UNIVERSITY OF TARTU
FACULTY OF MEDICINE
DEPARTMENT OF GENERAL AND MOLECULAR PATHOLOGY
Molecular Pathology Study Group

Julia Maslovskaja

The autoimmune regulator protein participates in loop formation between epidermal differentiation complex gene family members

Master thesis

Supervisor: PhD Pärt Peterson

TARTU 2007

CONTENTS

ABBREVIATIONS

3C – chromosome conformation capture

4C – (circular) chromosome conformation capture (on chip)

5C – chromosome conformation capture carbon copy

ACH – active chromatin hub

ad – adenovirus

AIRE – autoimmune regulator

APECED – autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy

bp – base pair

CD – cluster of differentiation

CBP – CREB (cAMP response element-binding) binding protein

CE – cornified envelope

cTEC – cortical thymic epithelial cell

DAPI – 4',6-diamidino-2-phenylindole

DMEM – Dulbecco's Modified Eagle's Medium

EDC – epidermal differentiation complex

EDTA – ethylenediamine tetraacetic acid

EpCAM – epithelial cell adhesion molecule

FC – fold change

FISH – fluorescence *in situ* hybridization

FITC – fluorescein isothiocyanate

GAL4 – galactose 4

Gapd – glyceraldehyde-3-phosphate dehydrogenase

GFP – green fluorescence protein

Hbb – hemoglobin beta

HS – hypersensitive site

HSR – homogenously staining region

kb – kilobase

Klf4 – Krüppel-like factor 4

KO – knockout

LacZ – beta-D-galactosidase

LCR – locus control region

LEP – late envelope protein

Mb – megabase

MHC – major histocompatibility complex

mTEC – medullary thymic epithelial cell

NLS – nuclear localization signal

nt – nucleotide

PBS – phosphate buffered saline

PCR – polymerase chain reaction

PE – phycoerythrin

PHD – plant homeodomain

PRR – proline rich region

SAND – Sp100, AIRE, NucP41/75, DEAF-1

SDS – sodium dodecyl sulfate

SPRR – small proline-reach repeat

TBE – Tris borate EDTA

TCR – T cell receptor

TG – transglutaminase

T_H1, T_H2 – T helper type1/type 2

qPCR – quantitative polymerase chain reaction

WT – wild-type

INTRODUCTION

Thymic medullary epithelial cells have an extraordinary ability to express large number of genes, which otherwise have tightly regulated expression profile in peripheral organs. The purpose of such 'promiscuous' expression is to present the self-molecules to maturating T lymphocytes and to eliminate those cells that are capable of recognizing autoantigens. In that way, the attack of immune system against its own tissues is avoided in organism.

It is shown clearly that the autoimmune regulator (AIRE) protein is responsible for this unusual expression pattern. However, despite the extensive research of AIRE function in recent years, the complete understanding of how exactly AIRE induces so many genes still lacks. As gene number activated by AIRE reaches hundreds, it was suggested that regulation does not include interaction with specific promoters but epigenetic mechanisms are involved. Microarray data comparing Aire-deficient mice to their normal counterparts indicate that many Aire-regulated genes are localized in genomic clusters, including major histocompatibility complex, chemokine clusters and epidermal differentiation complex (EDC). EDC combines several gene families involved in epidermal barrier formation and the fact that patients with AIRE mutations often develop ectodermal deficiencies without known etiology makes this cluster an interesting target for investigation.

As Aire-regulated genes in clusters alternate with genes that are not influenced by AIRE, it was proposed that this effect is achieved by loop formation around active chromatin hub, where activated genes are brought together in close proximity in the center and repressed genes loop out to the periphery. To test this hypothesis we applied chromosome conformation capture (3C) technique that enables to determine the interaction frequency of genomic loci. We examined if AIRE triggers loop formation between EDC cluster members.

1. LITERATURE REVIEW

1.1 IMMUNE TOLERANCE

The receptors of T and B cells, the functional cells of immune system, are assembled by a process of random rearrangement. As a result of this process, T and B cells produce two receptor chains, α and β chain, which both are combined by recombination of highly variable genetic loci. Such recombination in T cells yields an enormous number (approximately 10^{12}) of various receptor protein complexes with different binding affinities for diverse peptide molecules. The cost for this diversity is high, as in only small proportion of T cells these receptors are functionally active and capable of recognizing foreign antigens. The majority of receptors have truncated or non-functional form, and certainly, there are proteins that have strong affinity for epitopes of their own organism (Chaplin, 2006). To enable normal functioning of immune system, such self-reactive cells must be eliminated or their activities should be suppressed. If this elimination does not occur properly, the consequences will appear as attack against self proteins in the form of organ infiltrates and damaging autoantibodies, which subsequently give rise to autoimmune diseases (Rioux and Abbas, 2005).

The ability of organism to discriminate self from non-self and eliminate the autoreactive cells is called tolerance. Immune tolerance can be divided into two branches: central or peripheral, as respect to the location where the tolerance induction or maintenance occurs (Pugliese, 2004). Central immune organs are thymus and bone marrow, the tissues where lymphocyte maturation take place, so the checking process is done on the early stages of cell life (Mathis and Benoist, 2004). If some of the self-reactive cells still succeed to escape from central immune organs, the peripheral tolerance will be mobilized to ensure inactivation of hazardous cells outside of thymus or bone marrow (Walker and Abbas, 2002).

1.1.1 Selection in the thymus

Central tolerance for T cells occurs in the thymus, where T cells complete their maturation progressively passing through positive and negative selection (Fig. 1). In the early stages of development T cell precursors migrate from bone marrow to thymus, where they go through different stages of differentiation. During this process, developing T cells, also called

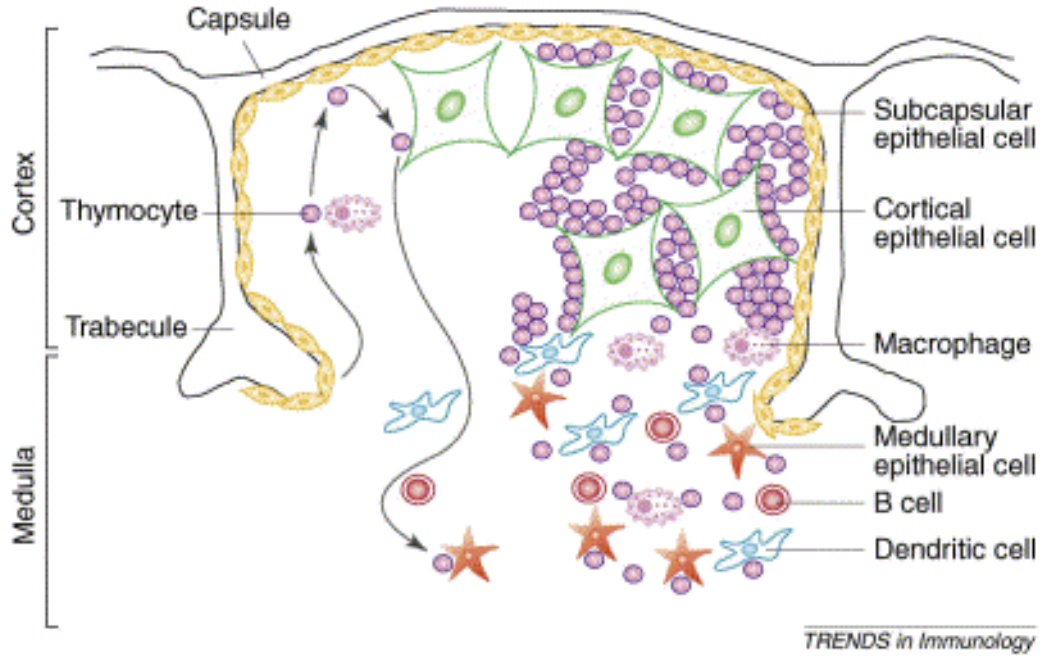


Figure 1. Histological structure of the thymus (Adapted from Kyewski et al., 2002). The main cell types populating in the thymus are shown and the route thymocytes complete during differentiation process is outlined.

thymocytes, upregulate the expression of different cell surface molecules. The most important molecule complexes that T cells start to express on their membrane surface during its development are a) T cell receptor (TCR), which interacts with major histocompatibility complex (MHC) loaded with peptide, b) CD3, which is responsible for directing the signal, received upon the binding, inside the cell, c) CD4 and CD8, which interact with MHC constant region and distinguish between MHC II and I complex, respectively. In addition, T cells activate expression of the adhesion molecules needed for cell-cell contacts and the cytokine receptors, which are necessary for T cell migration (Gray et al., 2005).

Several morphological regions can be distinguished within the thymus: the most important areas among them are the thymic cortex and the medulla. T cell progenitors, originating from bone marrow, enter the thymus at corticomedullary junction, which is rich in blood vessels, and start their migration through the different thymic areas constantly making cell contacts with thymic stroma cells (Takahama, 2006). The classical view of this process is that in the cortex densely

populated with early thymocytes, positive selection is predominant process, which ensures the survival and further maturation of T cells that recognize the MHC complex of their own organism. Upon successful accomplishment of positive selection the thymocytes proliferate and move further to medulla. In the medulla region, thymocytes interact with medullary thymic epithelial (mTECs) and dendritic cells (Starr et al., 2003). Dependent on the strength of interaction between TCR on thymocyte surface and MHC presented by antigen-presenting cells, they follow three possible routes of differentiation: if MHC II-peptide-TCR interaction is weak, the thymocytes are safe for organism and they differentiate into effectory T cells; if the binding is intermediate, cells are fated to die; the thymocytes that bind strongly, become regulatory T cells (Hori et al., 2003; Palmer, 2003).

1.1.2 Negative selection by medullary TECs

It was long thought that only thymocytes reacting against ubiquitous or blood-contained antigens can be eliminated during negative selection in the thymus, for the simple reason that the antigens of other tissues are inaccessible for developing thymocytes in the thymus. The breakthrough came in late 90s, when it was discovered that many tissue-restricted genes are expressed in certain type of thymic epithelial cells (TEC) – medullary TECs (Derbinski et al., 2001; Farr and Rudensky, 1998). The scope of the genes expressed by mTECs is large, the genes which are expressed under precise regulation in specific sites of the body or during certain developmental stages are activated in these cells. It is estimated that the number of genes expressed in that way in mTECs reaches thousands, covering 10% of the genome and for that reason, this type of expression was named ‘promiscuous expression’ (Derbinski et al., 2005). Although it is now accepted that the extent of central tolerance is large, it still does not diminish the importance of peripheral tolerance, which stays to be important mechanism for several antigens that T cells encounter in periphery (Pugliese, 2004).

1.1.3 AIRE and APECED

In 1997, autoimmune regulator (*AIRE*) gene was identified (The Finnish-German Consortium, 1997; Nagamine et al., 1997). *AIRE* gene consists of 14 exons localized on chromosome 21 in the human genome (Villasenor et al., 2005). AIRE expression is predominantly restricted to thymus and lymph node tissue, with the strongest expression in mTECs (Klamp et al., 2006). Defects in the AIRE gene cause the rare autosomal-recessive systemic autoimmune disease termed

autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED, OMIM #240300) (Kyewski and Klein, 2006).

APECED is characterized by the presence of multiple symptoms, the most frequent of which are chronic mucocutaneous candidiasis, hypoparatyroidism and Addison's disease. In addition to the latter more rare symptoms can arise, including type 1 diabetis, gastritis, hepatitis, pernicious anemia, vitiligo and alopecia, gonadal failure and ectodermal dysplasia (Peterson et al., 2004; Villasenor et al., 2005). Even though for many conditions autoimmune etiology was shown and specific autoantibodies were found in patients' blood (Perniola et al., 2000), the reason why APECED patients are susceptible to *Candida* infection and ectodermal deficiencies still remains a mystery. AIRE expression was also detected in peripheral monocyte-dendritic cells by some research groups, possibly, ensuring resistance against *Candida* (Kogawa et al., 2002).

Already after initial analyzes of the AIRE protein sequence, it was proposed that AIRE regulates the expression of other genes on transcription level. The major hint was the presence of several protein motifs found in other transcriptional regulators (Fig. 2). Among them, two PHD (plant homeodomain) zinc fingers, the domains that have recently been shown to interact with differentially modified histone variants (Jia et al., 2007; Lan et al., 2007). The AIRE protein also contains four LXXLL stretches, which within other nuclear proteins are known to mediate binding to nuclear receptors, proteins that function as transcription co-activators. In addition, AIRE holds a potential DNA-binding domain SAND (Sp100, AIRE, NucP41/75, DEAF-1) and, accordingly, AIRE binding to DNA has also been proposed (Kumar et al., 2001). Among other protein motifs identified in AIRE, NLS (nuclear localization signal) indicating the higher concentration of protein inside the nucleus the HSR (homogenously staining region) domain is responsible for oligomerization and nuclear dots formation, as several HSR mutations disrupt the characteristic pattern (Bjorses et al., 2000). In overexpression conditions AIRE can be also detected in the cytoplasm following the pattern of cytoplasmic filaments. The function of PRR (proline rich region) is not exactly known.

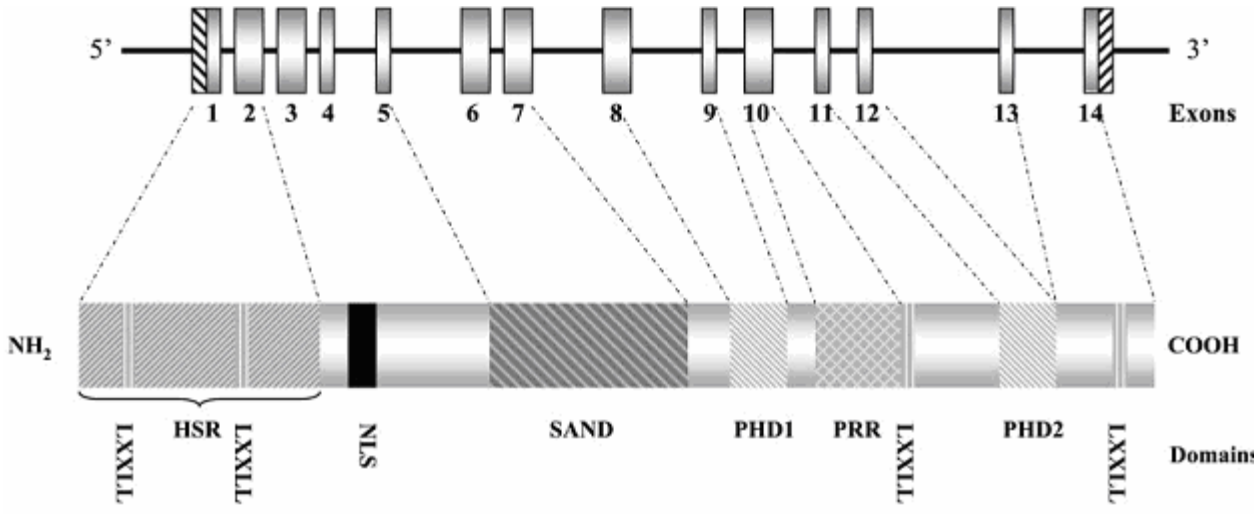


Figure 2. The human *AIRE* gene and the protein structure (Adapted from Meloni et al., 2002). Protein domains of AIRE are depicted: HSR – homogenously stained region, LXXLL – nuclear receptor interaction motif, NLS – nuclear localization signal, SAND – Sp100, AIRE, NucP41/75, DEAF-1, PHD – plant homeodomain, PRR – proline rich region.

In reporter system, where AIRE is tethered to the reporter via GAL4 DNA-binding domain, AIRE strongly activates the transcription from a reporter gene (Halonen et al., 2004; Pitkanen et al., 2000). It was also found that *in vitro* AIRE interacts with ubiquitous co-activator CBP (CREB-binding protein). Further studies showed that AIRE and CBP interaction has co-activating influence on interferon-β minimal promoter (containing only TATA box) (Pitkanen et al., 2005). AIRE oligomerazition is necessary for its transcriptional activity, it was suggested that AIRE phosphorylation triggers its oligomerizition (Kumar et al., 2001). So far AIRE dimers and tetramers have been detected by Western blot analysis of human thymus extract (Peterson et al., 2004).

Generation of Aire-deficient mice demonstrated that similar to APECED disease but milder phenotype develops in murine model with autoantibodies and infiltrates detected in several organs (Anderson et al., 2002; Ramsey et al., 2002). However, the autoantibodies' pool present in Aire-deficient mice differs from those produced in APECED patients (Pontynen et al., 2006). Also, the severity of symptoms and the group of organs targeted by autoimmune reactions depends on the genotype of the mouse strain (Jiang et al., 2005).

The high expression of AIRE in the thymus, autoimmune symptoms that develop in the absence of it and its transcriptional activator properties have lead to the conclusion that AIRE is a major

regulator of promiscuous expression in mTECs (Kyewski and Klein, 2006). However, until recent time there was no direct evidence that the self antigens, which are attacked by autoantibodies in APECED patients and Aire-deficient mice are also regulated by AIRE in the thymus. In case of two antigens, eye- and stomach-specific, a connection between the absence of AIRE protein in the thymus and the development of an autoimmune symptoms in a specific organ was demonstrated (DeVoss et al., 2006; Gavanescu et al., 2007).

Microarray experiments with Aire-knockout mice confirmed that Aire plays a main role in promiscuous expression, as over thousand gene transcripts are expressed at lower level in Aire-knockouts compared to wild-type mice. From microarray data analysis, it also infers that many Aire-regulated genes form clusters in the genome (Gotter et al., 2004; Johnnidis et al., 2005).

1.2 STRUCTURE OF EPIDERMIS

Epithelial tissues form a barrier between the body and the environment. Epithelia can be divided into different types based on the number of layers and the degree of keratinization. Among other functions of the epithelium, the most important one is the protection against environmental stress, noxious chemical substances and various infections (Segre, 2006).

The skin epidermis is an example of stratified squamous epithelium, the toughest and the most protective type. Four layers can be distinguished in epidermis, starting from the inner surface: stratum basale, spinosum, granulosum and corneum (Fig. 3). In normal skin, dividing cells can be found exclusively in basal layer, from where they migrate towards the outer surface accomplishing maturational changes, mainly concerned with the production of keratin (Presland and Jurevic, 2002). The stratum corneum consists of cells where the complex differentiation program is terminated. By that time, the epidermis cells lose their organelles and nucleus and they are almost entirely filled with keratin filaments and are enclosed within a highly specialized structure termed the cornified envelope (CE) (Presland and Dale, 2000). Additionally, these cornified cells are coated from outside with an extracellular layer of lipids (Hardman et al., 1998). This architecture can be imagined as bricks and mortar system, for cell remnants standing as bricks glued together by lipids – the mortar (Nemes and Steinert, 1999). The proteinous insoluble cornified envelope on the inner surface of the plasma membrane and a lipid layer on its outer surface are the major constitutes of the water barrier, which prevents an organism from excessive water loss therefore enabling terrestrial life.

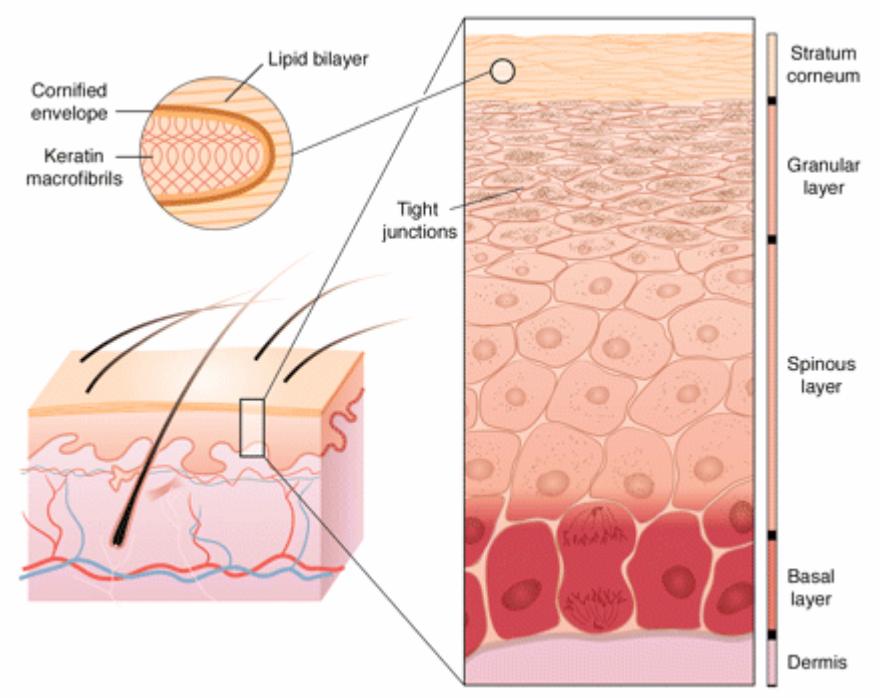


Figure 3. The layers of the epidermis (Adapted from Segre, 2006). Four layers of epidermis with magnified image representing detailed composition of stratum corneum are displayed.

1.2.1 Cornified envelope

The CE is composed of more than 20 proteins and has complex structure. The distinguishable property of CE is its exceptional insolubility. CE is formed by crosslinking small proline rich proteins (SPRR) with larger structural proteins. Crosslinks are mediated by enzymes named transglutaminases, which connect the proteins by N^{ε}-(γ-glutamyl) lysine isopeptide bonds. This special type of connections is responsible for strong mechanical properties of CE. Several temporal stages can be discriminated in epidermal permeability barrier assembly (Fig. 4). The process is dependent on cellular calcium ion concentration and begins with scaffold formation, composed of three proteins – envoplakin, periplakin and involucrin, joined by TG-1. The next step is the fusion of the lipid-containing vesicles produced in Golgi apparatus with plasma membrane releasing molecules on the outer plasma membrane. At the end, other structural proteins, among them loricrin and SPRR proteins, are attached to the scaffold to harden the structure and to complete the envelope formation. TG-3 is responsible for creating the bonds between loricrin and SPRR proteins.

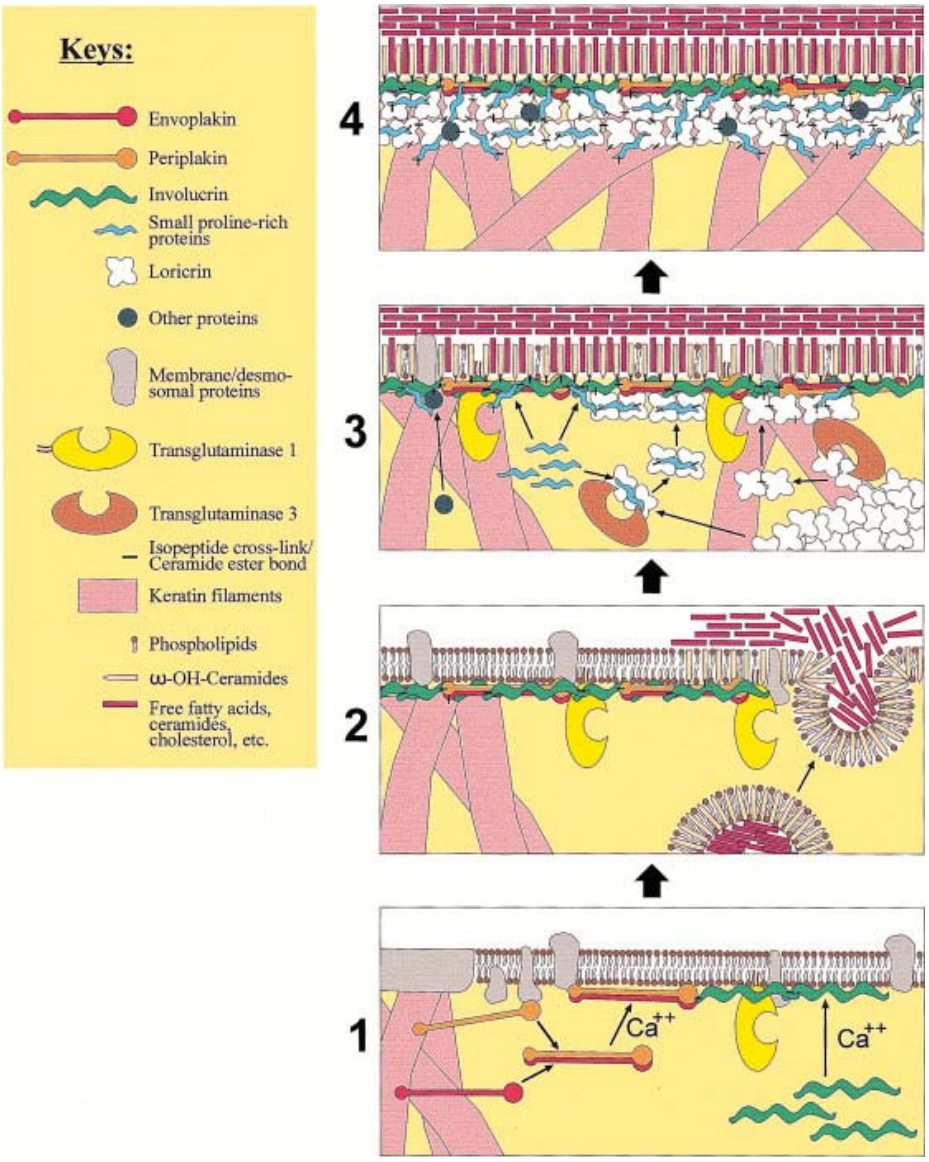


Figure 4. The assembly of cornified envelope (Adapted from Kalinin et al, 2002). Calcium-dependent signal induces envoplakin, periplakin and involucrin deposition on the inner side of the cell membrane. Lipids are transported on the outer side. Activated transglutaminases generate crosslinks between small proline-rich proteins and the major structural protein – loricrin.

1.2.2 CE deficiencies

In attempt to explore in details the functions of proteins associated with CE assembly, several knockout mouse models were created (Aho et al., 2004; Djian et al., 2000; Maatta et al., 2001). Mice with deficiencies in CE components show mild, if any, phenotype changes. Usually, some skin abnormalities are observed in newborns, but they ameliorate quickly after birth. This information implies the existence of effective compensatory mechanisms. The most striking example of such compensation is loricrin-deficient mice (Koch et al., 2000). Although in normal

organisms this protein comprises 80% of cornified envelope, the protein loss does not appear to affect the appearance of skin, suggesting that other structural proteins substitute loricrin.

The most severe consequences due to gene loss develop in Tg1-defiecent mice, the animals die postnatally because of extensive water loss (Matsuki et al., 1998). Loss of TG1 enzyme in humans results in lamellar ichtyosis in newborns. In case of this disease, the body surface is covered by large, plate-like scales that exhibit increased water loss (Russell et al., 1995). Another form of ichtyosis – ichtyosis vulgaris – is possibly connected with mutations in filaggrin. This protein is contained in granules of cells in stratum granulosum, and when discharged, it crosslinks keratin filaments to reinforce the fibers (Nomura et al., 2007). In light of this knowledge, it appears that CE structural proteins, although important, are easily interchangeable. At the same time, there is no acceptable alternative for proteins with enzymatic properties.

1.2.3 Epidermal differentiation complex

Surprisingly, the majority of genes of proteins composing cornified envelope, including involucrin, small proline-rich proteins and loricrin, mentioned in the previous paragraphs, are brought together in the genome in a cluster, located at locus 1q21 in human and 3F2 in mouse and named epidermal differentiation complex (EDC) gene cluster (Fig. 5).

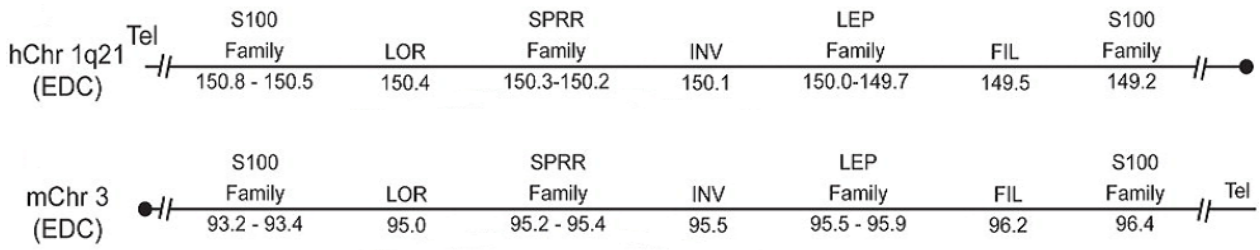


Figure 5. Epidermal differentiation complex (Adapted from Martin et al., 2004). The simplified representation of EDC in human and mouse genome. The position of three large gene families (S100, small proline-rich and late envelope protein) and three individual proteins (loricrin, involucrin and filaggrin) are indicated. Human EDC order is reversed in respect to mouse sequence.

The cluster can be subdivided onto three large subclusters containing following gene families: small proline-rich proteins (SPRR), late envelope proteins (LEP) and Ca^{2+}-binding S100 proteins. The cluster composition is very similar between human and mice, the homology is shared not

only by protein-coding sequences, but is also present in intergenic areas, pointing to possible regulatory sequences conserved between genomes (Martin et al., 2004).

Many EDC proteins share significant sequence similarities, especially the stretches of glutamine and lysine residues. These amino acids are substrates for TGs, which form crosslinks between proteins, giving the necessary endurance and insolubility of epidermal tissues. Taking into account high similarity between EDC genes, it is believed that they have evolved from a common ancestral gene that have underwent many duplications during mammalian evolution (Marenholz et al., 1996). This extensive duplication was necessary to produce proteins with slightly different structures to modulate the properties of epidermal barrier in various sites of body and so far to cope with dry medium of terrestrial life.

1.2.2.1 SPRR family

The *Sprr* genes located tandemly within EDC encode for highly homologous proteins many of them just few amino acids different. Thus, the diversity in regulatory regions seems to be higher than in protein coding sequences of Sprr genes. Apparently, creation of regulatory rather than structural diversity has been the driving force behind the evolution of the SPRR gene family. All Sprr genes consist of two exons with the protein-coding sequence encoded only from the second exon. Sprr and other EDC genes can be coordinately regulated with high expression in papillomas or lower expression in newborn skin (Patel et al., 2003). Also, some local control mechanisms exist upregulating just some of Sprr genes in situations where more rigid barrier is needed. Small bowel resection, bile duct ligation, or introduction of commensal bacteria into germ-free mice causes the increase in SPRR2a protein amount in digestive system (Nozaki et al., 2005). SPRR proteins expression is also detected in reproductive system organs, like in ovarian epithelial cells, uterine lining epithelium at certain stages of the estorous cycle, during development of prostate gland and also in regenerating peripheral neurons (Bonilla et al., 2002; Tan et al., 2006).

Several investigative efforts were made to figure out what mechanisms are responsible for coordinate regulation of Sprr gene family. Three hypersensitive sites were discovered in this area; the enhancing activity of one of them was proved (Martin et al., 2004). Among transcription factors guiding coordinate regulation of this family genes, Krüppel-like factor 4 (Klf4) appears to

have pivotal role. The loss of this factor leads to complete loss of epidermal barrier (Segre et al, 1999).

1.2.2.2 S100 family

The expression of Ca^{2+}-binding S100 proteins is not limited to epidermis, they can be found in different tissues (but only in vertebrate organisms) yet saving cell specific pattern of expression (Donato, 2003). As Ca^{2+} is a widely prevalent secondary messenger, it is clear that S100 proteins are engaged in various processes inside the cell by interacting with target proteins in a calcium-dependent manner, but they also have cytokine-like activities in extracellular space (Eckert et al., 2004). In epidermal tissues, these proteins have potential roles in wound repair, cancer, differentiation, and response to stress. The S100 proteins form homo-, or more rarely, heterodimers, which upon calcium ion binding adopt more relaxed structure exposing binding sites for target proteins (Donato, 1999). The most remarkable protein of this family is S100A7, the overexpression of which is a marker for psoriasis and is also detected in other inflammatory skin diseases (Eckert and Lee, 2006).

S100A8 and S100A9 expression levels, otherwise low in epidermis, increase in response to a wide variety of stresses, such as UV exposure, tape stripping, vaseline application and detergent treatment (Marionnet et al., 2003). S100A8/S100A9 heterodimer is found in high concentration in cytoplasm of neutrophils and exhibits fungistatic activity inhibiting *Candida albicans* growth, probably by chelating zinc, which deprives the pathogens of essential nutrients (Kleinegger et al., 2001; Murthy et al., 1993).

1.2.2.3 LEP family

These proteins are induced during keratinocyte terminal differentiation and relocate to the envelope very late in terminal differentiation where they are substrates for TG-mediated envelope incorporation. During epidermal terminal differentiation, LEP family proteins locate to the envelope much later than loricrin, which is among the last proteins to be incorporated (other proteins on figure) (Marshall et al, 2001). The production of LEP family protein is calcium-dependent, whereby calcium ion concentration increase triggers signal that induces the *LEP* genes activation. It is possible that differential expression of multiple *LEP* genes modulates barrier quality over the animal surface, as has been proposed for SPRR members (Jackson et al., 2005).

1.3 GENE EXPRESSION CONTROL

Although almost all cells in an organism are identical in genetic content, they differ largely in their structure and function. The main reason of such variety is that only certain set of genes is expressed in a specific cell type. Therefore, regulation of gene expression is extremely important to assure the specificity and normal functioning of the cell. Multiple mechanisms exist to complete this task. Every step in transcription and translation processes, including transcription initiation and elongation, mRNA and protein transport, processing and degradation can be decisive for functional product synthesis (Orphanides and Reinberg, 2002). Despite the plethora of possibilities, the most frequent level of control is transcription initiation, where the majority of regulatory events take place (Ogbourne and Antalis, 1998).

1.3.1 Transcriptional *cis*-regulators

Whether transcription will happen depends on the presence of many factors. Two mechanisms important for gene activation can be distinguished: modification of the chromatin structure of a gene locus and the activity of *trans*-acting factors on local *cis*-regulatory sequences. These elements are tightly intermingled (McBride and Kleinjan, 2004). Nucleosome position and conformation, which reflect protein binding to critical *cis*-regulatory, changes in the histone content of nucleosomes and changes in modification status of histones and DNA influence gene expression amounts (Ansel et al., 2003).

In eukaryotes the interplay of several *cis*-elements is required for precise transcription (Fig. 6). Some of them locate far apart from the gene. Those distal *cis*-elements may confer positive (enhancers, LCRs) or negative (silencers) effect on gene activation. Insulator protect genes from the transcriptional activity of neighboring genes (Maston et al., 2006).

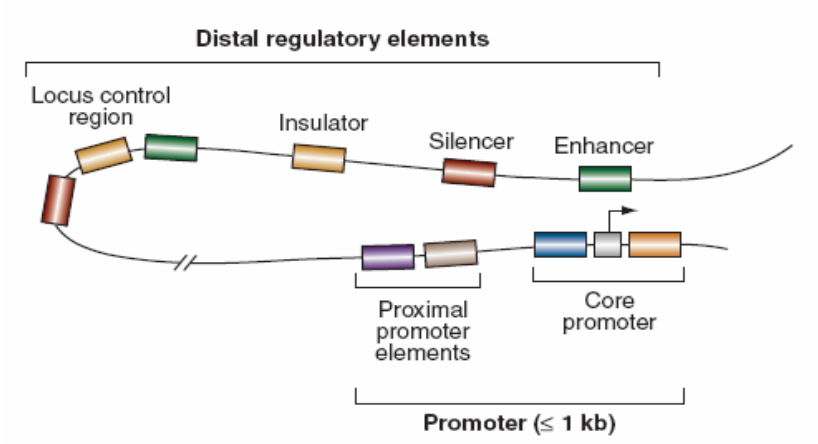


Figure 6. Gene regulatory region (Adapted from Maston et al., 2006). Proximal regulatory elements are situated within 1 kb near the gene. Distal regulatory elements including enhancers, locus control regions, insulators, silencers can be located up to 1 Mb.

1.3.1.1 Locus control regions

Locus control regions (LCR) are regulatory elements involved in regulation of an entire locus or gene cluster. They are operationally defined as elements that direct tissue-specific, physiological expression of a linked transgene in a position-independent and copy-number-dependent manner (Li et al., 2002). They are often marked by the presence of multiple Dnase I hypersensitive sites (HS) and may combine different *cis*-regulatory elements (Maston et al., 2006), however, the only activity they show is strong enhancing effect. Human β-globin locus was first where LCR was identified (Grosveld et al., 1987), as γβ-thalassemia developed in patients who had all functional globin genes, but distal upstream region was deleted. Since that time β-globin served as a paradigm for higher-order chromatin structure and nuclear organization during the course of sequential gene activation events (Ragoczy et al., 2006). By now several others LCRs are identified in other loci in human and mice, including immunoglobulin, growth hormone gene, apolipoprotein (Li et al., 2002). The property that distinguishes LCRs from other regulatory sequences is there ability to drive expression of a linked transgene from any position in the genome (West and Fraser, 2005). Analyzing β-globin locus, it has been shown that direct genomic contacts between activated genes and LCRs are created (Tolhuis et al., 2002). The model of a spatial unit of regulatory DNA elements was proposed termed as an active chromatin hub (ACH) (de Laat and Grosveld, 2003). The formation of ACH through interactions of *cis*-regulatory sequences and gene promoters ensures specific gene expression.

Tissue-specific genes are not always clustered but their transcription should be induced also. It has been shown that active genes co-localize into the sites of ongoing transcription sharing transcription factors and enzymes (Osborne et al., 2004). These subnuclear structures were called transcription factories.

It has been long argued how distal regulatory elements achieve their effect being located tens to hundreds kilobases away from their cognate genes. Three hypothetical models have been proposed: tracking, linking and looping (Li et al., 2006) (Fig. 7). Tracking model suggested that the transcription-activating complex recruited by an enhancer moves along chromatin until it recognizes a competent promoter. In case of the linking model, the binding of facilitator factors between an enhancer and its cognate gene defines the activated domain, so that activating effect spreads from an enhancer to a promoter. Linking is the only model where a direct contact between two *cis*-elements is not required. The enhancer and the gene can be brought together by loop formation. The initial contact between two sequences is made by facilitated diffusion.

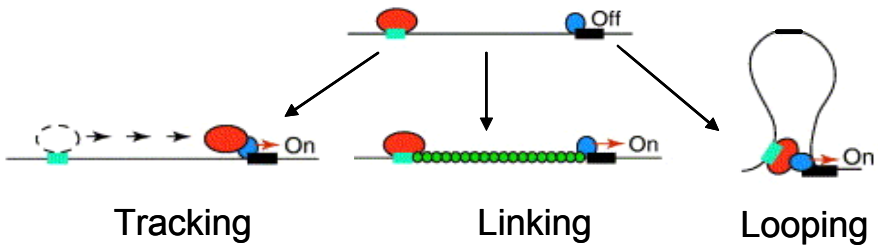


Figure 7. Regulatory sequences interplay models (Modified from Li et al., 2006). Three possible models of how an enhancer interacts with a promoter to initiate gene expression. Sky blue rectangles represent an enhancer, the activation complex recruited by the enhancer is shown in red. The genes are represented in black and the complex formed on the gene promoter is represented by the blue circles. The linking proteins are shown in green.

Although looping model is widely accepted now, there was no direct evidence that distant sequences physically interact *in vivo* until methods enabling to study contacts between chromosomal sequences were devised.

1.3.2 Methods for DNA interaction detection

Until recent time, the number of molecular biology techniques suitable for identifying DNA sequences that interact with each other was very limited. For long time fluorescence *in situ*

hybridization (FISH) was the only option to study DNA-DNA interactions. Although FISH method was successfully applied to analyze the frequency of juxtaposition of fluorescently labeled DNA sequences (Kosak et al., 2002; Sayegh et al., 2005), it has some serious limitations. It is time-consuming, low-throughput method, as each single cell needs to be analyzed and limited number of loci can be analyzed simultaneously, and it does not provide high resolution when *cis*-sequences are analyzed (Skok et al., 2007).

In 2002 two new methods appeared allowing to explore DNA interactions. One of them is RNA TRAP (tagging and recovery of associated proteins) developed by (Carter et al., 2002) consists of the deposition of a biotin tag in the immediate vicinity of chromatin interacting with the nascent primary transcript, marking *cis*-elements needed for gene activation. The other method that was developed was chromosome conformation capture (3C).

1.3.2.1 Chromosome conformation capture

3C enables high-resolution analysis of intra- and interchromosomal interactions at the cell-population level. Initially, it was established for yeast cells (Dekker et al., 2002), some modifications were introduced to adapt the protocol for mammalian cells (Tolhuis et al., 2002). In this procedure, cells are first treated with formaldehyde to crosslink proteins to DNA and themselves, then incubated in lysis buffer to disrupt cell membranes. Released nuclei are harvested and suspended in restriction buffer containing SDS, to eliminate non-crosslinked proteins of the chromatin, and Triton X-100, to sequester SDS, and the genomic DNA is digested with excessive amount of a selected restriction enzyme. The resulting restriction fragments, held together in close spatial proximity, are ligated at low DNA concentration to favor only intramolecular ligation reactions. After that the crosslinks are reversed and the ligation products are detected with primers located near the ends of the restriction fragments of interest (Fig. 8). The crosslinking frequency, measured by quantification of PCR signal, reflects the frequency with which two genomic fragments interact in the nucleus.

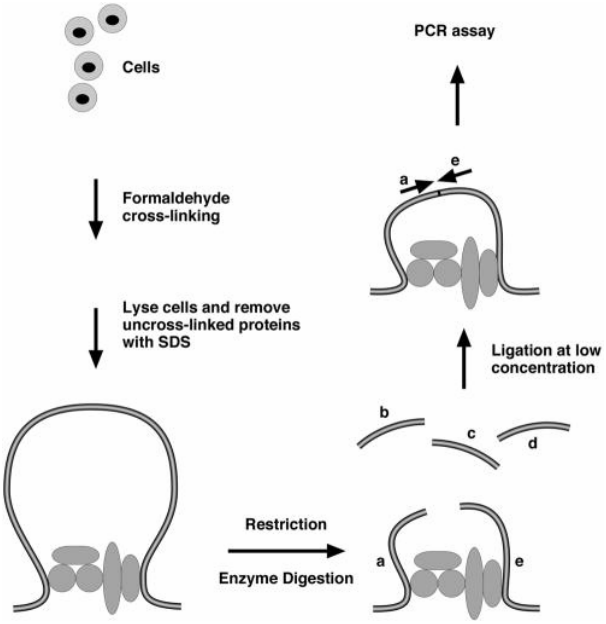


Figure 8. Chromosome conformation capture scheme (Adapted from Liu and Garrard 2005). The major steps of 3C protocol are shown. The fixation of cells 'freezes' the existing network of protein and DNA interactions. Restriction and successive ligation at low concentration promote bond synthesis only between closely located DNA pieces. The produced fragments are detected by PCR.

As a technique protocol of 3C is relatively straight-forward and does not demand sophisticated technical facilities, it has been applied by several research groups to examine possible interactions in many loci. The preferred sites being studied were β-globin cluster, *Igf2-H19* imprinted region, cytokine clusters of T helper cells and V-DJ rearrangement domains of lymphocytes (Table 1).

Table 1. Chromosomal interactions revealed by 3C (Modified from Fraser, 2006).

Gene loci	Type of interaction	Distance	Tissue/cells	Reference
Hbb	LCR-gene	50 kb	Fetal liver	(Tolhuis et al., 2002)
Igf2-H19	DMR-DMR	90 kb	Liver	(Murrell et al., 2004)
Dlx5-Dlx6	Silenced genes	60 kb	Brain	(Horike et al., 2005)
Igk	Enhancer-gene	22 kb	B cells	(Wei et al., 2005)

Il4-Il5-Il13	LCR-genes	150 kb	T_H2 cells	(Spilianakis and Flavell, 2004)
Ifng	Gene-distal region	24 kb	T_H1 cells	(Eivazova and Aune, 2004)
Tcrb, Tcra	V, D, J and C segments	665 kb	DN thymocytes	(Skok et al., 2007)
Ifng-T_H2 LCR	LCR-gene	interchromosomal	Naive $CD4^+$ cells	(Spilianakis et al., 2005)
Hbb-Eraf-Igf2	Transcription factory	25-40 Mb, interchromosomal	Fetal liver	(Osborne et al., 2004)

The main reason why those sites were chosen is that the crucial *cis*-acting elements for those genes were known before. The major limitation of 3C it is necessary to know the sequences of potential interacting genomic fragments for primer design. Thus, the 3C method allows determination of the interaction frequency between identified fragments in different cell types but it is impossible to find new *cis*-elements with 3C. This drawback was overcome by introducing 4C technique by four independent laboratories in 2006 (Ling et al., 2006; Lomvardas et al., 2006; Simonis et al., 2006; Zhao et al., 2006). The key modification was the generation of circular molecule that allows the amplification of unknown sequence with inverse PCR primers (Fig. 9). The pool of fragments created by PCR can be analyzed by sequencing or microarray analysis.

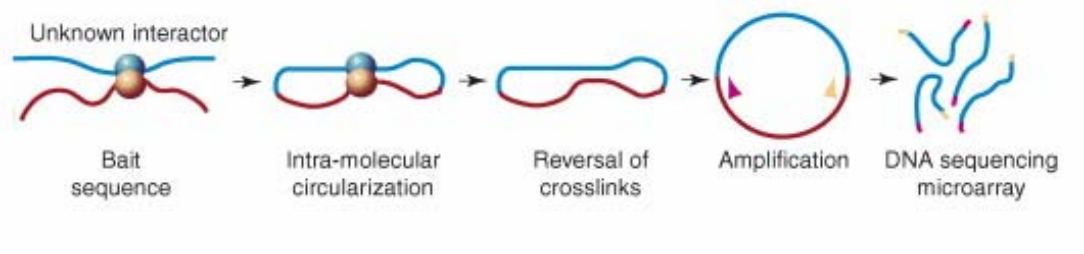


Figure 9. Chromosomal interactions disclosed by 4C (Modified from Ohlsson et al.). Only one regulatory sequence, termed bait sequence, is known. The circular molecule forms under appropriate ligation conditions. The reverse primers bind at the ends of bait sequence amplifying unknown fragment. Unexplored interactor is identified by sequencing.

The research results soon revealed that although the interactions within the same chromosome are more frequent, numerous contacts with other chromosomes also exist (Simonis et al., 2006). Moreover, genes with similar transcription status tend to co-localize. This fact supports transcription factory formation hypothesis.

Table 2. Chromosomal interactions disclosed by 4C (Modified from Lanctot et al., 2007)

Starting locus	Captured interacting locus	Intra- or interchromosomal	Cell type	Reference
Rad23a (active)	Multiple active gene regions	Intra > inter	Fetal liver and brain	(Simonis et al., 2006)
Hbb (active)	Multiple active gene regions including those containing *Eraf*, *Uros* and *Kcnq1*	Intra > inter	Fetal liver	
Hbb (inactive)	Silent gene clusters: for example, olfactory-receptor genes	Intra	Fetal brain	
H19 imprinting control region (maternal allele)	*Osbpl1a, Impact*	Inter	Neonatal liver	(Zhao et al., 2006)
H19 imprinting control region	*Wsb1, Nf1*	Inter	Fibroblasts	(Ling et al., 2006)
HoxB1 promoter	Large diversity of genomic loci	Intra > inter	Embryonic stem cells	(Wurtele and Chartrand, 2006)
H enhancer	Expressed olfactory-receptor allele	Intra > inter	Olfactory sensory neurons	(Lomvardas et al., 2006)

Two additional methods, termed associated chromosome trap (ACT) (Ling et al., 2006) and 5C (Chromosome Conformation Capture Carbon Copy) (Dostie and Dekker, 2007; Dostie et al., 2006), were also developed to study unknown interactions from known baits. ACT method failed to detect numerous intra-chromosomal interactions that both 3C (Kurukuti et al., 2006) and 4C (Zhao et al., 2006) analyses were able to find.

Despite the power of 3C based methods, it is useful to keep in mind their possible disadvantages. To date, 3C sensitivity remains low: at least one million cells are required to get reproducible results. As a result, an accumulated picture is derived, interactions detected by the 3C technique may not occur in the same cells. Brief contacts can be missed, as well as the method does not reflect the actual situation in every single cell. The purity of cell population analyzed by 3C should be high. FISH method is of choice to determine the exact state of the inter- or intrachromosomal interactions within single cell.

The use of restriction enzyme digestions in 3C method sets the limits for resolution and coverage. The digestion with two restrictases with short recognizing sites helps to avoid the bias. 4C resolution also depends on the length of the fragment amplified and, consequently, the effectiveness of reversed PCR. The 3C method requires the generation of crosslinks, which is done by formaldehyde treatment on living cells. It is efficient in creating covalent bonds between proteins and DNA-protein complexes, however, the adjacent amino-groups must be in very close juxtaposition. The use of other fixative reagents could be considered.

2. AIMS OF THE STUDY

1) To determine the expression pattern of EDC genes, a potential gene cluster regulated by AIRE, in thymus, skin and thymic epithelial cell line 1C6

2) To examine whether AIRE triggers conformational changes during gene regulation

3) To apply 3C technique to analyze potential interactions between members of EDC and regulatory sequences in C6 cell line in the presence or absence of AIRE

3. MATERIALS AND METHODS

3.1 MICE

Aire $^{-/-}$ mice on C57BL/6J background were generated at the Walter and Elisa Hall Institute (Melbourne, Australia) and maintained at the mouse facility of the Institute of Molecular and Cell Biology, Tartu University. The targeting construct containing LacZ gene was inserted into exon 8 of *Aire* gene. For genotyping, the genomic DNA was extracted from mouse tails using JetQuick Tissue DNA Spin Kit (Genomed) and analyzed by PCR using following primers: 5'-cagaagaacgaggat-3' and 5'-ctgtcttctgtgaaggcttctagg-3' that bind to exon 8 and detect wild-type allele. Primers 5'-tcgccattcaggctg-3' and 5'-cagactgccttggga-3' are selection cassette specific and in combination with primers from exon 8 amplify knockout allele (Fig.). Thymuses from 4- to 7-week old wild-type (WT) or knockout (KO) mice were used in expression and 3C experiments. All experiments were performed in accordance with EU guidelines (directive 86/609/EEC) on the ethical use of animals using the experimental protocol approved by Estonian Ministry of Agriculture.

3.2 CELL CULTURE

Mouse thymic epithelial 1C6 cell line (Mizuochi et al., 1992) was kindly provided by G. Holländer (University of Basel, Switzerland). Cells were maintained in Dulbecco's modified Eagle's medium (PAA Laboratories) supplemented with 10% fetal calf serum (PAA Laboratories) and 100 U/ml penicillin, 100 μg/ml streptomycin and 0.25 μg/ml amphotericin B (Gibco BRL). Cell line was grown at 37°C in 5% CO_2.

3.3 THYMIC STROMAL CELL ISOLATION

Eight to twelve thymuses were pooled together and small cuts were made into the capsules to release thymocytes by repetitive pipeting. The remaining thymic fragments were treated with 0.5 mg/ml collagenase/dispase (Roche) and 1.5 μg/ml DNase I (Applichem) solution in PBS at 37°C for 20 min. The treatment was repeated three times, each time supernatant with released cells was collected and fresh enzyme solution was added. The cells were resuspended in 5 mM EDTA in RPMI-1460. All supernatants were collected and cells were resuspended in Percoll (Fluka) solution with 1.08 density. TECs were concentrated with discontinuous Percoll density gradient (ρ 1.07; 1.045; 1.03; 1.0) at 350g for 30 min at 4ºC. The three upper interfaces enriched with

light-density cells were collected. The thymocytes were removed by depletion with CD45 MicroBeads (Miltenyi Biotec) according to manifacture's instructions. The cell composition was analyzed by staining with anti-I-Ab PE (AF6-120.1, BD Biosciences) and G8.8 FITC using FACSCalibur Flow Cytometer (BD Biosciences). Around 400 000 mTECs were obtained from 10 thymuses with purity of 40%.

3.4 CELL SORTING

The cells released after collagenase/dispase treatment were counted and $2x10^8$ cells were used for further cell sorting. To eliminate the rest of thymocytes CD45 depletion with CD45 MicroBeads was performed. For cortical epithelial cell (cTEC) isolation, the CD45- cells were stained with H213-HB antibody (anti-CDR1) followed by incubation with goat anti-rat IgG MicroBeads Miltenyi Biotec) and AutoMACS separation (isolation mode: Possel-S). The positive fraction contained magnetically bound cTECs and the negative fraction was used for mTECs isolation. The procedure was the same as with cTECs, but G8.8 antibody (anti-EpCAM, generated from a G8.8 hybridoma line) was used. The purity of sorted cells was >80% as assesed by staining with anti-CD45 FITC and anti-I-Ab PE (AF6-120.1, BD Biosciences). Between 250 000 and 500 000 mTECs were derived from 10 thymuses.

3.5 EPIDERMIS ISOLATION

Epidermal layer was derived from mouse ears. First of all, skin layer was removed from cartilage part of the ear and incubated with 0.5 mg/ml collagenase/dispase in PBS at 4ºC overnight. Then, the thin epidermal layer was detached from dermis and subjected to RNA isolation.

3.6 QUANTITATIVE REAL-TIME PCR

RNA was isolated with TRIzol (Invitrogen, Life Technologies) and reverse-transcribed to cDNA with SuperScriptTM III Reverse Transcriptase (Invitrogen, Life Technologies) and oligo$(dT)_{18}$ primer. Real-time PCR was performed with the ABI Prism 7900HT Fast Real-Time PCR System instrument (Applied Biosystems) using qPCR SYBR Green Core Kit (Eurogentec) according to the manufacture's instructions except that 2mM $MgCl_2$ concentration was used. The amplification program included an initial denaturation step at 95 °C for 10 min, followed by denaturation at 95 °C for 15 s, and annealing and extension at 60 °C for 1 min, for 45 cycles.

At least one primer from the pair was designed to span exon-exon junction to prevent unspecific amplification from genomic DNA. Because Sprr genes have only two exons and own high

homology, we did not succeed to design primers for Sprr2f that would give unique product in qPCR. To ensure that obtained qPCR product is not contaminated with primer dimers or unspecific products both the melting curve analysis in the end of the reaction and gel analysis were performed. The relative gene expression levels were calculated using the comparative C_t ($\Delta\Delta C_t$) method (according to Applied Biosystems), where the relative expression is calculated as $2^{-\Delta\Delta Ct}$, and where C_t represents the threshold cycle. Every sample was run in three parallel reactions. The primers used are given in the table 3.

Table 3. Primers for quantitative PCR analysis.

Gene name	Forward	Reverse
K2-8	AGGAGCTCATTCCGTAGCTG	TCTGGGATGCAGAACATGAG
β2microglobulin	TGAGACTGATACATACGCCTGCA	GATGCTTGATCACATGTCTCGATC
Aire	CCCCGCCGGGGACCAATCTC	AGTCGTCCCCTACCTTGGCAAGC
Sprr1a	CTCTGAGTATTAGGACCAAGTGC	CAGGGATCCTTGGTTTTGG
Sprr2a	CGGAGAACCTGATTCTGAGAC	ACATGACAGGACGACAAGG
Sprr2d	TTTGGAGAACCCGATCCTGAG	GACACTTTAGAGGAGGACAAGG
Sprr2i	TTCGGAGAAGTAGATCCTGAG	GAGGAGGACAAGGTTCAGG
S100a8	CTTCAAGACATCGTTTGAAAGG	TCATTCTTGTAGAGGGCATGG
S100a9	CCATCAATACTCTAGGAAGGAAGG	CTTCTCTTTCTTCATAAAGGTTGC
S100a10	ACTAGCCTCATCGTGGT	GGTCCTCCTTTGTCAAGTG

3.7 ADENOVIRUS INFECTION

The construction of Aire adenovirus is described in Kont et al, 2007. The infection was conducted on 100mm 70% confluent dishes with 150 MOI in 2 ml serum-free DMEM for one hour with gentle mixing after every 15 min. After that, the medium containing virus particles was replaced with 10 ml fresh DMEM. The cells were analyzed after 24 hours of incubation. To get the microscope picture of the infected cells, a coverslip was placed on the bottom of a dish before seeding the cells. After the virus infection, the coverslip was removed from the bottom and placed upside down on a slide with DAPI (Roche) and Fluorescent Mounting Medium (DakoCytomation) mixture dripped before. The images were acquired by fluorescence microscopy (Eclipse TE2000-4; Nikon, Melville, NY).

3.8 CHROMOSOME CONFORMATION CAPTURE

3.8.1 Primer design

The sequences of the necessary regions were downloaded from GenBank (NCBI) and restriction sites were found with Clone Manager Software (SciEdCentral). The primers were designed with Primer3 program to recognize the 5'- and 3'-ends of *Bgl II* restriction fragments. Primers' length varied from 17 to 27 nucleotides with melting temperatures around 59°C. The final PCR product length was 100-150 nt. The genomic uniqueness of all primers was verified with BLAST. The primers were synthesized by TAG Copenhagen A/S.

3.8.2 Probe preparation

In case of EDC cluster, 1C6 cell with and without adenoAire treatment were used. 1C6 cells (5 mln) were trypsinized and fixed with 1% formaldehyde solution in DMEM for 10 min at room temperature. The reaction was quenched with 0.125 M glycine. The cells were spinned down and resuspended in ice-cold lysis buffer (10mM Tris HCl, 10 mM NaCl, 0.2% NP-40) with freshly added protease inhibitors (100mM AEBSF, 80 μM aprotinin, 2mM leupeptin) and lysed on ice for 10 min. The cell membranes were disrupted with Dounce homogenizer (Bellco, pestle B). The cell lysis efficiency was checked by staining the cells with trypane blue. After the recovery of crosslinked nuclei by centrifugation at 2500g for 5 min, the pellet was washed twice with 1x B3 restriction buffer (Jena Bioscience GmbH) and resuspended in the same buffer with 0.1% SDS added. The solution was incubated at 65ºC for 10 min and Triton X-100 to the final concentration of 1% was added and thoroughly mixed to sequester SDS. The mixture was incubated overnight with 400 units of *Bgl II* (Jena Bioscience GmbH) at 37ºC with shaking. The restriction enzyme was inactivated at 65ºC for 20 min with 1.6% SDS. The digestion mixture was highly diluted (17 times) with 1x T4 ligase buffer (Fermentas), including 1% Triton X-100, and 500 Weiss units of T4 DNA ligase (Fermentas) was added. The incubation at 16ºC for 5 hours proceeded. To de-crosslink DNA 500 μg of proteinase K was added and incubated at 65ºC overnight. One additional proteinase K treatment at 42ºC for 2 hours was performed. DNA was purified by multiple phenol/chloroform extraction and ethanol precipitation. The quantity of DNA was analyzed by gel visualization on 0.7% agarose 0.5x TBE, the marker FastRulerDNALadder, High Range (Fermentas) was used. For β-globin locus analysis, E14.5 livers and brains were analyzed. Two livers and four brains were homogenized and fixed with 2% formaldehyde in PBS. Restriction enzyme and ligation reactions were performed as described with 1C6 cells.

3.8.3 PCR conditions and primers for EDC cluster analysis

The DNA samples were amplified by nested PCR. The first round PCR conditions were: 15 min heat start at 96ºC and 35 cycles of denaturation at 95ºC for 20 s, annealing at 57ºC for 40 s and extension at 72ºC for 1 min. For the second round of PCR, the first round PCR product was diluted 10-fold and the same cycling protocol was used except that extension step was shortened to 40 s and cycle number was 40. For both rounds, HotStart Taq (Solis Biodyne) was used.

Table 4. Nested PCR primers for Sprr gene family

Name	External	Internal
324 HS1 S	GAGGACCTCTGGGGTTGAAT	AAGCTGGAAGAAACTGAAGAGT
Sprr1a 10329 S	GCCTGGCCTGTCTTGTGGTGG	CAAGATGAATCTGTGAAGTAACAGTAA
Sprr2a 7018 AS	GGGAATGCTCTTGTGTCAGCAGTGT	TTTCTATCATAACCTGAATTTGACC
Sprr2f 5415 AS	ACCCAGTACCATTTGGCCTACAAGA	TGTTCCTGGATATTATCATTGTTGA
Sprr2i 15191 S	CCTGCTGCTGTAAAGGGCCTGC	TCTTGCATGGTTCAGTAGACC

The PCR products were analyzed on 2% agarose 0.5x TBE gel with ethidium bromide staining. To estimate the siza of the products O'GeneRuler™DNALadder, LowRange (Fermentas) was applied. The positive bands were cut out and DNA was extracted from the gel with UltraClean 15 DNA Purification Kit (MoBio Laboratories, Inc.).

3.8.4 PCR conditions and primers for β-globin locus

For β-globin analysis the DNA was amplified by conditional PCR with one round, the conditions were identical with the first round of nested PCR. Primers highlighted in bold were used in 3C assay and the combination of bold and corresponding regular primer was used for control preparation (Table 5).

Table 5. β-globin locus primers

Name	Forward	Reverse
5'HS2	**GCCCAGCTAACCTCCTCTG**	TTTTTCCAGCTTCCATCTATGA
Hbb-bh1	**TCATCTTTAACTTAGCTCTTTCCCTAA**	TTGATTCTTAATAACATTGACTTTTTG
Hbb-b1	**TCTTAGGGACACTTGCCAAA**	CACACATACCCATAAAGAGCCATA
Hbb-b2	GCTGGATACTATGGGTTTTGC	**TGCATTAAACATTGCTGATAACCT**
3'HS1	**AACAAACACCATGGATATTCAGA**	TGATTTTCAGGTGTGTTCGTG

In the *Gapd* locus, we used two *Bgl*II restriction fragments (A and B) that are near to each other (559 bp apart) in chromosome conformation capture analyses with primers 2, 3, 4 and 5. In quantative PCR experiment the mixture of primer 3 and 5 was used.

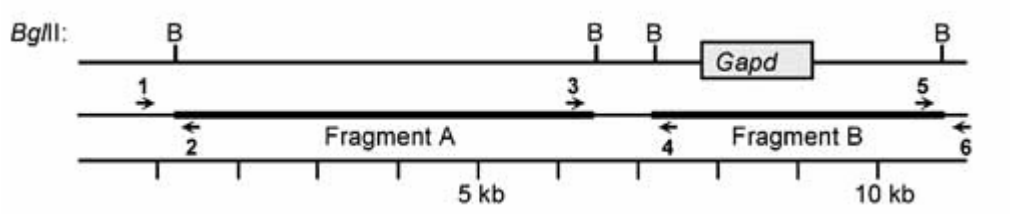


Figure 9. Schematic representation of *Gapd* gene locus, restriction fragments and control primers.

Table 6. Gapd primers

Name	Forward	Reverse
Gapd 1/2	CTTCATCTGCCTCCCTAAG	**ACACAGGCAAAATACCAATG**
Gapd 3/4	**CTGCGCCTCAGAATCCTG**	**GAATGCTTGGATGTACAACC**
Gapd 5/6	**CAAAACTCCTGGGTGCAAG**	CAGACAGGAATGCGTACAG

Control template was created to avoid bias because of primer efficiencies. For that, 5' and 3' ends of each restriction fragment were amplified in such a way that *Bgl II* restriction site position was in the middle of the fragment. PCR products were purified and the concentration was carefully measured with NanoDrop spectrophotometer. Obtained fragments were mixed in equimolar proportions and restriction digestions with successive ligation reactions were performed. After the enzyme inactivation, the DNA fragments were ethanol precipitated and 10 ng was added as a template to PCR reactions. The principle of control template generation is shown in Fig. 10.

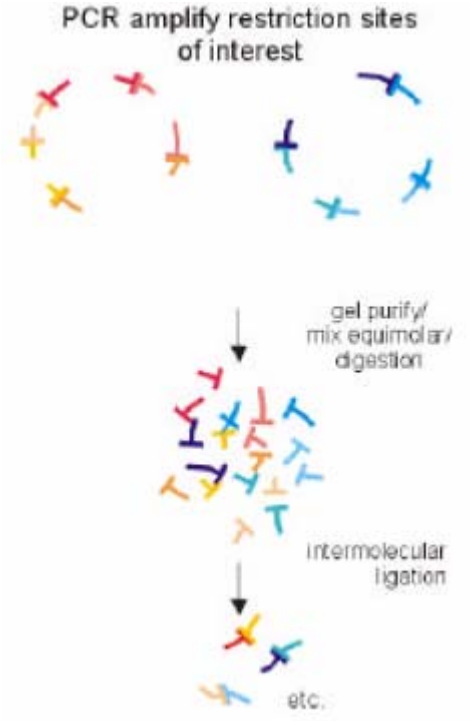


Figure 10. Control template preparation with PCR products (Adapted from Splinter et al., 2004).

3.9 SEQUENCING

To verify that PCR products obtained in conventional were created by intramolecular ligation, all PCR products were sequenced. Before sequencing, the products were re-amplified after gel extraction. The reactions were set with BigDye Terminator Cycle Sequencing Ready Reaction Kit according to manufacturer's instructions; each reaction contained 60 ng of DNA. Using ABI Prism 310 apparatus the sequence reads were obtained and analyzed with BioEdit software.

4. RESULTS

4.1 ESTABLISHING 3C TECHNIQUE

4.1.1 3C with β-globin locus

First, we decided to test if 3C protocol works in our hands. For this purpose we chose the β-globin locus, which was previously thouroughly analyzed by other research groups (Tolhuis et al., 2002). In previously published works, others than C57Bl/6J mouse strain were used. For this reason, we could not use the same primers as binding sites tended to differ between mouse strains. We designed primers for each end of five restriction fragments that corresponded to a gene or regulatory sequence. Since it was earlier shown that hypersensitive site 5'HS2 has the prominent activation effect, it was used as an 'anchor' fragment against which all other fragments were tested. For normalization, *Gapd* (see schematic in Materials and Methods section, Fig. 9), a gene constitutively expressed in all cell types without any known secondary chromatin conformation, was used. From 20 possible primer combinations, 5 reactions that gave the most intensive and background-free bands were selected for further analysis by quantative PCR. To avoid the bias in interaction frequency because of different primer efficiencies, control templates were generated. In control template, all possible ligation products are represented in equal amounts. The crosslinking efficiency of LCR and Hbb-b2 in control template was arbitrarily set as a value of 1.

The aim was to verify if we get the same interaction frequency pattern between the locus genes and LCR. Indeed, as seen in Fig. 11, the interaction between LCR and Hbb-b1/b2 was stronger in fetal liver cells. This correlates nicely with data that in early fetal liver definitive red cells already primarily express adult globin genes (Trimborn 1999). Overall, contacts within β-globin locus in liver cells were more frequent than in brain tissue (Fig. 11), where the locus is sustained in closed conformation, altogether showing that 3C method was working.

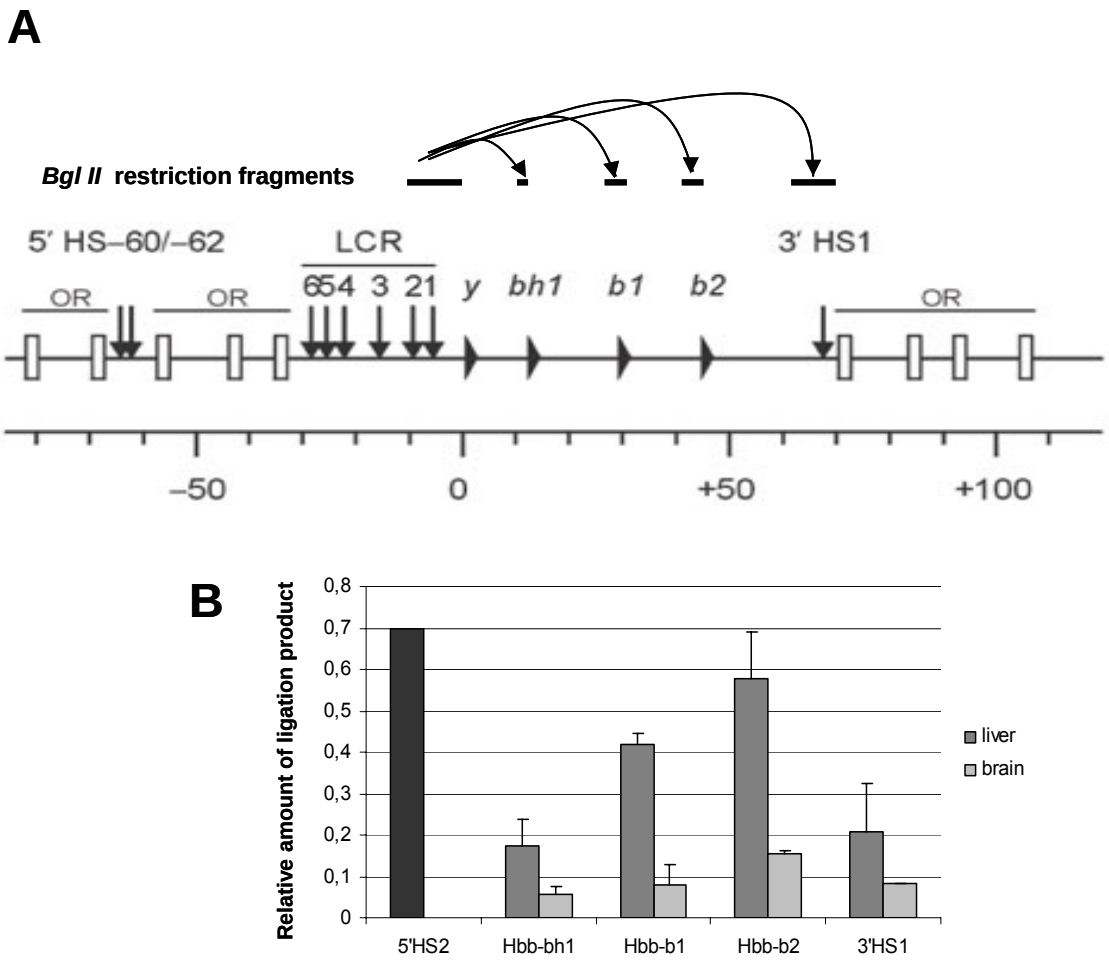


Figure 11. β-globin locus organization and locus control region interactions with gene promoters. (A) β-globin locus upstream (5'HS-60/-62 and LCR) and downstream (3'HS1) regulatory regions are indicated. Hemoglobin Y and Z (Hbb-y and Hbb-bh1) beta-like embryonic chains, and beta adult major (Hbb-b1) and minor (Hbb-b2) chains are depicted. β-globin locus is surrounded by olfactory receptors (OR) genes. The distance is given in kb starting from Hbb-y start site. The position and size of *Bgl II* restriction fragments is shown above. (B) Quantative RT-PCR analysis of ligation products using LCR 5'HS2 as an anchor fragment is shown. The relative amount of ligation products in liver and brain tissue are shown as indicated left, the value 1 corresponds to 5'HS2-Hbb-b1 ligation product quantity in control template. The reaction values are normalized against the amount of *Gapd* gene in each sample.

4.1.1 3C protocol optimization for 1C6 cell line

For our research purposes, we chose mouse thymic epithelial cell line 1C6. Although in earlier published protocols the widely used formaldehyde concentration was 2%, we first tested if this concentration gives the best results with 1C6 cell line. We fixed the cells with three different formadehyde solutions: 0,5%, 1% and 2%. We examined the yield and the quality of DNA obtained starting from the same cell number (Fig. 12). While the DNA quantity obtained by 0,5% and 1% fixing was comparable to each other, the yield with 2% formaldehyde was substancially

lower (Fig. 12A). We also observed that DNA purity was very low despite the multiple phenol extractions, that is why the gel visualization is preferrable to spectrophotomer measurement. We performed PCR with *Gapd* primers to see if there are any differencies in PCR efficiency due to the fixation protocol (Fig. 12B). The same template DNA amount was added to the reactions. The fixation with 0,5% was the least efficient since no products were synthesized with some *Gapd* primer combinations. Thus, we concluded that 1% formaldehyde is the most optimal in our conditions.

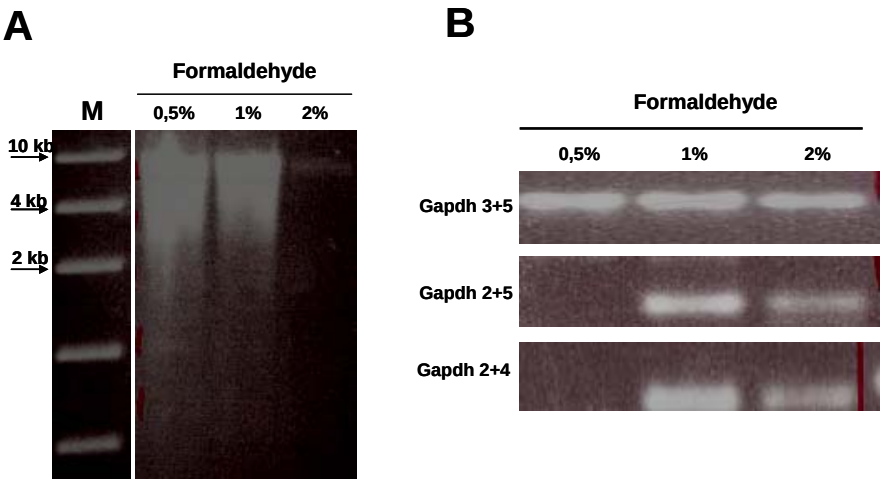


Figure 12. Optimizing formaldehyde concentration in fixative solution. Agarose gels showing the quantity of 3C DNA (A) and PCR products (B) amplified with *Gapd* primers combinations. DNA marker (A) and *Gapd* primer combinations (B) are shown left, used formaldehyde concentrations are indicated above. The schematic of *Gapd* locus restriction fragments and positions of the praimers are shown in Materials and Methods section (Fig. 9).

To check if the restriction and ligation reactions were appropriately completed with the fixated cells, we subjected DNA to agarose gel analysis after every step of protocol (Fig. 13).

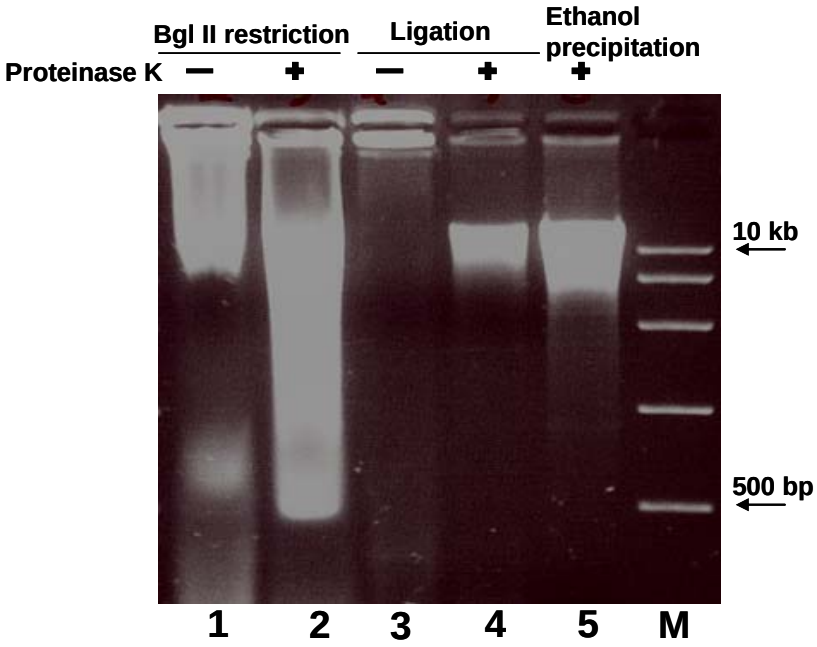


Figure 13. Restriction and ligation efficiency in 3C protocol.
Because of an extensive crosslink net, the DNA cannot migrate further from the well if proteinase treatment is not performed (lane 1 and 3). After restriction, DNA is fractionated with fragments lengths variyng from 10 kb to 500 nt (lane 2). The average *Bgl II* restriction fragment size among EDC cluster is 5.5 kb. After ligation, DNA size increases considerably, indicating a good performance of the enzyme (lane 4). DNA is effectively recovered after 3C procedure (lane 5).

4.2 GENE EXPRESSION ANALYSES OF EDC GENES IN THYMUS EPIDERMIS AND IN 1C6 CELL LINE

Previously, several EDC cluster genes have been sown to be downregulated in Aire knockout mouse [Derbinski et al 2005 GEO database (available under accession no. GSE2585) at http://www.ncbi.nlm.nih.gov/projects/geo/, Hamish Scott (unpublished data)]. To study whether Aire can cause conformational changes in EDC gene cluster, we first wanted to confirm that Aire indeed is able to regulate EDC genes. Thus, we first wanted to confirm the Aire knockout data, and to test whether the adenoviral Aire infection changes expression pattern of EDC genes in thymic epithelial cell line 1C6. Next, to illustrate the expression data, we also compared thymic expression of EDC genes with expression in skin.

4.2.1 Thymic stromal cell isolation

To study the expression of EDC genes in the thymus, we first isolated thymic medullary epithelial cells from other cells. The cells expressing AIRE are very scarce, we tried two methods

to obtain the highest content of mTECs by cell isolation. First, cell density fractionation with Percoll gradient was employed, which relies upon the fact that epithelial cell are lighter than thymocytes and erythrocytes. However, it does not enables to distinguish between cortical and medullary epithelial cells. Usually, around 40% of derived cells expressed mTEC specific markers (Fig. 14).

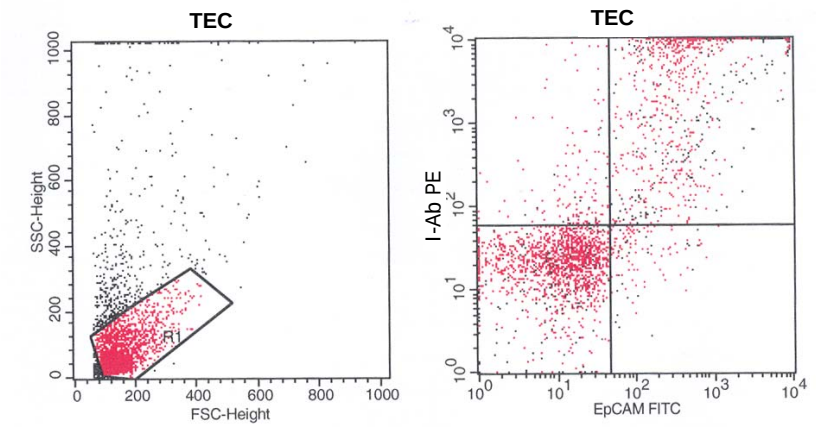


Figure 14. Cell population obtained with Percoll gradient density franctionation. Cell size and granularity is estimated by forward and side scatter. Medullary TECs with high MHC II (I-A^b) and EpCAM (epithelial cell adhesion molecule) expression are in the upper right corner.

As our laboratory facilities improved, we were able to sort cells with autoMACS (automated magnetic cell sorter), which implies using an antibody recognizing a specific marker on the cell surface. The latter method gave significantly better results, the mTEC content in final cell population reached to 80% (Fig. 15). Since cell sorting provides remarkably higher cell purity, we proceeded with this technique and in futher expression analysis experiments cell sorted with autoMACS were used.

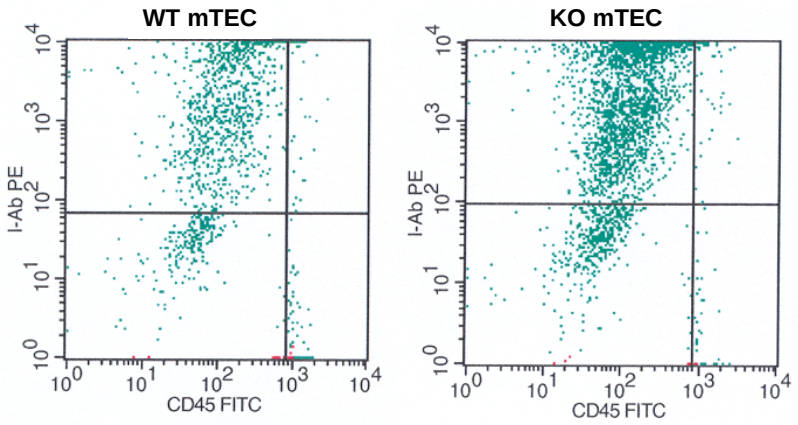


Figure 15. Cell composition after cell sorting with MACS. MTECs with high MHC II (I-A^b) and low CD45 (general leukocyte marker) expression are shown in the upper left corner.

To re-check that mice with suitable genotype were used in sorting protocol, the released thymocytes from the pooled 10 thymuses were re-genotyped. Figure 16 confirms that all mice used in analysis had expected genotypes.

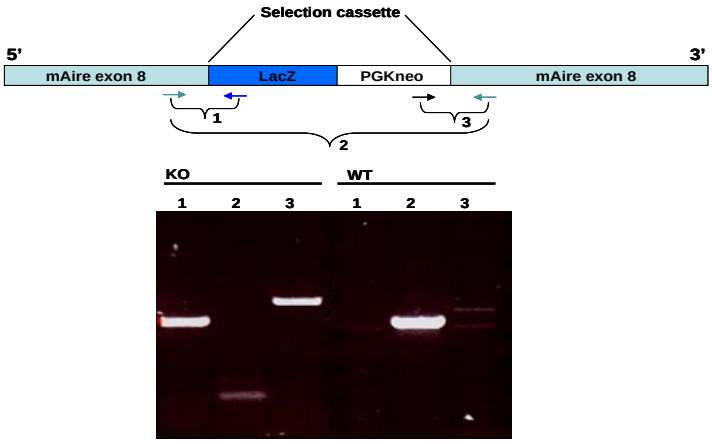
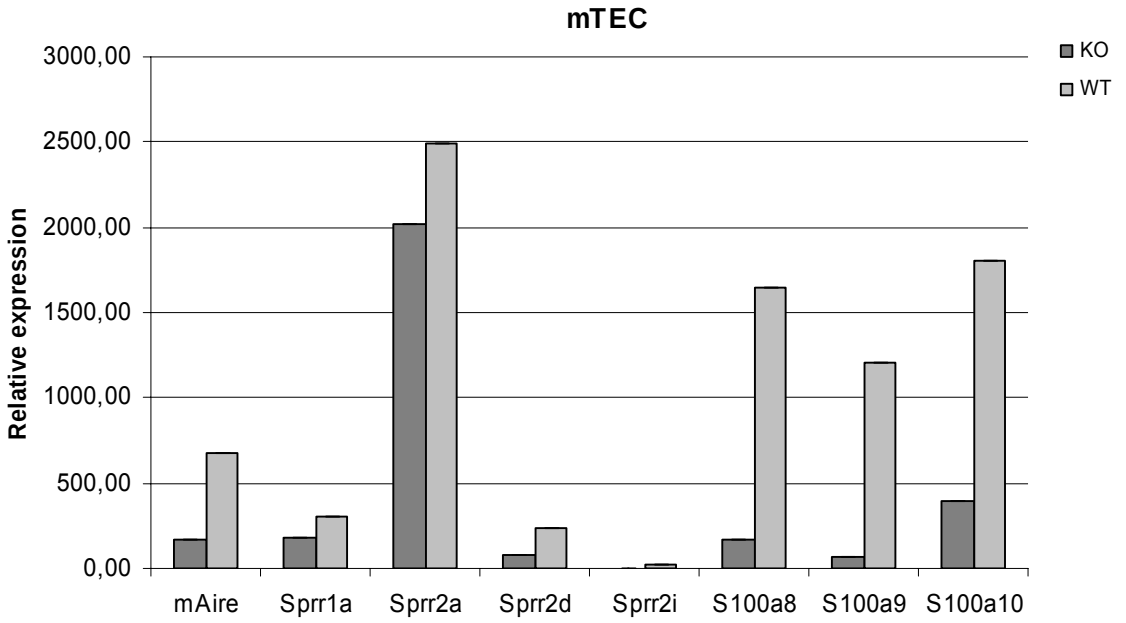


Figure 16. Genotyping of wild-type and Aire knockout mice. In Aire-deficient mice, exon 8 is disrupted by selection cassette. Primer pairs where one primer is insert-specific and the other one binds to exon 8 give product from KO allele. The combination of both exon 8-specific primers amplifies WT allele. Reactions 1 and 3 are KO specific, reaction 2 gives product only in case of WT, as insert is too long to be amplified in our PCR conditions.

4.2.2 EDC gene expression in mTECs

Next, we investigated how the expression of EDC genes changes in the absence of Aire gene and performed expression analysis to above described sorted mTECs derived from wild-type and Aire knockout mouse thymuses. Figure 17 shows that in the absence of Aire, a slight downregulation of expression occurs in Sprr gene family members except for sharp drop in Sprr2i expression. More pronounced changes appear in S100a gene family. The results are comparable with data obtained from array expression experiments. The selection cassette with LacZ gene inserted into exon 8 of Aire invokes frame shift mutation leading to the absence of functional protein, but it did not affect transcript production. Even though further exon 12 of Aire is amplified in qPCR (Table 3), remarkable quantity of mRNA is detected in KO mTECs. *Western blot* analysis (data not shown) confirms the lack of Aire protein.

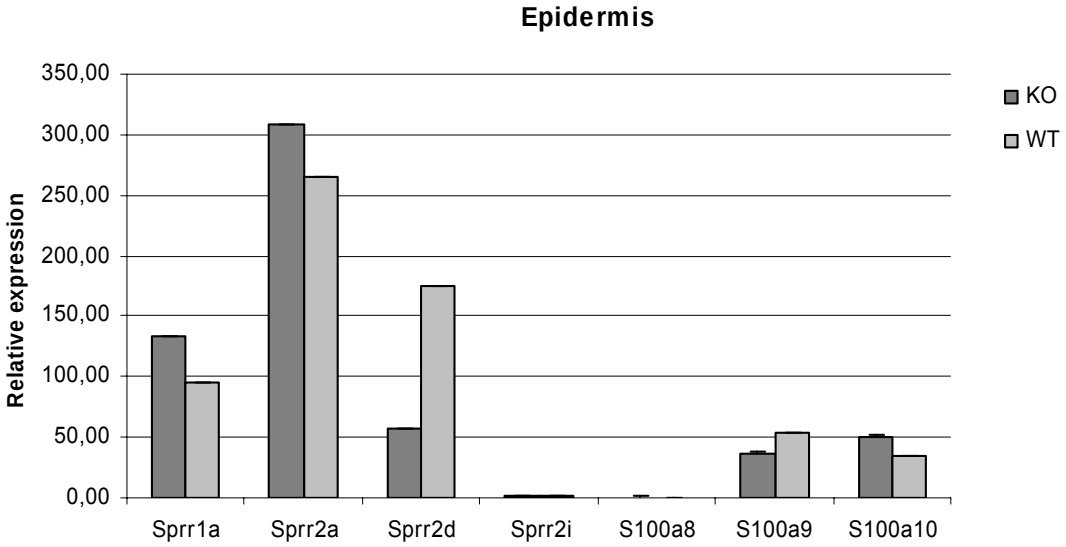


	mAire	Sprr1a	Sprr2a	Sprr2d	Sprr2i	S100a8	S100a9	S100a10
FC WT/KO	4,04	1,64	1,24	2,86	23,51	9,43	17,29	4,63

Figure 17. Relative expression of Aire and several EDC genes in wild type and Aire knockout mouse mTECs. Bars represent the relative amount of mRNA. Gene expression was normalized to keratin K2-8, which expression is the strongest in the medulla. Data are presented compared to Sprr2i expression in KO (=1). Fold changes are presented on table below the graph. Data are one out of three representative experiments.

4.2.3 EDC gene expression in epidermis

The APECED patients develop ectodermal dystrophy including dental enamel (Lukinmaa et al., 1996) and nail pitting defects in with unknown etiology. Recent research shows these dermatological manifestations could be secondary phenomena of recurrent *Candida* infections. Although there is none skin defects observed in mice, we decided to determine the expression level in epidermal tissue of KO and WT mice to see if there still can be some changes of autoimmune or other destructive mechanism that are not detectable morphologically. Figure 18 shows that the same genes in epidermal tissue did not show any clear trend in expression difference between Aire KO and WT mouse. While the expression of some of genes were upregulated in wild type epidermis (Sprr2d, Sprr2i, S100a9), the drop in expression was observed for the others (Sprr1a, Sprr2a, S100a8, S100a10), showing the existence of the compensatory mechanism. Concordantly, no detectable expression of Aire was observed in epidermis. One KO and one WT mouse was analysed for EDC gene expression in skin and the obtained result was predicted, no further experiments were considered necessary.



	Sprr1a	Sprr2a	Sprr2d	Sprr2i	S100a8	S100a9	S100a10
FC WT/KO	0,71	0,86	3,05	1,62	0,24	1,42	0,69

Figure 18. Quantative RT-PCR analysis of Aire and EDC genes mRNA extracted from epidermis. The data was normalized to $\beta_2$microglobulin, the value of 1 on the graph is assigned to Sprr2i expression in KO. One experiment in triplicate is shown.

4.2.4 Comparison of expression values in the thymus and the epidermis

The general concept is that tissue specific antigen expression is lower in the thymus compared to its original expression site in a peripheral organ. We compared EDC gene expression in the thymus and in the epidermis. Interestingly, we found that the difference in expression between thymic and epidermal tissue differs only several-fold. Surprisingly, S100a8 and S100a9 have even higher expression in the thymus. One possible explanation is a low level of these proteins in the healthy epithelium, whereas only inflammation reactions cause noticeable increase in their expression (Ryckman et al., 2003). The keratinocytes begin CE formation only when they arrive at granular layer (Kalinin et al., 2002). Because the whole epidermal tissue was used in the experiment and is not a homogenous cell population, it might to explain the difference in relation to thymic tissue, where cells were sorted for specific markers.

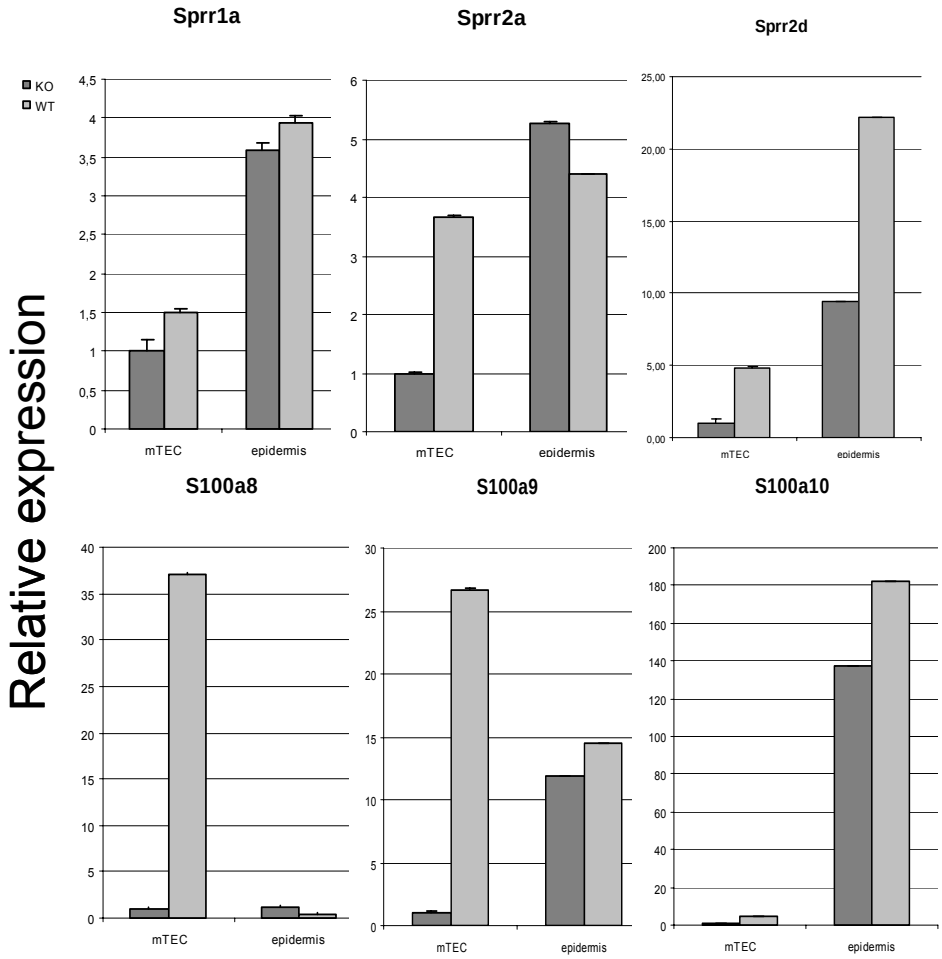


Figure 19. Comparison of the expression levels in the sorted mTECs and in the epidermis. The signals are normalized against the amount of $\beta_2$microglobulin in each sample; the relative expression levels are shown compared to KO thymus mTECs (= 1). Data are mean with S.E.M. of triplicate qPCR measurements of one representative experiment.

4.2.5 EDC gene expression in 1C6 cell line infected with adAireGFP

As the number of Aire expressing cells in the thymus is very limited, and the quantity of initial material needed for 3C is rather large, we tried to find *in vitro* system mimicking the changes in EDC gene expression in the presence of Aire. Thus, we next tested if adenoviral infection with Aire of mouse thymic epithelial 1C6 cell line produces the same changes. As the adenoviral construct in addition to Aire gene contains a GFP sequence, the cells in which the infection is performed can be easily visualized under fluorescent microscope (Fig. 20), and the efficiency of the infection can be estimated, which usually exceeded 80%.

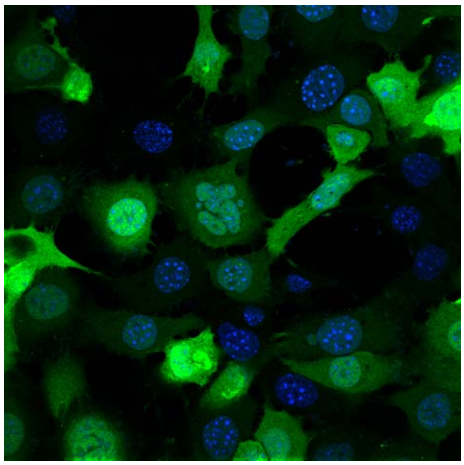
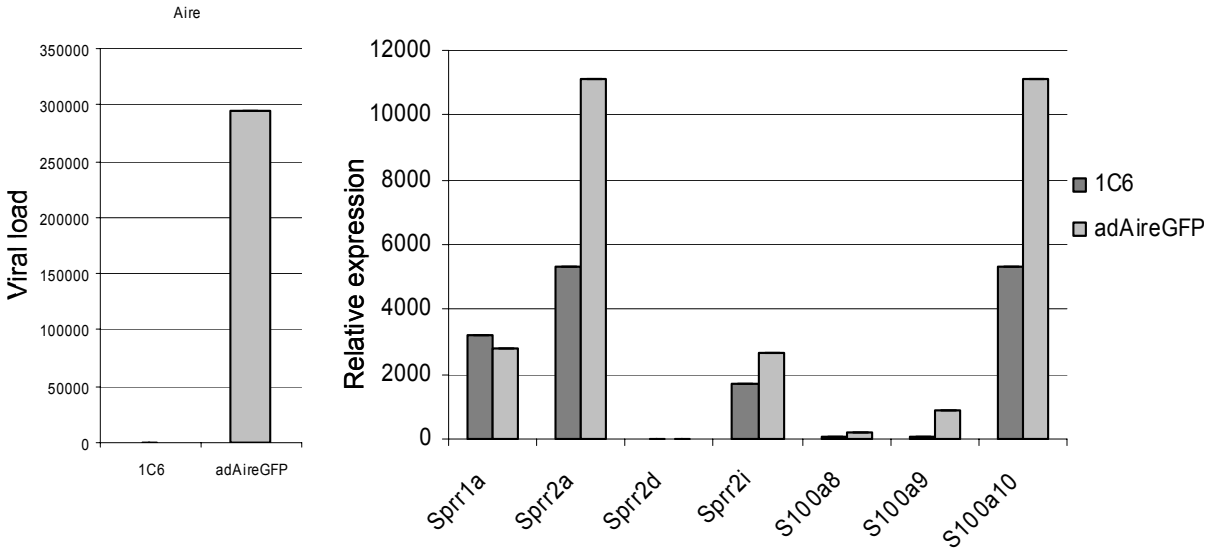


Figure 20. The adenoviral infection of 1C6 cell line. The GFP fluorescence indicates infected cells, the DAPI staining shows nuclei. Magnification x60.

Despite the 20 000-fold increase in Aire expression, Sprr genes expression rise was in the same range of WT versus KO mice. Although, the increase of S100a gene expression was slightly stronger, it did not reach the level of *in vivo* changes being 2-3 times lower.



	Aire	Sprr1a	Sprr2a	Sprr2d	Sprr2i	S100a8	S100a9	S100a10
FC adAireGFP/1C6	20658,16	0,88	2,07	4,83	1,56	3,00	9,06	2,07

Figure 20. The expression analysis of EDC gene in 1C6 cell line infected with adenovirus containing Aire gene. The signals are normalized against the amount of K2-8 in each sample; the relative expression levels are shown compared to Sprr2d expression level in control-infected 1C6 cells (= 1). Data are one out of two representative experiments.

4.3 EDC CLUSTER INTERACTIONS

Firstly, on the basis of microarray data (Johnnidis et al., 2005) 12 genes from EDC cluster with noticeable expression change and additional hypersensitive site were chosen for 3C analysis. The studied genes were Sprr1a, 2a, 2i, 2f, 2d, S100a8, a9, a10, Lor, Ivl, Selenbp1, 2300002Rik and Sprr family HS, that was identified by the comparison of *Sppr* gene family sequence in human and mouse genome (Martin et al., 2004). From three identified HS, one showed the activating properties. For primer design, *Bgl II* restriction fragments, that encompass gene promoter regions were chosen, however, as EDC cluster genes are short and *Bgl II* produces several kilobase-long fragments, usually the whole gene fitted into the restriction fragment. The primers were designed to 5' and 3' end of *Bgl II* restriction fragment, in some cases the primer design was not possible because of high content of repetitive sequences, and then only one end of restriction fragment was analyzed. Altogether 202 primer combinations were possible. 202 reactions were performed with infected and non-infected 1C6 cell line. Some primer combinations gave positive bands that were sequenced and proved to originate from ligation product, however, these bands were not reproducible. The reason for the poor reproducibility could be very low template concentration in reaction mix. This forced us to use nested PCR. Nested PCR consists of two rounds of PCR, the first one with external primers, and the second with the primers that bind within the first PCR product, these two successive runs ensure the increased specificity and sensitivity. Since it doubles the number of required reactions, we decided to narrow our analysis down to Sprr and S100 gene families.

Figure 21 demonstrates, that multiple contacts of Sprr genes with HS site aroused, when Aire expression was boosted in 1C6 cell line. Similarly, several Sprr genes interacted with each other. Although the majority of contacts were created in the presence of Aire, Sprr2a interaction with HS and Sprr1a was present even in uninfected cells. This phenomenon can be explained by high expression of Sprr2a, which was the highest among Sprr genes as indicated by our expression analysis (Fig. 18, 17, 20).

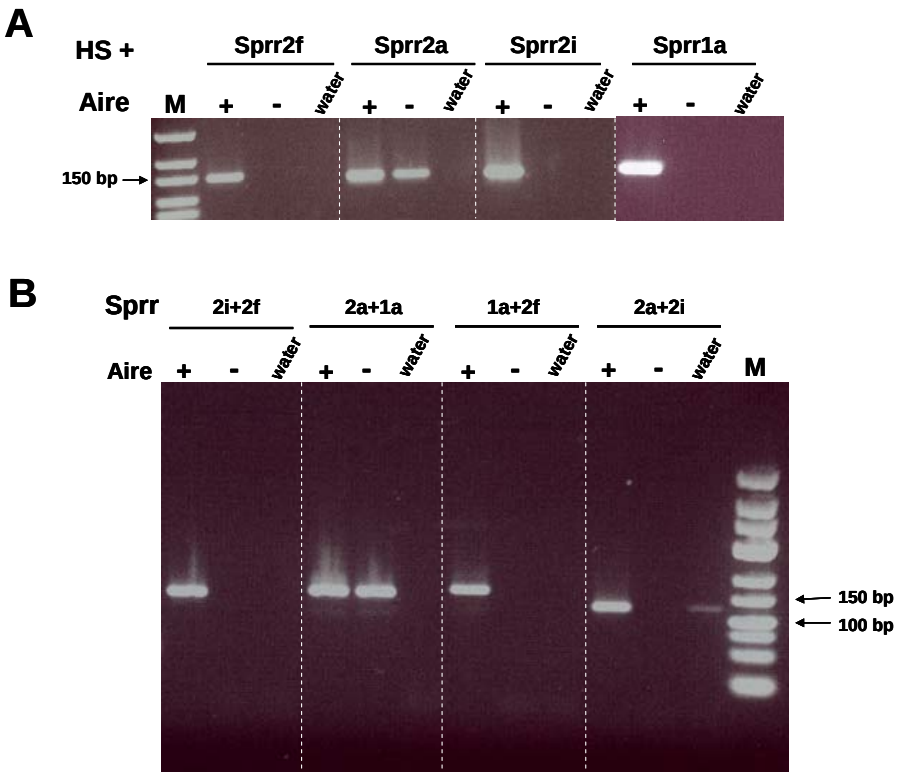


Figure 21. The interactions detected by 3C in 1C6 line in the presence or absence of Aire. Agarose gels pictures presenting ligation products amplified by specific Sprr gene and hypersensitive site primers (A). Panel B represents the interactions between different Sprr genes as indicated. The specificity of all nested PCR products obtained was confirmed by sequence analysis. Positive bands from two different experiments are presented.

As mentioned above, we also analyzed the presence of interactions between S100a8, S100a9 and S100a10 gene promoters and Sprr family HS site. However, no interactions were detected neither in uninfected nor in adAireGFP infected 1C6 cell line which shows that either we were not able to detect interactions or no coordinative regulation of S100 family genes occurs in 1C6 cell line, that could be the case as expression increase *in vivo* was higher compared to infected cell line. It is possible that Aire and other factors' co-influence is required for S100a family interactions. Positions of S100a genes that we analyzed are shown at Fig. 22. While Sprr family lies between loricrin and involucrin genes and covers approximately 490 kb, S100a family members are located more sparcely with longer distances between them. The longest contact we detected in Sprr family was 256 kb long between Sprr2a and Sprr1a genes. The lack of known regulatory site specific for S100a family and the scattered distribution of its members could prevent interaction detection between them.

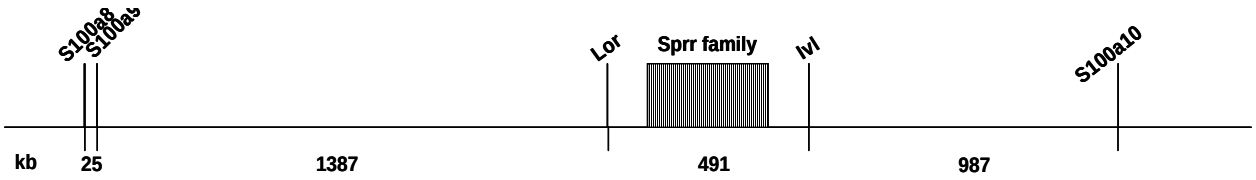


Figure 22. Schematic representation of position of S100a genes regulated by Aire. S100a8, S1000a9 and S100a10 show downregulation in Aire-deficient mice. S100a8 and S100a9 are closely positioned upstream from Sprr family, S100a10 locates downstream from Sprr family. The distance in kb between genes is marked below.

To illustrate the 3C results about Sprr family genes, the converged scheme of the detected interactions is shown in Figure 23.

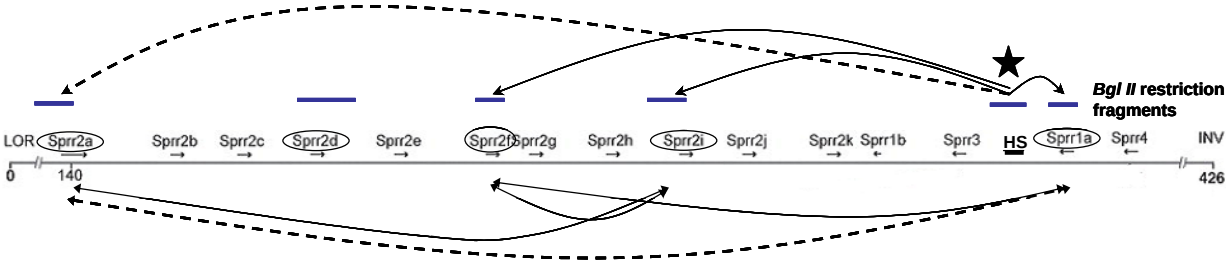


Figure 23. The interactions detected across Sprr gene family. Above, *Bgl II* restriction fragments and contacts with hypersensitive site are depicted, below intergenic interactions are shown. Solid lines represent interactions induced by Aire; dashed lines show interactions also existing without Aire.

From the 3C information, the hypothetical model of loop formation can be deduced (Fig. 24). In the absence of Aire, HS contacts predominantly with Sprr2a inducing its high protein production. When Aire is introduced into the cells, the expression of Sprr family genes is higher and number of interactions between Sppr family genes and HS increases.

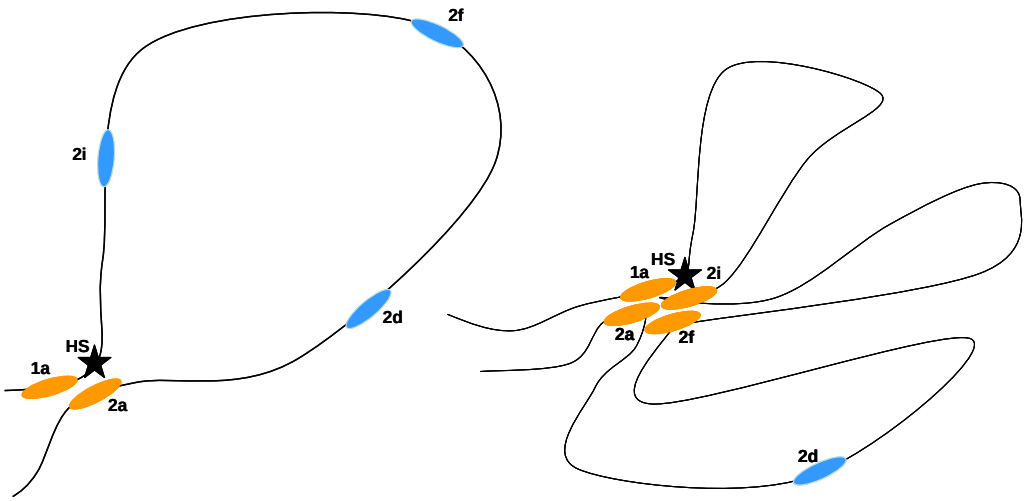


Figure 24. Loop formation model between Sprr gene family members induced by Aire. The left schematic picture describes the state in non-infected 1C6 cells, the only contacts we succeeded to detect were Sprr2a with HS and Sprr1a. When Aire was introduced into the cells, two other genes co-localized in the proximity of HS, generating conjunction with it and among each other. We were not able to show Sprr2d interaction with other genes.

Thus, our results suggest that Aire activates Sprr family genes via formation of active loops between Sprr gene family members and HS site, or in other words, Aire induces the state when at least some Sprr genes are brought together and interact with the regulatory site.

5. DISCUSSION

It has been confirmed that AIRE has crucial importance in establishing T cell central tolerance (Kyewski et al., 2002). Earlier it was generally accepted that only free circulating or ubiquitously expressed antigenes could be presented to maturating thymocytes in the thymus, nowadays it is proved that due to AIRE activity the expression of many tissue-specific genes is induced in mTECs, ensuring the elimination of autoreactive thymocytes. However, there are tissue-restricted proteins that are expressed in the thymus but are not influenced by the presence of AIRE (Sospedra et al., 1998) suggesting that other transcriptional regulators may also contribute. The recent microarray experiments indicate AIRE-regulated gene clustering, interestingly, some genes are downregulated in the presence of AIRE (Johnnidis et al., 2005), suggesting that AIRE may have dualistic effect on gene expression. EDC cluster spans 3 Mb in mouse genome and includes multiple genes functionally related to epidermal barrier formation. The expression of several genes of this cluster is downregulated in Aire-deficient mice. It is not conceivable that Aire affects the global opening of chromatin, as only subset of the genes is regulated by Aire.

In the current study the expression of Sprr and S100a gene family members in both thymic and epithelial tissues, as well as in 1C6 cell line was analyzed. We were interested to see if Aire can upregulate EDC gene expression, as previously suggested from microarray analysis of Aire KO and WT mouse thymus material. Indeed, AIRE causes gene upregulation of several studied EDC genes, whereas, as shown by adenoviral transfection of 1C6 cell line, the excessive amount of Aire will not result in much higher expression of studied genes. EDC genes in infected cell line rise to a certain level, comparable to the *in vivo* changes. S100 family genes show more pronounced changes compared to Sprr family genes that were just weakly upregulated. A similar relatively mild changes are visible when reported Aire-regulated genes like Mup1, Spt1, Ins2, Tff3 were studied or when adenoviral Aire was introduced into 1C6 cell line or into primary thymic epithelial cells (Kont et al., 2007). Thus, this study together with previous studies indicates that AIRE might be transcriptional regulator that enhances transcription to certain extent, but never causes very strong upregulation.

Futhermore, we also cannot exclude that the increased mRNA production in the presence of AIRE can be due to indirect effects. AIRE can have direct effect on other transcriptional

regulators and through this action influence gene activity. It has been demonstrated that Klf4 plays significant role on the genes involved in epidermal barrier formation (Segre et al., 1999), in Aire knockout mice slight reduction in Klf4 mRNA production was also observed (data not shown). In addition, there is also the theory that the presence of AIRE affects maturation process of mTECs. When AIRE is absent, mTECs do not reach the terminal differentiation stage and for this reason are not able to express tissue-specific proteins (Gillard et al., 2007). An interesting speculation is that, EDC cluster gene expression reflects the maturation phase of mTECs achieving their peak in terminal phase, similarly with MHC II and CD80/86 molecules. In case of abnormal differentiation process, in Aire deficient organism, EDC protein production cannot be fully achieved.

To study further whether Aire can be involved directly in regulation of EDC genes, we applied 3C technique. With that we discovered that *Sprr* family specific hypersensitive site brings together several *Sppr* genes when AIRE is expressed in 1C6 cells. However, there was no definite connection between expression activation and interaction with HS. For instance, Sprr2d does not show contact with regulatory sequence, however, its the increase of expression is still remarkable. In similar fashion, Sprr2a maintains contact with HS in non-infected cells too; still the presence of Aire intensifies slightly its expression. It is very likely that with 3C method we were able to detect only some but not all occurring interactions. We did not succeed to trace any contacts within S100A family. Probably, the lack of known regulatory sequence responsible for coordinate regulation of S100, makes this task difficult.

All the interactions we detected occurred in infected cell line. Despite many attempts to analyze whether the same interactions are present *in vivo* with sorted thymic medullary epithelial cells, we did not obtain any positive results. Most likely, there are some technical problems. Since we usually get only half a million cells in the process of cell sorting, and a minimum required quantity for 3C right now is a million (Anita Göndor, personal communication), and usual number of the cells for one 3C experiment is 5 millions. Another obstruction could be a long enzymatic treatment during the cell sorting before formaldehyde fixation. If the contacts are not strong enough, they could be disrupted.

Although, we see several contacts with HS, it is an averaged picture of 5 million cells. It would be interesting to know what the situation on the individual cell level is, whether there is only one

gene activated at a time or simultaneous induction of several genes of the locus occurs. It would be interesting to know if the same loop formation mechanism functions in the skin or is it somehow modified according to the external stimuli and the presence of different transcription regulators.

Although several laboratories have attempted to uncover AIRE functioning mechanism, the precise picture is not accomplished yet. The variety and the big number of AIRE-regulated genes imply that AIRE activity involves epigenetic mechanisms and/or regulation of whole gene clusters occurs. The result of this study proposes the hypothesis that at least for some AIRE-regulated gene families, gene activation loop formation is required, so that the activated genes are concentrated in center and silenced genes extrude outside. Although AIRE seems to upregulate its target genes in various systems, suggesting its ability to directly regulate so many genes, further studies are needed to describe the mechanism behind this regulation.

SUMMARY

Autoimmune regulator (AIRE) protein is a transcriptional regulator, which expression is the strongest in medullary thymic epithelial cells. The thymus is an organ, where T cells accomplish their maturation interacting with thymic stromal cells. AIRE mutations provoke a rare organ-specific autoimmune disease named autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), indicating that AIRE plays a major role in establishing central tolerance. APECED is one of few autoimmune disease that are caused only by genetic component. For this reason unraveling functioning mechanisms of AIRE is especially important.

It has been shown that AIRE upregulates the expression of many tissue-specific genes in medullary thymic epithelial cells. As the number of Aire-regulated genes is up to thousand, it is assumed that its regulation is achieved through epigenetic mechanisms. Microarray data with Aire-deficient mice show that genes influenced by AIRE tend to co-localize in clusters. Usually clustered genes are subjected to coordinate regulation. Locus control regions are *cis*-regulatory elements, that directly contact with gene promoters by looping out intervening sequences and in such way induce chromatin modifications in activated genes.

One of Aire-regulated clusters is epidermal differentiation complex (EDC) that unifies gene families involved in epidermal barrier synthesis. We chose 12 genes and one earlier described hypersensitive site (HS) to test hypothesis that Aire promotes loop formation between HS and upregulated genes. First, we analyzed EDC gene expression in thymic, epidermal and cell line materials. We confirmed that Aire presence elevates transcript production of these genes. After that, we applied chromosome conformation capture technology (3C), the method that enables to detect direct contacts between chromosomal regions, to investigate if EDC genes and HS interact with each other and whether such interaction is modulated by Aire.

We were able to ascertain that several contacts within Sprr gene family. We found that without Aire only Sprr2a shows interaction with HS and Sprr1a, however, introduction of Aire induces HS interaction with Sprr2i and Sprr2f too.

Autoimmuunregulaator valk osaleb lingude tekkimises epidermaalse diferentseerumiskompleksi geeniperekondade liikmete vahel

KOKKUVÕTE

Autoimmuune regulaator (AIRE) on transkriptsiooni faktor, mille kõrge ekspressioon on näidatud tüümuse epiteeli medullaarsetes rakkudes. Tüümus on organ, kus T rakud küpsevad luues kontakte tüümuse strooma rakkudega. Haruldane autoimmuunhaigus APECED (autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy) tekib AIRE mutatsioonide tagajärjel, mis näitab AIRE osalust tsentraalse tolerantsuse tekkes. APECED on üks vähestest autoimmunhaigustest, mis on tingitud ainult geneetilisest komponendist. See asjaolu teeb AIRE funktsioneerimismehhanismi väljaselgitamist eriti tähtsaks.

On tuvastatud, et AIRE on võimeline aktiveerima mitmete geenide ekspresiooni tüümuse epiteeli rakkudes. Kuna Aire poolt reguleeritud geenide arv ulatub tuhandeni, siis ilmselt Aire ei mõjuta individuaalselt proomotoreid, vaid rolli mängivad üldisemad epigeneetilised mehhanismid. Mikrokiibi eksperimedid võrreldes Aire *knockout* ja metsiktüüpi hiiri näitasid paljude Aire poolt reguleeritud geenide paigutust klastrites. Tavaliselt geenid, mis asuvad genoomis klastrites, alluvad ka ühisele regulatoorsele mehhanismile.

Üks klastritest, mis allub AIRE mõjule, on epidermaalse diferentseerumise kompleks (EDC), kuhu on koondunud epidermise arengu eest vastutavad geenid. Käesolevas töös valiti sellest klastrist 12 geeni ja ühe varem kirjedatud hüpersensitiivse saidi, et kontrollida hüpoteesi, et AIRE mõjutab selle klastri regulatsiooni tekitades lingusid hüpersensitiivse saidi ja aktiveeritud geenide vahel.

Kõigepealt, sai teostatud EDC geenide analüüs tüümuses, epidermises ja rakuliinis ja sai kinnitatud, et Aire võimendab EDC geenide ekspressiooni. Me kasutasime kromosomaalse konformatsiooni vangistamist (3C, chromosome conformation capture), et testida kas EDC geenide ülesekspression on seotud kontakti loomisega aktiveeritud geeni ja hüpersensitiivse saidi (HS) vahel. Meil õnnestus detekteerida interakstioonid Sprr geeniperekonna sees. Rakuliinis Aire puudumisel ainult Sprr2a geen interakteerus HSi ja Sprr1a geeniga, Aire valgu olemasolu raku sees põhjustas lisaks uute kontaktide tekkimise, kus HS interakteerus täiendavalt Sprr2i ja Sprr2f geenidega.

ACKNOWLEDGMENTS

I am very grateful to my supervisor Pärt Peterson, who gave me the opportunity to work in his amazing group and offered this challenging project. His patient and persistent guidance helped to overcome the difficulties that inevitably arise in laboratory work.

I would like warmly thank Ana Rebane for her inspiration and fruitful discussions. With her huge experience Ana was able to grasp any problem and propose possible solutions. The most amazing that despite plenty experiments Ana always managed to find the time to deal with other people problems.

My sincere thanks go to Kaidi Möll, who introduced me to the laboratory and rendered a lot of assistance on early stages of the project. I really missed her help and motivation later on.

Kai Kisand with her endless joyfullness aquainted me with the cell sorting, Vivian Kont created the viral construct and demonstrated how the right virus infection should be done, Martti Laan offered emotional and physical support during animal experiments, Annika Häling provided precise and careful technical assistance and brought much amusement to our routine laboratory life.

I also want to thank all the members of our research group for their consant willingness to help and creating very friendly atmosphere inside and outside the lab:

Tõnis Org

Ingrid Liiv

Mario Saare

Rainis Venta

Liina Tserel

Uku Haljasorg

Maire Pihlap

Inga Rander

The work was supported by the grants from Wellcome Trust Senior Fellowship grant, EU Framework program 6 (Thymaide and Euraps) and by Estonian Science Foundation (6663).

REFERENCES

The Finnish-German Consortium (1997) An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat Genet*, **17**, 399-403.

Aho, S., Li, K., Ryoo, Y., McGee, C., Ishida-Yamamoto, A., Uitto, J. and Klement, J.F. (2004) Periplakin gene targeting reveals a constituent of the cornified cell envelope dispensable for normal mouse development. *Mol Cell Biol*, **24**, 6410-6418.

Anderson, M.S., Venanzi, E.S., Klein, L., Chen, Z., Berzins, S.P., Turley, S.J., von Boehmer, H., Bronson, R., Dierich, A., Benoist, C. and Mathis, D. (2002) Projection of an immunological self shadow within the thymus by the aire protein. *Science*, **298**, 1395-1401.

Ansel, K.M., Lee, D.U. and Rao, A. (2003) An epigenetic view of helper T cell differentiation. *Nat Immunol*, **4**, 616-623.

Bjorses, P., Halonen, M., Palvimo, J.J., Kolmer, M., Aaltonen, J., Ellonen, P., Perheentupa, J., Ulmanen, I. and Peltonen, L. (2000) Mutations in the AIRE gene: effects on subcellular location and transactivation function of the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy protein. *Am J Hum Genet*, **66**, 378-392.

Bonilla, I.E., Tanabe, K. and Strittmatter, S.M. (2002) Small proline-rich repeat protein 1A is expressed by axotomized neurons and promotes axonal outgrowth. *J Neurosci*, **22**, 1303-1315.

Carter, D., Chakalova, L., Osborne, C.S., Dai, Y.F. and Fraser, P. (2002) Long-range chromatin regulatory interactions in vivo. *Nat Genet*, **32**, 623-626.

Chaplin, D.D. (2006) 1. Overview of the human immune response. *J Allergy Clin Immunol*, **117**, S430-435.

de Laat, W. and Grosveld, F. (2003) Spatial organization of gene expression: the active chromatin hub. *Chromosome Res*, **11**, 447-459.

Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002) Capturing chromosome conformation. *Science*, **295**, 1306-1311.

Derbinski, J., Gabler, J., Brors, B., Tierling, S., Jonnakuty, S., Hergenhahn, M., Peltonen, L., Walter, J. and Kyewski, B. (2005) Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. *J Exp Med*, **202**, 33-45.

Derbinski, J., Schulte, A., Kyewski, B. and Klein, L. (2001) Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nat Immunol*, **2**, 1032-1039.

DeVoss, J., Hou, Y., Johannes, K., Lu, W., Liou, G.I., Rinn, J., Chang, H., Caspi, R.R., Fong, L. and Anderson, M.S. (2006) Spontaneous autoimmunity prevented by thymic expression of a single self-antigen. *J Exp Med*, **203**, 2727-2735.

Djian, P., Easley, K. and Green, H. (2000) Targeted ablation of the murine involucrin gene. *J Cell Biol*, **151**, 381-388.

Donato, R. (1999) Functional roles of S100 proteins, calcium-binding proteins of the EF-hand type. *Biochim Biophys Acta*, **1450**, 191-231.

Donato, R. (2003) Intracellular and extracellular roles of S100 proteins. *Microsc Res Tech*, **60**, 540-551.

Dostie, J. and Dekker, J. (2007) Mapping networks of physical interactions between genomic elements using 5C technology. *Nat Protoc*, **2**, 988-1002.

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., Green, R.D. and Dekker, J. (2006) Chromosome

Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*, **16**, 1299-1309.
Eckert, R.L., Broome, A.M., Ruse, M., Robinson, N., Ryan, D. and Lee, K. (2004) S100 proteins in the epidermis. *J Invest Dermatol*, **123**, 23-33.
Eckert, R.L. and Lee, K.C. (2006) S100A7 (Psoriasin): a story of mice and men. *J Invest Dermatol*, **126**, 1442-1444.
Eivazova, E.R. and Aune, T.M. (2004) Dynamic alterations in the conformation of the Ifng gene region during T helper cell differentiation. *Proc Natl Acad Sci U S A*, **101**, 251-256.
Farr, A.G. and Rudensky, A. (1998) Medullary thymic epithelium: a mosaic of epithelial "self"? *J Exp Med*, **188**, 1-4.
Fraser, P. (2006) Transcriptional control thrown for a loop. *Curr Opin Genet Dev*, **16**, 490-495.
Gavanescu, I., Kessler, B., Ploegh, H., Benoist, C. and Mathis, D. (2007) Loss of Aire-dependent thymic expression of a peripheral tissue antigen renders it a target of autoimmunity. *Proc Natl Acad Sci U S A*, **104**, 4583-4587.
Gillard, G.O., Dooley, J., Erickson, M., Peltonen, L. and Farr, A.G. (2007) Aire-dependent alterations in medullary thymic epithelium indicate a role for Aire in thymic epithelial differentiation. *J Immunol*, **178**, 3007-3015.
Gotter, J., Brors, B., Hergenhahn, M. and Kyewski, B. (2004) Medullary epithelial cells of the human thymus express a highly diverse selection of tissue-specific genes colocalized in chromosomal clusters. *J Exp Med*, **199**, 155-166.
Gray, D.H., Ueno, T., Chidgey, A.P., Malin, M., Goldberg, G.L., Takahama, Y. and Boyd, R.L. (2005) Controlling the thymic microenvironment. *Curr Opin Immunol*, **17**, 137-143.
Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, **51**, 975-985.
Halonen, M., Kangas, H., Ruppell, T., Ilmarinen, T., Ollila, J., Kolmer, M., Vihinen, M., Palvimo, J., Saarela, J., Ulmanen, I. and Eskelin, P. (2004) APECED-causing mutations in AIRE reveal the functional domains of the protein. *Hum Mutat*, **23**, 245-257.
Hardman, M.J., Sisi, P., Banbury, D.N. and Byrne, C. (1998) Patterned acquisition of skin barrier function during development. *Development*, **125**, 1541-1552.
Hori, S., Nomura, T. and Sakaguchi, S. (2003) Control of regulatory T cell development by the transcription factor Foxp3. *Science*, **299**, 1057-1061.
Horike, S., Cai, S., Miyano, M., Cheng, J.F. and Kohwi-Shigematsu, T. (2005) Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet*, **37**, 31-40.
Jackson, B., Tilli, C.M., Hardman, M.J., Avilion, A.A., MacLeod, M.C., Ashcroft, G.S. and Byrne, C. (2005) Late cornified envelope family in differentiating epithelia--response to calcium and ultraviolet irradiation. *J Invest Dermatol*, **124**, 1062-1070.
Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A. and Cheng, X. (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*.
Jiang, W., Anderson, M.S., Bronson, R., Mathis, D. and Benoist, C. (2005) Modifier loci condition autoimmunity provoked by Aire deficiency. *J Exp Med*, **202**, 805-815.
Johnnidis, J.B., Venanzi, E.S., Taxman, D.J., Ting, J.P., Benoist, C.O. and Mathis, D.J. (2005) Chromosomal clustering of genes controlled by the aire transcription factor. *Proc Natl Acad Sci U S A*, **102**, 7233-7238.
Kalinin, A.E., Kajava, A.V., Steinert P.M. (2002) Epithelial barrier function: assembly and structural feautures of the cornified cell envelope. *BioEssays*, **24**, 789-800.

Klamp, T., Sahin, U., Kyewski, B., Schwendemann, J., Dhaene, K. and Tureci, O. (2006) Expression profiling of autoimmune regulator AIRE mRNA in a comprehensive set of human normal and neoplastic tissues. *Immunol Lett*, **106**, 172-179.

Kleinegger, C.L., Stoeckel, D.C. and Kurago, Z.B. (2001) A comparison of salivary calprotectin levels in subjects with and without oral candidiasis. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, **92**, 62-67.

Koch, P.J., de Viragh, P.A., Scharer, E., Bundman, D., Longley, M.A., Bickenbach, J., Kawachi, Y., Suga, Y., Zhou, Z., Huber, M., Hohl, D., Kartasova, T., Jarnik, M., Steven, A.C. and Roop, D.R. (2000) Lessons from loricrin-deficient mice: compensatory mechanisms maintaining skin barrier function in the absence of a major cornified envelope protein. *J Cell Biol*, **151**, 389-400.

Kogawa, K., Nagafuchi, S., Katsuta, H., Kudoh, J., Tamiya, S., Sakai, Y., Shimizu, N. and Harada, M. (2002) Expression of AIRE gene in peripheral monocyte/dendritic cell lineage. *Immunol Lett*, **80**, 195-198.

Kont, V., Laan, M., Kisand, K., Merits, A., Scott, H.S. and Peterson, P. (2007) Modulation of Aire regulates the expression of tissue-restricted antigens. *Mol Immunol*.

Kosak, S.T., Skok, J.A., Medina, K.L., Riblet, R., Le Beau, M.M., Fisher, A.G. and Singh, H. (2002) Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science*, **296**, 158-162.

Kumar, P.G., Laloraya, M., Wang, C.Y., Ruan, Q.G., Davoodi-Semiromi, A., Kao, K.J. and She, J.X. (2001) The autoimmune regulator (AIRE) is a DNA-binding protein. *J Biol Chem*, **276**, 41357-41364.

Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W. and Ohlsson, R. (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci U S A*, **103**, 10684-10689.

Kyewski, B., Derbinski, J., Gotter, J. and Klein, L. (2002) Promiscuous gene expression and central T-cell tolerance: more than meets the eye. *Trends Immunol*, **23**, 364-371.

Kyewski, B. and Klein, L. (2006) A central role for central tolerance. *Annu Rev Immunol*, **24**, 571-606.

Lan, F., Collins, R.E., De Cegli, R., Alpatov, R., Horton, J.R., Shi, X., Gozani, O., Cheng, X. and Shi, Y. (2007) Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature*, **448**, 718-722.

Lanctot, C., Cheutin, T., Cremer, M., Cavalli, G. and Cremer, T. (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet*, **8**, 104-115.

Li, Q., Barkess, G. and Qian, H. (2006) Chromatin looping and the probability of transcription. *Trends Genet*, **22**, 197-202.

Li, Q., Peterson, K.R., Fang, X. and Stamatoyannopoulos, G. (2002) Locus control regions. *Blood*, **100**, 3077-3086.

Ling, J.Q., Li, T., Hu, J.F., Vu, T.H., Chen, H.L., Qiu, X.W., Cherry, A.M. and Hoffman, A.R. (2006) CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science*, **312**, 269-272.

Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J. and Axel, R. (2006) Interchromosomal interactions and olfactory receptor choice. *Cell*, **126**, 403-413.

Lukinmaa, P.L., Waltimo, J. and Pirinen, S. (1996) Microanatomy of the dental enamel in autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED): report of three cases. *J Craniofac Genet Dev Biol*, **16**, 174-181.

Maatta, A., DiColandrea, T., Groot, K. and Watt, F.M. (2001) Gene targeting of envoplakin, a cytoskeletal linker protein and precursor of the epidermal cornified envelope. *Mol Cell Biol*, **21**, 7047-7053.

Marenholz, I., Volz, A., Ziegler, A., Davies, A., Ragoussis, I., Korge, B.P. and Mischke, D. (1996) Genetic analysis of the epidermal differentiation complex (EDC) on human chromosome 1q21: chromosomal orientation, new markers, and a 6-Mb YAC contig. *Genomics*, **37**, 295-302.

Marionnet, C., Bernerd, F., Dumas, A., Verrecchia, F., Mollier, K., Compan, D., Bernard, B., Lahfa, M., Leclaire, J., Medaisko, C., Mehul, B., Seite, S., Mauviel, A. and Dubertret, L. (2003) Modulation of gene expression induced in human epidermis by environmental stress in vivo. *J Invest Dermatol*, **121**, 1447-1458.

Martin, N., Patel, S. and Segre, J.A. (2004) Long-range comparison of human and mouse Sprr loci to identify conserved noncoding sequences involved in coordinate regulation. *Genome Res*, **14**, 2430-2438.

Maston, G.A., Evans, S.K. and Green, M.R. (2006) Transcriptional Regulatory Elements in the Human Genome. *Annu Rev Genomics Hum Genet*, **7**, 29-59.

Mathis, D. and Benoist, C. (2004) Back to central tolerance. *Immunity*, **20**, 509-516.

Matsuki, M., Yamashita, F., Ishida-Yamamoto, A., Yamada, K., Kinoshita, C., Fushiki, S., Ueda, E., Morishima, Y., Tabata, K., Yasuno, H., Hashida, M., Iizuka, H., Ikawa, M., Okabe, M., Kondoh, G., Kinoshita, T., Takeda, J. and Yamanishi, K. (1998) Defective stratum corneum and early neonatal death in mice lacking the gene for transglutaminase 1 (keratinocyte transglutaminase). *Proc Natl Acad Sci U S A*, **95**, 1044-1049.

McBride, D.J. and Kleinjan, D.A. (2004) Rounding up active cis-elements in the triple C corral: combining conservation, cleavage and conformation capture for the analysis of regulatory gene domains. *Brief Funct Genomic Proteomic*, **3**, 267-279.

Meloni, A., Perniola, R., Faa, V., Corvaglia, E., Cao, A. and Rosatelli, M.C. (2002) Delineation of the molecular defects in the AIRE gene in autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy patients from Southern Italy. *J Clin Endocrinol Metab*, **87**, 841-846.

Mizuochi, T., Kasai, M., Kokuho, T., Kakiuchi, T. and Hirokawa, K. (1992) Medullary but not cortical thymic epithelial cells present soluble antigens to helper T cells. *J Exp Med*, **175**, 1601-1605.

Murrell, A., Heeson, S. and Reik, W. (2004) Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat Genet*, **36**, 889-893.

Murthy, A.R., Lehrer, R.I., Harwig, S.S. and Miyasaki, K.T. (1993) In vitro candidastatic properties of the human neutrophil calprotectin complex. *J Immunol*, **151**, 6291-6301.

Nagamine, K., Peterson, P., Scott, H.S., Kudoh, J., Minoshima, S., Heino, M., Krohn, K.J., Lalioti, M.D., Mullis, P.E., Antonarakis, S.E., Kawasaki, K., Asakawa, S., Ito, F. and Shimizu, N. (1997) Positional cloning of the APECED gene. *Nat Genet*, **17**, 393-398.

Nemes, Z. and Steinert, P.M. (1999) Bricks and mortar of the epidermal barrier. *Exp Mol Med*, **31**, 5-19.

Nomura, T., Sandilands, A., Akiyama, M., Liao, H., Evans, A.T., Sakai, K., Ota, M., Sugiura, H., Yamamoto, K., Sato, H., Palmer, C.N., Smith, F.J., McLean, W.H. and Shimizu, H. (2007) Unique mutations in the filaggrin gene in Japanese patients with ichthyosis vulgaris and atopic dermatitis. *J Allergy Clin Immunol*, **119**, 434-440.

Nozaki, I., Lunz, J.G., 3rd, Specht, S., Stolz, D.B., Taguchi, K., Subbotin, V.M., Murase, N. and Demetris, A.J. (2005) Small proline-rich proteins 2 are noncoordinately upregulated by IL-6/STAT3 signaling after bile duct ligation. *Lab Invest*, **85**, 109-123.

Ogbourne, S. and Antalis, T.M. (1998) Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J*, **331 (Pt 1)**, 1-14.

Ohlsson, R., C., Gondor, A., T., Cremer, M., Cavalli, G. and Cremer, T.

Orphanides, G. and Reinberg, D. (2002) A unified theory of gene expression. *Cell*, **108**, 439-451.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W. and Fraser, P. (2004) Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet*, **36**, 1065-1071.

Palmer, E. (2003) Negative selection--clearing out the bad apples from the T-cell repertoire. *Nat Rev Immunol*, **3**, 383-391.

Patel, S., Kartasova, T. and Segre, J.A. (2003) Mouse Sprr locus: a tandem array of coordinately regulated genes. *Mamm Genome*, **14**, 140-148.

Perniola, R., Falorni, A., Clemente, M.G., Forini, F., Accogli, E. and Lobreglio, G. (2000) Organ-specific and non-organ-specific autoantibodies in children and young adults with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED). *Eur J Endocrinol*, **143**, 497-503.

Peterson, P., Pitkanen, J., Sillanpaa, N. and Krohn, K. (2004) Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED): a model disease to study molecular aspects of endocrine autoimmunity. *Clin Exp Immunol*, **135**, 348-357.

Pitkanen, J., Doucas, V., Sternsdorf, T., Nakajima, T., Aratani, S., Jensen, K., Will, H., Vahamurto, P., Ollila, J., Vihinen, M., Scott, H.S., Antonarakis, S.E., Kudoh, J., Shimizu, N., Krohn, K. and Peterson, P. (2000) The autoimmune regulator protein has transcriptional transactivating properties and interacts with the common coactivator CREB-binding protein. *J Biol Chem*, **275**, 16802-16809.

Pitkanen, J., Rebane, A., Rowell, J., Murumagi, A., Strobel, P., Moll, K., Saare, M., Heikkila, J., Doucas, V., Marx, A. and Peterson, P. (2005) Cooperative activation of transcription by autoimmune regulator AIRE and CBP. *Biochem Biophys Res Commun*, **333**, 944-953.

Pontynen, N., Miettinen, A., Arstila, T.P., Kampe, O., Alimohammadi, M., Vaarala, O., Peltonen, L. and Ulmanen, I. (2006) Aire deficient mice do not develop the same profile of tissue-specific autoantibodies as APECED patients. *J Autoimmun*, **27**, 96-104.

Presland, R.B. and Dale, B.A. (2000) Epithelial structural proteins of the skin and oral cavity: function in health and disease. *Crit Rev Oral Biol Med*, **11**, 383-408.

Presland, R.B. and Jurevic, R.J. (2002) Making sense of the epithelial barrier: what molecular biology and genetics tell us about the functions of oral mucosal and epidermal tissues. *J Dent Educ*, **66**, 564-574.

Pugliese, A. (2004) Central and peripheral autoantigen presentation in immune tolerance. *Immunology*, **111**, 138-146.

Ragoczy, T., Bender, M.A., Telling, A., Byron, R. and Groudine, M. (2006) The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes Dev*, **20**, 1447-1457.

Ramsey, C., Winqvist, O., Puhakka, L., Halonen, M., Moro, A., Kampe, O., Eskelin, P., Pelto-Huikko, M. and Peltonen, L. (2002) Aire deficient mice develop multiple features of APECED phenotype and show altered immune response. *Hum Mol Genet*, **11**, 397-409.

Rioux, J.D. and Abbas, A.K. (2005) Paths to understanding the genetic basis of autoimmune disease. *Nature*, **435**, 584-589.

Russell, L.J., DiGiovanna, J.J., Rogers, G.R., Steinert, P.M., Hashem, N., Compton, J.G. and Bale, S.J. (1995) Mutations in the gene for transglutaminase 1 in autosomal recessive lamellar ichthyosis. *Nat Genet*, **9**, 279-283.

Ryckman, C., Vandal, K., Rouleau, P., Talbot, M. and Tessier, P.A. (2003) Proinflammatory activities of S100: proteins S100A8, S100A9, and S100A8/A9 induce neutrophil chemotaxis and adhesion. *J Immunol*, **170**, 3233-3242.

Sayegh, C., Jhunjhunwala, S., Riblet, R. and Murre, C. (2005) Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes Dev*, **19**, 322-327.

Segre, J.A. (2006) Epidermal barrier formation and recovery in skin disorders. *J Clin Invest*, **116**, 1150-1158.

Segre, J.A., Bauer, C. and Fuchs, E. (1999) Klf4 is a transcription factor required for establishing the barrier function of the skin. *Nat Genet*, **22**, 356-360.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B. and de Laat, W. (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*, **38**, 1348-1354.

Skok, J.A., Gisler, R., Novatchkova, M., Farmer, D., de Laat, W. and Busslinger, M. (2007) Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. *Nat Immunol*, **8**, 378-387.

Sospedra, M., Ferrer-Francesch, X., Dominguez, O., Juan, M., Foz-Sala, M. and Pujol-Borrell, R. (1998) Transcription of a broad range of self-antigens in human thymus suggests a role for central mechanisms in tolerance toward peripheral antigens. *J Immunol*, **161**, 5918-5929.

Spilianakis, C.G. and Flavell, R.A. (2004) Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol*, **5**, 1017-1027.

Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R. and Flavell, R.A. (2005) Interchromosomal associations between alternatively expressed loci. *Nature*, **435**, 637-645.

Splinter, E., Grosveld, F. and de Laat, W. (2004) 3C technology: analyzing the spatial organization of genomic loci in vivo. *Methods Enzymol*, **375**, 493-507.

Starr, T.K., Jameson, S.C. and Hogquist, K.A. (2003) Positive and negative selection of T cells. *Annu Rev Immunol*, **21**, 139-176.

Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., Pant, V., Tiwari, V., Kurukuti, S. and Ohlsson, R. (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*, **38**, 1341-1347.

Takahama, Y. (2006) Journey through the thymus: stromal guides for T-cell development and selection. *Nat Rev Immunol*, **6**, 127-135.

Tan, Y.F., Sun, X.Y., Li, F.X., Tang, S., Piao, Y.S. and Wang, Y.L. (2006) Gene expression pattern and hormonal regulation of small proline-rich protein 2 family members in the female mouse reproductive system during the estrous cycle and pregnancy. *Reprod Nutr Dev*, **46**, 641-655.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F. and de Laat, W. (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell*, **10**, 1453-1465.

Walker, L.S. and Abbas, A.K. (2002) The enemy within: keeping self-reactive T cells at bay in the periphery. *Nat Rev Immunol*, **2**, 11-19.

Wei, G.H., Liu, D.P. and Liang, C.C. (2005) Chromatin domain boundaries: insulators and beyond. *Cell Res*, **15**, 292-300.

West, A.G. and Fraser, P. (2005) Remote control of gene transcription. *Hum Mol Genet*, **14 Spec No 1**, R101-111.

Villasenor, J., Benoist, C. and Mathis, D. (2005) AIRE and APECED: molecular insights into an autoimmune disease. *Immunol Rev*, **204**, 156-164.

Wurtele, H. and Chartrand, P. (2006) Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosome Res*, **14**, 477-495.