presented on HLA molecules. TCRs get selected through interactions with APCs and thymic epithelial cells (TECs) in order to eliminate TCRs that do not bind HLA molecules or react to self-peptides. The main goal of the selection is to establish central-tolerance, in which T-cells do not recognise self as a threat. The T-cell exits the thymus as either CD4 or CD8 positive T-cells into the periphery to continue their development or is signalled to undergo apoptosis and die in the thymus.

Immature T-lymphocytes migrate from the bone marrow to the thymus. The thymus gland is highly active from fetal development to age 2-3 years when it reaches its peak weight of 30-40 g (Ohigashi, Kozai, & Takahama, 2016). During life, the thymus starts to shrink, and the TECs, responsible for most of thymocyte differentiation and development, become replaced by fat cells (Haroun, 2018; Haynes, Sempowski, Wells, & Hale, 2000). Therefore, the development of T-lymphocytes is age-dependent, and fewer T-cells undergo maturation later in life.

In the thymus, immature T-lymphocytes (thymocytes) develop into fully mature functional T-cells. Thymocytes differentiate by migrating through the structural compartments of the thymus, cortex and medulla, that provide specific factors needed for their development (Figure 1.4). When first entering, the thymocytes are termed early thymic progenitor (ETP) cells. These will undergo a round of division and develop into double negative cells (DN2), that do not express CD4 or CD8 receptors. As they mature, they develop their invariant pre-TCR, this is the DN3 stage, that will rearrange and develop into fully functional TCRs specific to a particular antigen.

Similarly to the BCR, the TCR is encoded by V, D, J and C genes. TCR has two chains, α- and β-chains, and in some cases γ- and δ-chains, with a constant and variable region, held together by a disulphide bond. A thymocyte can either express αβ-chains and later be selected to generate $CD4^+$ or $CD8^+$ T-cells, or γδ-chains and generate immuno-regulators or surveillance cells. The γδ-TCRs are invariant and do not recognise HLA-presented peptides.

T-cell selection takes place in the peripheral cortex and central medulla and involves two phases: the positive and negative selection. Positive selection which happens in the cortex, ensures that only the T-cells that have a functional receptor will migrate into the medulla, where they undergo negative selection. The negative selection will then ensure that the T-cells that respond to the individual's own cells or tissue (self-antigen) will be removed. Thus positive and negative selection critically examine the ability of the receptors to select the useful pathogen recognising antigen receptor that will not react to self peptides.

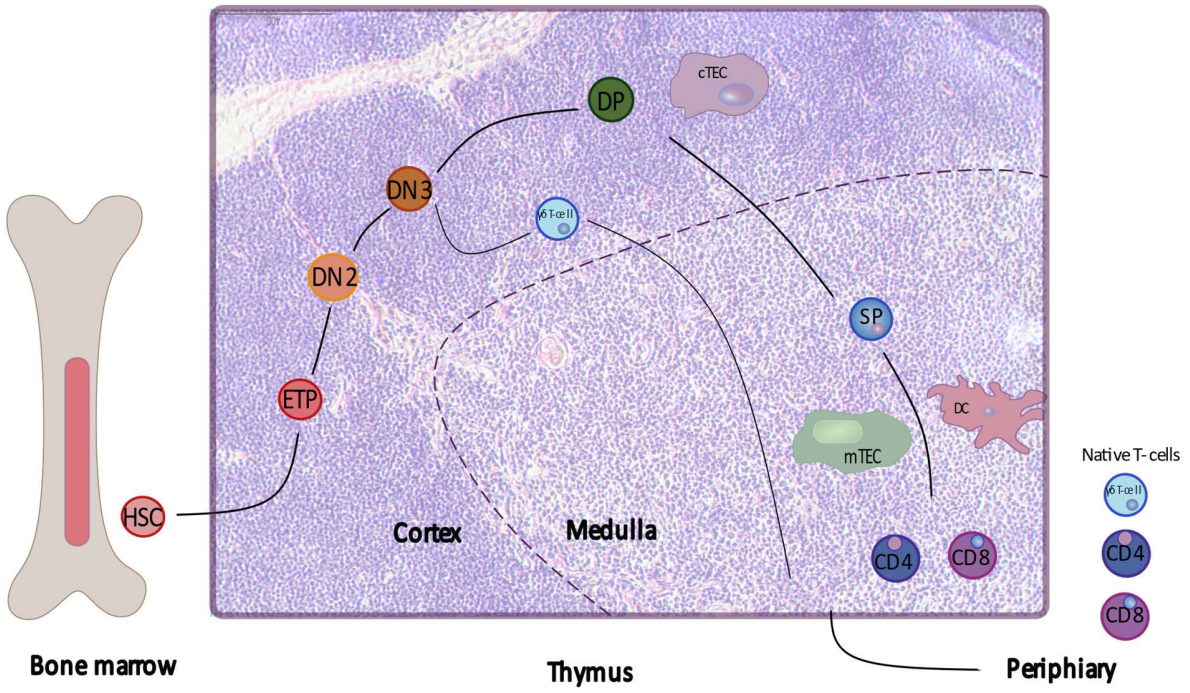

Figure 1.4. T-cell development in the thymus. Hematopoietic stem cells (HSC) from the bone marrow evolve into early thymic progenitors (ETP) when entering the thymus. In the thymus, these ETPs transform into double negative (DN) thymocytes in the cortex. The DN thymocytes develop αβ or δγ TCR and turn into γδT-cells or double positive thymocytes. The DP-thymocyte react with the cTECs and turn into single positive (SP) T-cells. SPT-cells turn into CD4 or CD8 by the help of mTECs or dendritic cells. CD4, CD8, and γδT-cells are exported out into the periphery. Illustration based on (Cano, Lopera, Anaya, Shoenfeld, & Rojas-Villarraga, 2013). Image of thymus was kindly provided by Marthe Heimli.

1.2.1 Positive Selection

Positive selection proves thymocytes based on their successful rearranged receptors and signals to the αβ thymocyte to mature further if they are capable of recognising peptide-HLA molecule with appropriate avidity. The thymocytes develop in the cortex to express both $CD4^+$ and $CD8^+$, now called double positive (DP) T-cells. The positive selection happens when a DP-thymocyte binds to a self-antigen presented by HLA molecules on cortical thymic epithelial cells (cTECs). The cTECs make contact with the DP-thymocyte, and at the regions of contact, tests the binding between HLA molecules and DP-thymocyte receptor. If the TCRs interact with the HLA molecules and the avidity is low (fewer contact points), the DP-thymocyte will receive a positive signal, and the DP-thymocyte will continue its maturation (van den Boorn, Le Poole, & Luiten, 2006). If the TCR do not interact strongly enough (less than 3-4 days) they will lose contact and

undergo apoptosis and die (Egerton, Scollay, & Shortman, 1990). This process ensures that the selected T-cells will be able to interact with HLA molecules in the body. Only about 10 % of the developed DP-thymocytes are selected and further matured (Klein, Kyewski, Allen, & Hogquist, 2014).

When selected by the positive selection, the DP-thymocyte receives a signal to mature into a single positive (SP) T-cell, either as CD4+ or CD8+ (Murphy et al., 2008). Whether the T-cell commits to be a CD4+ or a CD8+ T-cell depends on which class of the HLA molecules that interacts with the TCR (Janeway et al., 2001). When interacting with HLA class I molecule, the DP-thymocyte will mature into $CD8^+$ thymocyte, and halt the expressing of CD4 molecules. Correspondingly, HLA class II molecule- recognizing thymocytes will halt the expression of CD8 molecules and solely express CD4. After the positive selection, SP-thymocyte enter the medulla to undergo negative selection as a second checkpoint in their education (Decker, 2012).

1.2.2 Negative Selection

T-cells that survive positive selection migrate further into the medullary junction of the thymus where they encounter medullary thymic epithelial cells (mTECs) and thymic APCs all of which present HLA-loaded self-peptide molecules on their surface. mTECs express all the major proteins that are found in most peripheral tissues in the body. These proteins that are expressed by mTECs are called tissue-restricted antigens (TRA).When a T-cells TCR binds with high affinity and high avidity (more peptides bind) to self-peptide:HLA molecules presented by mTECs and APCs the T-cell receive a signal to undergo apoptosis (van den Boorn et al., 2006). This process, called negative selection, serves to eliminate self-reactive T cells before they enter the circulation. The transcription factors that control the expression of TRA are therefore important to ensure proper education of the T-cells (Takaba et al., 2015).

Out of all αβ T-cells, only 5 % survive both positive and negative selection and leave the thymus (Figure 1.5) to circulate in blood and reach secondary lymphoid organs where they await activation by APCs, mainly dendritic cells (Klein et al., 2014). A dendritic cell will transport pathogenic antigens from infected tissues into lymphoid organs and present the antigen to T-cells (Dieli, 2003). The presentation of pathogenic antigens and activation of T-cells requires three signals: peptide-HLA and TCR interaction, co-stimulation through dendritic cell receptor B7 and T-cell receptor CD28 (B7:CD28), and the release of cytokines (such as IL-2) from the T-cell

causing activation of the T-cell through autocrine signalling (Xing & Hogquist, 2012). Activated T-cells differentiate further into effector T-cells of various types.

$CD8^+$ T-cells kill their targets by cytotoxicity (attaching and releasing toxic proteins that induce apoptosis in target cell) whereas $CD4^+$ T-cells (also called T helper cells) have regulatory functions through secreting cytokines. $CD8^+$ T-cells also produce the cytokines lymphotoxin and IFNγ, which induce inflammation and activates macrophages to clean up cell debris and inhibits replication of virus in infected cells (Param, 2009). Once the $CD4^+$ T-cell is activated, it differentiates into either a T_{reg}-cell, a T_h1-cell, or a T_h2-cell that have different functions, depending on what factors they are exposed to. T_{reg}-cells regulate the other effector T-cells and has an important role in preventing autoimmunity. T_h1-cell activates tissue macrophages to enhance the phagocytosis. T_h2-cell activates B-cells specific for the same antigen for the B-cells to differentiate into plasma cells.

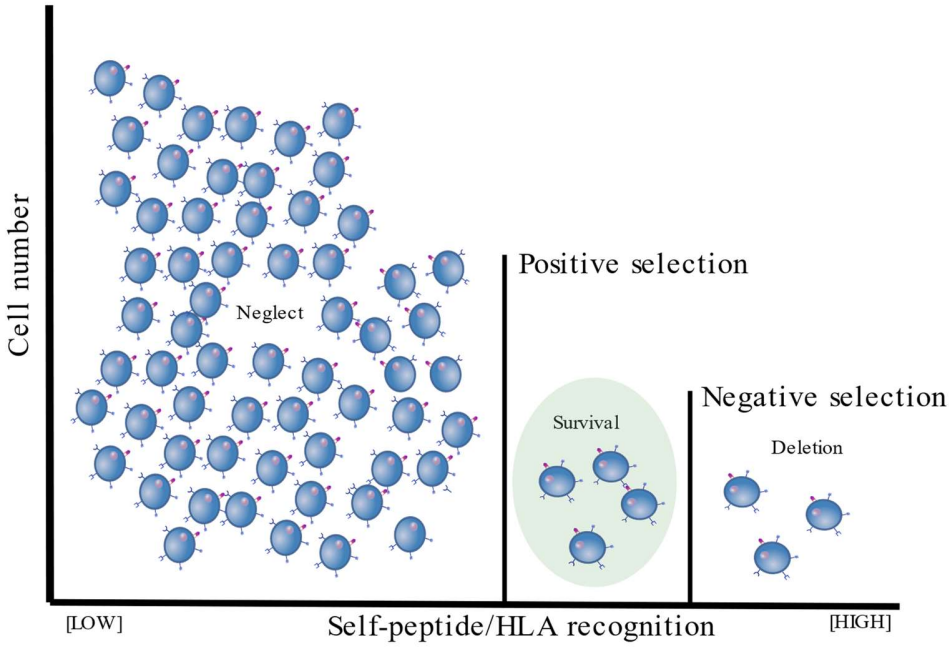

Figure 1.5. The T-cell repertoire is determined by positive and negative selection. Around 95 % of the thymocytes die from neglection by positive selection or are signalled to undergo apoptosis by negative selection. The selection is based on self-peptide/HLA recognition and avidity. Illustration based on (James & Kwok, 2008).

1.2.3 Peripheral tolerance

Central tolerance mechanisms are efficient in removing most self-reactive T-cells. However, some self-reactive T-cells escape into the periphery partly because not all self-antigens are expressed in the thymus. Peripheral tolerance ensures that T-cells that escaped central tolerance do not cause autoimmune disease through either peripheral clonal deletion (apoptosis upon activation) or anergy (making the T-cell unresponsive to antigen) (Xing & Hogquist, 2012). Peripheral tolerance is mediated by dendritic cells, lymph node stromal cells (LNSC), extrathymic AIRE-expressing cells (eTAC) in lymph nodes. The lymph nodes aids in the activation of immune responses facilitates migration of lymphocytes, cytokines and chemokines. The LNSC and eTACs express TRAs that are transferred to and presented by dendritic cells. Clonal deletion is carried out by dendritic cells that present self-antigens on the HLA molecule without or with low levels of the B7 co-stimulatory molecule (Xing & Hogquist, 2012). If a self-reactive T-cell recognises and bind to the self-peptide, the T-cell will first receive a signal for activation, but as co-stimulation is absent, the T-cell will undergo apoptosis. Some auto-reactive T-cells survive, but remain inactivated and cannot respond to antigenic stimuli. Anergy is characterized by repressed TCR and lack of IL-2 expression (Iberg & Hawiger, 2020). This presentation in the lymph nodes broadens the tolerance of T-cells (Fuhlbrigge & Yip, 2014; Hirosue & Dubrot, 2015).

1.2.4 Antigen-Presenting Cells in the Thymus

Antigen presenting cells located in the thymus are cortical and medullary TECs, dendritic cells ($CD141^+$ and $CD123^+$ dendritic cells) and B-cells ($CD19^+$ B-cells). The thymus contains at most around one million APCs and TECs (under 5 % of the total cell number in the thymus) (Sakata, Ohigashi, & Takahama, 2018). Dendritic cells and B-cells contribute to antigen presentation either by presenting unique TRAs obtained from TECs, or transferred antigens from circulation (Gies et al., 2017; Hadeiba et al., 2012; Perry et al., 2018; Yamano et al., 2015). However, TECs are seen as the driver of antigen presentation, and here we focus on the promiscuous gene expression by mTECs.

Until recently, it was not known how the T-cells could learn to recognize “self” peptides in the thymus as peripherally expressed proteins, such as pancreatic insulin, do not have a function in the thymus. However, the unique capacity of the mTEC population to collectively express almost

all peripheral transcripts, ensures full coverage of self being presented to developing thymocytes. Medullary TECs express more than 18 000 genes, approximately 85-90 % of the protein-coding genome, while other cell types typically express 60-65 % of the protein-coding genome (Abramson & Anderson, 2017). Each TRA is expressed by a minor fraction (1–3%) of mTECs at any given time (Klein et al., 2014). This is called promiscuous gene expression and is a process specific to the thymus and to establishing central tolerance.

Thymic TRA expression is not fully understood. The process is highly regulated and has been explained as an ordered yet stochastic process. The expression of TRAs in a single mTECs do not necessarily mirror individual periphery organ cells as co-expression of highly correlated genes is rare in individual mTECs, and TRA co-expression varies between individuals, hence stochastic (Derbinski, Pinto, Rösch, Hexel, & Kyewski, 2008; Meredith, Zemmour, Mathis, & Benoist, 2015; Passos, Speck-Hernandez, Assis, & Mendes-da-Cruz, 2018). However, the process is not completely random, and in mice, full diversity of self-antigens represented by mTECs is obtained by assembling at least two-thirds of the TRA population in co-expression clusters (Brennecke et al., 2015; Dhalla et al., 2019).

1.3 AIRE, FEZF2, and DEAF1 Functions in APCs

The promiscuous gene expression of the protein-coding genome in mTECs is partly under the control of the Autoimmune Regulator (AIRE) and Forebrain Embryonic Zinc Finger-Like Protein 2 (FEZF2) (Takaba et al., 2015). Loss of key transcription factors that regulate gene expression in peripheral organs has no effect on TRA gene expression in mTECs (Danso-Abeam et al., 2013). Recent research has revealed that FEZF2 and AIRE collectively control most of the TRA expression in mTECs (Takaba & Takayanagi, 2017). Deformed Epidermal Autoregulatory Factor 1 Homolog (DEAF1) control TRA expression in the pancreatic lymph node, but the transcription factor has also been identified in the thymic APCs (Gabrielsen et al., 2019; L. Yip et al., 2009).

1.3.1 AIRE

AIRE is an essential regulator for negative selection and autoimmune disease and has been extensively studied in both mice (assigned as Aire/*Aire*) and humans (assigned as AIRE/*AIRE*). Both mice without the *Aire* gene and patients with *AIRE* mutations develop autoimmune disease (Akirav, Ruddle, & Herold, 2011; Takaba et al., 2015). Patients may develop autoimmune-

polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED). APRCED causes inflammation and cell infiltration into a wide variety of tissues and produces self-antibodies against various cytokines (Browne, 2014). The repertoire of auto-antibodies (antibody reacting to self) is unique to each patient with AIRE-deficiency (Meyer et al., 2016). This suggests that AIRE-deficiency causes unpredictive and abnormal T-cell selection and disrupts self-tolerance.

AIRE does not have an obvious DNA binding domain and is regarded as a transcriptional co-regulator interacting with nuclear factors and protein complexes. In a study by Bansal et al., involving ChIP-seq on Aire in mTECs from mice, it was found that Aire binds to 42,124 genomic sites, including super-enhancers (regions in the genome containing multiple enhancers and bound by an array of transcription factors) (Bansal, Yoshida, Benoist, & Mathis, 2017). Here, Aire interacts with histones to open chromatin and locates on super-enhancers to induce the TRA expression with many regulatory proteins such as Top1, Atf7ip-MBD1, Irf8, and Cbp (Bansal et al., 2017).

AIRE is expressed in mature mTECs and a subset of B-cells ($CD19^+$ B-cells) in the thymus (in addition, some extrathymic cells (eTACs) in secondary lymphoid organs) (Gies et al., 2017). Little research is performed on B-cells in the thymus and specifically the role of AIRE in B-cells. However, AIRE positive B-cells have shown to express a low number of TRAs, including TRA genes connected to autoimmune disease (Gies et al., 2017). In mice mTECs, Aire plays a role in driving the expression of 3,793 TRAs (Sansom et al., 2014). Around 533 of these are entirely dependent on Aire for their expression (Aire-dependent), and Aire enhances expression of the remaining 3,260 genes (Aire-enhanced). Aire-independent mechanisms control the promiscuous expression of 3,947 TRAs (Sansom et al., 2014). Most of the Aire-dependent TRAs are secretory proteins. Examples of AIRE regulated TRAs are insulin. In the pancreas, where insulin is created and released, insulin expression is dependent on pancreas/duodenum homeobox protein 1 (PDX1). This protein is also expressed in the thymus; however, here insulin expression is AIRE dependent, proven by the transcription of insulin occurring in the thymus of *Pdx1* knock-out mice, and lacking in *Aire* knock-out mice (Danso-Abeam et al., 2013).

1.3.2 FEZF2

Studies on FEZF2 in human mTECs and thymus are sparse, and so far, most knowledge is based on studies on the Fezf2 homolog in mice. In humans, transcriptomes from thymic APCs has

shown that FEZF2 are expressed in mTECs (Gabrielsen et al., 2019). Otherwise, studies performed with mouse cells and tissue have revealed that Fezf2 is a critical regulator of autoimmune responses. A loss or mutation of Fezf2 leads to autoantibody production in mice (Takaba & Takayanagi, 2017). Although FEZF2 mutations are not directly linked to autoimmune diseases, several studies have linked FEZF2 dependent genes to different autoimmune diseases such as rheumatoid arthritis (*Ttr)*, autoimmune pancreatitis and type 1 diabetes (*Amy2a)* (Fatourou & Koskinas, 2009; Sharma et al., 2014; Takaba & Takayanagi, 2017).

FEZF2 expression in mTECs is regulated by the LTβR pathway (Takaba et al., 2015), an essential signalling pathway within immune development and host defence (Norris & Ware, 2013). FEZF2 contains six zinc-finger (C2H2-ZF) domains and engrailed homology 1 (Eh1) domain that requires open chromatin and directly recognises specific DNA motifs (Takaba & Takayanagi, 2017).

Fezf2 directly regulates a unique set of TRA genes independently of Aire (Takaba et al., 2015). As opposed to Aire-dependent TRAs, who mostly include secretory proteins, Fezf2-dependent TRAs are mostly intracellular or membrane proteins (Takaba & Takayanagi, 2017). Aire regulates 28.9 % of TRAs in mTECs, and together with Fezf2, they control 61.2 % of the expression of TRAs (Derbinski et al., 2005) suggesting there are additional transcription factors regulating the remaining expression of TRAs in mTECs (Takaba & Takayanagi, 2017) (Figure 1.6.).

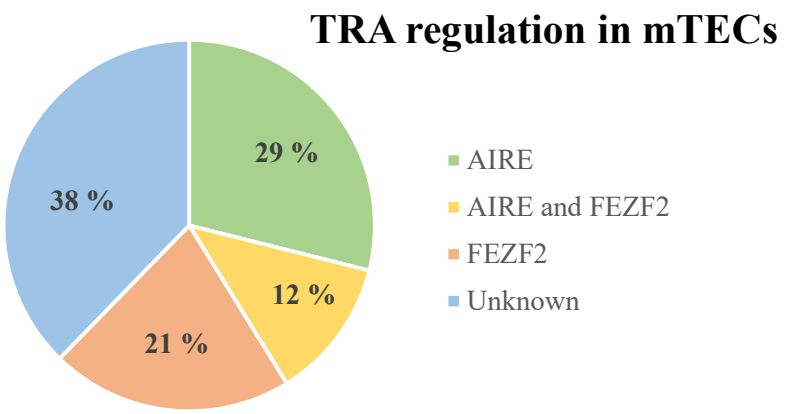

Figure 1.6. The ratio of TRAs regulated by Aire and Fezf2 in mice. Illustration based on (Takaba et al., 2015).

1.3.3 DEAF1

DEAF1 control the expression of peripheral tissue antigens genes in the pancreatic lymph nodes (L. Yip et al., 2009). Deaf1 in mice LNSCs control the expression and processing of around 600 TRAs presented to T-cells (Fuhlbrigge & Yip, 2014). Human DEAF1 have shown to regulate the translation of the gene *EIF4G3* that encodes eIF4GII, an important part of the pre-initiation complex that enables translation of genes involved in antigen presentation on MCH class II molecules (Linda Yip, Creusot, Pager, Sarnow, & Fathman, 2013), making DEAF1 an important protein for the peripheral tolerance mechanism. Gabrielsen et al. theorised that DEAF1 might control TRA expression in the thymus as well, and examined *DEAF1* expression in APCs from the human thymus. They found the transcription factor was expressed in four major APC types (Gabrielsen et al., 2019), suggesting that DEAF1 may take part in controlling TRA expression or presentation in the thymus, thus also contributing to central tolerance.

1.4 Autoimmune Diseases

Through the production of effector cells, the adaptive immune response terminates infection, and clonal expansion of pathogen-specific T-cells and B-cells produces long-lived clones of memory cells that enable immunological memory. Adaptive immunity evolves throughout an individual's lifetime, and the lymphocyte repertoire is widened for each infection. As opposed to innate immunity, adaptive immunity is not inherited; however, failures to develop proper response can come from inherited deficiencies. These failures can disrupt self-tolerance and create an overactive response to otherwise harmless self-proteins causing autoimmune disease.

Autoimmune diseases (AIDs) are chronic inflammatory disorders caused by the immune system attacking host-cells and -tissues. A critical function of the immune system is to discriminate self from non-self. AID occurs when the immune system nevertheless attacks self-antigens. Since such antigens are virtually impossible to fully eliminate the consequence is a chronic inflammatory injury to the affected tissues. Autoimmunity is believed to commence by the activation of antigen-specific T-cells that are becoming autoreactive effector cells. These T-cells will provoke immune events leading to the production of antibodies against self-antigens, called auto-antibodies. When self-reactive T-cells are not eliminated during the central tolerance development and not controlled by peripheral tolerance, the consequence will be an abnormal immune system response causing pathologic inflammation and injures to otherwise healthy tissues. (Bolon, 2012).

There are two main types of autoimmunity, organ-specific and systematic. Organ-specific autoimmunity affects a single organ and produces autoantibodies towards components of that organ. Systematic autoimmunity affects more than one tissue. Animal models have taught us about how tolerance in self-reactive B- and T-cells fail, and the lymphocytes become self-reactive (Abbas, Litchman, & Pillai, 2012).

AIDs are a collection of over 80 different diseases affecting approximately 5-10 % of the world population with increasing incident rates (Cooper, Bynum, & Somers, 2009). Some individuals develop multiple AIDs, and certain AIDs commonly occur together (e.g. Rheumatoid Arthritis (RA) and Systemic Lupus Erythematosus (SLE)) (Bolon, 2012). Development of multiple AIDs can cause overlap of symptoms and make diagnosing difficult.

Why the immune system attacks self-antigens is still not known. Both genetic and environmental factors are thought to contribute to autoimmunity. Women have a higher chance of developing AIDs (Bolon, 2012) and children with AID parents have a higher risk of developing any AID later in life (Hemminki, Li, Sundquist, & Sundquist, 2009). Environmental factors such as chemical pollution and over-protective environments including the hygiene hypothesis (leading to individuals not being introduced to a large enough repertoire of pathogens) have been theorized as causes, supported by the fact that the AID incident rate is increasing in developed countries (Bach, 2001; Patrick, 2009). A deeper understanding of immunological processes and how the immune system distinguishes self from non-self is important to acquire more knowledge about the mechanisms of AID.

2 Aims

In order to understand the mechanism behind autoimmune disease, we need to understand how the body is taught to protect and recognize self. More specifically, how the T-cells learn to distinguish self from foreign through TRA presentation by the epithelial cells in the thymus, mTECs. Research on the role of mTEC for cell education has mostly been performed on rodents while the human thymus is underrepresented and still little is known about the promiscuous gene expression in human mTECs.

The main aim of this thesis was to locate the genomic binding sites of known and potential transcription factors AIRE, FEZF2, DEAF1 in the thymic APCs and mTECs.

The study aimed to:

- Establish a Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) strategy to locate genomic binding sites of transcription factors for AIRE, FEZF2, and DEAF1 in mTECs and APCs using human thymic tissue through the use of ChIP-qPCR.
- Analysis of ChIP-seq sample data for AIRE, FEZF2, and DEAF1 to find binding sites.
- Generate a list of TRA genes in the proximity of the three transcription factor binding sites that can be used for further studies.

3 Materials and Methods

All operations performed at room temperature unless described otherwise. Buffer recipes and information about commercial kits, reagents, instruments, software, primers and thymic tissue is given in the appendix page I-II. Thymus collection, frozen thymus tissue preparation, and fresh thymus tissue dissociation were performed by members of the thymus research group.

3.1 Thymus collection

The thymus of children (age 0-3 years) undergoing cardiac surgery, where the thymus is otherwise discarded, was collected upon obtaining a written consent by the parents. The project was approved by the regional ethical committee (REK). All samples were anonymized. The collected tissue was kept on ice during transport to the laboratory at Oslo University Hospital (OUS), Ullevål, submerged in RPMI-1640 cell culture medium (Cat. No. R7509, Sigma Aldrich, USA) supplemented with 10 % heat-inactivated FBS (Cat. No. F4135, Sigma Aldrich, USA), that was filtered through 0.2 µm filter.

3.2 Frozen thymus tissue preparation

Deep frozen (in liquid nitrogen tank) thymic tissue used for Chromatin Immunoprecipitation followed by Sequencing (ChIP-Seq) experiments were stored in tubes with 1-4 g tissue pieces per cryotube. The pieces were cut off from fresh tissue and snap-frozen in liquid nitrogen (- 140 °C) before storage. The tissues were homogenized to powder using mortar and pestle, adding liquid nitrogen as needed to avoid thawing of the tissue. The homogenized tissue powder was stored at -80 °C, aliquots of which were used for chromatin preparation. Thymic samples from three donors were utilized in the optimisation of the ChIP-Seq protocol.

3.3 Fresh thymus tissue dissociation

While working on a cold (boxes of ice placed under the working surface) LAF-bench, the freshly acquired thymus tissue was cleaned from blood clots, connective-, necrotic-, and fat-tissue while keeping the tissue soaked in RPMI 1640 medium. The thymus piece was cut into 2-4 mm pieces and washed three times by resuspending in RPMI using a wide pipette tip, discarding the liquid after each wash. The tissue was dissociated in 10 mL Liberase TM buffer (0.17 U/ml Liberase TM (Cat. No. 5401119001, Sigma Aldrich) and 0.1% w/v DNaseI (Cat. No. 11284932001,

Sigma-Aldrich, Germany) diluted in RPMI-1640. The tissue-solution was dissociated in GentleMACS C tubes (Cat. No. 130-093-237, Miltenyi) using 4-6 g tissue per tube, on GentleMACS Octo Dissociator for 15 minutes at 37 °C and slow swirling for five consecutive dissociation cycles, dissociating for 75 minutes in total. After each cycle, an equal volume of Solution C (PBS, 5% FBS, 5mM EDTA, 0.1%wv DNase I) was added, and the sample was centrifuged at 100 x g for 30 seconds to pellet undissolved tissue. The supernatant was filtered through a 70 µm + 30 µm filter stack then the cells present in the supernatant was stored on ice, and the undissociated tissue on the filter was transferred back to the GentleMACS C tube. After the first cycle, 10 mL Liberase TM buffer was added to the GentleMACS C tube and 5 mL Liberase TM buffer for the following four cycles. While the next dissociating cycle was running the supernatant was centrifuged at 340 x g for 10 minutes at 4 °C, decanted and the cell pellet was resuspended in Solution C and kept on ice, swirling the samples every 5-10 minutes to avoid clumping. This was done for all five baths.

After filtering the cells from the fifth round of dissociation, remaining residue on the filter was washed out with PBS and centrifuged at 340 x g for 10 min at RT. The pellet was resuspended in 2.5 mL Liberase TM buffer with 0.05 % trypsin-EDTA (Cat. No. 25200056, Thermo Fisher), and incubated on the GentleMACS Octo Dissociator with a slow swirl for 45 minutes at 37 °C. The suspension was then filtered through a 70 µm + 30 µm filter stack, and 2.5 ml Solution C was added. The suspension was centrifuged at 340 x g at 4 °C for 10 min, and the pellet was resuspended in 20 mL Solution C.

The cells were centrifuged at 340 x g for 10 minutes at 4 °C, the supernatant was discarded, and the cells were resuspended in 1.07 g/mL Optiprep solution (5 mL per $1*10^9$ cells) and transferred to 15 mL Falcon tubes (5 mL per tube) carefully adding 5 mL of 1.061 g/mL Optiprep solution on top of the 1.07 g/mL OptiPrep solution, and 2.5 mL FBS on top of the 1.061 g/mL OptiPrep solution. The tubes were centrifuged using a swing-out rotor 1700 x g for 30 min at 4 °C using brake one and acceleration one. The band at the top of the 1.061 g/mL OptiPrep gradient, containing smaller cells and making up the APC enriched part of the sample, was transferred to a clean 15 mL tube, topped with PBS, and centrifuged at 340 x g for 10 min at 4 °C. The supernatant was discarded, and the pellet was resuspended in 1.5 mL PBS and kept on ice, while the cell concentration was determined.

The cell suspension was treated with EasyStep Human CD45 Depletion Kit II (Cat. No. 17898, STEMCELL, Canada) to remove unwanted CD45 positive cells. Producers instructions were followed using STEMCELL EasyStep magnet and STEMbuffer (PBS, 2 % FBS, 1 mM EDTA). The unwanted cells were tagged by antibodies recognising CD45 with magnetic particles and separated from the other cells by using EasyStep magnet, leaving the unwanted cells in the tube and decanting the wanted cells into a fresh tube. After depletion, the TEC enriched cell suspension was stored in liquid nitrogen (-200 °C).

3.4 Chromatin Immunoprecipitation followed by Sequencing

To locate binding sites of the transcription factors AIRE, FEZF2, and DEAF1, Chromatin immunoprecipitation (ChIP, Figure 3.1.) followed by sequencing was performed using human thymic tissue. Thereby, the chromatin (complexes of protein, RNA and DNA) is isolated and bound on protein G coated magnetic beads with antibodies specific to the three different transcription factors to enrich the complexes containing the respective factor. After this, the associated DNA was isolated to be analysed by qPCR and sequencing. Biologically, the binding of a transcription factor to DNA is transient - meaning that protein interactions are formed and broken easily. Therefore, there is a need to stabilise these interactions between the proteins in order to detect them. A low (1 % v/v) concentration of formaldehyde (FA) is often used in ChIP to form crosslinks between proteins and DNA. However, transcription factors often work in clusters with other transcription factors, meaning these interactions has to be fixed as well. Disuccinimidyl glutarate (DSG) is a membrane preamble protein crosslinking reagent and is used here, in addition to FA, to create crosslinks between proteins. The strength of fixation (number of crosslinks) is proportionate to the concentration of the fixation agent and the time of fixation. While sufficient fixation is necessary to stabilize all of the interactions, excessive fixation (high crosslinking reagent concentration and more prolonged fixation) results in epitope masking, reduced antigen accessibility and reduced fragmentation efficiency. Therefore it was necessary to experimentally determine the optimal fixation conditions for a particular amount (20 or 50 mg) of tissue fixed. Fixation time had been optimised prior to the beginning of this thesis and the optimal fixation time was 5 minutes, as the short fixation time improved sonication efficiency.

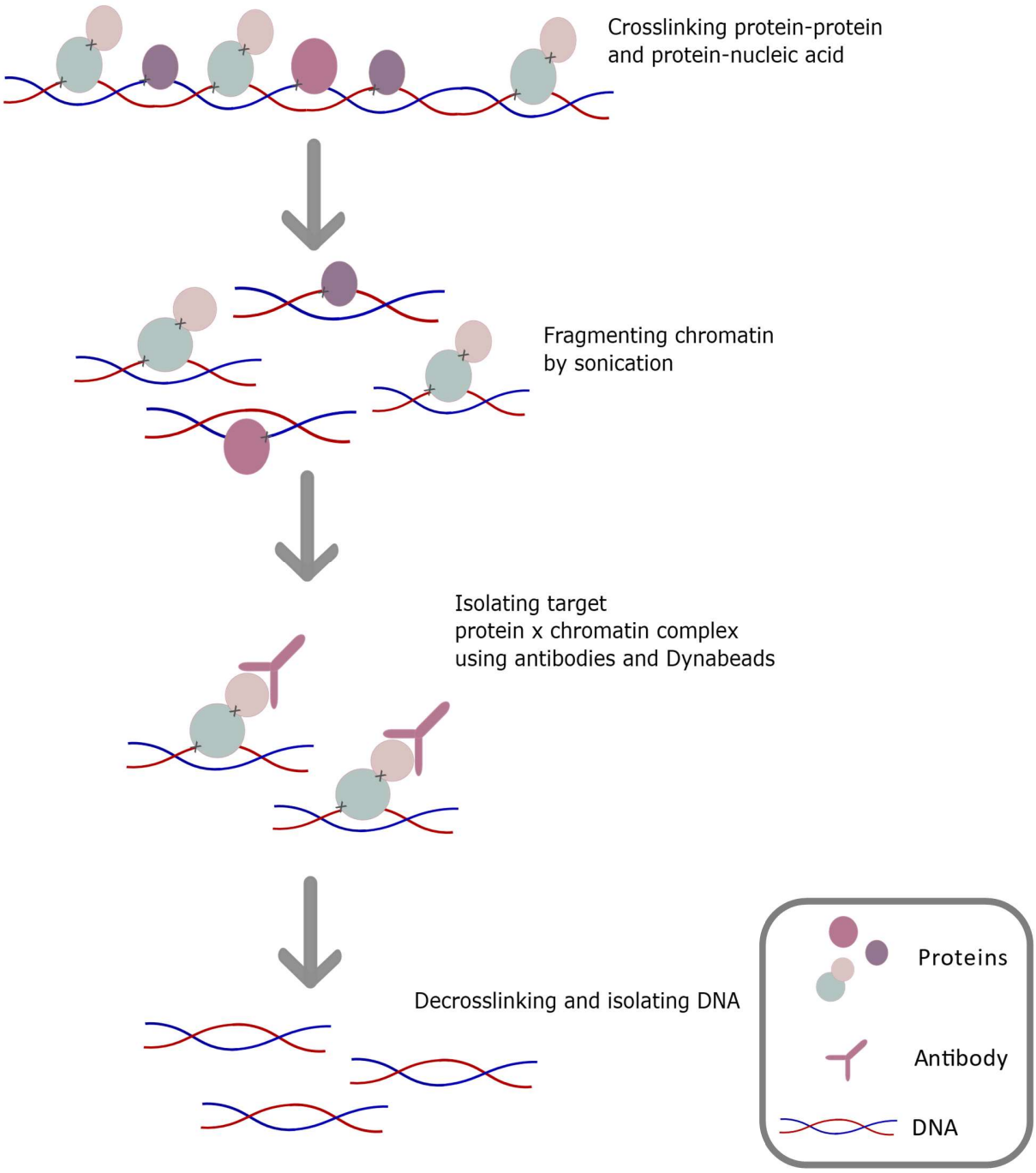

Figure 3.1. Schematic illustration of ChIP protocol steps. Protein-protein and protein-nucleic acids are crosslinked. Chromatin is isolated and fragmented. Target protein-bound chromatin fragments are enriched with protein-specific antibodies and magnetic beads. Upon reversal of crosslinks, the enriched DNA is isolated and subjected to qPCR to evaluate the specific recovery of each target gene, or to sequencing to identify the target genes genome-wide.

Initially, we analysed the binding to DNA of CCCTC-binding factor (CTCF) as a positive control during IP, since it is relatively abundant and ubiquitously expressed and its target genes are known. CTCF is a multi-functional protein that binds to ca. 55,000 DNA sites (Wang et al., 2012). Therefore, we expected a high recovery of selected CTCF target sites (Table 3.5. confirmed by (ENCODE & Consortium, 2004)) in ChIP. As a negative control, we used for the IPs either rabbit or goat IgG, the same origin as the transcription factor specific antibodies used. The target loci that we chose to test for AIRE, FEZF2 and DEAF1 are potential target genes, being published in studies involving ChIP on sorted mice mTECs (AIRE and FEZF2 target genes), overexpression studies of DEAF1, or studies involving microarray to examine expression patterns in mice pancreatic lymph node tissue (DEAF1 target genes), all compared to knock-out models (Bansal et al., 2017; Jensik et al., 2014; Takaba et al., 2015; L. Yip et al., 2009). There are no reports of target loci of these proteins from human thymus, and therefore we cannot be sure that they are actual target genes in the human thymus. The selected target genes (Table 3.5.) had high expression driven by either AIRE, FEZF2, or DEAF1. The primers for qPCR were designed to cover a 100-200 bp region in the promoters (100-1000 bp upstream) of the reported target genes.

3.4.1 Chromatin isolation

Chromatin was prepared from 20-50 mg of frozen thymus tissue powder that was dissolved in 2 mL Fixation buffer (Table 3.1), containing either 1 % Formaldehyde (FA) (Cat. No. 28906, Thermo Fisher, USA), 1 % FA + 1 mM DSG (disuccinimidyl glutarate) (Cat. No. 20593, Thermo Fisher, USA) or 1 % FA + 2.5 mM DSG, and incubated for exactly 5 min at RT. During this time the solution was transferred into a Dounce homogenizer, and the tissue was dissociated with the loose-fitting pestle A to get a homogeneous suspension. The solution was then transferred to a 2 mL Eppendorf tube and placed on a rotator at RT for the remaining of the total 5 min. To quench crosslinking, 200 µL of 1.25 M (final conc. 0,125 M) Glycine was added to the tissue solution and incubated for another 5 min on the rotator at RT. The tissue solution was then transferred to a 15 mL Falcon tube and centrifuged at 3000 rpm for 5 min at 4˚C. The supernatant was gently discarded, while the pellet was resuspended in 10 mL ice-cold PBS to wash off the crosslinking agents and was centrifuged at 3000 rpm for 5 min at 4˚C. The supernatant was gently discarded, and the cells were resuspended in 10 mL ice-cold complete Lysis buffer 1 (Lysis buffer

containing 1:100 Proteinase Inhibitor Cocktail (PIC), EDTA-free (Cat. No. 78439, Thermo Scientific, Netherlands)) and gently mixed 4˚C for 20 min on a rotator. After centrifugation at 3000 rpm for 5 min at 4˚C, the supernatant was gently discarded, the pellet was re-suspended in 10 mL complete Lysis buffer 2 (containing 1:100 PIC) and incubated with gentle mixing at 4˚C for 10 min on a rotator. Upon 5 min centrifugation at 3000 rpm at 4˚C, the supernatant was gently discarded, and the pellet was re-suspended in 1 mL of Sonication buffer (containing 1:100 PIC) and transferred to a 2 mL Dounce homogenizer. The sample was homogenized using tight pestle B and incubated on ice for 10 min. The chromatin solution was transferred to a 1 mL Covaris sonication tube (milliTUBE 1ml AFA Fiber, Cat. No. 520130, Covaris, USA) in a 24 tube rack (Rack 24 Place milliTUBE 1 ml, Cat. No. 500368, Covaris, USA). The chromatin was sonicated on Covaris Sonicator LE220 using the following sonication conditions: PIP: 450; Duty 20 %; CBP 300, for a variable time (7-25 min) until optimal DNA fragment length in the range of 200-700 bp fragments was achieved. To determine DNA fragment length, 25 µl aliquots were taken at different sonication times. They were first treated with diluted RNase cocktail at concentration RNase A = 0.04 U/mL; RNase T1 = 1.6 U/mL (RNase Cocktail™ Enzyme Mix, Cat. No. AM2286, Invitrogen, Lithuania) before being de-crosslinked by addition of 25 µl Elution buffer and 5 µl 5 M NaCl and incubated for 4 h at 65 °C with vigorous shaking. The DNA was isolated using QIAquick PCR purification kit (Cat. No. 28104, QIAgen, Germany) and analysed on Tapestation 2200 using DNA ScreenTape Analysis (D1000 ScreenTape/Reagents, Cat. No. 5067-5582/5067-5583, Agilent Technologies, USA). Chromatin with optimal fragment size was centrifuged at 4 ˚C, 10 000 rpm for 10 min to collect the supernatant containing soluble chromatin, which was stored -20 ˚C short term (up to a month), or at -80 ˚C, for longer-term storage.

Table 3.1. Crosslinking fixation buffers - The table contains an overview of reagents and volumes used for each crosslinking

Reagent	**1 % FA**	**1 % FA + 2.5 mM DSG**	**1 % FA + 1 mM DSG**
	Volume [µL]	**Volume [µL]**	**Volume [µL]**
16% FA	125	125	125
DSG (5M)	0	100	50
Fixation buffer	1870	1770	1830
Total	2000	2000	2000

3.4.2 Chromatin Immunoprecipitation

Chromatin immunoprecipitation was performed with consideration of lowering the background noise created by potential non-specific binding of DNA to an antibody or magnetic beads or antibody to non-target proteins. The measures made was to include a pre-clearing step to reduce the non-specific binding of chromatin to beads, include BSA in the ChIP reaction mix to prevent non-specific binding of antibodies, and wash the ChIP-bead complexes with high salt buffers to remove contaminants and non-specific binding. With ChIP-seq, it is recommended to use ChIP-validated antibodies. However, such antibodies were not available (ENCODE & Consortium, 2004). The α-AIRE antibodies tested in this study (AIRE polyclonal Cat. No. GTX13573, Genetex, AIRE monoclonal Cat. No. SC-37303, Santa Cruz Biotechnology, AIRE monoclonal Cat. No. 14953480, Thermo Fisher) were tested by Western blot prior to this study and yielded protein at the expected molecular weight. The antibody against FEZF2 (Cat.No Ab214186, Abcam) was the only antibody available that were predicted to react with human FEZF2 with no mismatch within the immunogen sequence. The antibodies against DEAF1 (DEAF1 polyclonal, Cat.No. LS-B10862, LSBio, DEAF1 monoclonal, Cat.No. MA5-21070, Invitrogen) were chosen based on immunogen recognition (the antibodies recognised amino acid sequence in the middle of the protein).

Before the chromatin was used in immunoprecipitation (IP) reactions, it was pre-cleared by incubating samples with protein G-coated magnetic Dynabeads™ (Cat. No. 520130, Invitrogen, Norway) (100 µL beads/1 mL chromatin solution) for 1-2 hours at 4˚C while rotating, after which the chromatin was recovered by placing the tube on a magnet and transferring the supernatant to a fresh tube. Pre-clearing aims to reduce background caused by unspecific binding between magnetic beads and chromatin.

ChIP reaction mixes were prepared as stated in Table 3.2, including chromatin solution, 3x Covaris dilution buffer, PIC, BSA and antibody. In case more than 100 µl chromatin solution was used, the other reagents were up-scaled proportionally. Antibodies and antibody amount per IP was found by optimisation. One per cent of the ChIP reaction was reserved as Input control before IP and stored at -20˚C until it was reverse crosslinked and purified together with the ChIP reactions.

Table 3.2. ChIP reaction mix - The table contains an overview of reagents, antibodies, and volumes used during IP.

Reagent	Volume [µL]		Antibody	Amount [µg]
Chromatin solution	100		α-AIRE (GTX/TF/CZ)	1-10
3x Covaris Dilution buffer	200		α-DEAF1 (LSbio/TF)	1-10
Protein inhibitor cocktail (PIC)	1.5		α-FEZF2	1-10
5% BSA	6		α-CTCF	2.5
Antibody	X		IgG (Goat/Rabbit)	1-10
Final Volume	**307.5**			

The assembled ChIP reactions were incubated overnight at 4˚C on a rotator. On the next day, antibody-bound chromatin fragments were isolated after the reactions were incubated at 4˚C on a rotator for 1-2 h with magnetic protein G-coated Dynabeads (30 µL per reaction). The protein G bind to the Fc domain on immunoglobulins, leaving the Ig binding site open to bind to DNA-bound protein. The beads were washed with 350 µL Wash buffer 1, resuspended by pipetting, incubated at 4˚C for 5 min on a rotator, followed by placing the tubes on a magnet and discarding the supernatant. The same procedure was repeated with Wash buffers 2, 3 and 4. After the last wash, the samples were eluted from the beads by resuspending the beads in 100 µL Elution buffer and incubating for 30 minutes on a rotator at RT. The samples were placed on a magnet, and the eluate was transferred to a new 1.5 mL Eppendorf tube. After adding 4 µL 5M NaCl the samples were incubated at 65˚C for 4 hours or overnight shaking to reverse the crosslinks. Input samples were filled up to 100 µL with Elutionbuffer and were supplemented with 4 µL 5 M NaCl, after which they were de-crosslinked together with the ChIP DNA samples. DNA was isolated with Zymo ChIP DNA and concentrator kit (Cat. No. D5205, Zymo Research, USA) and eluted in 25-50 µL nuclease-free H2O. The concentration was measured using Qubit® 2.0 Fluorometer (Cat. No. Q32866, Invitrogen, USA) and Qubit® dsDNA HS Assay Kit (Cat. No. Q32851, Invitrogen, USA) using 1 µL eluted ChIP DNA.

3.4.3 Quantitative PCR (qPCR)

Quantitative PCR (qPCR) analysis was performed with the ChIP DNA and Input DNA using primers specific for AIRE, DEAF and FEZF2 target genes, as found in the literature (see Table 3.5). The primers were designed by NCBI Primer-BLAST online tool (www.ncbi.nlm.nih.gov/tools/primer-blast) using genomic regions extracted from the UCSC

Genome browser (www.genome.ucsc.edu) with human genome version GRCh38/hg38. Quantitative RT-PCR reactions were assembled as described in Table 3.3., following the thermocycler program in Table 3.4. The PCR reactions were performed on QuantStudio™ 12K Flex Real-Time PCR System (Cat. No. 4471134, Applied Biosystems™, USA) machine using 384-well plates (MicroAmp™ Optical 384-Well Reaction Plate with Barcode, Cat. No. 4326270, Applied Biosystems™, USA) and foil (MicroAmp™ Optical Adhesive Film, Cat. No. 4360954, Applied Biosystems™, USA), applying 40 cycles and melting curve analysis.

Table 3.3. Reaction mix for qPCR

Reagent	Volume [µl]
2x PowerUp™ SYBR™ Green Master Mix	5
10 pM Primer mix (F+R)	1
Nuclease free H_2O	2
Diluted ChIP/Input DNA sample (1:5/ or 1:10)	2
Final Volume	**10**

Table 3.4. Quantitative PCR thermocycler program

Stage	Temperature [˚C]	Time	Cycles
Activation	50	2 min	1x
	95	2 min	1x
Amplification cycle	95	15 sec	40x
	60	1 min	

The CT output values after qPCR were used to analyse and confirm the recovery of target genes specific to the studied transcription factor. The input sample represents the amount of DNA used in the ChIP before IP. Here 1% of chromatin solution per IP is used as input.

$$\text{Recovery (\% of input)} = 100 * 2^{((Ct\ input - log_2 100) - Ct\ IP)}$$

Table 3.5. Primers used in qPCR

Transcription factor	Target Gene	Forward primer (5' -> 3') (F)	Reverse primer (5' -> 3') (R)	Product length (bp)
DEAF1	*EIF4G3***	AACGAGCAGAGCATCCAACA	AAAGAGCCGAAGAGGCTTCC	128
	*HTRA1****	CGGCGATTTGCAGGAACTTT	GGACTCAGTTTCCCCGTCTG	151
	*HOXD4****	ATTCTCCCTTGAGCTGTGCC	GTAGTTCCAGTCCCCAGCAC	152
	*PCK1****	ACCAGCAGCTCTTGGTACAC	GACTTCGAGCCCTCAACCAA	180
	*AMBP****	ATTACAGCCCCAAGGACAGC	TGACTCAGCAGAAAGGCCAG	203
	*ANAPEP****	CCACATCTGGACGGCATCTG	TTATCTGCACCACCCGTCTG	125
FEZF2	*FN1**	GAGTGGCTGGACTTGTGTGA	CAGCTGCAAGGTCGTGGATA	192
	*COL19A1**	CCTACTCGTGAACGTTTGCC	GATGTCGCCTTAGAGCTGGG	160
	*KRT10**	CAGAGCTCCCACGGCTAAAA	CCCTGGGCTAAACAGCATCA	209
	*RESP18**	AGCGATGAAACAGTAGGCACTC	AGTGGTTTGGACCAAGGCGA	205
	*MAOA**	AGCTTGGACGACACCTCCTA	TGGTAGACTTGGGGATCCGA	247
	*TIMD2**	GGGATGGCGTGGACATACTT	GCTTTGGCCCCATGAGGAAT	213
	*CALB1**	GCGAGAGTTTGGGCTTGTTG	CCGTCATCAGCTTTCCCCTT	194
	*NOL4**	GTGACCAGGACCATCCCAAC	TCTAGGGACCACCCCAATTCA	115
	*FABP9**	GAGACCAGCTTCCAGGTTCC	TGAGTCAAATGCTGGTGCCT	210
	*ALDHLA2**	GTCCTCCGTTCTCCTCCTCC	GGATCCTCATTCTGAGCCCG	170
	*LRAT Short**	ACGGCCATAAAAAGTCGCAG	CACAATGGGGAATGAGACGGA	164
	*LRAT Long**	TGGTGACGCATTCGCATGTA	GGTTTCCCGGACCGTTAAGA	87
	*TLCD2**	GGATGTTCGCCAATAGGGGT	GATGACGCTCAGCAGGTGAA	160
	*LONRF3**	CCTCCCAACCTGCTTCCTTT	GGGTGGGGAGAAACTCAGTG	240
AIRE	*S100A8**	GCATTGGTGGTCAGGCCATA	CTTTTGCGGTCTTTGGACCC	132
	*S100A8 Proximal enhancer**	GTTGGTAGCTTTCCCTCCCC	AATCTGGACTTGGGCAGTCG	200
	*S100A8 Distant enhancer**	GGGTTGAGGAGGCTACACAC	GATGGCTTCCTGCCTCAAGT	132
	*INS**	TTGGTCGTCAGCACCTCTTC	CCTCCAGCTCTCCTGGTCTA	162
	*AIRE*****	CTGCCAAGGATGACACTGCCA	CGAAGGTGTGCTCGCTCAGAAG	81
	*Enhancer******	AGACGCCGGCATTATTCCTC	CCGCTTCGTACTGTCGCTTA	138
DEAF1 And AIRE	*GRIN2C****	CAAAGTCACTAGCACCGCCA	CCTCGGATGGTGGCGTTATC	142
	*DTX4****	CAATGCGTGTCTCCAGTTCG	CGTGTGTTTCCCTGGTTCCT	187
	*BCL9L****	CACACTCCTGATGCTCCCTG	CCTCTGCTCTTCCCTGGTTG	243
Negative control	*PHF11******	CGGCTAGTGACTCAAAGCGA	ATAGCGAGATCCCGGAAGTG	80
CTCF	*H-19(1)******	GAGTGTCTATCTCTGACAACC	TGCAGGCTCACACATCACAGC	118
	*H-19(2)******	TCTCGGCCTAAGCGTGAGACC	CACAACTACAACCGATTCTGC	141
	*HTR7p1******	AGGGCAGGTAAGGGAACAGA	CCTGAGAGGACAGCTTGCTT	216

*(genes found in (Takaba et al., 2015))

**(genes found in (Jensik et al., 2014))

***(genes found in (L. Yip et al., 2009))

****(genes found in (Org et al., 2009))

*****(genes found in (Bansal et al., 2017))

Primers from Eurofins Genomics and Integrated DNA Technologies (IDT)

3.4.4 Library preparation and sequencing

Before library preparation, the concentration of DNA in the samples was measured using Qubit® 2.0 Fluorometer (Cat. No. Q32866, Invitrogen, USA) and Qubit® dsDNA HS Assay Kit (Cat. No. Q32851, Invitrogen, USA) using 1 µL ChIP DNA.

The library was prepared using Accel-NGS® 2S PLUS DNA Library Kit with Dual Indexing from Swift Bioscience (Cat. No. 21096, Cat. No. 29096, Swift Bioscience, Sweden). The library preparation was performed following the kit workflow. For each sample listed in Table 3.6., 10 ng DNA was used in the library preparation. As the FEZF2 and AIRE samples had less than 10 ng, all DNA available was used. The sequencing libraries were prepared with the 12 ChIP samples. TapeStation analysis was performed to determine the amount of DNA present after library preparation. The samples that produced libraries were selected for sequencing.

Table 3.6.: **Sample information for library preparation**

Tissue	Anti-X	Cons. [ng/µL]	Total [ng]	Produced libraries
3005	FEZF2 1µg	0,566	9,056	Yes
	AIRE 5µg	0,6	9,6	Yes
	DEAF1 5µg	4,44	71,04	No
	INPUT 1%	6,52	104,32	Yes
	IgG Goat 5µg	N.D.	N.D.	No
	IgG Rabbit 5µg	0,18	2,88	Yes
3005	DEAF1 5µg	2,22	48,84	Yes
	INPUT 1%	5,76	126,72	Yes
3001	DEAF1 5µg	6,24	137,28	Yes
	INPUT 1%	8,76	201,48	Yes
3006	DEAF1 5µg	7,6	136,8	Yes
	INPUT 1%	9,92	198,4	Yes

N.D. – Not detectable

Sequencing of the libraries was performed on NovaSeq™ 6000 Sequencing System (Illumina), SP flow cell, 150 bp paired-end reads, with 1% PhiX spike-in sequencing control. Sequencing performed by Norwegian sequencing centre (NSC) at OUS.

3.5 ChIP DNA analysis

ChIP-seq data in the FASTQ format were analysed using the Galaxy web platform (www.usegalaxy.org). The ChIP-seq data were loaded into Galaxy and mapped to the human reference genome (GRCh38) using Bowtie 2 (Langmead, Trapnell, Pop, & Salzberg, 2009), allowing no mismatches and including trimming of adapters and low-quality reads. Identification of enriched regions (peaks) for each ChIP-seq Transcription Factor IP or IgG IP file compared to ChIP-seq Input DNA file was performed using Model-based Analysis of ChIP-Seq (MACS) (Y. Zhang et al., 2008). MACS peaks are called based on ChIP-seq tags that represent the 5' end reads of ChIP-DNA fragments sequenced. ChIP-DNA fragments are sequenced on both DNA-strands, creating a tag density on both sides on the binding site. MACS shift these tags towards the 3' direction to find the midpoint between the tags that represent the protein-DNA interaction site. To call a peal MACS uses a dynamic parameter, λ_{local}, that is defined for each peak. The lambda parameter is estimated from the input sample and measures maximum value across three windows in the genome, 1 kb, 10 kb, and 15 kb.

$\lambda_{local} = \max(\lambda_{background}, \lambda_{1k}, \lambda_{5k}, \lambda_{10k})$

The default minimum fold-enrichment (*mfold)* is 35, meaning the peaks with fold-change over 35 are used to build the shift-model. MACS peaks are called when regions with more tags relative to the random tag distribution, represented by the input (1% chromatin solution), have a higher enrichment than the set *mfold* (Y. Zhang et al., 2008). MACS output (.bed format) files were exported, and peaks were visualized using Integrated Genome Viewer (IGV, http://software.broadinstitute.org/software/igv/) with GRCh38 set as the reference genome.

3.6 CUT&Tag

As an alternative to the ChIP-seq method Cleavage Under Targets and Tagmentation (CUT&Tag) was tested (Henikoff et al., 2019). The CUT&Tag method was created to detect the epigenomic profile of small cell samples with reduced background compared to conventional ChIP-seq. The method uses a transposome consisting of hyperactive Tn5 transposase-protein A (pA-Tn5) fusion protein loaded with sequencing adapters. These get inserted into the DNA in the proximity of the transcription factor, being previously bound by a specific antibody by in-situ

antibody incubation in permeabilized cells. The target adapter-containing DNA fragments can then become a library by PCR using complementary primers that complete the sequencing adapter ends of the fragments (Henikoff et al., 2019). In more detail, the procedure followed was the following:

Cells were thawed in 37 °C for 30 seconds to 1 minute, transferred to 10 mL RPMI-1640 medium containing 20% FBS and centrifuged at 600 x g for 5 minutes using a swinging bucket rotor at RT. The supernatant was removed, and cells were resuspended in 10 mL PBS to wash and centrifuged at 600 x g for 5 minutes using a swinging bucket rotor at RT. The supernatant was removed, and cells were resuspended in 1 mL PBS, counted, and diluted to 500 000 cells/mL. In order to test the tagmentation in the presence of crosslinking, four reactions with 500 000 cells each were set up as follows: 1) unfixed: 2) fixed with 0,1 % FA;3) fixed with 1 % FA; 4) fixed with 1 % FA and 2.5 mM DSG. The crosslinking reactions incubated for 2 minutes before being stopped with 0.125 M glycine. The samples were centrifuged using a swinging bucket rotor at 1300 x g for 4 minutes at 4 °C. The supernatant was removed, and cells were resuspended in Wash buffer to a concentration of 1 million cells per mL.

Concanavalin A-coated beads (Cat. No. 86057, Polysciences, Inc, Germany) (10 µL per 500 000 cells) were washed with 1 mL Binding buffer. Ten microliter beads were added to each cell solution, incubated for 5-10 minutes while on rotator at room temperature and placed on a magnet. The supernatant was withdrawn and discarded. The cells, now bound on ConA-beads, were resuspended in 800 µL ice-cold Antibody buffer. Primary antibody anti-H3K27me3 (Cat. Nr. C15410195, Diagenode) was added (1:100) and incubated overnight with gentle mixing at 4 °C. On the next day, the samples were quickly spun and placed on a magnet to withdraw the liquid. The secondary antibody, guinea pig α-Rabbit IgG (Cat. No. ABIN101961, antibodies-online) was mixed 1:100 in 100 µL Dig-wash buffer and was added to each sample. The samples were incubated with gentle mixing for 30 – 60 min at RT. After a quick spin, the tubes were placed on a magnet stand to clear, and the supernatant was withdrawn and discarded. The beads were washed with 1 mL Dig-wash buffer twice by resuspending beads, placing on a magnet and removing the supernatant. After the second wash, the beads were resuspended in 100 µL pA-Tn5 mix. The pA-Tn5 adapter complex was diluted in Dig-300 buffer to a final concentration of 1:250 and was incubating with the beads-cell suspension for 1 hour at RT. After a quick spin, the bead-bound cells were washed with 1 mL Dig-300 buffer twice by resuspending beads, placing on a

magnet and removing the supernatant. After the second wash, the beads were resuspended in 300 µL Tagmentation buffer (Dig-300 buffer, 10 mM MgCl2) and incubating for 1 hour at 37 °C. To stop tagmentation and solubilize DNA, 10 µL 0.5 M EDTA, 3 µL 10 % SDS and 2.5 µL 20 mg/mL Proteinase K were added to each sample, and the samples were incubated for 1 hour at 55 °C. DNA was extracted with the Phenol-Chloroform isolation and Ethanol precipitation method as follows. To each sample 300 µL phenol:chloroform:isoamyl alcohol (25:24:1 v/v, pH = 8.05, Cat. No. 15593031, Invitrogen) was added, the sample was mixed and centrifuged at 16,000 x g for 3 minutes at RT. Then 300 µL Chloroform was added, mixed and centrifuged at 16,000 x g for 3 minutes at RT. The aqueous layer was transferred to a fresh 1.5 mL Eppendorf tube containing 750 µL 100% Ethanol. The samples were placed in a -80 °C freezer overnight to allow DNA to precipitate. The samples were centrifuged for 30 minutes to 1 hour at 4 °C before removing the supernatant and washing the pellet with 1 mL 100% Ethanol and centrifuged at 16,000 x g for 1 min at 4 °C. The ethanol was removed, and the pellets were left to air dry. DNA was dissolved in 25-30 µL low EDTA-TE buffer (1 mM Tris-HCl pH 8, 0.1 mM EDTA).

The eluted DNA was subjected to PCR, by being mixed with indexed primers and NEBNext HiFi 2x PCR Master mix (see Table 3.7.) in a thin-wall 0.5 ml PCR tube, using a different indexed primer for each sample. The PCR samples were mixed, spun and placed in Thermocycler to begin the cycling program with a heated lid (Table 3.8.).

Table 3.7. PCR reaction mix

Reagent	Concentration [ng/µl]	Volume [µL]
DNA Sample		21
i5 primer	10 µM	2
Indexed i7 primer	10 µM	2
NEBNext HiFi 2x PCR Master mix		25
Total volume		50

Table 3.8. PCR thermocycler program

Stage	Temp (˚C)	Time	Cycles
Activation	72	5 min	1x
Denaturation	98	30 sec	1x
	98	10 sec	13x
Annealing/Extension	63	10 sec	
	72	1 min	1x
Hold	8	∞	

The libraries (PCR products) were cleaned up by adding 1.3x volumes (65 µL) of SPRI bead slurry, mixing by pipetting up and down and incubating for 5 min. The tubes were placed on a magnet and allowed to clear before carefully withdrawing the liquid. While on the magnet two rounds of 200 µL, 80% ethanol were added without disturbing the beads. The liquid was withdrawn, the tubes were removed from magnet stand and resuspended in 25 µL 10 mM Tris-HCl, pH 8, to elute the DNA from the beads. After 2 minutes incubation at RT, the tubes were placed on a magnet stand and allowed to clear. The eluate was transferred to a fresh 1.5 ml Eppendorf tube. Agilent 4200 TapeStation with D1000 reagents was used as previously described to determine the size distribution and concentration of libraries (section 3.2).

4 Results

4.1 AIRE, FEZF2, and DEAF1 expression in thymus

The research group (Gabrielsen et al., 2019) has formerly performed RNA-seq of the gene expression from different types of APCs sorted from six dissociated human thymi. The data was re-analysed by plotting expression level of each transcription factor for each biological replicate (each individual indicated by number nX in Figure 4.1.) and cell-type (cTECs, mTECs, CD1c dendritic cells, CD123 dendritic cells, CD141 dendritic cells, and CD19 B-cells) to observe the expression levels and in order to assess the abundance of the transcription factors FEZF2, AIRE and DEAF1 in each cell type. AIRE and FEZF2 were expressed in cTECs and mTECs while DEAF1 were expressed in cTECs, mTECs, dendritic cells and B-cells (Figure 4.1.). AIRE and FEZF2 were almost exclusively expressed at a high level in mTECs. Between each replicate of an enriched cell sample, there was variance in expression of the transcription factors. Because AIRE and FEZF2 are only expressed in TECs, and there is a variance in expression between each sample, sequencing of ChIP samples should include as many biological replicates as possible, at least five.

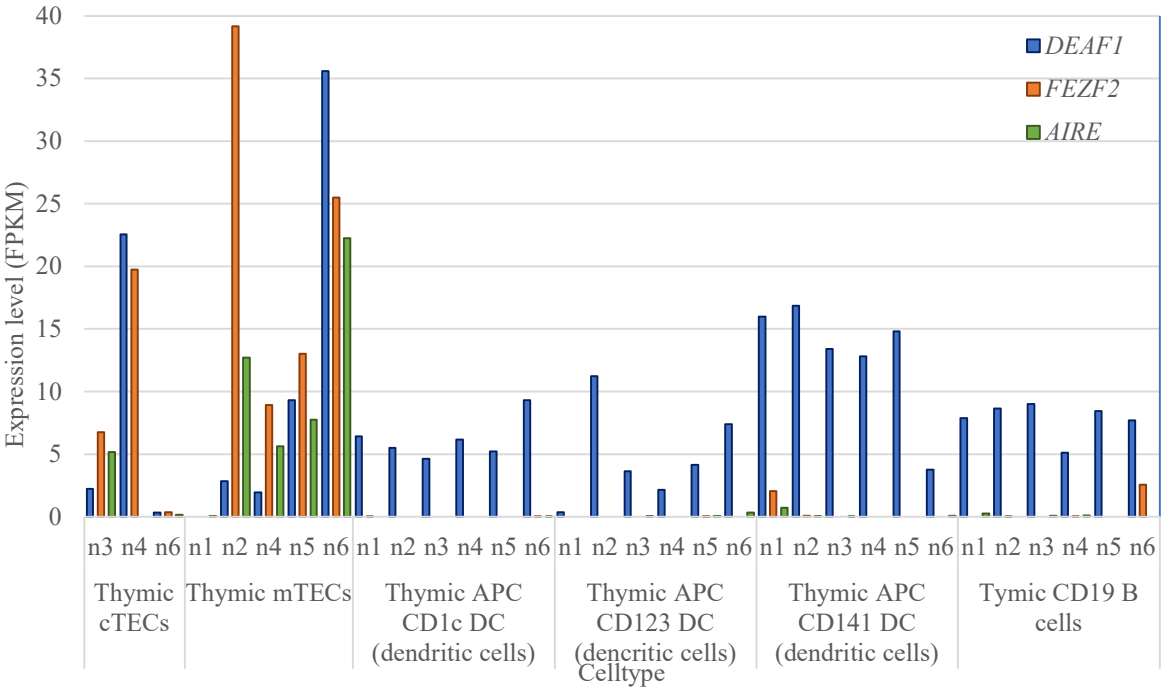

Figure 4.1. DEAF1 (blue), FEZF2 (orange), and AIRE (green) expression level assessed by RNA-sequencing of different APCs. Expression levels are measured in Fragments Per Kilobase of transcript per Million mapped reads (FPKM). Data points are the expression level of the three transcription factors in the different cells extracted from different individuals (nX). Abbreviations: Cortical thymic epithelial cell (cTEC), medullary thymic epithelial cell, dendritic cell (DC), antigen-presenting cell (APC). RNA-sequencing data collected from (Gabrielsen et al., 2019).

4.2 Chromatin Immunoprecipitation Optimisation

The aim of this study was to construct a ChIP-seq strategy to investigate transcription factor AIRE, FEZF2 and DEAF1 in mTECs and APCs using human thymic tissue. The optimised parameters were crosslinking, sonication time, antibody clone, and amount antibody per IP.

4.2.1 Crosslinking and Sonication Optimisation

AIRE, FEZF2, and DEAF1 have different protein domains where AIRE binds primarily to protein, acting as a transcriptional coactivator/corepressor (Kumar et al., 2001), and both DEAF1 and FEZF2 bind directly to DNA (DNA-binding proteins) (Jensik et al., 2014). FA crosslinks protein to DNA, and DSG crosslinks protein to protein. The chromatin crosslinking had to be optimised for the two DNA-binding methods as well as to reach a balance in which the transcription factors are bound to DNA, but the crosslinking does not affect fragmentation or

cause epitope masking. Inefficient crosslinking and fragmentation can increase the background signal, which leads to a decrease in the signal to background ratio of an experiment, decreasing the sensitivity of the ChIP-seq assay. The signal fold-enrichment should reach five-fold in order to get significant results through sequencing (ENCODE & Consortium, 2004). Strong crosslinking (high FA concentration, in the presence of protein-protein crosslinker, longer crosslinking time) increases chromatin rigidity, and thus it becomes more resistant to sonication and less accessible to antibodies (epitope masking). Excessive fragmentation leads to protein degradation and dissociation of protein from the DNA, which would impair the efficiency of immunoprecipitation. Therefore, we optimised the combined crosslinking and sonication conditions aiming to capture both direct and indirect interactions with DNA and to be able to reach chromatin fragment lengths (of ca. 300 bp on average) that would enable target region enrichment with good resolution and high specificity. Several rounds of ChIP optimisation were conducted using tissue sample TC3005. The optimised strategy was used to perform sequencing of ChIP samples from three different thymic tissues (TC3005, TC3001, and TC3006). The crosslinking and sonication conditions during chromatin preparation were first optimised, aiming at yielding the highest specific recovery of target genes by the protein of interest.

For determining the optimal crosslinking conditions, three tissue aliquots (20 mg each) were fixed in three different crosslinking buffers containing 1 % FA; 1 % FA and 2.5 mM DSG; and 1 % FA and 1 mM DSG. We optimised the chromatin fragmentation by establishing the minimum sonication time for achieving an average fragment length of 300 bp. The samples were sonicated for a total of 15 minutes in three 5-minute intervals.

TapeStation analysis showed that the chromatin samples required ca. 15 min of sonication to reach an average fragment size of 300 bp (Figure 4.2.). After 5 min of sonication, the 1 % FA and 1 mM DSG fixed chromatin had started to fragment and the sample had an average fragment length of 621 bp (Figure 4.2., right corner), while the two other chromatin samples had fragments over 1500 bp. After 10 min, the chromatin fixed with 1 % FA had an increasing number of fragments from 100 bp to over 1500 bp. After 15 min sonication, the chromatin fixed with 1 % FA + 1 mM DSG reached an average fragment length of 306 bp with a DNA concentration of 2.16 ng/µL, while chromatin crosslinked with 1 % FA and 1 % FA + 2.5 mM DSG required additional 5 min of sonication finally yielding 1.93 ng/µL and 2.19 ng/µL DNA at an average of 300 bp, respectively. Based on these results, the sonication strategy for all following ChIP

experiments was to sonicate the chromatin for a minimum of 10 min, analyze the fragment size and sonicate additionally in 5-min intervals as needed until reaching an average fragment size of 300 bp.

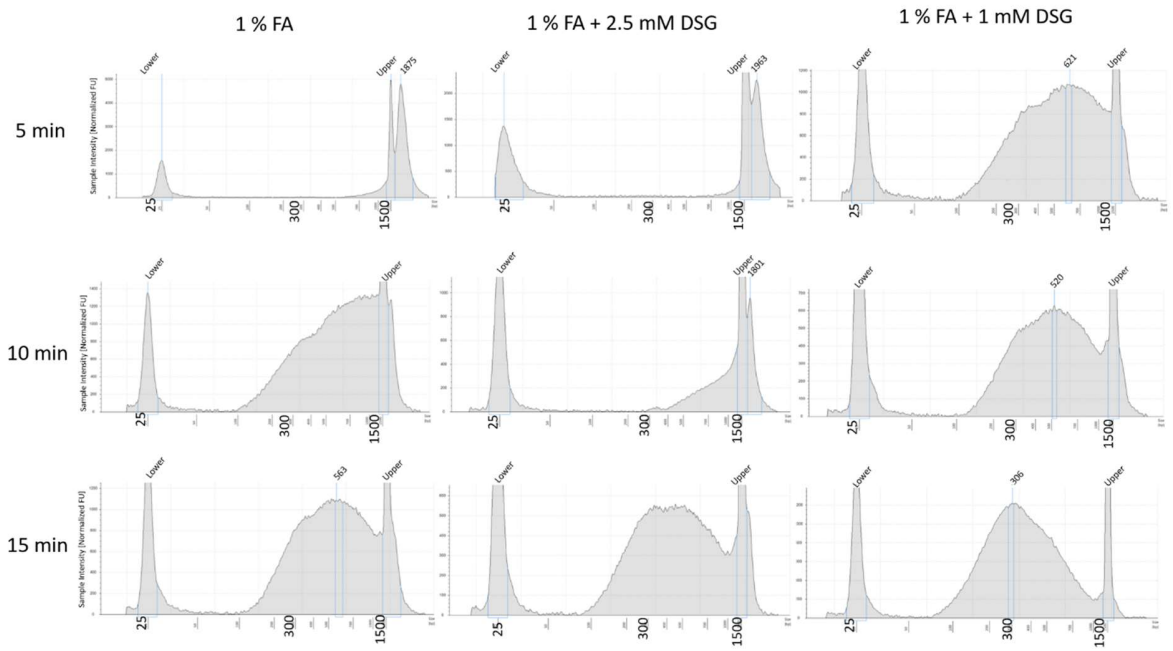

Figure 4.2. Fragment size of chromatin after sonication in three 5 minute intervals. Three tissue samples fixated in one crosslinking buffer each containing: 1 % FA, 1 % FA + 2.5 mM DSG, 1 % FA + 1 mM DSG. Samples are sonicated for 15 minutes total in three 5 minute intervals. Abbreviations: Formaldehyde (FA), Disuccinimidyl glutarate (DSG), base pair (bp).

4.2.2 Chromatin Immunoprecipitation optimisation using α-CTCF

The first step of this study was to ensure that the ChIP strategy was plausible with thymic tissue and determine the optimal crosslinking reagent that allowed for sufficient fixation of the transcription factor to the DNA while maintaining the epitopes available for binding to antibodies used for the immunoprecipitation. This was tested by targeting a ubiquitously expressed transcription factor, CTCF, as a positive control using an α-CTCF antibody, and identify the fixation buffer that recovered the highest amount of chromatin. We have tested the binding to the following target genes; H19-1, H19-2, and HTR7p1, as they are known targets of CTCF.

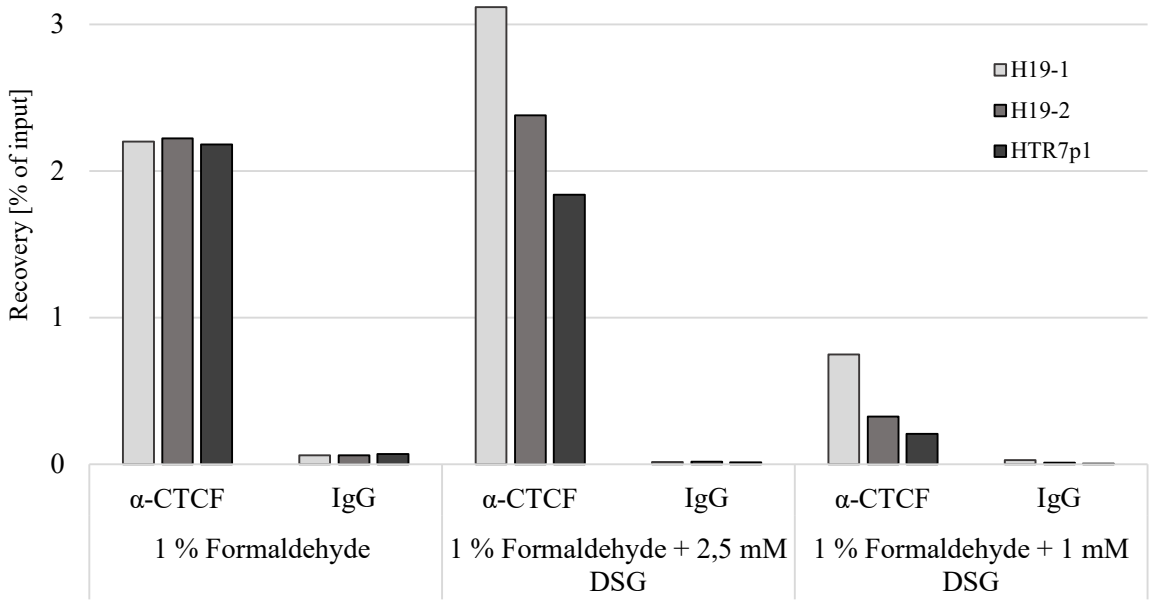

Figure 4.3. ChIP-qPCR with optimisation of crosslinking reagents with the α-CTCF antibody. Crosslinking conducted with three tissue samples crosslinked with one fixation buffer each: 1 % FA, 1 % FA + 2.5 mM DSG, and 1 % FA + 1 mM DSG. The genes H19-1, H19-2, and HTR7p1 are targets for CTCF. IgG Rabbit used as a negative control. Abbreviations: Disuccinimidyl glutarate (DSG), Formaldehyde (FA), CCCTC-binding factor (CTCF).

Crosslinking with 1 % FA and 1 mM DSG fixation buffer was not optimal as the target gene recovery overall tested regions were lower as compared to the other crosslinking conditions (Figure 4.3.). The crosslinking with 1 % FA + 1 mM DSG appear to not stabilise the interactions necessary to efficiently immunoprecipitate the targeted chromatin regions. Crosslinking with 1% FA gave an even recovery of all gene targets. However, it also exhibited a slightly higher background (IgG signal) than the two other crosslinking conditions, implying slightly more insufficient crosslinking and fragmentation than with 1 % FA and 2.5 mM DSG. Crosslinking with 1 % FA and 2.5 mM DSG gave a high and more variable signal across the three gene targets while showing the lowest IgG background signal, suggesting more specific enrichment. Based on these results, we concluded that the optimal crosslinking condition for human thymic tissue was to use either 1 % FA alone or 1 % FA+ 2.5 mM DSG containing buffers.

4.2.3 Chromatin Immunoprecipitation optimisation with α-AIRE

For optimisation of AIRE ChIP, we determined the optimal combination of antibody amount, the antibody clone and the crosslinking conditions. To test the crosslinking reagents for AIRE, the chromatin solutions from section 4.2.1 were used in an IP with 1 µg α-AIRE antibody and IgG Rabbit using 200 µL chromatin solution and scaling the other reagents as needed. We decided to use a polyclonal antibody AIRE antibody produced by rabbit immunization (GeneTex (GTX)) in this first optimisation step as it was polyclonal α-AIRE antibody, which would reduce the risk of epitope masking, and was confirmed by the manufacturer to perform well with human thymi.

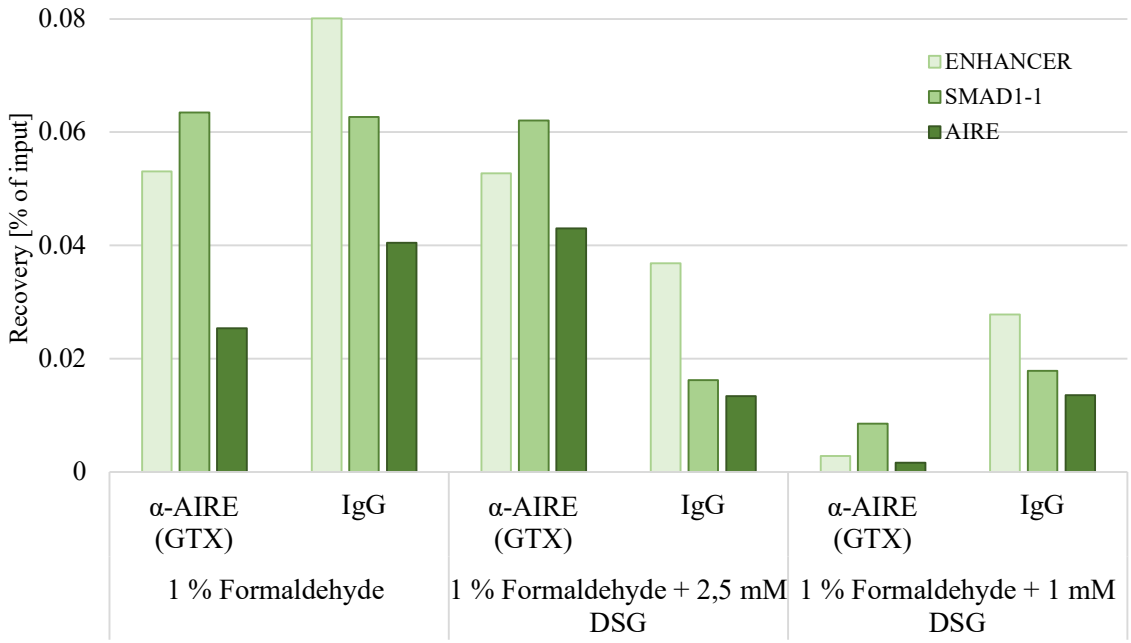

Figure 4.4. Optimisation of crosslinking reagents with α-AIRE (GTX) antibody by ChIP-qPCR. ChIP with anti-AIRE (GeneTex, USA) to test crosslinking conducted with three different samples containing 20 mg aliquot of frozen tissue TC3005 and one crosslinking condition each: fixation buffer containing 1 % FA, 1 % FA + 2.5 mM DSG, and 1 % FA + 1 mM DSG. The genes SMAD1-1, AIRE, and gene region called "ENHANCER" are targets for AIRE. IgG Rabbit used as a negative control. Abbreviations: Autoimmune regulator (AIRE), GeneTex (GTX), Disuccinimidyl glutarate (DSG), Formaldehyde (FA).

High IgG background signal was present for all crosslinking conditions compared to the signal from AIRE GTX (Figure 4.4.). *AIRE* is not ubiquitously expressed as *CTCF*, but only expressed in mTECs. The recovery of AIRE target regions was, therefore, lower compared to CTCF target regions (below 1 %). Crosslinking with 1 % FA and 1 mM DSG resulted in lower signal for AIRE GTX than IgG. Both AIRE GTX and CTCF had low recovery overall genes with crosslinking using 1 % FA and 1 mM DSG compared to IgG signal and the other two crosslinking conditions. This fixation method is not optimal and is not included in the ChIP-qPCR crosslinking optimisation with the α-FEZF2 antibody. Crosslinking with 1 % FA and 2.5 mM DSG resulted in the highest α-AIRE to IgG signal ratio. The rabbit IgG had a high enrichment for all gene loci compared to AIRE GTX over all genes with 1 % FA fixation. Given the high enrichment for IgG, the experiment with 1 % FA fixation with AIRE and IgG was conducted one more time to control for possible contaminants that might result in high background. For this optimisation step, we tested the GTX antibody (α−AIRE GTX) one more time and added two new AIRE antibodies from other producers, Santa Cruz Biotechnology (α−AIRE CZ), and ThermoFisher (α−AIRE TF). The α−AIRE CZ and α−AIRE TF were both monoclonal antibodies. qPCR was applied to measure the amount of precipitated AIRE genes and two other AIRE dependent genes; ENHANCER and the SMAD1-1 gene.

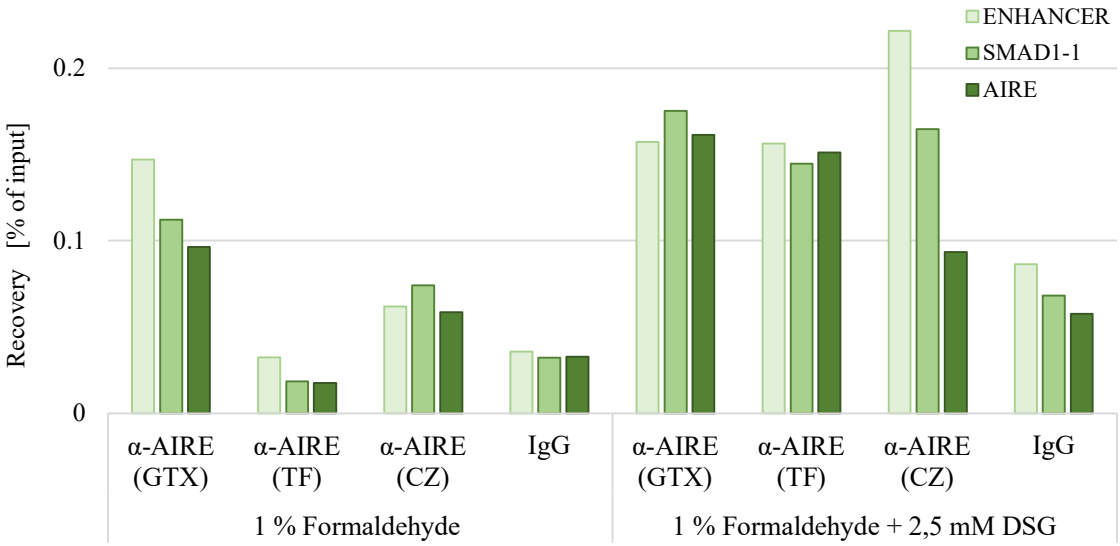

Figure 4.5. ChIP-qPCR with optimisation of AIRE antibody and fixation method for AIRE. Two tissue samples fixated in one fixation buffer each: 1 % FA or 1 % FA + 2.5 mM DSG. Sheared chromatin from each tissue sample split into four immunoprecipitation reactions. Antibodies used are AIRE from GeneTex (USA) (AIRE GTX), AIRE TH724 from ThermoFisher (USA) (AIRE TF), and AIRE from Santa Cruz (USA) (AIRE CZ). *ENHANCER*, *SMAD1-1* and *AIRE* are AIRE dependent genes. IgG Rabbit used as a negative control. Abbreviations: Autoimmune regulator (AIRE), Disuccinimidyl glutarate (DSG), Formaldehyde (FA).

The AIRE GTX antibody performed the best for both fixation conditions (Figure 4.5.). AIRE TF and AIRE GTX have higher expression levels with all AIRE dependent genes, while AIRE CZ had varied expression levels over all genes using 1 % FA + 2.5 mM DSG fixation compared to 1 % FA fixation. AIRE GTX had the highest recovery with both 1 % FA fixation and 1 % FA and 2.5 mM DSG fixation. As crosslinking with 1 % FA and 2.5 mM DSG resulted in the highest gene recovery for all three AIRE antibodies and was concluded as the optimal crosslinking solution for AIRE. The AIRE GTX was selected for further ChIP experiments.

Further α-AIRE antibody optimisation included antibody titration in determining the amount of AIRE GTX per IP. The antibody amounts included 1 µg, 5 µg, and 10 µg α-AIRE GTX. The recovery increased proportionally to the antibody amount from 1 µg to 5 µg α-AIRE antibody per IP, indicating that more chromatin was captured during IP (Figure 4.6.) Increasing the α-AIRE antibody amount further from 5 µg to 10 µg antibody decreased the recovery of the genes. Based on these results, the optimal amount of α-AIRE antibody per IP was decided to be 5 µg α-AIRE GTX.

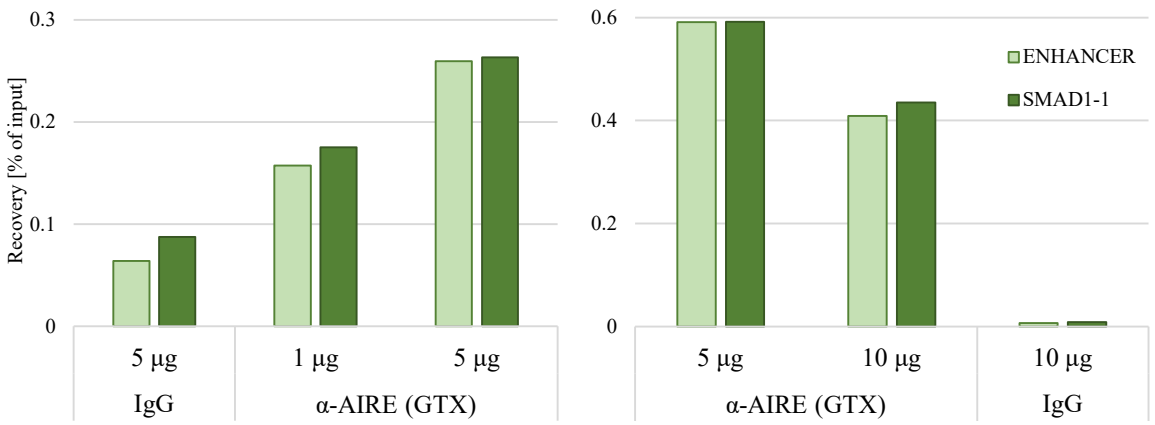

Figure 4.6. ChIP-qPCR with optimisation of the amount of α-AIRE GTX antibody per IP. Two separate ChIP-qPCR experiments testing IP's conducted with three different antibody amounts: 1 µg, 5 µg, and 10 µg. IgG Rabbit (5 µg) and IgG Goat (10 µg) is used as a negative control. The genes ENHANCER and SMAD1-1 are gene targets for AIRE. Abbreviations: Autoimmune regulator (AIRE), GeneTex (GTX).

4.2.4 Chromatin Immunoprecipitation optimisation with α-FEZF2

In order to optimise the ChIP strategy for FEZF2, crosslinking and antibody amount were determined. The crosslinking optimisation experiment was performed with two crosslinking conditions: 1 % FA fixation and 1 % FA and 2.5 mM DSG fixation. Immunoprecipitation with the two crosslinking condition samples was set up with different amounts of the antibody for each IP including 1 µg, 5 µg, and 10 µg α-FEZF2 antibody and 10 µg IgG Rabbit.

FEZF2 IP showed the highest recovery with 1 % FA and 2.5 mM DSG and using 1 µg α-FEZF2 antibody per IP (Figure 4.7.). Fixation with 1 % FA alone resulted in the high background (rabbit IgG signals), and reduced enrichment for FEZF2 dependent genes. Fixation with 1 % FA and 2.5 mM DSG resulted in lower background signal and increased recovery across all genes for the IP with 1 µg and 5 µg α-FEZF2 antibody. The IP with 10 µg α-FEZF2 antibody showed reduced recovery and was not optimal for ChIP against FEZF2. For further testing, 1 µg α-FEZF2 antibody was used as it showed sufficient recovery, and as 5 µg or 10 µg α-FEZF2 antibody did not capture more chromatin.

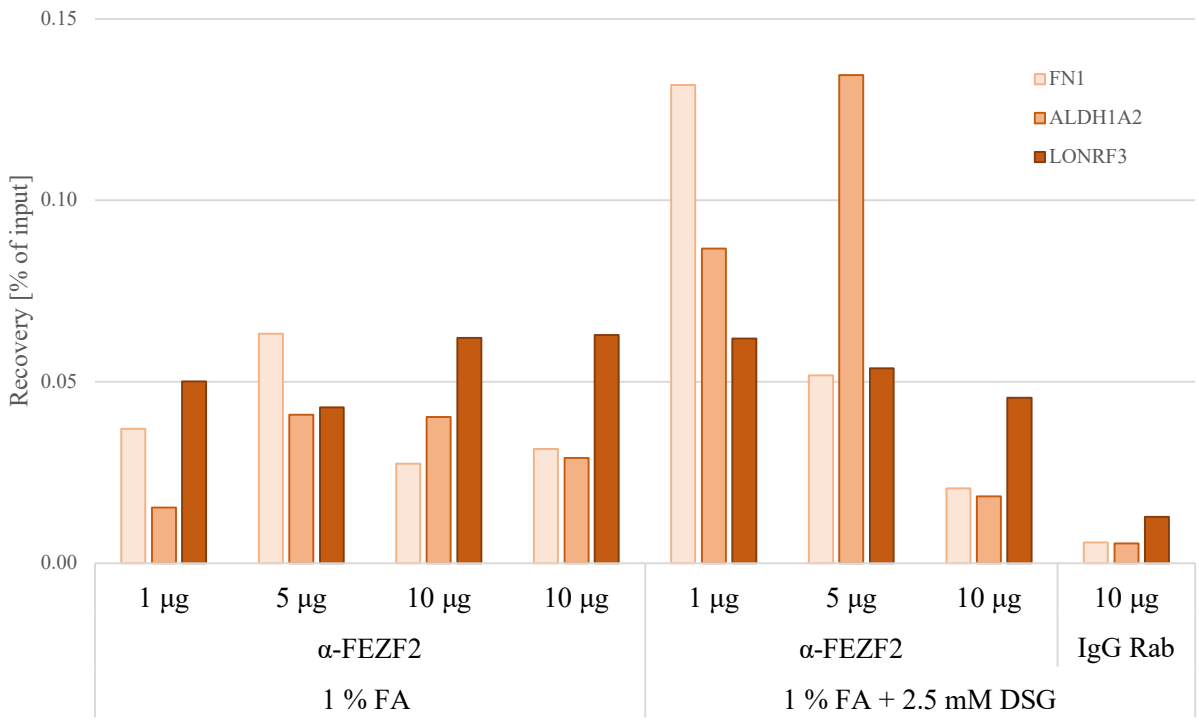

Figure 4.7. ChIP-qPCR with optimisation for α-FEZF2 to test crosslinking reagents and amount of antibody per IP. Crosslinking conducted with two different conditions: fixation buffers containing 1 % FA and 1 % FA + 2,5 mM DSG. IP conducted with three different antibody amounts per crosslinking condition: 1 µg, 5 µg, and 10 µg. The genes FN1, ALDH1A2, and LONRF3 are targets for FEZF2. IgG Rabbit used as a negative control. Abbreviations: Forebrain Embryonic Zinc Finger-Like Protein 2 (FEZF2), Disuccinimidyl glutarate (DSG), Formaldehyde (FA), immunoprecipitation (IP).

The optimal crosslinking conditions for AIRE, FEZF2 and CTCF were with 1 % FA and 2,5 mM DSG based on the results and are used for all following crosslinking steps. This crosslinking condition performed the best across all proteins. Similar to FEZF2, DEAF1 binds directly to DNA; therefore crosslinking optimisation step was not performed for DEAF1.

4.2.5 Chromatin immunoprecipitation optimisation with α-DEAF1

To optimise the ChIP strategy against DEAF1, type of antibody and antibody amount were determined. Two antibodies were tested, a monoclonal α-DEAF1 antibody produced by ThermoFisher, and a polyclonal α-DEAF1 antibody produced by LSBio. For these antibodies, three IP reactions were set up with different amounts of the antibody for each IP, including 1 µg, 5 µg, and 10 µg α-DEAF1 antibody.

Polyclonal α-DEAF1 antibody produced by LSBio had higher enrichment for all DEAF1 dependent genes than monoclonal α-DEAF1 antibody produced by ThermoFisher (Figure 4.8.), indicating that α-DEAF1 (LSBio) captured more chromatin than its monoclonal counterpart. One observation made was that the gene *EIF4G3* had a higher recovery than the genes *HOXD4, AMBP,* and *BCL9L.* When increasing the antibody amount from 5 µg to 10µg, the enrichment of this gene was recused. This may imply that 5 µg α-DEAF1 antibody per IP is sufficient, and 10 µg is too much, based on this gene alone. The other genes had increased enrichment when using 10 µg antibody; however, 5 µg antibody gave satisfactory results compared to IgG, and yielded more specific enrichment. Based on these results, it was concluded that the optimal DEAF1 ChIP strategy was to use 5 µg polyclonal α-DEAF1 antibody produced by LSBio.

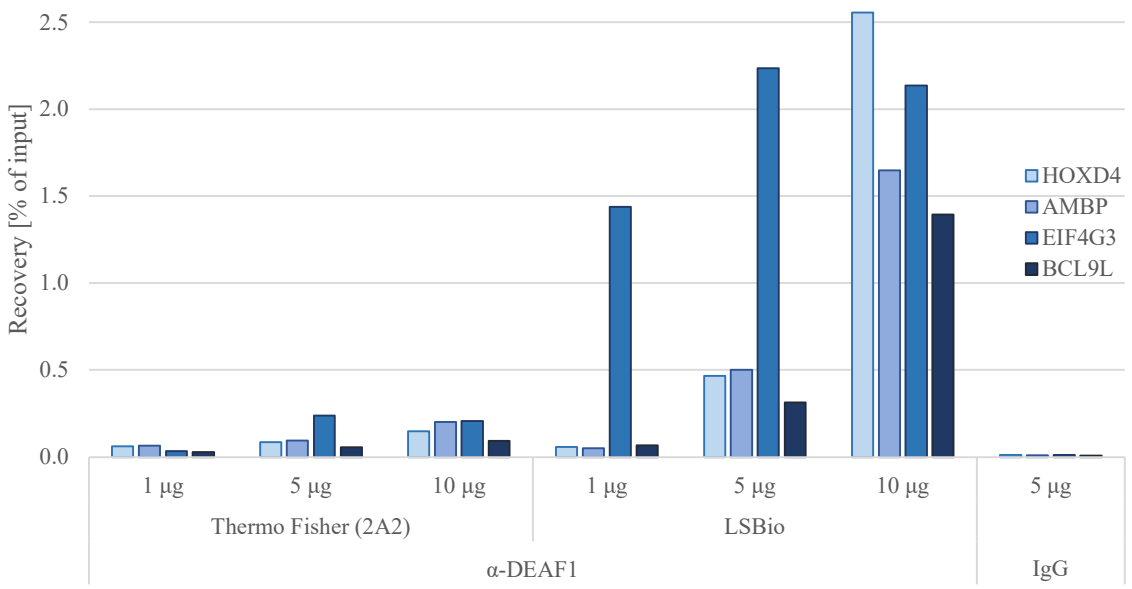

Figure 4.8. ChIP-qPCR with optimisation of α-Deaf1 antibody and the amount of antibody per IP. Two antibodies are tested: Thermofisher (2A2) monoclonal DEAF1 antibody, and LSBio polyclonal DEAF1 antibody. Antibody amounts tested are 1 µg, 5 µg, and 10 µg per IP. Negative control was IP with 5 µg IgG antibody. The genes HOXD4, AMBP, EIF4G3, and BCL9L are gene targets for DEAF1. Abbreviations: Deformed Epidermal Autoregulatory Factor 1 Homolog (DEAF1), immunoprecipitation (IP).

4.2.6 Test of Background Signal Produced During IP

The gene *PHF11* is not a target for AIRE, FEZF2 or DEAF1 and was therefore used as a negative control gene in the qPCR to control for unspecific binding by the antibodies during IP. An experiment was set up to illustrate the background signal produced in the ChIP samples with AIRE, FEZF2, and DEAF1 antibody caused by unspecific binding of the antibodies. New chromatin preparation and IP was conducted with α-AIRE GTX, α-DEAF1 LSbio and α-FEZF2 antibodies and analysed by qPCR with negative control gene *PHF11*. The non-target gene had 2-3 fold change over IgG (Figure 4.9.). The presumed target genes have a fold change of 2-4 over the non-target gene. Implying that the antibodies against AIRE, DEAF1, and FEZF2 bind non-specifically to DNA or other proteins, causing unwanted regions to be captured. Based on these results, a high background signal is expected during sequencing as the antibodies binding seems to be unspecific and likely to yield high background noise during sequencing.

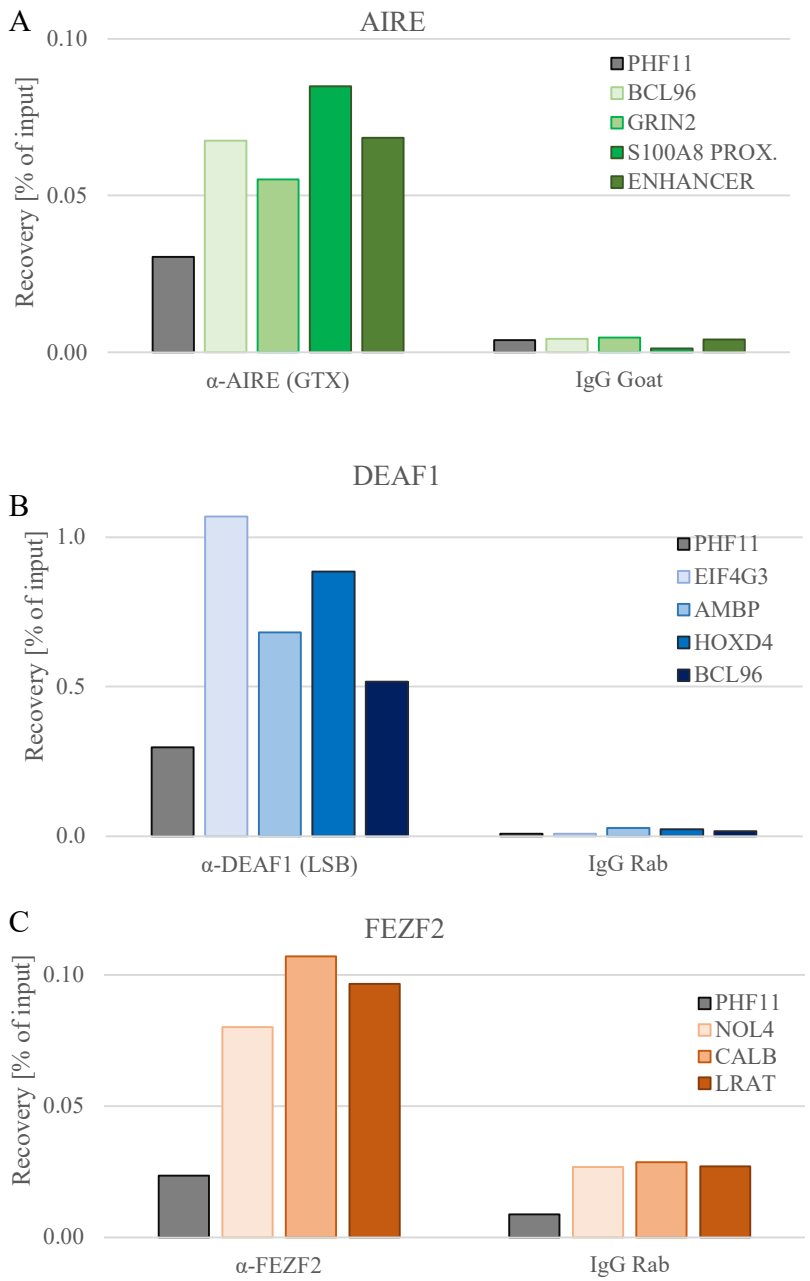

Figure 4.9. AIRE, FEZF2 and DEAF1 Enrichment of potential target and non-target genes. The gene PHF11 is not a target for AIRE, FEZF2 or DEAF1 and is used as a negative control. **A)** α-AIRE GTX antibody The genes BCL96, GRIN2, S100A8 PROXIMAL ENHANCER GENE (PROX.), and ENHANCER are gene targets for AIRE. Negative antibody control was IP with 5 µg IgG Goat antibody. **B)** ChIP-qPCR with α-FEZF2 antibody comparing target genes to a negative control gene. The genes NOL4, CALB, LRAT are gene targets for FEZF2. Negative antibody control was IP with 5 µg IgG Rabbit antibody. **C)** ChIP-qPCR with α-Deaf1 LBS antibody comparing target genes to a negative control gene. The genes HOXD4, AMBP, EIF4G3, and BCL9L are gene targets for DEAF1. Negative antibody control was IP with 5 µg IgG Rabbit antibody. Abbreviations: Deformed Epidermal Autoregulatory Factor 1 Homolog (DEAF1), immunoprecipitation (IP), Forebrain Embryonic Zinc Finger-Like Protein 2 (FEZF2), Autoimmune regulator (AIRE), LSBio (LSB), GeneTex (GTX).

4.3 AIRE, FEZF2, DEAF1 ChIP-seq in different Human Thymus Tissue samples

After we have optimised the ChIP procedure, we set out to apply it to three different thymic tissue samples (TC3005, TC3006, and TC3001) as a pilot sequencing experiment, after which, if successful, would lead to performing ChIP-seq on a higher number of thymic tissue samples subsequently.

ChIP was performed on the three tissues with IP against AIRE, FEZF2 and DEAF1 for all three tissue samples. Before sequencing the ChIP samples were analysed by qPCR in order to determine if the signal to noise ratio is sufficient for sequencing (over three-fold). AIRE IP with three thymic tissues showed low enrichment with less than two fold-change over IgG for tissue TC3001 and TC3006, and between 2-3 fold change over IgG for tissue TC3005 (Figure 4.10.). FEZF2 also showed low enrichment over all three tissues with a fold-change of less than two to four over IgG. DEAF1 IP yielded high enrichment for all three tissues with the highest enrichment for tissue TC3001, with over three-fold for all tissues. Based on these results, only the DEAF1 IP samples with three thymic tissues were used for library preparation and sequencing. The FEZF2 and AIRE samples used for library preparation were selected based on the highest transcription factor signal to background (IgG) ratio from earlier ChIP-qPCR experiments.

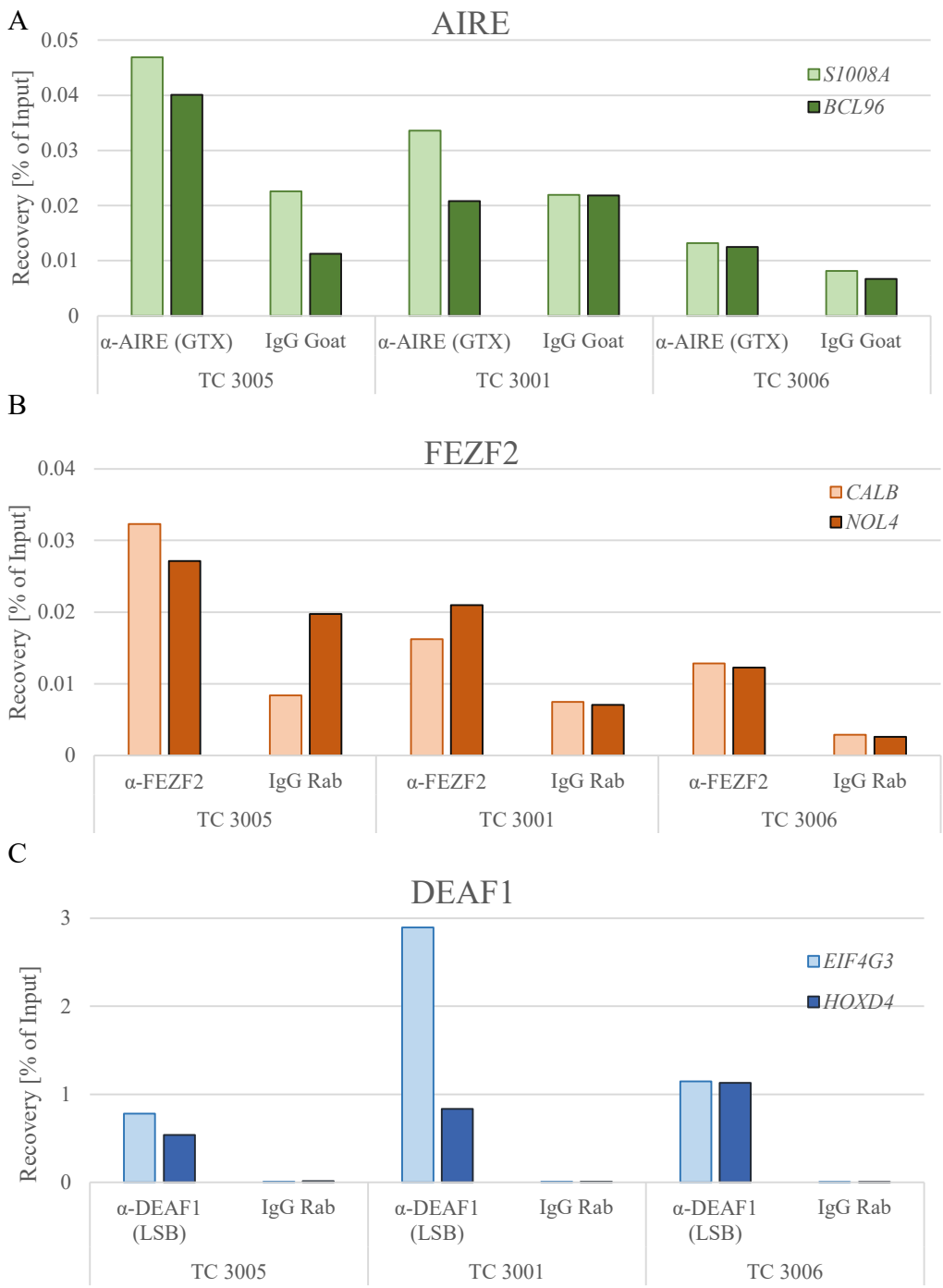

Figure 4.10. ChIP-qPCR with three biological samples. A) ChIP-qPCR with α-AIRE antibody with three biological samples. S1008A and BCL96 are targets for AIRE**. B)** ChIP-qPCR with α-FEZF2 antibody with three biological samples. CALB and NOL4 are target genes for FEZF2. **C)** ChIP-qPCR with α-DEAF1 antibody with three biological samples. EIF4G3 and HOXD4 are targets for DEAF1. Abbreviations: Deformed Epidermal Autoregulatory Factor 1 Homolog (DEAF1), immunoprecipitation (IP), Forebrain Embryonic Zinc Finger-Like Protein 2 (FEZF2), Autoimmune regulator (AIRE), LSBio (LSB), GeneTex (GTX).

The sequencing libraries were prepared with the 12 ChIP samples. TapeStation analysis was performed to determine the amount of DNA present after library preparation. Based on the TapeStation results (Table 3.6.), DEAF1 3005.1 was excluded form sequencing, as no DNA could be retrieved after the IP (not visualized). The samples that produced libraries were selected for sequencing by NSC at OUS.

The sequencing data was analysed in Galaxy using Bowtie2 and MACS analysis. Bowtie2 aligned the sequences to the human genome, and MACS localized enriched areas visualized as peaks in the IGV. MACS with the default 35 *mfold* called no peaks for any of the ChIP-seq samples. The *mfold* was adjusted to 5, which gave between 100 – 600 peaks per ChIP-seq sample (Figure 4.12.). A majority of the AIRE, FEZF2 and DEAF1 peaks were accumulated around the centromere, (see details in Table 4.2. and Figure 4.13.), and therefore align to IgG peaks (Figure 4.12.). Since this seems to be non-specific enrichment, we removed all peaks overlapping with the centromeres and with IgG peaks in an attempt to find possible specific enrichment peaks. The non-centromeric peaks make up less than 50 % of the peaks called for all samples, except for AIRE. However, AIRE still has a low average fold-enrichment for non-centromeric peaks, as does the other ChIP-seq samples, similar to IgG background peaks. For the three DEAF1 ChIP-seq samples there was little overlap in peaks between the samples and no areas where all three overlap, excluding centromeres. This indicated lacking enrichment in our ChIP-seq samples leading to background signal making up a majority of the called peaks.

Table 4.1. ChIP-seq enrichment peaks statistics

Peaks \ Sample	IgG Rab 3005.1	FEZF2 3005.1	AIRE 3005.1	DEAF1 3005.2	DEAF1 3001	DEAF1 3006
All	138	116	186	133	580	214
Fold-enrichment	5.78	6.43	7.12	6.20	8.02	6.96
Non-centromeric	56	51	111	48	156	82
Centromeric	82	65	75	85	424	132
% Non-centromeric	**41**	**44**	**60**	**36**	**27**	**38**
Highest fold-enrichment*	9.31	11.69	11.56	9.79	12.37	25.41
Average fold-enrichemnt*	5.49	6.2	6.38	5.65	6.17	6.16
Length of peaks* [bp]	754	1127	1256	670	1493	686

*of non-centromeric

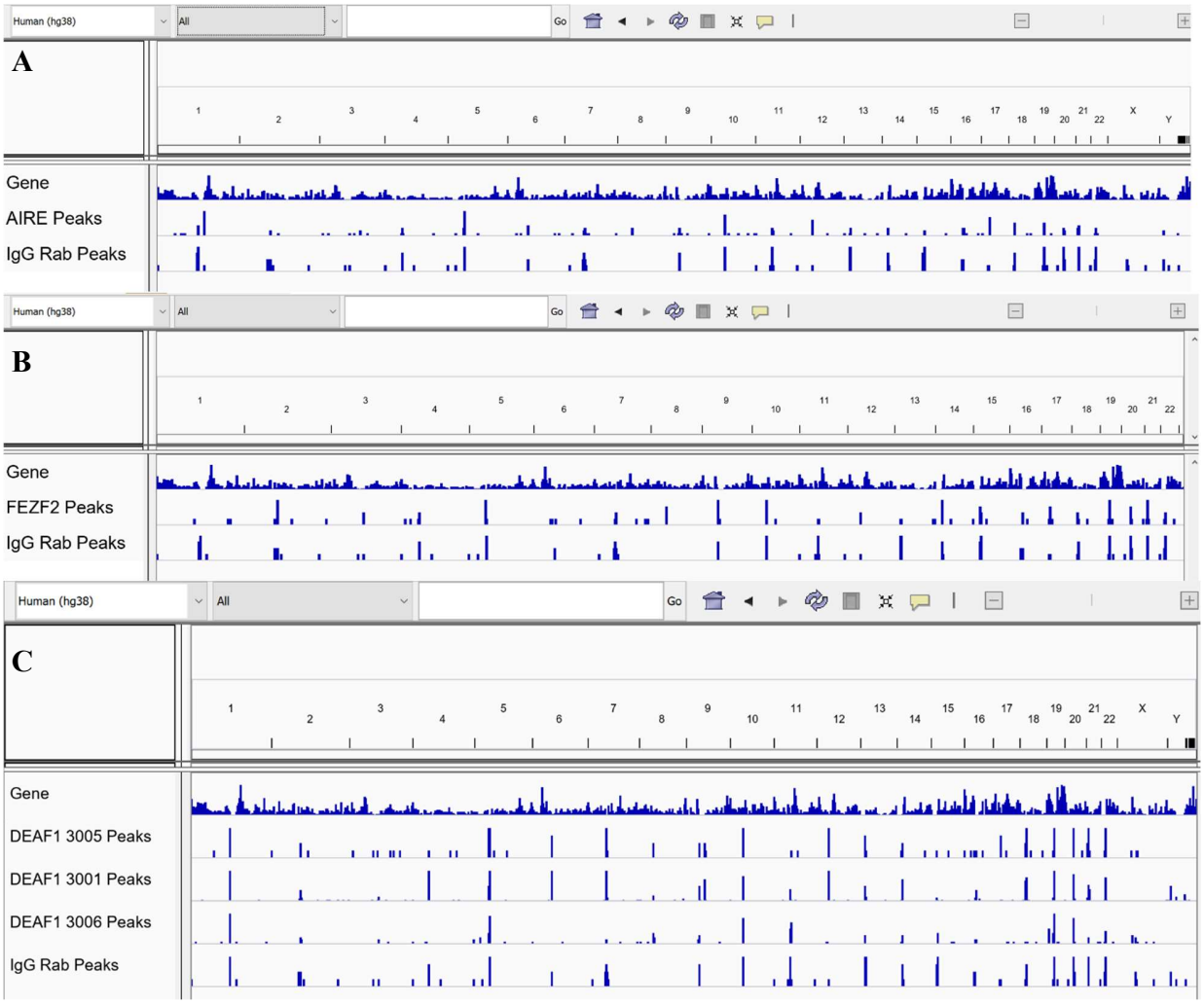

Figure 4.11. Whole genome distribution of ChIP-seq peaks. A) AIRE and IgG ChIP-seq peaks. B) FEZF2 and IgG-ChIP seq peaks. C) DEAF1, for tissue TC3005, TC3001 and TC3006, and IgG ChIP-seq peaks. Abbreviations: Deformed Epidermal Autoregulatory Factor 1 Homolog (DEAF1), Forebrain Embryonic Zinc Finger-Like Protein 2 (FEZF2), Autoimmune regulator (AIRE), Rabbit (Rab).

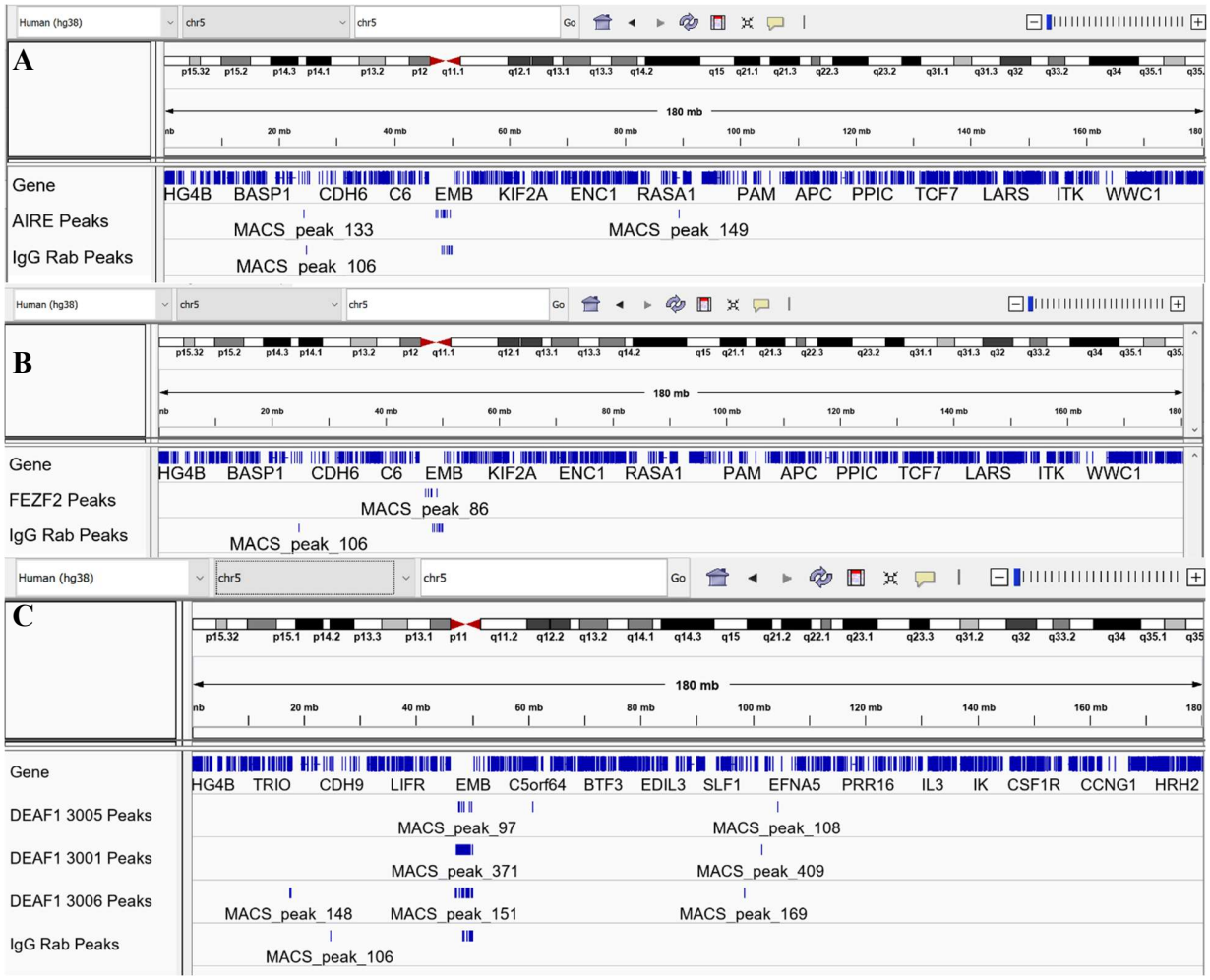

Figure 4.12. AIRE-, FEZF2- and DEAF1 ChIP-seq peaks accumulate in centromeric regions. **A**) View of peaks on chromosome 5 for AIRE and IgG. **B**) View of peaks on chromosome 5 for FEZF2 and IgG. **C**) View of peaks on chromosome 5 for DEAF1, for tissue TC3005, TC3001 and TC3006, and IgG. No overlap between the three tissue samples. Abbreviations: Deformed Epidermal Autoregulatory Factor 1 Homolog (DEAF1), Forebrain Embryonic Zinc Finger-Like Protein 2 (FEZF2), Autoimmune regulator (AIRE), Rabbit (Rab), Chr (chromosome).

As a quality control, all the presumed target genes used in qPCR (Table 3.5) were checked for MACS peaks in the proximity (in 100k bp up-and down-stream of the gene) of the reported target genes. For this purpose, we gathered all MACS data into IGV, locating the gene based on gene location from UCSC Genome Browser (GRCh38/hg38) and checking for enrichment peaks upstream of the gene. None of the transcription factors had MACS peaks close to any of their presumed target genes. Furthermore, we looked if any MACS peaks overlapped with TRA genes (the gene list was generated by (Gabrielsen et al., 2019)). This TRA list had been generated by extracting "tissue enriched" genes from Human protein atlas (genes with mRNA levels five-fold higher in one tissue than any other tissue) and comparing them to genes expressed by thymic APCs (Gabrielsen et al., 2019). Three peaks in total were identified with overlap (Table 4.3.). DEAF1 with tissue TC3001 had a MACS peak overlap on chromosome 2 right after exon 1 of gene *MAP2*, and overlap with gene RAB5B in the promoter region. DEAF1 with tissue TC3006 had a MACS peak overlap on chromosome 4 in the middle of exon 3 and 4 at the end of gene *GRXCR1*.

Table 4.3. Fold-change and location of DEAF1 peaks overlapping with TRA genes.

DEAF1 sample	Chr	Start peak	End peak	Fold-change	Gene	Start gene	End gene
TC3006	4	42,990,254	42,990,562	5.11	GRXCR1	42,883,266	43,040,658
TC3001	12	55,973,809	55,974,116	6.78	RAB5B	55,976,022	55,944,104
TC3001	2	209,428,941	209,429,213	6.18	MAP2	209,414,057	209,740,518

We examined more closely some of the MACS peaks in IGV to visualise enrichment levels in each sample in order to determine if the MACS peaks found were true binding sites or excessive background noise, illustrated by mock IP (with IgG). The enrichment of AIRE peak_74 (Figure 4.13.) and FEZF2 peak_2 (Figure 4.14.) was, respectively, three- and four-fold over IgG, indicating a true binding site. The DEAF1 peaks peak_124 and peak_540 (Figure 4.15) were aligned with an IgG peak that showed high enrichment without a peak called. This indicated that

the input for the IgG must have had high enrichment in that same area. The inputs for tissue TC3005, TC3001, TC3006 showed enrichment in the same area, with enrichment in input TC3006 being higher than that of DEAF1 TC3006. As illustrated by the chromatin view at the top of Figure 4.15., is this in a peri-centromeric area. Based on this result, the peaks were determined as false positives and not a true binding site of DEAF1.

Due to the and lack of enrichment compared to the background, no conclusive results regarding the binding sites of the transcription factors using ChIP-seq method with frozen tissue could be made, and the strategy was deemed as unsuccessful. Due to the low number of relevant cells in the thymus, using aliquots of frozen tissue could result in too few AIRE, FEZF2, and DEAF1 expressing cells per IP, further yielding low enrichment. Before sequencing another ChIP-seq sample for AIRE, FEZF2 and DEAF1 from the thymus, further optimisation to increase enrichment and reduce background noise was needed.

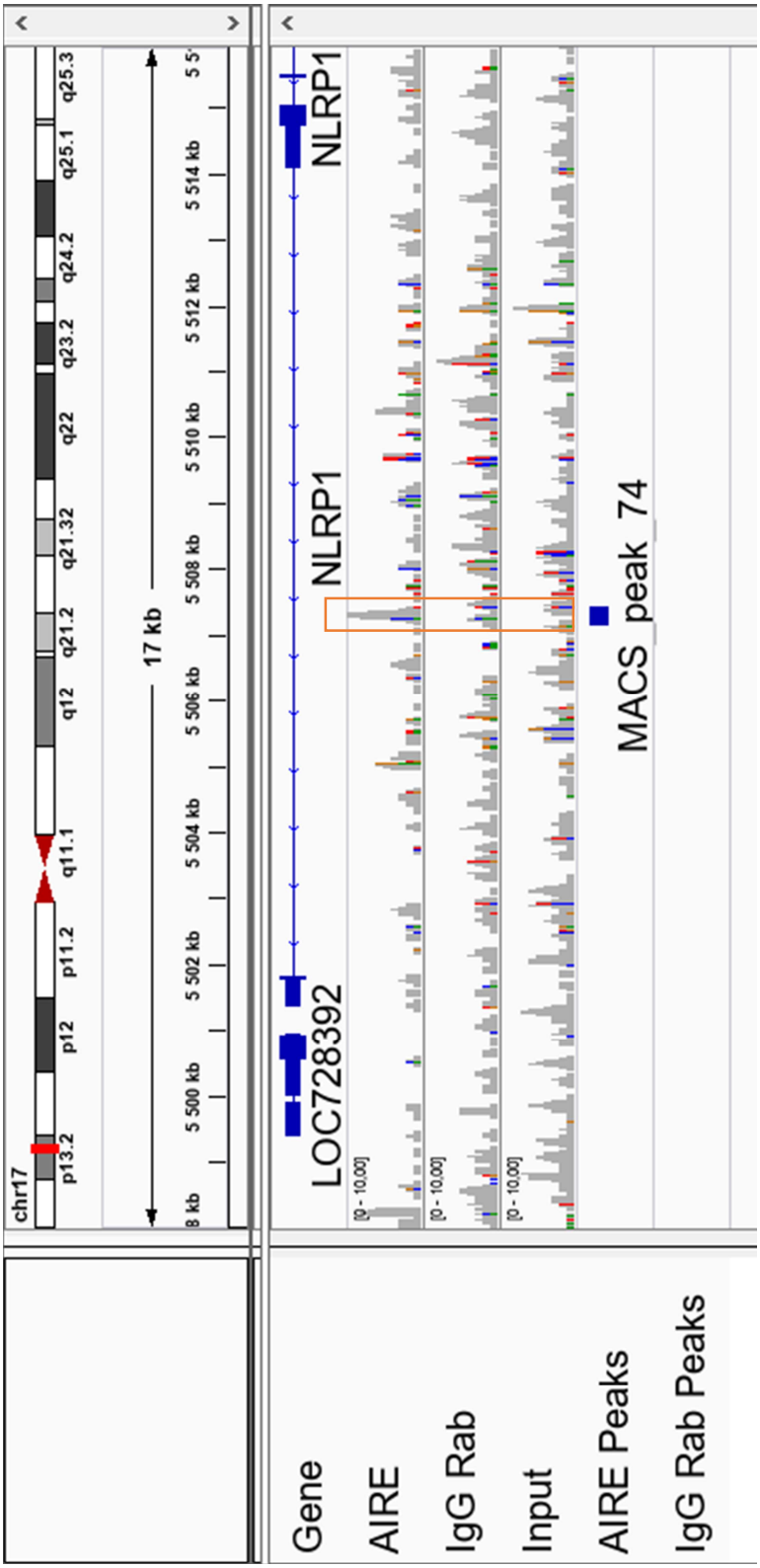

Figure 4.13. AIRE peak "MACS_peak_74" is a true peak with a high enrichment relative to IgG. Aligned sequence against human genome (hg38) viewing MACS peak in AIRE ChIP-seq sample. view of enricment level for AIRE, IgG and Input alligned to chromosome 17: 5 498 000-5 516 000 in the end of the *NLRP1* gene. Abbreviations: Autoimmune regulator (AIRE), Rabbit (Rab), Chr (chromosome), Kilo base pairs (kb).

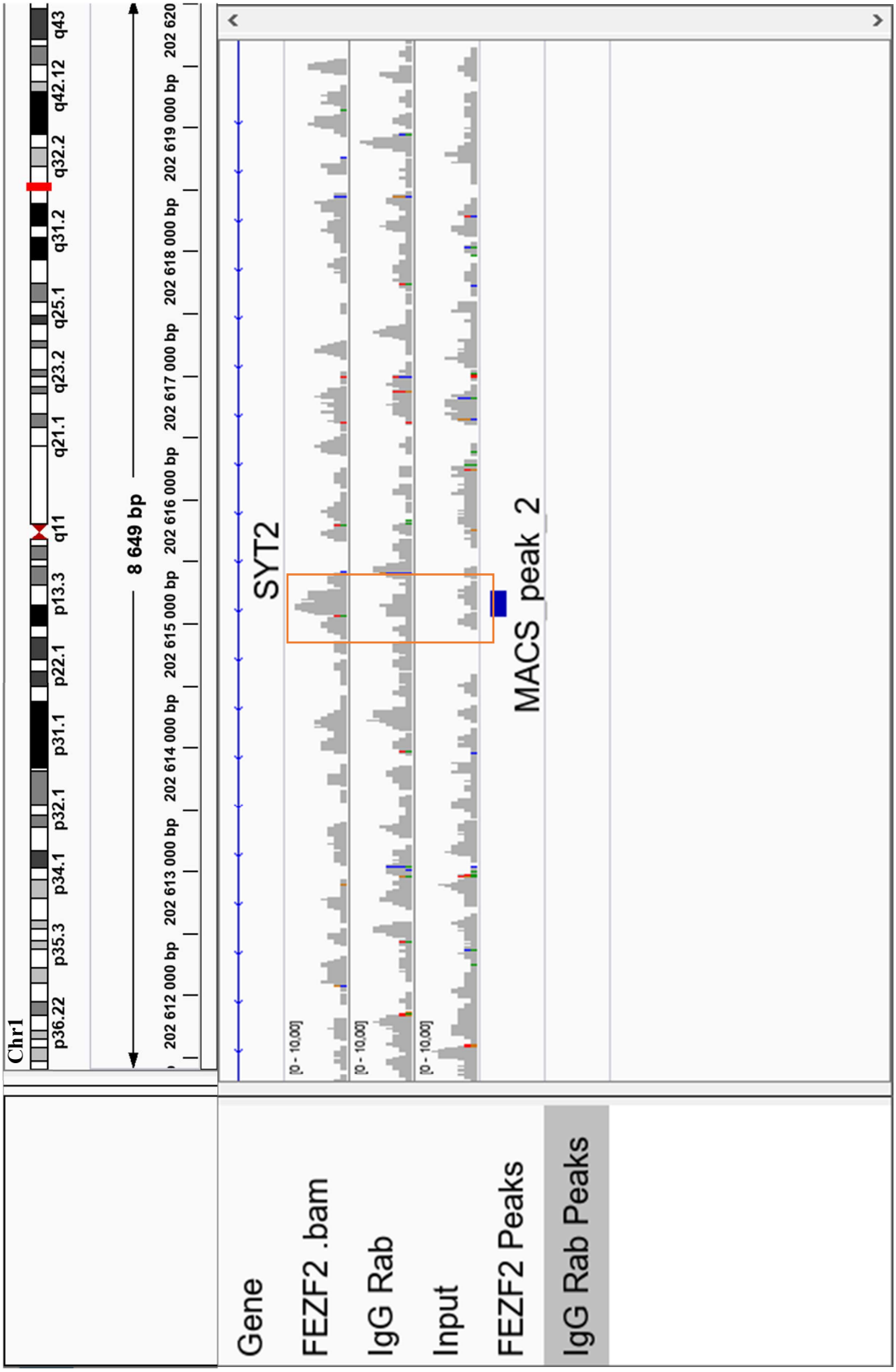

Figure 4.14. FEZF2 peak "MACS_peak_2" is a true peak with a enrichment higher than IgG. Aligned peaks against human genome (hg38) viewing enriched areas in FEZF2 and IgG Rabbit samples. View of enricment level for FEZF2, IgG and Input alligned to chromosome 1: 202,611,000-202,620,000 in the middle of the gene *SYT2*. Abbreviations: Forebrain Embryonic Zinc Finger-Like Protein 2 (FEZF2), Rabbit (Rab), Chr (chromosome), base pair (bp).

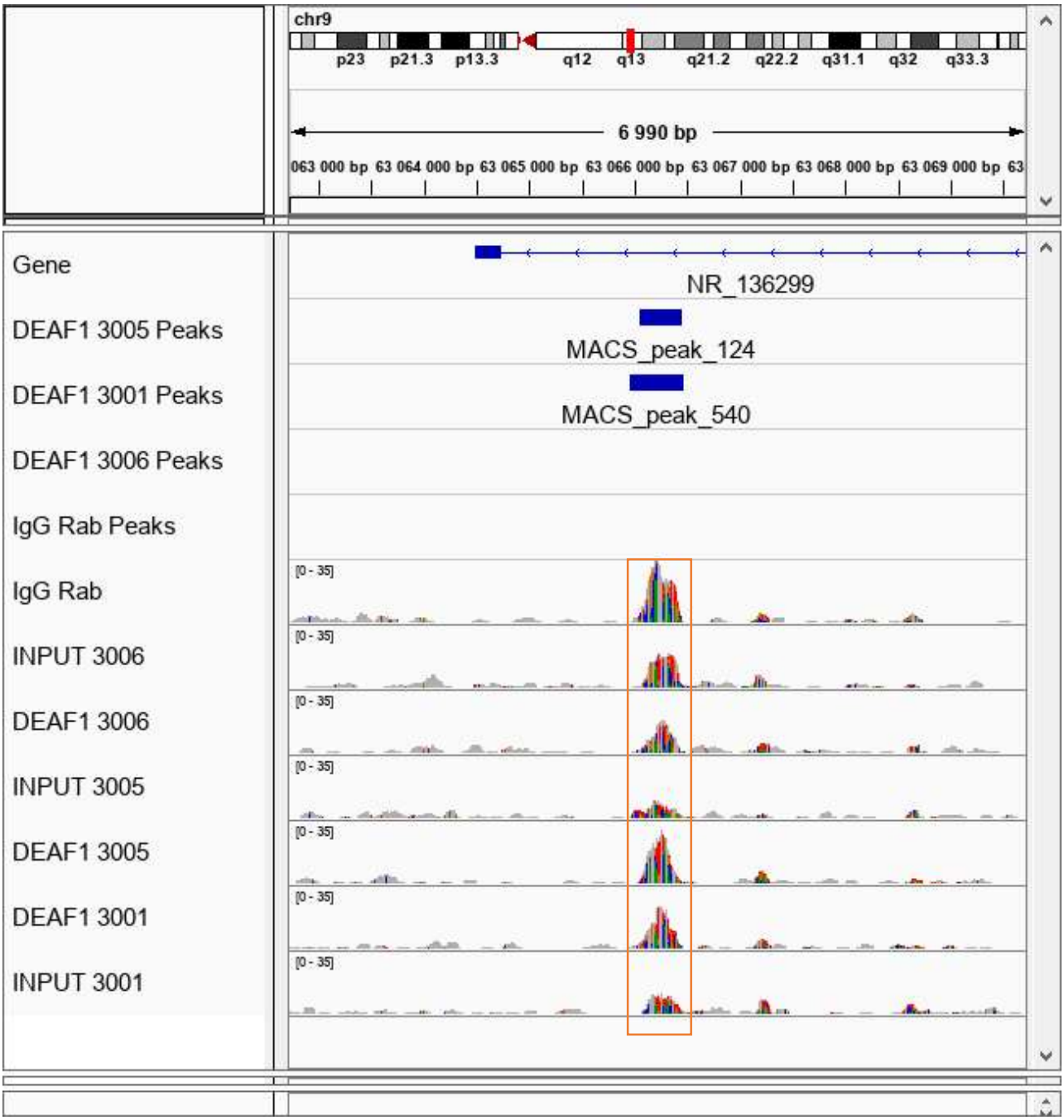

Figure 4.15. Aligned peak against the human genome (hg38) viewing enriched area in DEAF1 ChIP-seq sample. View of enrichment level for DEAF1, IgG and Input under DEAF1 peaks called MACS_peak_124 and MACS_peak_540 aligned to chromosome 9: 63,063,000-64,070,000 in the end of a long-non-coding RNA gene. Abbreviations: Deformed Epidermal Autoregulatory Factor 1 Homolog (DEAF1), Rabbit (Rab), Chr (chromosome), base pair (bp)

4.4 Chromatin Immunoprecipitation with TECs and APCs

Since the low number of cells expressing the three transcription factors could be the reason for the low enrichment we observed with ChIP-seq on tissue samples, we set out to perform ChIP-seq on enriched TECs and APCs extracted from half of a fresh thymus. The goal was to increase the concentration and number of relevant TECs and possibly remove noise created by irrelevant cells and connective tissue. Chromatin immunoprecipitation was performed following the same protocol as for the frozen tissues excluding the Dounce homogenizers (as it was not needed), and fixating with 1 % FA and 2.5 mM DSG. The chromatin was sonicated for 8 minutes and analysed with TapeStation. The results after sonication showed fragments around 300 bp (Figure 4.16) with a DNA concentration of 6.62 ng/µL. Using the pure cell sample lead to faster fragmentation of the chromatin and yielded more DNA than using thymic tissue homogenate.

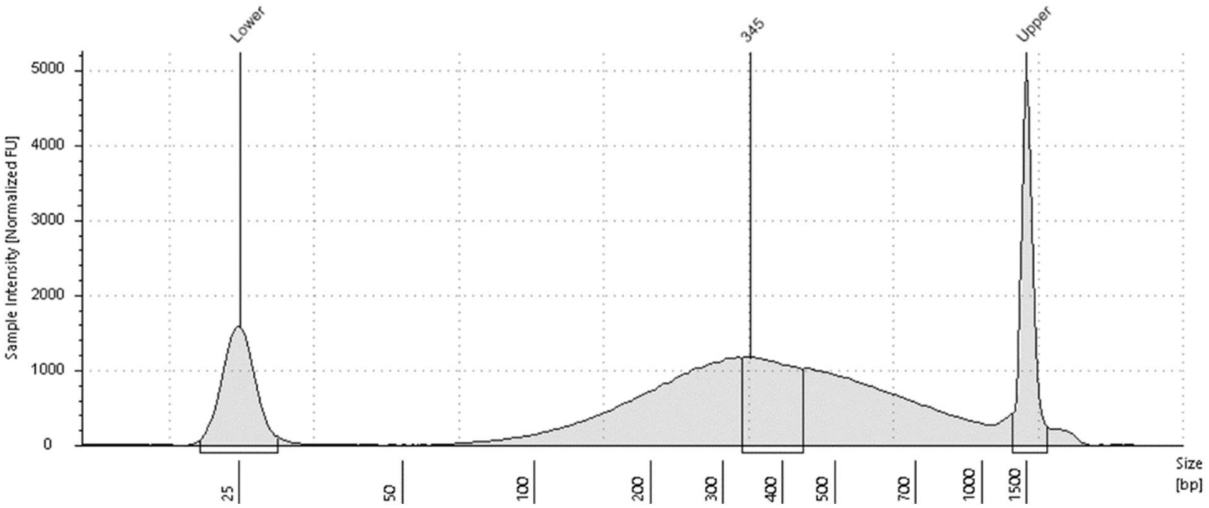

Figure 4.16. TapeStation image of APC chromatin sample after 8 minutes sonication. Chromatin fragments between 100 to 1000 bp, with an average of 345 bp. Abbreviations: Base pair (bp).

The TEC and APC enriched cell solution was split into five IP reactions with α-CTCF, α-AIRE GTX, α-FEZF2, α-DEAF1 LSBio, and IgG using the optimised conditions. The ChIP samples were analysed with qPCR.

The qPCR results from performing ChIP with TEC and APC enriched cell samples showed similar background to signal ratio as ChIP with frozen thymus tissue for all protein targets (Figure 4.17). Chromatin immunoprecipitation with CTCF showed that the ChIP method was successful with thymic APCs and TECs. However, as the goal was to reduce the background signal and increase the ChIP sample signal, ChIP on thymus cells and tissue was concluded as an insufficient method, and further testing using other methods were conducted.

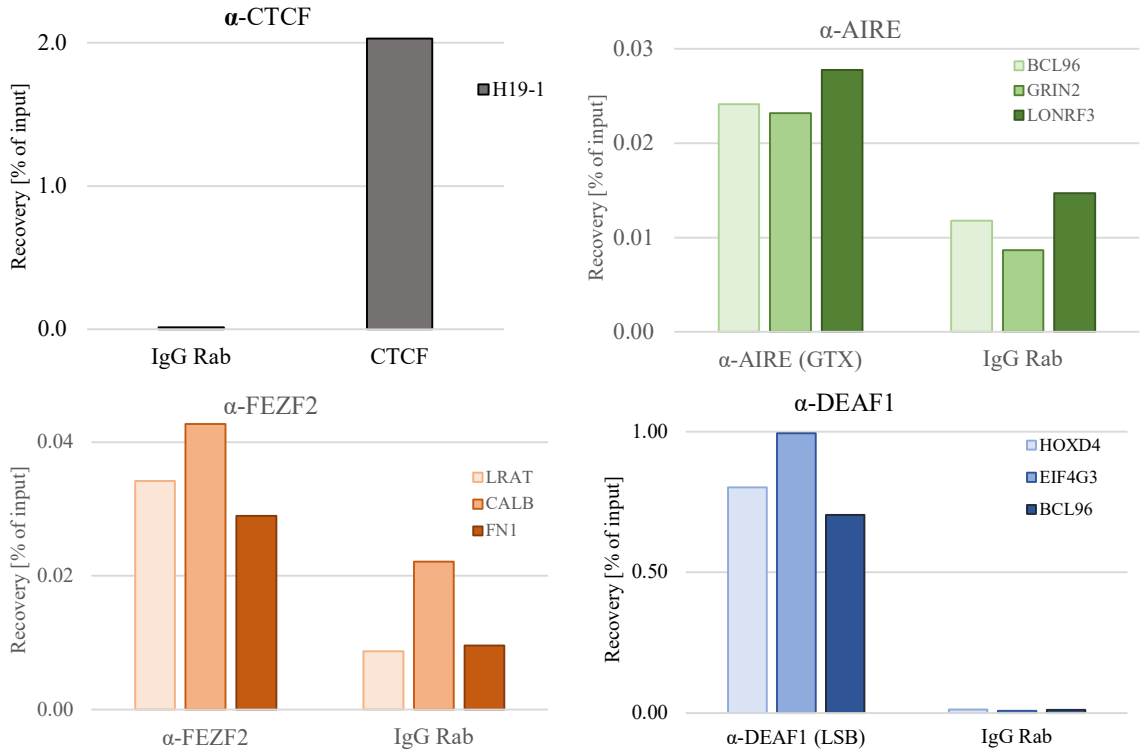

Figure 4.17. ChIP-qPCR with APC enriched cell suspension. Immunoprecipitation was set up using 250 µL chromatin solution per IP reaction, scaling the other ChIP reagents up proportionally. Five ChIP reactions were set up: 1 µg α-FEZF2 antibody, 5 µg α-AIRE GTX antibody, 5 µg α-DEAF1 LSB antibody, 5 µg IgG Rabbit, and 1 µg α-CTCF antibody. IgG Rabbit used as a negative control. ChIP with APC enriched cell suspension and α-CTCF as a positive control to ensure ChIP with APC enriched cell sample works. H19-1 is a target gene for CTCF. BCL96, GRIN2, and LONRF3 are target genes for AIRE. LRAT, CALB, and FN1 are target genes for FEZF2. HOXD4, EIF4G3, BCL96, and GRIN2 are target genes for DEAF1. Abbreviations: Deformed Epidermal Autoregulatory Factor 1 Homolog (DEAF1), immunoprecipitation (IP), Forebrain Embryonic Zinc Finger-Like Protein 2 (FEZF2), Autoimmune regulator (AIRE), LSBio (LSB), GeneTex (GTX), CCCTC-binding factor (CTCF), Rabbit (Rab).

4.5 CUT&Tag on Thymic cells

In an aim to reduce background noise and increase enrichment, a new method was tested. Cleavage Under Targets and Tagmentation (CUT&Tag), published in 2019, allows for epigenomic profiling on chromatin through the use of a protein A-Tn5 transposase fusion protein and activation of the transposase that efficiently generates sequencing libraries (Henikoff et al., 2019). It promises lower background and needs for fewer cells (100-500k) than for ChIP-seq. It was needed to control that this method was efficient for thymic cells. To do this, we used thymocytes isolated during dissociation of thymic tissue and an antibody against the abundant histone modification mark- H3K27me3, as a positive control (that was used in the publication as well). We tested four fixation reactions: one sample without fixation, and samples with fixation reactions containing 0.1 % FA, 1 % FA, and 1 % FA + 2.5 mM DSG. The CUT&Tag PCR amplified samples were run on TapeStation 2200 (Table 4.4.). No product was visible on the TapeStation, and the DNA concentration is low, indicating that the DNA was not tagmented. Based on the results, the method was not successful as is, and further optimisation of the CUT&Tag protocol for thymocytes was needed.

Table 4.4. The average fragment size, size distribution and concentration of DNA in CUT&Tag thymocyte libraries.

Sample	From [bp]	To [bp]	Average Size [bp]	Conc. [pg/µl]
No fixation	50	869	196	43.5
0.1 % FA	50	3500	527	60.8
1 % FA	50	3500	584	61.2
1 % FA + 2.5 mM DSG	50	4291	808	60.8

5 Discussion

In order to assess the function of transcription factors important for the establishment of self-tolerance, we optimised the chromatin immunoprecipitation procedure using homogenised frozen thymic tissue aiming to reach highly sensitive and specific enrichment.

5.1 ChIP optimisation

To reach the highest specific enrichment possible, we optimised crosslinking, sonication and antibody used in IP for each transcription factor. We established the conditions under which the chromatin had both the shortest fragment size and highest yield of protein complexes on chromatin, as was monitored by the recovery of target genes after IP compared to non-target genes and/or "mock" IP (background, IgG). As transcription factors bind to DNA in various ways (AIRE acts as a co-regulator, and FEZF2 and DEAF1 are direct DNA-binding proteins), we included the crosslinking optimisation experiments in AIRE and FEZF2 IP to identify the optimal crosslinking buffer accordingly (Chen, Zheng, Yang, Li, & Guo, 2011; Jensik et al., 2014; Takaba et al., 2015).

Chromatin regions that have a lower occupation of proteins and a more accessible conformation are more easily sheared and solubilized, thus they will be more abundant in the chromatin. This is a general bias of sonication and is reflected in the input control sample- where such regions will have more reads. Prolonged crosslinking leads to more rigid chromatin that is more difficult to shear and solubilizes, leading to a lower amount of usable chromatin for IP. The IP with CTCF showed that a balance had been reached between crosslinking and fragmentation using 1 % FA and/or 2.5 mM DSG (Figure 4.3.) whereby high and specific enrichment of target regions were CTCF binds was yielded.

We saw inter-experiment variability, best illustrated in Figure 4.6. Here, using the same amount of α-AIRE antibody (5 µg), yielded twice the enrichment from one experiment to the other. This may be due to several factors, among others- a varying concentration of TECs and APCs in the different aliquots of thymus tissue used in each ChIP experiment and unproven target genes for the tested transcription factors in human mTECs.

5.1.1 Antibody Selection

Critical for the specificity of ChIP-seq is using the optimal antibody. ChIP-seq validated antibodies bind specifically the target protein in crosslinked chromatin, with minimal unspecific binding to other proteins and DNA. The antibodies used in this study are not validated for use in ChIP-seq (such antibodies were not available), and may not bind with high specificity to the studied proteins. The antibodies against AIRE that we used for ChIP detected a protein at the expected molecular weight (with some background) in Western blot experiments with chromatin lysates (preliminary work to this thesis), which indicated that the antibodies bind the protein and that we have detectable amounts of the protein in the thymus tissue homogenates. Antibodies are a source of background if they do not bind the target protein specifically (binding also other proteins) or if they non-specifically bind DNA. This would be indicated by similar enrichment for target and non-target genes. Unspecific binding is one of the most common problems in ChIP. We performed methods to reduce unspecific binding including pre-clearing of chromatin with Dyna beads, blocking of beads with BSA, and washing with high-salt buffers. However, the background that we observe seems high due to low enrichment, rather than the high background, as indicated by the recovery rates, which were much lower for AIRE and FEZF2 (between 0.1 and 0.6 %) and higher for DEAF1 (1-2 %) as compared to CTCF (ca. 3%).

Protein epitopes that are recognized by the respective antibodies may be masked during the crosslinking. Therefore, for both AIRE and DEAF1 antibodies, we tested polyclonal and monoclonal antibodies. Monoclonal antibodies recognise one epitope on the protein, while polyclonal antibodies recognize several and will increase the probability of binding an epitope not masked by fixation. Comparing the polyclonal antibodies, α-DEAF1 (LSBio) and α-AIRE (GTX), to their monoclonal counterparts, α-DEAF (TF) and α-AIRE (CZ/TF), confirmed that polyclonal antibodies perform better with ChIP, giving higher recoveries of target genes (Figure 4.8. and Figure 4.5., respectively).

The antibodies used did not yield a high enrichment of presumed target versus non-target gene, as assessed by qPCR (Figure 4.9.). The presumed target genes were only enriched two to four-fold over the non-target genes for all transcription factors. For α-DEAF1 and α-AIRE, the non-target gene recovery exceeded the recovery in the IgG mock IP. For FEZF2, the target and non-target recovery were similar to IgG signals.

Collectively, the results indicate that we had low enrichment with AIRE and DEAF1 of the presumed target genes, which was somewhat high for non-target genes as well, while FEZF2 CHIP experiments did not have specific enrichment. Thus, we can conclude that the low enrichment observed can be due to one (or several) of the following reasons: the proteins studied were not abundant enough in the chromatin lysates, the antibodies used are not suitable for ChIP, or that the target genes studied are not real targets in human thymus.

5.1.2 Target genes for AIRE, FEZF2 and DEAF1 used in qPCR

The target gene primers that we used in qPCR are not based on known TRA genes in the human thymus but are found in the literature where studies on native Aire, Fezf2 or Deaf1 were performed in mouse thymus or in human cells (of non-thymic origin) upon overexpression of the proteins. The selection of TRA primers is based on TRAs with the highest fold-change expression in ChIP-seq studies with sorted mice mTECS (for AIRE and FEZF2 target genes) and overexpression experiments of Deaf1 compared to knock-out models in their respective studies (Bansal et al., 2017; Jensik et al., 2014; Org et al., 2009; Takaba et al., 2015; L. Yip et al., 2009). The assumption made was that the target genes pick out from these studies in mice are targets for AIRE, FEZF2 and DEAF1 in the human thymus as well. Although we do not know for sure where exactly the transcription factors bind their target genes, we designed the primers to be in the promotor regions (0-1000 bp upstream of the gene), except “Enhancer”, since it is not a gene. As AIRE is reported to act as a co-activator, it can bind in complexes with other proteins at enhancer regions further from the target genes and still activate their expression. The primer for target site “Enhancer”, used in AIRE qPCRs, was designed to cover such a region (Bansal et al., 2017).

We cannot determine if low the enrichment we see for AIRE, FEZF2 and DEAF1 target genes is true enrichment or just background caused by unspecific DNA binding by the antibodies. It was evident that the enrichment was low in our ChIP-seq samples when comparing the presumed target genes to a negative control gene (Figure 4.9). With such low enrichment (less than 0.1-1% of input) the background appears high. The presumed target genes all had higher enrichment than the negative control gene by two- to four-fold. However, when performing ChIP-seq, it is desired to have at least five-fold enrichment over negative control before sequencing in order to have sufficient signal to noise ratio for reliable detection of enriched regions.

The EIF4G3 promotor region was most highly enriched in DEAF1 ChIP, similarly enriched as CTCF target genes upon CTCT ChIP (Figure 4.8. and Figure 4.3., respectively). Yip et al. in 2013 found that Deaf1 controls Eif4g3 expression, encoding the protein eif4gII (Linda Yip, Creusot, Pager, Sarnow, & Fathman, 2013). It is a part of a complex that initiates the expression of genes that enable antigen presentation on MHC/HLA class II molecules. Suggested by the high enrichment of EIF4G3 in our qPCR results (i.e. Figure 4.8.), DEAF1 may control the expression of EIF4G3 in thymic APCs as well. DEAF1 might, therefore, contribute to efficient antigen presentation on HLA class II molecules in thymic APCs as it does in APCs in the pancreatic lymph node. If antigen presentation is impaired, the consequences may be self-reactive T-cells being released into the periphery, potentially leading to AID. Mice studies have revealed that some of the 600 TRA genes controlled by Deaf1 in epithelial cells in the pancreatic lymph node are controlled by AIRE in mTECs (Fuhlbrigge & Yip, 2014). The role of DEAF1 in the thymus might strictly be enabling the processing and presentation of TRAs on the HLA class II complex.

As we only have moderate enrichment of all the presumed target genes for all transcription factors during qPCR, we can conclude that they may not be true target genes in the thymus, or that our tissue material contains too few cells expressing the transcription factors to be able to determine their targets with sufficient signal-to-noise ratio.

5.1.3 Transcription Factor Abundance

Analysis of the expression levels of the three studied transcription factors in the different sorted APCs in the human thymus (produced in a previous project) showed that DEAF1 was expressed in all APCs, while AIRE and FEZF2 were expressed at a high level only in mTECs (Figure 4.1.). FEZF2 was expressed at a low level in one (out of 6) sample of dendritic cells and B-cells and showed expression in cTECs as well. FEZF2 has not been detected in cTECs in earlier studies. Given that no other B-cell or CD141+ dendritic cell samples had FEZF2 expression, this may be contamination of mTECs during isolation of cells. DEAF1's reported role in antigen presentation may explain the DEAF1 expression in all APCs and TECs (L. Yip et al., 2009).

The thymus contains over 90 % thymocytes, and around 5 % APCs and mTECs (Sakata et al., 2018; Sansom et al., 2014). The tissue used in the ChIP experiments was from pieces of the frozen thymus. Whether they contain medulla, or what the ratio of the medulla to the cortex was, is not known. Therefore, it is possible that the APC and mTEC content in these pieces was lower

than the average content per milligram of the whole thymus. We have no cell or tissue sample with high expression of the studied proteins to use as a positive control. Therefore, we cannot be certain that our samples express sufficient levels of AIRE, FEZF2, and DEAF1, or that our antibodies are specific to these proteins. When performing ChIP with different tissues (Figure 4.10.), there was variable enrichment for the target genes of each transcription factor. DEAF1 had the seemingly highest enrichment of targets with tissue TC3001 and TC3006, while FEZF2 and AIRE had low enrichment for these tissues. This may suggest that in these samples there may be more APCs and fewer TECs than in sample TC3005, pointing to varying cell composition in tissue aliquots.

In order to increase the number of cells expressing AIRE, FEZF2 and DEAF1, we used TEC-and APC- enriched cell suspension from freshly dissociated half of a thymus. However, the ChIP experiment with these cells yielded even lower enrichment for each transcription factor than with ChIP using aliquots of frozen thymus tissue (Figure 4.17.). Although we attempted to enrich for mTECs, we did not assess the percentage of TECs or other APCs in the sample, and we may have not actually enriched TECs and APCs, which may explain our results. The mice thymus contains around one million mTECs, and the expected number of mTECs one is able to isolate is around 200-300k (according to reports (Sakata et al., 2018)). A ChIP-seq experiment requires around ten million cells to obtain 10-100 µg ChIP DNA when studying less abundant proteins (Kidder, Hu, & Zhao, 2011). Thus, performing ChIP on enriched for mTECs and APCs cell sample, containing sufficient numbers of transcription factor-expressing cells, is a prerequisite for successful ChIP-seq experiments. Future experiments using the whole thymus and confirming the mTEC and APC cell numbers may improve the outcome of ChIP experiments for these transcription factors.

5.1.4 Biological Specificities of Transcription Factor Binding

High expression of the TRAs is unnecessary to “educate” the maturing thymocytes, and each TRA is presented by 1-3 % of mTECs at any given time (Klein et al., 2014). Therefore, each TRA gene will only be activated in a small proportion of all mTECs and its promoter will be bound by transcription factors in a small proportion of the cells, in turn affecting the amount of detectable binding sites by ChIP. AIRE is reported to activate genes promiscuously – binding to different target genes in each mTEC cell (Passos et al., 2018). If AIRE only binds to a particular target region only in a small proportion of all mTECs, then to detect enrichment in this region by

ChIP, we would need a sufficient number of this cell type. This may be possible if we would use the whole thymus for one ChIP-seq sample, but even then it may not be enough to reach a sufficient signal-to-noise ratio. The ENCODE consortium, who set the worldwide standards for ChIP-seq experiments, recommend using two to three biological replicates in each ChIP-seq experiment to achieve reliable ChIP-seq data (ENCODE & Consortium, 2004). Since it may be too difficult (or practically impossible) to uncover every target region in each sequenced sample, it is necessary to ChIP-seq as many samples as possible or to pool several samples per ChIP reaction, in order to comprehensively identify all target regions.

5.1.5 ChIP-seq sequencing results

The enrichment of transcription factor-bound DNA regions that we observed from the ChIP-seq data was low (avg. 6.86 fold over input, Table 4.2.) with highest enrichment peaks mapping to centromeric regions, similar to IgG-ChIP peaks (Figure 4.12.). These peaks can be considered as nonspecific, since they are also found in IgG IP, and may be explained by background DNA captured non-specifically by the antibodies (Figure 4.15.). Centromeric regions, constituted of DNA repeats, frequently produce artefacts and noise in ChIP-seq and are on the blacklist of ENCODE consortium (Wimberley & Heber, 2019). Sonication itself is another source of background, resulting in a higher abundance of DNA from open chromatin as it is more susceptible to shearing than the more rigid heterochromatic regions. This bias is visible as regions with a higher number of reads in the input control sample, and therefore it is an important control accounting for sonication bias.

The α-DEAF1 antibody yielded a higher amount of peaks around the centromere and was the only antibody that had peaks aligned to IgG peaks outside of the centromere. This suggests that the α-DEAF1 antibody captured more non-target regions compared to α-AIRE and α-FEZF2 antibodies.

The expected results were to find more enriched areas and binding sites after sequencing for the DEAF1 ChIP sample as *DEAF1* is expressed in all APCs and TECs in the thymus and is reported to function as an important transcription factor within presentation and processing of TRAs (Figure 4.1., (Linda Yip et al., 2013). The presumed target genes for DEAF1 were not found close to any peaks. The enrichment in our qPCR analysis showed a high recovery of *EIF4G3* for all DEAF1 ChIP samples. This was not mirrored in any of the three DEAF1 sequenced samples,

where no DEAF1 peaks were close to the *EIF4G3* gene. We can conclude that the ChIP-qPCR result was non-specific enrichment of EIF4G3 gene. Including many different target and non-target regions in the ChIP-qPCR experiments is a good way to detect specific vs unspecific enrichment, but even then what looks like specific enrichment can also be false. Therefore, it is necessary to confirm the results with more biological samples.

After removing the centromeric peaks and the few peaks aligning to IgG peaks, remained 20-60 % of the original number of peaks. These had low fold-enrichment (average around 5-6 fold) over input for all transcription factors (and IgG) (Table 4.2.). The peaks found may be true binding sites of AIRE, FEZF2, and DEAF1, but due to the overall low enrichment, we cannot be confident about them. We analysed if any of the peaks corresponded to TRA genes from a list generated by Gabrielson et al. As this list is constructed by comparing tissue enriched genes and transcriptomes of mTECs and APCs from the human thymus (where human thymus is collected from 0-3 y.o. healthy patients, as in this study), would correlation mean we could further connect the transcription factors to TRA genes they regulate in the human thymus. Three DEAF1-enriched peaks overlapped with TRA gene regions (*GRXCR1*, *RAB5B*, and *MAP2*, Table 4.3.). DEAF1 peaks were found in intronic regions of two of the TRA genes, and in the promoter region of one. When transcription factors bind in the intronic region, it can acts as an enhancer or repressor (Fuxman Bass et al., 2014), while transcription factors in the promoter region may have a role in the formation of the transcriptional complex activating gene expression. DEAF1 may, therefore, have both co-regulator and direct-activator effect for expression of TRA genes in the thymus, and may also be an important regulator of the promiscuous gene expression alongside AIRE and FEZF2.

In general, our ChIP-seq, similar to the ChIP-qPCR experiments, exhibited low and mostly unspecific enrichment. We conclude that a larger number of cells expressing the transcription factors studied is necessary in order to retrieve significant enrichment of their target genes with this type of experiments.

5.2 CUT&Tag

A more suitable method for target site enrichment may be CUT&Tag (Kaya-Okur & Henikoff, 2020). The advantage of this novel epigenomic profiling method is the low background and the need for fewer cells. In order to resolve the problem of low enrichment caused by the need for

larger amounts of AIRE, FEZF2 and DEAF1 expressing cells, we attempted to perform Cut&Tag experiments using the H3K27me3 antibody used in the published protocol with thymocytes. Our first experiments did not retrieve any DNA as seen by TapeStation analysis (Table 4.4.), which could be attributed to insufficient permeabilization of the cells and inability of the antibody and the fusion protein A-tagmentase enzyme to enter the cells.

Since we used D1000 screen tape on TapeStation, we could only visualize fragments below 1000 bp. If the DNA was not tagmented, then the DNA will be of high molecular weight and not visible in the TapeStation. What we see as fragmented DNA is a very small amount and short fragments, which could be degraded DNA, and not tagmented DNA.

6 Conclusion

This research aimed to identify binding sites of AIRE, FEZF2, and DEAF1 in the genome of mTECs and APCs in the human thymus. The results indicate that DEAF1 controls TRA expression in thymic cells alongside AIRE and FEZF2, which will need to be confirmed through further studies. Based on the results it can be concluded that performing epigenomic profiling of transcription factors in the smaller cell populations in the thymus requires using cell suspensions from thymus enriched for mTECs and APCs and using a strategy with high sensitivity and specificity.

6.1 Future studies

To more efficiently and confidently characterise the binding sites of AIRE, FEZF2, and DEAF1 in TECs with CUT&Tag, we need to optimize the cell-permeabilization and apply it on enriched for mTEC and APC cell samples. CUT&Tag is applicable to perform on the single-cell level (Henikoff et al., 2019). This would give the opportunity to visualise individual binding sites in each cell and better understand the function of the transcription factors in the process of negative selection of T-cells in the thymus. Then we could determine if, i.e. DEAF1 have different target regions from APCs to mTECs. This would further connect transcription of TRA genes to different cells. As AIRE has been reported to have different targets from cell to cell, using single-cell analysis would yield a more precise analysis of AIRE function. It would also be interesting to see if FEZF2 has different targets for each cell, as it activates the promiscuous gene expression in mTECs alongside AIRE.

In order to assess the problem of not knowing whether or not our tissue sample contains cells expressing *AIRE*, *FEZF2* and *DEAF1*, there should be included a step in which mRNA for AIRE, FEZF2 and DEAF1 are detected and quantified in the CUT&Tag or ChIP-seq samples via qPCR. The protein could also be detected via western blot, which could also give insight into how specific our antibody binds to the studied protein based on the number of bands seen on the blot.

Any future study on AIRE, FEZF2, and DEAF1 using human thymus would require testing more antibodies, preferably ChIP-grade antibodies if they become available. The antibody should bind with high specificity to the protein in order to determine all binding sites in the genome. In

addition, further optimisation should be carried out to determine the need and strength of fixation and antibody titration.

When the binding sites of AIRE, FEZF2 and DEAF1 are found, the next step should include determining any correlation between binding sites and risk alleles and expression profiles of various T-cell mediated AIDs and explore alternative splicing and binding strength of the transcription factors. Knowing the functions of AIRE, FEZF2 and DEAF1 will greatly improve our understanding of the immune system and will be a stepping stone on the road to understanding the causes of autoimmune disease.

References

Abbas, Abul K., Litchman, Andrew H., & Pillai, Shiv. (2012). Basic Immunology: Functions and Disorders of the Immune System (4 ed.). Philadelphia Elsevier.

Abramson, Jakub, & Anderson, Graham. (2017). Thymic Epithelial Cells. Annual Review of Immunology, 35(1), 85-118.

Afgan, Enis, Baker, Dannon, Batut, Bérénice, van den Beek, Marius, Bouvier, Dave, Čech, Martin, . . . Blankenberg, Daniel. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Research, 46(W1), W537-W544.

Akirav, Eitan M., Ruddle, Nancy H., & Herold, Kevan C. (2011). The role of AIRE in human autoimmune disease. Nature Reviews Endocrinology, 7(1), 25-33.

Bach, J. F. (2001). Protective role of infections and vaccinations on autoimmune diseases. J Autoimmun, 16(3), 347-353.

Bansal, Kushagra, Yoshida, Hideyuki, Benoist, Christophe, & Mathis, Diane. (2017). The transcriptional regulator Aire binds to and activates super-enhancers. Nat Immunol, 18(3), 263-273.

Bolon, Brad. (2012). Cellular and Molecular Mechanisms of Autoimmune Disease. Toxicologic Pathology, 40(2), 216-229.

Brennecke, Philip, Reyes, Alejandro, Pinto, Sheena, Rattay, Kristin, Nguyen, Michelle, Küchler, Rita, . . . Steinmetz, Lars M. (2015). Single-cell transcriptome analysis reveals coordinated ectopic gene-expression patterns in medullary thymic epithelial cells. Nat Immunol, 16(9), 933-941.

Browne, Sarah K. (2014). Anticytokine Autoantibody–Associated Immunodeficiency. Annual Review of Immunology, 32(1), 635-657.

Cano, R Luz Elena, Lopera, H Damaris E, Anaya, J M, Shoenfeld, Y, & Rojas-Villarraga, A. (2013). Introduction to T and B lymphocytes Colombia: El Rosario University Press.

Chen, Lishan, Zheng, Jiashun, Yang, Nan, Li, Hao, & Guo, Su. (2011). Genomic Selection Identifies Vertebrate Transcription Factor Fezf2 Binding Sites and Target Genes. Journal of Biological Chemistry, 286(21), 18641-18649.

Choo, Sung Yoon. (2007). The HLA system: genetics, immunology, clinical testing, and clinical implications. Yonsei medical journal, 48(1), 11-23.

Cooper, G. S., Bynum, M. L., & Somers, E. C. (2009). Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. J Autoimmun, 33(3-4), 197-207. doi:10.1016/j.jaut.2009.09.008

Danso-Abeam, Dina, Staats, Kim A., Franckaert, Dean, Van Den Bosch, Ludo, Liston, Adrian, Gray, Daniel H. D., & Dooley, James. (2013). Aire mediates thymic expression and tolerance of pancreatic antigens via an unconventional transcriptional mechanism. European Journal of Immunology, 43(1), 75-84.

Decker, Janet M. (2012). T Cell Development Immunology.

DeFranco, A., Locksley, R.M., & Robertson, M. (2007). Immunity: The Immune Response in Infectious and Inflammatory Disease: OUP Oxford.

Dendrou, Calliope A., Petersen, Jan, Rossjohn, Jamie, & Fugger, Lars. (2018). HLA variation and disease. Nature Reviews Immunology, 18(5), 325-339.

Derbinski, Jens, Gäbler, Jana, Brors, Benedikt, Tierling, Sascha, Jonnakuty, Sunitha, Hergenhahn, Manfred, . . . Kyewski, Bruno. (2005). Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. The Journal of Experimental Medicine, 202(1), 33.

Derbinski, Jens, Pinto, Sheena, Rösch, Stefanie, Hexel, Klaus, & Kyewski, Bruno. (2008). Promiscuous gene expression patterns in single medullary thymic epithelial cells argue for a stochastic mechanism. Proceedings of the National Academy of Sciences, 105(2), 657.

Dhalla, Fatima, Baran‐Gale, Jeanette, Maio, Stefano, Chappell, Lia, Holländer, Georg A., & Ponting, Chris P. (2019). Biologically indeterminate yet ordered promiscuous gene expression in single medullary thymic epithelial cells. The EMBO Journal.

Dieli, F. (2003). Dendritic cells and the handling of antigen. Clinical and experimental immunology, 134(2), 178-180.

Egerton, M., Scollay, R., & Shortman, K. (1990). Kinetics of mature T-cell development in the thymus. Proc Natl Acad Sci U S A, 87(7), 2579-2582.

ENCODE, & Consortium, Project. (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science, 306(5696), 636-640.

Fatourou, Evangelia M., & Koskinas, John S. (2009). Adaptive immunity in hepatocellular carcinoma: prognostic and therapeutic implications. Expert Review of Anticancer Therapy, 9(10), 1499-1510.

Fuhlbrigge, Rebecca, & Yip, Linda. (2014). Self-Antigen Expression in the Peripheral Immune System: Roles in Self-Tolerance and Type 1 Diabetes Pathogenesis. 14(9).

Fuxman Bass, Juan I., Tamburino, Alex M., Mori, Akihiro, Beittel, Nathan, Weirauch, Matthew T., Reece-Hoyes, John S., & Walhout, Albertha J. M. (2014). Transcription factor binding to Caenorhabditis elegans first introns reveals lack of redundancy with gene promoters. Nucleic Acids Research, 42(1), 153-162.

Gabrielsen, I. S. M., Helgeland, H., Akselsen, H., HC, D. Aass, Sundaram, A. Y. M., Snowhite, I. V., . . . Lie, B. A. (2019). Transcriptomes of antigen presenting cells in human thymus. PLoS One, 14(7), e0218858.

Gies, Vincent, Guffroy, Aurélien, Danion, François, Billaud, Philippe, Keime, Céline, Fauny, Jean-Daniel, . . . Korganow, Anne-Sophie. (2017). B cells differentiate in human thymus and express AIRE. Journal of Allergy and Clinical Immunology, 139(3), 1049-1052.e1012.

Hadeiba, Husein, Lahl, Katharina, Edalati, Abdolhossein, Oderup, Cecilia, Habtezion, Aida, Pachynski, Russell, . . . Butcher, Eugene C. (2012). Plasmacytoid Dendritic Cells Transport Peripheral Antigens to the Thymus to Promote Central Tolerance. Immunity, 36(3), 438-450.

Haroun, Heshmat Sw. (2018). Aging of thymus gland and immune system. MOJ Anatomy & Physiology, 5(2).

Haynes, Barton F., Sempowski, Gregory D., Wells, Alvin F., & Hale, Laura P. (2000). The Human Thymus During Aging. Immunologic Research, 22(2-3), 253-262.

Hemminki, K., Li, X., Sundquist, K., & Sundquist, J. (2009). Shared familial aggregation of susceptibility to autoimmune diseases. Arthritis Rheum, 60(9), 2845-2847.

Henikoff, Steven, Kaya-Okur, Hatice S., Wu, Steven J., Codomo, Christine A., Pledger, Erica S., Bryson, Terri D., . . . Ahmad, Kami. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nature communications, 10(1), 1930-1930.

Hirosue, Sachiko, & Dubrot, Juan. (2015). Modes of Antigen Presentation by Lymph Node Stromal Cells and Their Immunological Implications. Frontiers in immunology, 6, 446-446.

Iberg, Courtney A., & Hawiger, Daniel. (2020). Natural and Induced Tolerogenic Dendritic Cells. The Journal of Immunology, 204(4), 733-744.

James, Eddie, & Kwok, William. (2008). Low-Affinity Major Histocompatibility Complex–Binding Peptides in Type 1 Diabetes. Diabetes, 57, 1788-1789.

Janeway, CA jr., Travers, P, & Walport, M. (2001). Immunobiology: The Immune System in Health and Disease (5 ed.). New York Garland Science

Jensik, Philip J., Vargas, Jesse D., Reardon, Sara N., Rajamanickam, Shivakumar, Huggenvik, Jodi I., & Collard, Michael W. (2014). DEAF1 Binds Unmethylated and Variably Spaced CpG Dinucleotide Motifs. PLoS One, 9(12), e115908.

Kaya-Okur, Hatice, & Henikoff, Steven. (2020). Bench top CUT&Tag. protocols.io. Protocols.io, Version 3

Kidder, Benjamin L., Hu, Gangqing, & Zhao, Keji. (2011). ChIP-Seq: technical considerations for obtaining high-quality data. Nat Immunol, 12(10), 918-922.

Klein, Ludger, Kyewski, Bruno, Allen, Paul M., & Hogquist, Kristin A. (2014). Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). Nature Reviews Immunology, 14(6), 377-391.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol, 10(3), R25.

Meredith, Matthew, Zemmour, David, Mathis, Diane, & Benoist, Christophe. (2015). Aire controls gene expression in the thymic epithelium with ordered stochasticity. Nat Immunol, 16(9), 942-949.

Meyer, Steffen, Woodward, Martin, Hertel, Christina, Vlaicu, Philip, Haque, Yasmin, Kärner, Jaanika, . . . Hayday, Adrian. (2016). AIRE-Deficient Patients Harbor Unique High-Affinity Disease-Ameliorating Autoantibodies. Cell, 166(3), 582-595.

Mogensen, Trine H. (2009). Pathogen recognition and inflammatory signaling in innate immune defenses. Clinical microbiology reviews, 22(2), 240-273.

Murata, Shigeo, Sasaki, Katsuhiro, Kishimoto, Toshihiko, Niwa, Shin-ichiro, Hayashi, Hidemi, Takahama, Yousuke, & Tanaka, Keiji. (2007). Regulation of CD8+ T Cell Development by Thymus-Specific Proteasomes. Science, 316(5829), 1349.

Murphy, Kenneth, Travers, Paul, & Waltport, Mark. (2008). Janeway's Immunobiology (7th ed.). New York and London Garland Science

Norris, Paula S, & Ware, Carl F. (2013). The LTβR Signaling Pathway Austin, TX: Landes Bioscience

Ohigashi, I., Kozai, M., & Takahama, Y. (2016). Development and developmental potential of cortical thymic epithelial cells. Immunological Reviews, 271(1), 10-22.

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(24), 4699-4710.

Param, Peter. (2009). The Immune System (3rd ed.). London and New York Garland Science

Passos, Geraldo A., Speck-Hernandez, Cesar A., Assis, Amanda F., & Mendes-da-Cruz, Daniella A. (2018). Update on Aire and thymic negative selection. Immunology, 153(1), 10-20.

Patrick, L. (2009). Thyroid disruption: mechanism and clinical implications in human health. Altern Med Rev, 14(4), 326-346.

Perry, Justin S. A., Russler-Germain, Emilie V., Zhou, You W., Purtha, Whitney, Cooper, Matthew L., Choi, Jaebok, . . . Hsieh, Chyi-Song. (2018). Transfer of Cell-Surface Antigens by Scavenger Receptor CD36 Promotes Thymic Regulatory T Cell Receptor Repertoire Development and Allo-tolerance. Immunity, 48(6), 1271.

Robinson, James T., Thorvaldsdóttir, Helga, Winckler, Wendy, Guttman, Mitchell, Lander, Eric S., Getz, Gad, & Mesirov, Jill P. (2011). Integrative genomics viewer. Nature Biotechnology, 29(1), 24-26.

Sadava, David, Hillis, David M., Heller, H. Craig, & Berenbaum, May R. (2014). LIFE (10th ed.): Sinauer Associates, Inc. .

Sakata, Mie, Ohigashi, Izumi, & Takahama, Yousuke. (2018). Cellularity of Thymic Epithelial Cells in the Postnatal Mouse. The Journal of Immunology, 200(4), 1382.

Sansom, S. N., Shikama-Dorn, N., Zhanybekova, S., Nusspaumer, G., Macaulay, I. C., Deadman, M. E., . . . Hollander, 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(12), 1918-1931.

Sharma, Saurabh, Ghosh, Sreejoyee, Singh, Lalit, Sarkar, Ashish, Malhotra, Rajesh, Garg, Onkar, . . . Biswas, Sagarika. (2014). Identification of Autoantibodies against Transthyretin for the Screening and Diagnosis of Rheumatoid Arthritis. PLoS One, 9, e93905.

Takaba, H., Morishita, Y., Tomofuji, Y., Danks, L., Nitta, T., Komatsu, N., . . . Takayanagi, H. (2015). Fezf2 Orchestrates a Thymic Program of Self-Antigen Expression for Immune Tolerance. Cell, 163(4), 975-987.

Takaba, H., & Takayanagi, H. (2017). The Mechanisms of T Cell Selection in the Thymus. Trends Immunol, 38(11), 805-816.

van den Boorn, Jasper G., Le Poole, I. Caroline, & Luiten, Rosalie M. (2006). T-cell avidity and tuning: the flexible connection between tolerance and autoimmunity. International reviews of immunology, 25(3-4), 235-258.

Wang, Hao, Maurano, Matthew T., Qu, Hongzhu, Varley, Katherine E., Gertz, Jason, Pauli, Florencia, . . . Stamatoyannopoulos, John A. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome research, 22(9), 1680-1688.

Wimberley, Charles E., & Heber, Steffen. (2019). PeakPass: Automating ChIP-Seq Blacklist Creation. Journal of Computational Biology, 27(2), 259-268.

Xing, Yan, & Hogquist, Kristin A. (2012). T-cell tolerance: central and peripheral. Cold Spring Harbor perspectives in biology, 4(6), a006957.

Yamano, Tomoyoshi, Nedjic, Jelena, Hinterberger, Maria, Steinert, Madlen, Koser, Sandra, Pinto, Sheena, . . . Klein, Ludger. (2015). Thymic B Cells Are Licensed to Present Self Antigens for Central T Cell Tolerance Induction. Immunity, 42(6), 1048-1061.

Yip, L., Su, L., Sheng, D., Chang, P., Atkinson, M., Czesak, M., . . . Creusot, R. J. (2009). Deaf1 isoforms control the expression of genes encoding peripheral tissue antigens in the pancreatic lymph nodes during type 1 diabetes. Nat Immunol, 10(9), 1026-1033.

Yip, Linda, Creusot, Remi J., Pager, Cara T., Sarnow, Peter, & Fathman, C. Garrison. (2013). Reduced DEAF1 function during type 1 diabetes inhibits translation in lymph node stromal cells by suppressing Eif4g3. Journal of molecular cell biology, 5(2), 99-110.

Zhang, Jun-Ming, & An, Jianxiong. (2007). Cytokines, Inflammation, and Pain. International Anesthesiology Clinics, 45(2), 27-37.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., . . . Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol, 9(9), R137.

Appendix I

Buffers used during thymus dissociation and mTEC cell enrichment:

DNase I stock

Reagent	Stock Concentration	End Concentration	Volume
DNase I	100 %	10 mg/mL	100 mg
Sterile distilled H_2O			10 mL
		Total	10 mL

Liberase TM stock

Reagent	Stock Concentration	End Concentration	Volume
Liberase TM	100 %	13 U/mL	10 mg
H_2O			10 mL
		Total	10 mL

DNase I [0.1 %], per C-tube dissociation

Reagent	Stock Concentration	End Concentration	Volume
DNase I stock	10 mg/mL	0.1 % w/v	100 µL
RPMI-1640			10 mL
		Total	10 mL

Liberase TM buffer 0.17U/ml, per C-tube dissociation

Reagent	Stock Concentration	End Concentration	Volume
Liberase TM	13 U/mL	0.17 U/mL	643.5 µL
DNase I [0.1 %]	0.1 % w/v	0.02 % w/v	9.9 mL
RPMI-1640			39.0 mL
		Total	50 mL

Trypsin-EDTA diluted in Liberase TM-buffer, per C-tube dissociation

Reagent	Stock Concentration	End Concentration	Volume
Trypsin-EDTA	0.25 %	0.05 %	500 µL
Liberase TM buffer	0.17 U/mL	0.14 U/mL	2 mL
		Total	2.5 mL

Solution C

Reagent	Stock Concentration	End Concentration	Volume
EDTA, pH 8 *	0.5 mM	0.005 mM	5 mL
FBS *	100 % v/v	5 % v/v	25 mL
PBS			470 mL
		Total	500 mL

* Filtered through 0.2 µm filter

Add DNaseI before use (1:1000 of 10mg/ml stock).

Solution D

Reagent	Stock Concentration	End Concentration	Volume
NaCl	100 %	137 mM	4 g
EDTA, pH 8	0.5 mM	0.005 mM	5 mL
Tricine	100 %	10 mM	0.9 g
		Total	500 mL

Adjust pH to 7.4 by adding drops of 5 M NaOH

Top up with MilliQ water up to 500 ml

Filter through a 0.2 µm filter using a 50 ml syringe.

STEMbuffer1 for CD45 depletion

Reagent	Stock Concentration	End Concentration	Volume
EDTA, pH 8 *	0.5 mM	0.001 mM	1 mL
FBS *	100 % v/v	2 % v/v	10 mL
PBS			489 mL
		Total	500 mL

*Filtered through 0.2 µm filter

30 % OptiPrep

Reagent	Stock Concentration	End Concentration	Volume
OptiPrep	60 %	30 %	20 mL
Solution D			20 mL
		Total	40 mL

1.07 g/ml OptiPrep gradient

Reagent	Stock Concentration	End Concentration	Volume
OptiPrep	30 %	12.2 %	20.4 mL
Solution D			29.6 mL
		Total	50 mL

Add DNaseI before use (1:1000 of 10mg/ml stock).

1.061 g/ml OptiPrep gradient

Reagent	Stock Concentration	End Concentration	Volume
OptiPrep	30 %	10.5 %	17.5 mL
Solution D			32.5 mL
		Total	50 mL

Add DNaseI before use (1:1000 of 10mg/ml stock).

Buffer recipes used for Chromatin isolation

Fixation Buffer

Reagent	Stock Concentration	End Concentration	Volume [mL]
EDTA	0,5M	1mM	0.5
EGTA	0,5M	0,5mM	0.25
NaCl	5M	100mM	1
HEPES, pH7,6	1M	50mM	2.5
H_2O			46.35
		Total	**50**

Lysis Buffer 1

Reagent	Stock Concentration	End Concentration	Volume [mL]
EDTA	0,5M	1mM	0.1
EGTA	0,5M	0,5mM	0.05
NaCl	5M	10mM	0.1
Igepal	10%	0,2%	1
Tris-HCl, pH 8	1M	10mM	0.5
Triton-X	10%	0,2%	1
H_2O			47.25
		Total	**50**

Lysis Buffer 2

Reagent	Stock Concentration	End Concentration	Volume [mL]
EDTA	0,5M	1mM	0.1
EGTA	0,5M	0,5mM	0.05
NaCl	5M	150mM	1.5
HEPES, pH7,6	1M	50mM	2.5
H_2O			45.85
		Total	**50**

Covaris Sonication Buffer

Reagent	Stock Concentration	End Concentration	Volume [mL]
EDTA	0,5M	3mM	0.05
SDS	10%	0,6%	0.25
Tris-HCl, pH 8	1M	30mM	0.625
H_2O			24
		Total	**25**

Elutionbuffer

Reagent	Stock Concentration	End Concentration	Volume [mL]
EDTA	0,5M	1mM	0.1
EGTA	0,5M	0,5mM	0.05
Tris-HCl, pH8	1M	10mM	0.5
H_2O			49.35
		Total	**50**

Buffer recipes used for IP

3xCovaris Dilution Buffer

Reagent	Stock Concentration	End Concentration	Volume [mL]
EDTA	0,5M	3mM	0,3
EGTA	0,5M	1,5mM	0,15
NaCl	5M	450mM	4,5
SDS	10%	0,6%	3
Tris-HCl, pH 8	1M	30mM	1,5
Triton-X	10%	3%	15
H_2O			25,55
		Total	**50**

Wash Buffer 1

Reagent	Stock Concentration	End Concentration	Volume [mL]
DOC	10%	0,1%	0,5
EDTA	0,5M	1mM	0,1
EGTA	0,5M	0,5mM	0,05
NaCl	5M	150mM	1,5
SDS	10%	0,1%	0,5
Tris-HCl, pH 8	1M	10mM	0,5
Triton-X	10%	1%	5
H_2O			41,85
		Total	**50**

Wash Buffer 2

Reagent	Stock Concentration	End Concentration	Volume [mL]
DOC	10%	0,1%	0,5
EDTA	0,5M	1mM	0,1
EGTA	0,5M	0,5mM	0,05
NaCl	5M	500mM	5
SDS	10%	0,1%	0,5
Tris-HCl, pH 8	1M	10mM	0,5
Triton-X	10%	1%	5
H_2O			38,35
		Total	**50**

Wash Buffer 3

Reagent	Stock Concentration	End Concentration	Volume [mL]
DOC	10%	0,5%	2,5
EDTA	0,5M	1mM	0,1
EGTA	0,5M	0,5mM	0,05
LiCl	4M	250mM	3,12
Igepal	10%	0,5%	2,5
Tris-HCl, pH 8	1M	10mM	0,5
H_2O			41,23
		Total	**50**

Wash Buffer 4

Reagent	Stock Concentration	End Concentration	Volume [mL]
EDTA	0,5M	1mM	0,1
EGTA	0,5M	0,5mM	0,05
Tris-HCl, pH 8	1M	10mM	0,5
H_2O			49,35
		Total	**50**

Buffer recipes used in CUT&Tag

Wash buffer

Reagent	Stock Concentration	End Concentration	Volume
HEPES, pH 7.5	1 M	20 mM	1 mL
NaCl	5 M	150 mM	1.5 mL
Spermidine	1 M	0.5 mM	25 µL
H_2O			47.5
		Total	**50 mL**

Added 1 Roche Complete Protease Inhibitor EDTA-Free tablet.

Binding buffer

Reagent	Stock Concentration	End Concentration	Volume
HEPES, pH 7.5	1 M	20 mM	200 µL
KCl	1 M	10 mM	100 µL
$CaCl_2$	1 M	1 mM	10 µL
$MnCl_2$	1 M	1 mM	10 µL
H_2O			9.68 mL
		Total	**10 mL**

Dig-wash buffer

Reagent	Stock Concentration	End Concentration	Volume
Digitonin	5 % w/v	0.05 % w/v	400 µL
Wash Buffer			40 mL
		Total	**40.4 mL**

Antibody buffer

Reagent	Stock Concentration	End Concentration	Volume
EDTA	0.5 M	2 mM	8 µL
BSA	30 % w/v	0.1 % w/v	6.7 µL
Dig-wash buffer			2 mL
		Total	**2 mL**

Dig-300 buffer

Reagent	Stock Concentration	End Concentration	Volume
HEPES, pH 7.5	1 M	0.25 M	1 mL
NaCl	5 M	3.75 M	3 mL
Spermidine	1 M	6.25 mM	25 µL
Digitonin	5 % w/v	0.01 % w/v	8 µL
		Total	**4 mL**

Added 1 Roche Complete Protease Inhibitor EDTA-Free tablet before use

Appendix II

Materials

Instruments and equipment:

Product name	Provider	Catalog nr.	Prod. Country
Dounce homogenizer	Sigma Aldrich	D8938-1SET	
Novaseq 6000	Illumina		
QuantStudio™ 12K Flex Real-Time PCR System	Applied Biosystems	4471134	USA
MicroAmp™ Optical 384-Well Reaction Plate with Barcode	Applied Biosystems	4326270	USA
MicroAmp™ Optical Adhesive Film	Applied Biosystems	4360954	USA
Qubit 2.0 Flurometer	Invitrogen	Q32866	USA
Tapestation 2200	Agilent Technologies		
D1000 ScreentTape	Agilent Technologies	5067-5582	USA
D1000 Reagents	Agilent Technologies	5067-5583	USA
GentleMACS C tubes	Miltenyi	130-093-237	
GentleMACS Octo Dissociator	Miltenyi Biotec	130-095-937	
milliTUBE 1ml AFA Fiber	Covaris	520130	USA
Rack 24 Place milliTUBE 1 ml	Covaris	500368	USA

Reagents:

Product name	Provider	Cat. nr.	Country
BSA	Sigma-Aldrich	A7906-100G	Germany
Concanavalin A-coated beads	Polysciences, Inc.	86057	Germany
DNase I	Sigma Aldrich	11284932001	Germany
Disuccinimidyl glutarate	Thermo Fisher	20593	USA
Dynabeads Protein G	Thermo Fisher	10004D	Norway
Digitonin	Norex	BN2006	USA
EDTA, pH8	Ambicon	Am9262	Lithuania
FBS	Sigma Aldrich	R7509	USA
Formaldehyde [16%]	Thermo Fisher	28906	USA
Glycine	Bio-Rad Labratories	161-0718	USA
Halt Protinase inhibitor cocktail, EDTA-free	Thermo Fisher	78439	USA
Liberase TM	Sigma Aldrich	5401119001	
NaCl	Sigma Aldrich	S5886-500G	
NEBNext HiFi 2x PCR Master mix	New England Biolabs	M0541	USA
OptiPrep	STEMCELL	07820	Canada

PBS	Gibco	14190-094	
Phenol:chloroform:isoamyl alcohol	Invitrogen	15593031	
Proteinase Inhibitor Cocktail (PIC), EDTA-free	Thermo Scientific	78439	Netherlands
RNase Cocktail™ Enzyme Mix	Invitrogen	AM2286	Lithuania
Complete Protease Inhibitor EDTA-Free tablet	Roche	11873580001	
RPMI-1640	Sigma Aldrich	R7509	USA
Sterile deatilled water	Qiagen	129114	
Tricine	Sigma Aldrich	T0377	
Tris pH 8.0	Thermo Fisher	AM9855G	
Trypsin-EDTA	Thermo Fisher	25200056	
2x PowerUp™ SYBR™ Green Master Mix	APBio	A25742	Lithuania

Software:

Product name	Version	Provider	Reference
2200 TapeStation	A.01.04	Agilent	
QuantStudio 12K Flex	V.1.2.2	Applied Biosystems	
Galaxy		Galaxy	(Afgan et al., 2018)
Integrative Genomics Viewer	2.8.0	IGV Broad Institute	(Robinson et al., 2011)
Microsoft Excel		Microsoft	

Antibodies:

Product name	Provider	Catalog nr.	Lot nr.	Host
AIRE (isoforms 1 and 2) Polyclonal antibody	GeneTex	GTX13573	821904782/ 822000528	Goat
AIRE-1 Antibody (C-2)	Santa Cruz Biotechnology	SC-373703	C0116	Mouse
AIRE Monoclonal Antibody (TM-724)	Thermo Fisher	14953480	4340082	Rat
Anti-ZNF312/FEZF2 Polyclonal antibody	Abcam	Ab214186	GR3177398-3/ GR3303245-1/ GR3303245-2	Rabbit
DEAF1 (2A2) Monoclonal antibody	Invitrogen	MA5-21070	UJ2864451	Mouse
IHCPlus™ NUDR / DEAF1 Polyclonal antibody	LSBio	LS-B10862	169557	Rabbit
H3K27me3 Antibody - ChIP-seq Grade	Diagenode	C15410195	A0824D	Rabbit
IgG Goat Isotype Control	Thermo Fisher	02-6202		Goat
IgG Rabbit Polyclonal	Diagenode	C15410206	RIG001AD	Rabbit
Guniea pig α-Rabbit IgG	Antibodies-online	ABIN101961		Guniea pig
CTCF Antibody - ChIP-seq	Diagenode	C15410210-10	A2359-00234P	Rabbit

Grade

Commercial kits:

Product name	Provider	Catalog nr.	Country
ChIP DNA clean and Concentrator Kits	Zymo Research	D5205	
Accel-NGS 2S, Plus DNA Library Kit (24 rxns)	Swift Biosciences	21024	
Accel-NGS 2S, Unique Dual Indexing Kit (24 indices, 96 rxns)	Swift Biosciences	29096	
EasyStep Human CD45 depetion Kit II	STEMCELL	17898	Canada
QIAquick PCR purification kit	QIAgen	28104	Germany
Qubit® dsDNA HS Assay Kit	Invitrogen	Q32851	USA

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway