Promiscuous gene expression in the thymic medulla – on regulation at the epigenetic and single cell level

INAUGURAL – DISSERTATION

zur Erlangung der Doktorwürde
der
Naturwissenschaftlich - Mathematischen Gesamtfakultät
der
Ruprecht – Karls – Universität
Heidelberg

vorgelegt von
Diplom-Biochemikerin Anna Sinemus
aus Göttingen

Gutachter: Prof. Dr. Günter Hämmerling
Prof. Dr. Bruno Kyewski

Die vorliegende Arbeit wurde angefertigt in der
Abteilung Entwicklungsimmunologie,
Leitung Prof. Dr. Bruno Kyewski,
im Deutschen Krebsforschungszentrum Heidelberg.

Hiermit erkläre ich, dass ich die vorgelegte Dissertation selbst verfasst und mich dabei keiner anderen, als der von mir ausdrücklich bezeichneten Quellen bedient habe.

Heidelberg,

Anna Sinemus

Contents

Zusammenfassung

Die Unterscheidung zwischen Selbst und Fremd (Selbsttoleranz) ist eine grundlegende Eigenschaft des Immunsystems. Die Induktion von Selbsttoleranz beruht auf verschiedenen Mechanismen, die sowohl im Thymus (zentrale Toleranz) als auch in peripheren lymphoiden und nicht-lymphoiden Organen (periphere Toleranz) wirksam sind. Die zentrale Toleranz, also die Selbsttoleranz des heranreifenden T-Zell Repertoires, wird durch negative Selektion sowie durch Induktion von regulatorischen T Zellen (Treg) im Thymus vermittelt. Dies geschieht durch TCR/MHC-Peptid Kontakte zwischen Thymozyten und antigenpräsentierenden Zellen (APC). Die Bandbreite der zentralen Toleranz wird unter anderem durch die Expression gewebsspezifischer Antigene (TRA) durch medulläre Thymusepithelzellen (mTEC) bestimmt. Dieser Prozess wird promiske Genexpression (pGE) genannt. Die pGE umfasst die ektopische Expression von TRA aus quasi allen Geweben des Körpers innerhalb der mTEC Population. Der Prozess der Toleranzinduktion für TRA wurde in den letzten Jahren besser verstanden, die molekularen Mechanismen jedoch, welche pGE in mTEC regulieren, sind noch weitgehend unbekannt. Der größte Anteil der TRA ist in der reifen $CD80^{hoch}$ mTEC Population ($mTEC^{high}$) exprimiert, allerdings sind die Expressionmuster in $mTEC^{high}$ auf Einzelzellebene stark heterogen. Nur 1-15% der $mTEC^{high}$ exprimieren ein bestimmtes Antigen und es wurde gezeigt, dass die Koexpressionsmuster auf Einzellzellebene stochastisch sind. Ein Charakteristikum promisk exprimierter Gene ist ihre Kolokalisation in chromosomalen Clustern, was für die Beteiligung epigenetischer Mechanismen an dieser unorthodoxen Genregulation spricht.

Um die Regulation solch geclusterter Genexpression besser zu verstehen wurde exemplarisch die epigenetische Regulation von pGE auf der Populations- und auf der Einzelzellebene in zwei verschiedenen Loci untersucht: im Kasein-Genlocus und im Gad67 Locus. In beiden Loci wurden die Regulationsmechanismen zwischen mTEC Subpopulationen und dem entsprechenden Gewebe verglichen. Der Fokus lag hierbei auf epigenetischen Regulationsmechanismen wie zum Beispiel Histonmodifikationen und bestimmten Aspekten der Kernstruktur. Beides wurde im Kaseinlocus analysiert. Im Gad67 Locus wurden Allel-spezifische und Gen-spezifische Ko-expressionsmuster an Hand einer Gad67/eGFP *knock in* Maus untersucht, zusätzlich wurden Analysen von epigenetischen Modifikationen durchgeführt.

Es konnte gezeigt werden, dass sowohl die Expression von Kasein beta (Csnb) als auch von Gad67 mit epigenetischen Parametern korreliert, jedoch spielen jeweils verschiedene aktive Histonmodifikationen eine Rolle. Auch eine Lockerung der Chromatinstruktur spielt für die Expression von Csnb, welches auf Grund seiner besonders häufigen Expression eine Ausnahmeposition innerhalb des Kaseinlocus einnimmt, eine Rolle. Im Gad67 Locus wurden zusätzlich Einzelzell- Expressionsanalysen durchgeführt, welche eine starke Abweichung zwischen mRNA- und Protein- Frequenz in TRA zeigen konnten.

Auf der Basis dieser Ergebnisse schlagen wir ein dreischrittiges Modell für die epigenetische Öffnung des Csnb Locus vor: Zuerst findet DNA Demethylierung statt, gefolgt von einer Dekompaktierung des Chromatins und der aktiven Modifikation von Histonen. Weiterhin legen die starken Abweichungen zwischen mRNA- und Protein-Frequenz für TRA auf Einzelzellebene nahe, dass das Repertoire der pGE auf Einzelzellebene fluktuiert. Diese Eigenschaft erhöht potentiell die Diversität der Antigenexpression im Mikroenvironment im Thymus und ist somit vermutlich essentiell für die Induktion von T Zell Toleranz.

Summary

The immune system is delicately balanced by self-antigen driven tolerance and pathogen-driven immunity. Self-tolerance of the T cell repertoire, which is an essential aspect of this balance, is mediated by multiple mechanisms operating both in the thymus (central tolerance) and in peripheral lymphoid and non-lymphoid organs (peripheral tolerance). Central tolerance, thus self-tolerance of the maturing T-cell repertoire in the thymus, is controlled by negative selection and the induction of regulatory T cells (Treg). These processes are mediated via TCR-MHC/peptide contacts between thymocytes and thymic antigen presenting cells (dendritic cells, thymic epithelial cells). The scope of central tolerance is to a large extent dictated by the expression of tissue-restricted antigens (TRA) by medullary thymic epithelial cells (mTEC), a process known as promiscuous gene expression (pGE). pGE encompasses the ectopic expression of TRA from virtually all tissues of the body in the mTEC population. While increasing insight into the tolerance modes linked to pGE has been gained in the last years, the molecular mechanisms involved in the regulation of pGE in mTEC remain largely obscure. The majority of TRA are expressed in the mature $CD80^{high}$ mTEC population ($mTEC^{high}$) whereby expression patterns of individual cells are highly heterogeneous. Only 1-15% of $mTEC^{high}$ express a given antigen and co-expression patterns have been characterized as highly stochastic. A conspicuous feature of promiscuously expressed genes is their co-localization in chromosomal clusters suggesting a regulation of pGE at the epigenetic level.

In order to gain insight into the regulation of such clustered gene expression, we exemplarily investigated the epigenetic regulation of pGE at the population and the single cell level of two genomic loci in mTEC: the casein gene locus and the Gad67 locus. For both loci, mechanisms of regulation were directly compared between mTEC subpopulations and the corresponding tissue. We focused on epigenetic regulation mechanisms such as histone tail modifications and certain aspects of nuclear structure, both were analyzed in the casein gene locus. In the Gad67 locus we analyzed allele specific and gene co-expression patterns in a Gad67/eGFP *knock in* mouse model in addition to epigenetic modifications.

We found the expression of Casein beta (Csnb) to correlate both with chromatin decompaction and active histone modifications. Gad67 expression equally correlated with active histone modifications. However, the types of histone modifications differed between Gad67 and Csnb, which is an unusual gene in the casein locus as it is expressed at a particularly high frequency. In the Gad67 locus we additionally performed single cell expression analysis. We found significant discrepancies between protein and mRNA frequencies in the case of TRA.
On the basis of these findings we propose a three-step model for the epigenetic opening of the Csnb gene: First, DNA demethylation takes place followed by chromatin decompaction and the introduction of active histone modifications. Furthermore, the discrepancies found between protein and mRNA expression frequency in the case of TRA let us assume that pGE fluctuates in individual cells. This concept potentially increases the diversity of antigen expression in the microenvironment in the thymus and thus may be crucial for the induction of T cell tolerance.

List of Abbreviations

Aire	autoimmune regulator
APC	antigen presenting cell
APECED	autoimmune polyendocrinopathy candidiasis ectodermal dystrophy
BAC	bacterial artificial chromosome
BSA	bovine serum albumin
C/EBPbeta	CCAAT-enhancer-binding protein beta
CARD	caspase recruitment domain
CChIP	carrier ChIP
CD	cluster of differentiation
cDNA	complementary DNA
CEA	carcinoembryonic antigen
ChIP	chromatin immunoprecipitation
CLSM	confocal laser scanning microscopy
CMJ	cortico-medullary junction
CpG	cytosine-guanine dinucleotide
CRP	C-reactive protein
Csn	casein
cTEC	cortical thymic epithelial cell(s)
CT-ITC	chromosome territory-interchromatin compartment
dATP	desoxyadenosine triphosphate
DC	dendritic cell(s)
dCTP	desoxycytosine triphosphate
dGTP	desoxyguanosine triphosphate
DMEM	Dulbecco's Modified Eagle Medium
DNA	desoxyribonucleic acid
dNTP	desoxynucleotide triphosphate (equal amounts of dATP, dCTP, dTTP, dGTP)
DP	double positive
DTT	dithiothreitol
dTTP	desoxythymidine triphosphate
dUTP	desoxyuridine triphosphate
EDTA	ethylenediaminetetraacetic acid
eGFP	enhanced green fluorescent protein
EL4	a mouse lymphoma cell line
FACS	fluorescence activated cell sorting
FCS	fetal calf serum
FISH	fluorescene in situ hybridization
FITC	fluorescein isothiocyanate
FoxP3	forkhead box P3
GABA	gamma-amino-butyric acid
GAD67	glutamate decarboxylase, 67 kD isoform
Gorasp2	golgi reassembly stacking protein 2
H3K27	histone 3 lysine 27
H3K4	histone 3 lysine 4
HAT	histone acetyl transferase
HDAC	histone deacetylase
IP	immunoprecipitation
kB	kilobase
kD	kilodalton
ko	knock out
LB	lysogeny broth
LPA	linear polyacrylamide
Lti	lymphoid tissue inducer cell
MAA	methanol acetic acid
MACS	magnetic cell separation
MB	megabase
MDB	methylcytosine domain binding
MEC	mammary gland epithelial cell(s)
MeCP	methyl CpG binding proteins
MHC	major histocompatibility complex
mTEC	medullary thymic epithelial cell(s)
$mTEC^{high}$	mTEC expressing high levels of costimulatory molecules

$mTEC^{low}$	mTEC expressing low levels of costimulatory molecules
MUC1	mucin 1
Myo3b	myosin III b
MΦ	macrophage
Ova	ovalbumin
PBS	phosphate buffered saline
PcG	polycomb group
PCR	polymerase chain reaction
PE	phycoerythrin
PerCP	peridinin chlorophyll protein
PFA	paraformaldehyde
pGE	promiscuous gene expression
PHD1	plant homeo domain 1
PI	propidium iodide
PLP	proteolipid protein (myelin) 1
PMSF	phenylmethylsulphonyl fluoride
PSF	point spread function
P-TEFb	positive transcription elongation factor b
Pth	parathyroid hormone
qPCR	quantitative PCR
RAG	recombinase activating genes
RANKL	receptor activator for nuclear factor κ B ligand
RIP	rat insulin promoter
RNA	ribonucleic acid
RPMI-1640	medium developed at Roswell Park Memorial Institute
RT	reverse trancriptase
sav	streptavidin
SC	single cell
SDS	sodium dodecyl sulfate
Sglt1	sodium/glucose cotransporter 1
SP	single positive
SPDM	spectral precision distance microscopy
SSC	saline-sodium citrate
TAE	tris-acetate EDTA
TCR	T cell receptor
TRA	tissue restricted antigen(s)
TEC	thymic epithelial cell(s)
Treg	natural regulatory T cell
WAP	whey acidic protein
wt	wildtype

1. Introduction

The immune system is crucial for the detection and elimination of pathogens which it accomplishes through the diverse mechanisms of innate and adaptive immune responses. T-cells, playing a central role in adaptive immunity, carry a highly diverse repertoire of T cell receptors (TCR) which they use to recognize foreign or self-antigens in combination with self major histocompatibility complexes (MHC) (self-restriction of the T cell repertoire). Generation and maturation of a highly diverse T cell repertoire occurs in the thymus. As auto-reactive T cells appear during maturation, the T cell repertoire undergoes a strict quality control process in the thymus, thus ensuring self-tolerance. The thymus confers the fundamental ability in this process to distinguish between foreign and self antigens. The crucial immunological function of the thymus was only discovered in the early 1960s by Jacques Miller [1].

1.1 Thymocyte maturation and central T cell tolerance

Thymocytes, which are immature T-cells in the thymus, undergo a complicated and highly ordered dynamic selection process during which they make contact with a large variety of cells of epithelial and mesenchymal origin. These cells together make up the thymic microenvironment and can be summarized as the thymic stroma [2]. The thymus consists of different, functionally distinct (but partially overlapping) compartments: the subcapsular zone, the cortex and the medulla which are separated by the cortico-medullary junction (CMJ). During the selection process, thymocytes migrate directionally through the different compartments of the thymus (Figure 1). Their maturation state is classified by expression of the co-receptors CD4 and CD8 which changes during maturation from double negative (DN) ($CD4^-CD8^-$) via double positive (DP) ($CD4^+CD8^+$) to single positive (SP) ($CD4^+$ or $CD8^+$).

1.1.1 T cell maturation and selection

T cell progenitors from the bone marrow entering at the cortical-medullary junction migrate as DN thymocytes along a chemokine gradient outwards to the subcapsular zone. Two weeks after entry thymocytes, which have successfully undergone somatic recombination of T cell receptor gene cassettes, become DP and express a functional and unique TCR on their surface. This somatic recombination is a stochastic process, which creates a TCR from a theoretical repertoire of 10^{15} possibilities. Two selection processes, positive and negative selection, ensure that thymocytes which are dysfunctional (binding peptide-MHC with insufficient avidity) or dangerous (autoreactive thymocytes) do not leave into the periphery but are removed from the repertoire [3, 4]. According to the avidity model by Jameson et al. [5], the combination of TCR avidity for peptide-MHC together with the peptide concentration on the antigen-presenting cell (APC) determines the fate of a thymocyte.

Positive selection occurs in the cortex and provides those thymocytes with a survival signal which are functional, i.e. which carry a TCR with sufficient avidity to bind peptide-MHC on cortical thymic epithelial cells (cTEC). The majority of thymocytes cannot bind with a sufficient avidity and undergoes death by neglect. Those thymocytes which are positively selected downregulate one of their coreceptors, become SP and migrate on a chemokine CCR7 gradient [6, 7] towards the medulla (Figure 1).

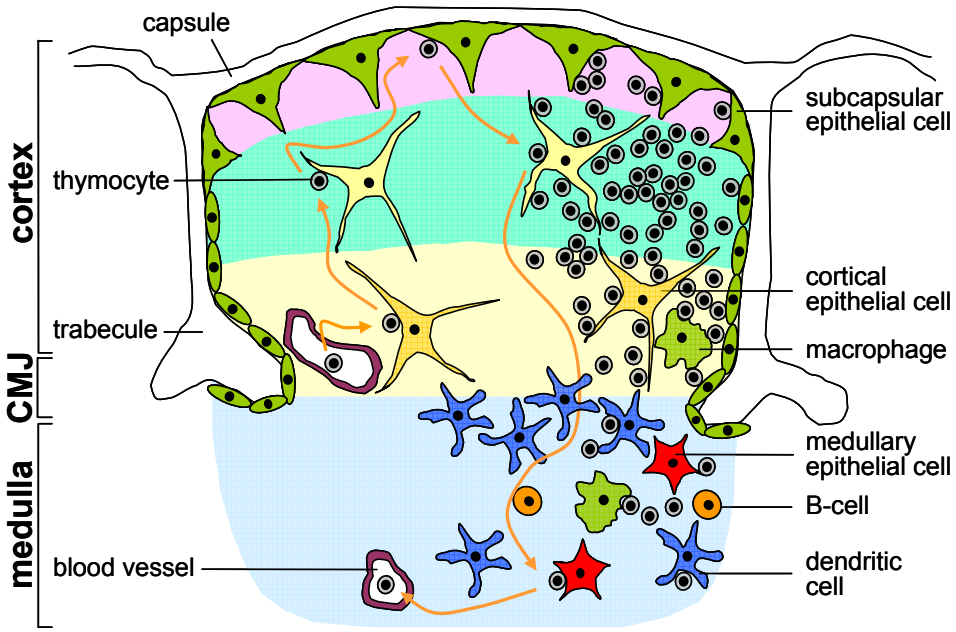


Figure 1
Cellular composition of the thymus. Multipotent lymphoid progenitors enter the thymus via endothelial venules at the cortico-medullary junction (CMJ). They commit to the T cell lineage, migrate outwards to the subcapsular zone and are positively selected in the cortex upon TCR- MHC/peptide interactions with cTEC. Positively selected thymocytes migrate back to the medulla, where they undergo negative selection or Treg induction upon interaction with different types of APCs (primarily mTEC and DC). mTEC, highlighted in red, play an essential role in self-tolerance induction toward tissue-restricted self-antigens. Shaded areas depict functionally distinct stratified microenvironments (adapted from [8]).

The second selection step, negative selection, is the crucial step in the induction of central T cell tolerance: Autoreactive thymocytes are removed from the repertoire, rendering it self- tolerant.
SP thymocytes migrating from the cortex to the medulla are negatively selected upon high-avidity interactions with peptide-MHC presenting APC. Engagement in high-avidity interactions leads to apoptosis and their removal from the repertoire (Figure 2). Negative selection has for long been believed to only occur in the medulla or at the cortical medullary junction [9, 10], but a recent study showed negative selection to occur also in the cortex [11], thus almost in parallel with positive selection. Whether positive selection is a prerequisite for negative selection is still unclear[12, 13].
The window of avidity between positive and negative selection is very small (Figure 2). It is still unclear how only one type of signal, successful binding of the TCR, can determine opposite cell fates (positive and negative selection or Treg induction). Hoquist et al. proposed that positively

and negatively selecting ligands are of different quality giving different signals on the level of the individual TCR-MHC-peptide interaction [14, 15].

Thymocytes with low affinities will preferentially become positively selected and develop to naïve T cells while thymocytes with higher affinities will preferentially differentiate to become natural regulatory T cells. Very high- affinity TCRs will lead to the negative selection of thymocytes, i.e. thymocytes receive a signal to undergo apoptosis (clonal deletion).

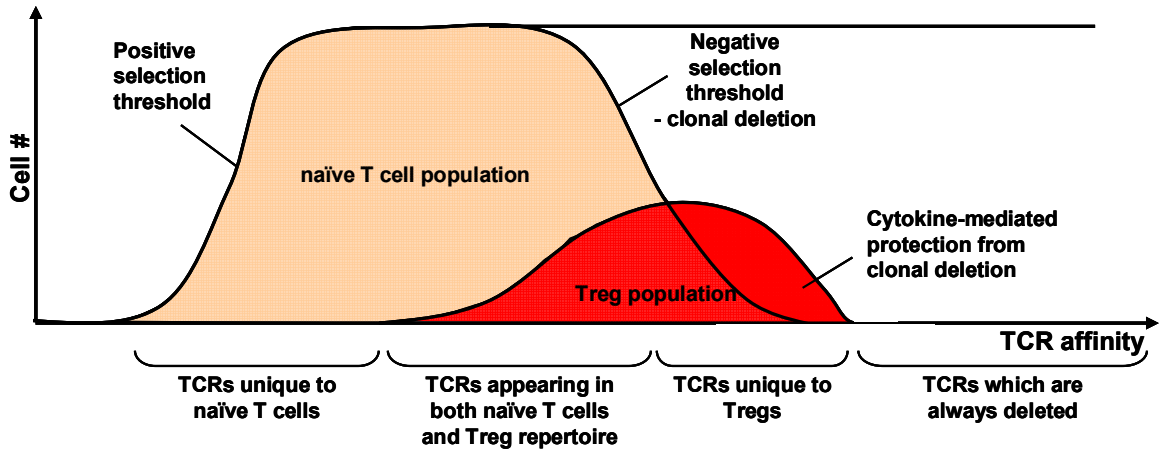


Figure 2

A model for thymocyte fate during maturation according to TCR affinity. The fate of a developing thymocyte is primarily determined by the affinity of its T-cell receptor (TCR) for self-peptide–MHC ligands. No or very low affinity will lead to death by neglect which is the fate of the large majority of progenitors. If the TCR has a low affinity for self-peptide–MHC, then the progenitor is positively selected into the naïve T cell lineage. Higher affinities for self-peptide-MHC may lead to naïve T cell or Treg generation while high TCR affinities lead to clonal deletion in the naïve T cell repertoire. Treg with high TCR affinities may nevertheless survive due to protective cytokine signaling. TCRs with very high affinity are always deleted. Adapted from [16].

Which exact mix of signals can induce which cell fate in a developing thymocyte is still an issue of debate. The type of APC inducing negative selection – cTEC, mTEC or DC – was thought to have an effect on the outcome of negative selection[17]. In the cortex, both cTEC and cortical DC can mediate clonal deletion [11]. In the medulla, a strict division of labor has been postulated: DC were held to be responsible for clonal deletion while mTEC were thought to induce differentiation into the Treg lineage [17]. Clonal deletion of thymocytes with high affinity to TRA was believed to be only possible in co-operation with DC [18-20]. In this case, expression of TRA by mTEC alone would allow for a cross-presentation of TRA on DC[20]. This strict "division of labor" turned out not to be true: Gallegos et al. [21] could show that presentation of Ova antigen exclusively by mTEC allowed deletion of cognate $CD8^+$ T cells. Only for deletion of Ova-specific $CD4^+$ T cells, crosspresentation of Ova on DC was indispensable. Clonal deletion can thus be induced by cTEC, mTEC and DC, but different thresholds for deletion may exist.

In conclusion, TCR avidity seems to be a major determinant of cellular fate, but additional factors such as the type of interacting APC and the presence of costimulatory molecules or cytokines may also play a role for the fate decision of individual thymocytes (Figure 2).

1.1.2 Dominant Tolerance

Apart from so-called recessive tolerance induction, which was described in the previous paragraphs, i.e. clonal deletion, dominant tolerance mechanisms are crucial for the induction and maintenance of tolerance. Natural regulatory T cells (Treg), characterized by expression of the FoxP3 transcription factor, are generated in the thymus and can suppress activation, expansion and effector functions of autoreactive T cells in the periphery [22 23]. Therefore, dominant tolerance mechanisms are in part induced in the thymus (natural regulatory T cells) but excert their function in the periphery (Peripheral tolerance). (Peripheral tolerance mechanisms will not be discussed in this study.).

Tregs, making up 5-10% of the peripheral $CD4^+$ Tcell repertoire, are selected just like other T cell populations in the thymus upon TCR-MHC-self peptide interactions with an APC, thus their repertoire is similarly diverse as the conventional T cell repertoire. The avidity hypothesis holds that Treg are generated from high-avidity interactions with self-antigen, just below the threshold for negative selection [16, 24]. Recently, the TCR variable regions from Tregs and conventional T cells were sequenced in a model of polyclonal T cell receptors: TCR avidities as well as TCR repertoires between Treg and conventional CD4 T cells were found not to differ significantly [25, 26], which contradicts the avidity hypothesis. According to a recently proposed model[16], TCR avidity is not directly linked to Treg identity and other signals seem to play additional roles for the induction of the Treg lineage in a developing thymocyte (Figure 2). Additional signals relevant for Treg induction include signalling through CD40-CD154 which was shown to be indispenable in the induction process[27].

Lineage induction of natural regulatory T cells was believed to occur in the medulla for spatio – temporal reasons [28], therefore mTEC were hypothesized to be the primary Treg-inducing cells [29, 30]. Aschenbrenner et al. showed for the first time the generation of Treg directly by mTEC [31]. Today it is known that different APCs can be responsible for the induction of the Treg lineage: Thymic epithelial cells as well as different thymic DC-subtypes can efficiently induce Treg development of immature thymocytes, albeit strikingly different optimal doses of cognate antigen were needed in *in vitro* studies[32]. Different APCs can thus be involved in Treg induction, but the exact timepoint, the role of TCR avidity and the combination of factors necessary for the induction of the regulatory T cell lineage are still a matter of controversial debate.

1.2 Promiscuous gene expression

In order to render the T cell repertoire tolerant to most potential self-antigens, which a mature T cell may encounter later on in the body, a comprehensive repertoire of antigens is presented to developing thymocytes by APC, including spatially or temporally restricted antigens.

It was long believed that central tolerance mechanisms would only induce tolerance towards ubiquitously expressed antigens and antigens from the epithelial and lymphoid cell lineages, peripheral tolerance mechanisms were believed to account for tolerance towards spatially or temporally restricted antigens. This clear division of tolerance induction dependent on the site of

antigen expression was questioned in 1989, when Linsk et al. postulated the presence of a larger number of gene products, including tissue-restricted genes, present in the thymus [33]. In the following years, a number of experiments could show the ectopic expression of transgenes in the thymus, i.e. under the rat-insulin-promoter, which lead to tolerance induction [34 35 36]. Evidence for ectopic expression of insulin in wild type mice was first shown in 1994 [36]. The expression of many other promiscuously expressed antigens was discovered in the following years [37-49].
pGE is a characteristic feature of mTEC and was attributed to this cell type by gene expression analysis of highly pure TEC populations in mouse and human [50, 51]. Lower amounts of tissue-specific transcripts could be detected in other thymic APCs. Yet, mTEC show the highest degree of pGE, cTEC show reduced numbers of promiscuously expressed genes and pGE in thymic DC or MΦ is less pronounced.
pGE in mTEC is characterized as the low-level ectopic expression of tissue-restricted genes from virtually all tissues of the body [52]. mTEC are a highly hetereogeneous population with regard to ectopic gene expression as only 1-3% of all mTEC express a given antigen at the protein level [42, 50, 53-58].

1.2.1 Regulation of pGE at the cellular level

MTEC can be further subdivided into two subpopulations by their expression of costimulatory molecules (CD80, CD86, MHCII), which are upregulated upon mTEC maturation. The more mature population shows high expression of costimulatory molecules ($mTEC^{high}$) and expresses a larger and more diverse pool of antigens [52]. The Autoimmune Regulator (Aire), the only mTEC specific transcriptional regulator found to date to be directly involved in pGE affects the expression of 200-1200 genes [59] and is only expressed in $mTEC^{high}$.
MTEC differentiation has been a controversial process in the past years and was described by contrasting models. A number of recent studies on mTEC differentiation and turnover now speak in favor of the terminal differentiation model. The model postulates that 1) mTEC co-express an increasing number of genes from different tissues at the single cell level and do not emulate tissue-specific expression patterns. 2) $mTEC^{low}$ and $mTEC^{high}$ belong to a single lineage. $mTEC^{low}$ differentiate to $mTEC^{high}$ and concomitantly upregulate promiscuous gene expression. 3) High levels of pGE and Aire expression are features of of $mTEC^{high}$ only [17].
Ad 1) Every tissue restricted antigen (TRA) is expressed at the protein level on average by only 1-3% of complete mTEC and gene expression on the single cell level seems to underlie a stochastic mechanism. Individual $mTEC^{high}$ were found to co-express antigens from different tissues and expression patterns do not necessarily follow the same regulatory pathways as in the peripheral tissues [60] (L. Tykocinski, unpublished). This finding contradicts the idea of tissue emulation [61] by mTEC.

Ad 2) Thymic epithelial cells, long believed to be postmitotic, have now been shown to be an actively cycling population [62-65]. $mTEC^{high}$ were found to turn over in around 3 weeks[64] which concurs well with previous estimates[63]. $mTEC^{low}$ are heterogeneous with respect to mitotic activity and contain actively dividing as well as non-diving cells.

In the ontogeny of mTEC, mTEC^{high} appear at E16-17 concomitantly with the upregulation of pGE, two days after the appearance of committed mTEC ^{low}. The conversion of mTEC^{low} to mTEC^{high} has been shown to be dependent on cell division *in vitro* and *in vivo* [64, 65]. RANKL signals, provided by lymphoid-tissue-inducer cells (Lti), are essential for this conversion and for the emergence of Aire^{+} mTEC [66]. Lymphotoxin signals and CD40 ligand are likewise important for mTEC maturation; they seem to be primarily involved in expansion and/or homeostasis of postnatal mTEC populations but dispensable for the initial development of Aire^{+}mTEC^{high} [67, 68] (). Aire^{+} mTEC^{high} cells are postmitotic and were shown to undergo apoptosis within a few days [65]. Aire^{-} mTEC^{high} on the other hand were found to be strongly cycling and replenish the pool of Aire^{+} cells (Figure 3). These data strongly argue for the model of terminal differentiation as they show a precursor-product relationship from mTEC^{low} via Aire^{-}mTEC^{high} to Aire^{+} mTEC^{high} cells.
Ad 3) pGE increases concomitantly with mTEC maturation. Aire^{+} cells, being postmitotic, are thus the most mature cell type which has been shown to exhibit the maximum amount of pGE before undergoing apoptosis (Figure 3).

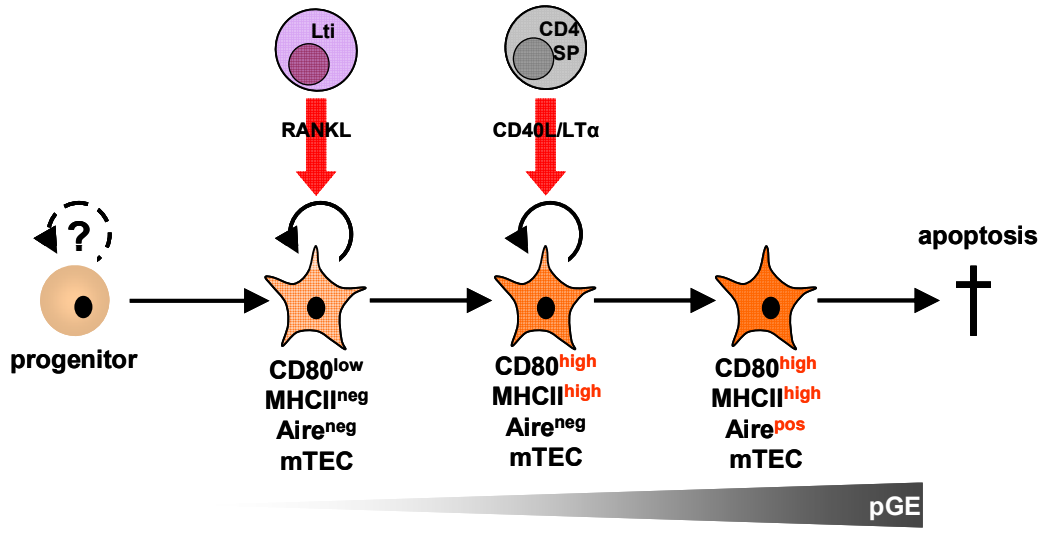


Figure 3
Development and turnover of different TEC subpopulations. MTEC, derived from a common TEC progenitor, develop into the mTEC lineage. Upon further differentiation, mTEC first receive critical RANKL signals from lymphoid tissue inducer cells (Lti) and later lymphotoxin and CD40L signals from CD4 SP thymocytes. Upon differentiation, mTEC upregulate CD80 and Aire protein. During the same differentiation process, the level of pGE increases and finally drives terminally differentiated Aire^{+} mTEC into apoptosis.

Aire$^{-/-}$ mice have increased numbers of mTEC^{high}, of which a subpopulation, just as the Aire^{+} mTEC in wt mice, is postmitotic. In Aire$^{-/-}$ mice this postmitotic population was shown to have a slightly longer half-life than the wt Aire^{+} mTEC population. Therefore, either an overload of the transcription machinery, instability of the genome due to genome-wide deregulation via an epigenetic mechanism [69] or even the Aire protein itself have been speculated to be implicated in the induction of apoptosis in postmitotic mTEC [65]. The possibility that Aire itself may induce apoptosis was supported by *in vitro* data and the identification of a domain homologous to the Apaf-1 CARD-domain present in the Aire protein [65, 70]. Which mechanism finally drives Aire^{+} mTEC into apoptosis still needs to be resolved.

Considering the short lifespan of just a few days of mature Aire^{+} mTEC^{high} [65] as the only cell type in the thymus producing Aire-dependent TRA, the system of antigen presentation to thymocytes, negative selection and Treg induction has to work extremely efficiently in order to ensure tolerance to TRA. Medullary thymocytes were recently shown to move with about the same speed as lymph node T cells which would allow for extensive scanning of DC and mTEC presenting self-peptides [17]. Rapid apoptosis and high turnover of mature mTEC may therefore provide a means to constantly feed TRA into the crosspresentation pathway of dendritic cells and thus spread the promiscuously expressed antigens. It is still unclear whether an individual Aire^{+} mTEC^{high}, once fully differentiated, expresses a single, fixed pool of promiscuous antigens or whether a single cell changes its expression during maturation, thus increasing the pool of expressed genes over time. Such a mechanism would introduce additional antigen diversity within a confined microenvironment.

1.2.2 Regulation of pGE at the molecular level

The scope of pGE has been characterized already in great detail, nevertheless the mechanisms regulating this type of gene expression are still largely unclear.

Genes expressed promiscuously can be classified into four pools which contain increasing percentages of tissue-restricted genes:

i) Genes expressed by cTEC and mTEC (i.e. PLP)
ii) Genes upregulated in mTEC (independently of maturation state) (i.e. CRP)
iii) Genes upregulated in mTEC^{high} (i.e. Csnb)
iv) Genes upregulated in mTEC^{high} and dependent on the transcriptional regulator AIRE (i.e. Insulin).

The expression of these pools is additive so that the terminally differentiated Aire^{+} mTEC^{high} can express genes from all four pool with some exceptions (i.e. Pth is expressed preferentially in mTEC^{low}) [17].

Although pGE can be divided into these distinct pools at the population level, the single cell level displays a completely different picture: It was found by protein staining that a particular antigen is only expressed in 1-3% of all mature mTEC [42, 50, 53-58]. Single cell analysis in the casein gene locus could show expression frequencies from approx. 2-15% of mTEC^{high} expressing a particular antigen. Csnb represented a singular exception with a frequency of over 80% [60].

Therefore, at least two levels of transcriptional regulation of pGE can be presumed: One at the level of the entire gene pool which sets the conditions necessary for the expression of entire antigen pools and a second layer of regulation which determines the expression of an individual antigen on the single cell level [52].

Aire plays a crucial role in the control of pGE as no other transcriptional regulator for this process has been identified to date: A large subgroup of promiscuously expressed genes, containing a particularly high percentage of tissue-restricted genes, was found to be dependent on Aire [52 71]. Absence or disruption of the Aire gene leads to defective central tolerance and results in humans in a rare monogenetic disease called autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED) with characteristic features of autoimmunity, primarily

hypoparathyroidism, primary adrenocortical failure and chronic mucocutaneous candidiasis [72]. Aire-KO mice show a very similar autoimmune disease where the severity and the affected organs are dependent on the mouse strain [73]. Expression of Aire is estimated to be indispensable for the expression of several 100 up to several 1000 promiscuously expressed antigens [74] but also has an effect on the maturation of mTEC: When the Aire gene was disrupted by a GFP-*knock in*, GFP^+ "wannabe" $Aire^+$ cells showed morphological changes and altered localization in the thymus [75]. $Aire^+$ $mTEC^{high}$ also express a high percentage of Aire-independent genes, thus clearly other mechanisms apart from Aire expression must have an influence on pGE [75]. Other transcriptional regulators may be involved which have not been identified up to now, but non-genetic or epigenetic regulation mechanisms have also been linked with the regulation of pGE [73] [17].

A number of examples of genes have been identified so far which are expressed independently of their tissue-specific transcription factors in mTEC. Regulation of these genes thus differs between the tissue and the ectopic expression in mTEC [60] [76] (L.Tykocinski, unpublished results). Whey acidic protein (WAP) for example is not co-expressed with its transcription factor Elf5 in single cell gene expression analysis in mTEC, although presence of Elf5 is indispensable for WAP expression in the mammary gland [77, 78] [79]. Different pieces of evidence indicate that pGE leads to stochastic expression patterns: mTEC were shown to use multiple alternative transcriptional start sites, partially different from those employed in the tissue in the cases of some antigens [76]. Single cell expression analysis proved pGE patterns to be overall stochastic and did not give evidence for particular preferential expression patterns [60, 76]. Along the same lines, single cells were shown to express particular promiscuously expressed genes either mono- or biallelically in a probabilistic way [76]. On the other hand, co-expression patterns, as for example for Muc1 and CEA, have also been observed by immunohistochemical analysis and SC PCR in human thymus[80] (S. Pinto, unpublished). Overall, regulation of pGE appears to be clearly different from gene regulation in the tissue and uses mechanisms allowing for stochastic expression patterns. This leads to the hypothesis that there might be a higher order mechanism of regulation which induces certain expression patterns but allows at the same time for stochastic fluctuation at the single cell level.

Epigenetic mechanisms have been implicated in the regulation of pGE by a number of studies but have not yet directly been proven to play a pivotal role. A role for epigenetic control is suggested by the following observations: (i) Promiscuously expressed genes are evenly distributed throughout the different chromosomes but highly clustered locally in clusters encompassing up to 100 genes [52] [59]. Clustering is a particular feature of pGE and is highly conserved over different species [51, 52] (S. Pinto unpublished). Genes within one particular cluster are not necessarily co-regulated but can be differentially regulated. Clusters may encompass different gene families all co-expressed in mTEC on the population level and may include Aire-dependent as well as Aire-independent genes [52].

ii) The imprinting status of Igf2 is lost in mTEC and Igf2 is expressed from both the paternal and maternal allele [52]. In imprinted genes typically only one allele is expressed (the paternal allele for Igf2) while the DNA of the other allele is methylated at CpGs and thus repressed [81].

(iii) Expression of certain TRA including cancer germ cell antigens (i.e. MAGE-A1 and –A3) correlates with promoter hypomethylation [17]. Likewise, this is the case for some other promiscuously expressed antigens (L.Tykocinski, unpublished results).
(iv) Aire was shown to preferentially bind non-methylated histone 3 at lysine 4 (H3K4me0) with its PhD finger, implying an effect of the state of histone lysine methylation on Aire activity and pGE[82, 83].

Altogether, this evidence lead to the following hypothesis: pGE is regulated at the level of gene clusters or groups of genes by epigenetic mechanisms. This will lead to the opening of larger stretches of chromatin in mTEC^{high} (in parallel with gene expression) or potentially already in mTEC^{low} (preceding gene expression). Once chromatin is accessible, Aire and potentially other factors may act and specifically turn on individual genes. This activation may be stochastic and thus guide different patterns of gene expression over time in individual mTEC [17], possibly dependent on the stoichiometry of transcription factors in each individual cell (Figure 4).

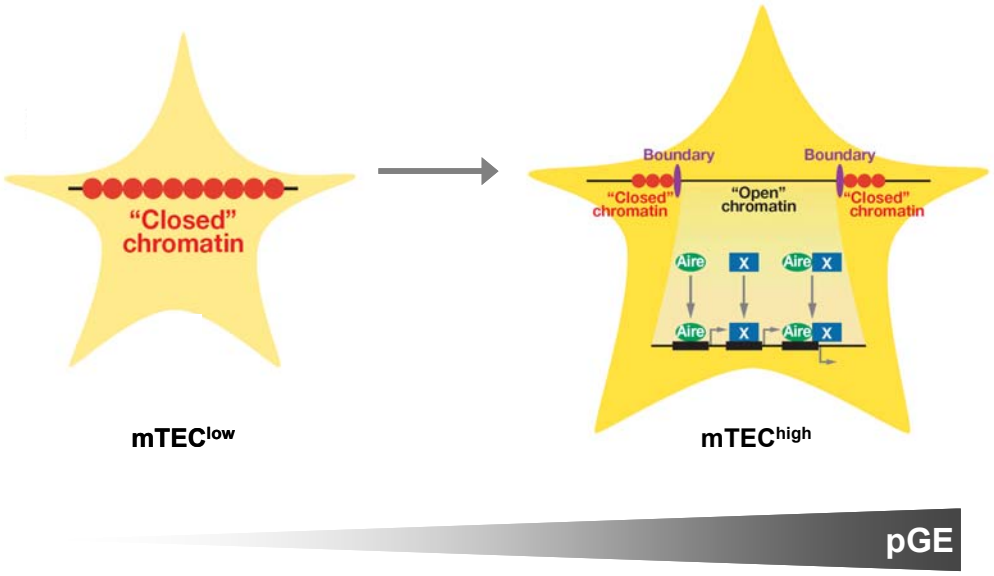


Figure 4
Model for clustered gene expression using epigenetic and genetic mechanisms leading to the full scope of pGE in mTEC^{high} according to the model of terminal differentiation (adapted from [17]). As a first layer of regulation, chromatin of larger gene clusters is made accessible either preceding gene expression – already in mTEC^{low} – or – concomitant with the full scope of pGE – in mTEC^{high}. Once chromatin is accessible, different factors can access and activate expression from individual genes, this pattern may vary from cell to cell according to the stoichiometry of transcription factors.

1.3 Mechanisms of epigenetic regulation

Epigenetic regulation of gene expression affects the local accessibility of the DNA for transcription factors and the transcription machinery. It comprises a number of different distinct marks and regulation mechanisms which could all potentially play a role in regulating promiscuous gene expression. Epigenetic marks can be passed on over multiple cell divisions, some modifications are even heritable.
The most frequently analyzed epigenetic modifications are DNA methylation at CpGs and the covalent modifications of histone tails. In comparison, other important mechanisms as the analysis of chromatin structure, packaging and localization in the nucleus are still a comparably young field of research where many aspects still require further clarification.
The major focus in this study will be placed on the modification of histone tails and the analysis of chromatin structure in the context of promiscuous gene expression, DNA methylation patterns in pGE have been analyzed in parallel in the laboratory (L.Tykocinski, unpublished results).

DNA in the nucleus is present in the form of nucleosomes which are connected by linker-DNA, a conformation which is named "beads on a string". A nucleosome consists of the DNA double helix wound around histone octamers consisting of two H2A, H2B, H3 and H4 proteins each. The "beads-on-a-string" conformation is the most accessible form of chromatin. Binding of H1 histones to the linker DNA results in a higher order chromatin structure, the so-called 30-nm-fiber, named according to its diameter. From the 30-nm-fiber onwards, chromatin can be packaged into larger, highly condensed structures of yet unclear organization. In the nucleus, one differentiates roughly two types of chromatin: Densely packed heterochromatin which is associated with repeat and satellite DNA and transcriptional inactivity and decondensed euchromatin which is associated with actively transcribed genes. Dense packaging of chromatin makes the chromatin fiber less accessible for transcription and thus correlates inversely with genetic activity. Different models have been proposed to explain higher order chromatin structure, but no definitive type of organization can explain the conformation of highly condensed chromatin to date.
Within the nucleus, chromosomes are localized into discrete "chromosome territories" which can be repositioned during cell division, thus the placement of chromosome domains may vary between different cell types [84, 85].

1.3.1 DNA methylation

Eukaryotic DNA contains about 2-7% methylated cytosines which are mostly part of a CpG-dinucleotide. Methylation of DNA is a very stable epigenetic mark generally associated with inactivation of gene expression and can lead to increased DNA condensation and heterochromatin formation. So-called DNA-methyltransferases (Dnmt) catalyze CpG methylation in mammals. Dnmt1 recognizes hemi-methylated DNA and can methylate the newly synthesized DNA strand according to the pattern on the parental strand, thus it is called

maintenance methylase [86]. As DNA methylation is preserved during cell division and DNA replication, it is a long-lived, heritable epigenetic mark. In contrast, DNA de novo methyltransferases Dnmt3a and Dnmt3b can methylate completely unmethylated double stranded DNA [87].

Several proteins have been implicated in DNA demethylation [88, 89], but a defined enzyme with DNA-demethylase activity has not been identified up to today. DNA demethylation can occur in a passive way, when Dnmt1 does not confer the methylation pattern to the newly synthesized DNA during replication [90]. DNA methylation inversely correlates with the expression of many tissue-restricted genes: It can inhibit transcription directly by precluding the binding of sequence-specific transcription factors or it can lead to recruitment of methyl-CpG binding proteins (i.e. the MBD or MeCP family). These will recruit further complexes and hinder transcription at target promoters [91].

1.3.2 The histone code

Histone tails can be modified in many different ways at varying amino acid residues, preferably at lysines. Due to the abundance of different modifications involved in the regulation of chromatin structure and recruitment of additional factors, crosstalk between modifications is very likely [92].

The most common marks are histone acetylation and histone methylation, but also other modifications such as phosphorylation or ubiquitination are possible [92]. Histone acetyltransferases (HATs) and Histone-deacetylases (HDACs) are the enzymes modifying the acetylation status of histone tails at distinct sites of histone 3 and 4 (H3: K9, K14, K18, K23 K27; H4: K5, K8, K12, K16). [93] [94]. Histone acetylation is a highly dynamic process which can change within minutes [92] and generally leads to a more open chromatin structure: On the one hand, acetylation of lysines at histone tails lowers the positive charge of the histone tails and thus decreases interactions with the negatively charged DNA which leads to chromatin opening. The acetylation of H4K16 was shown to have a particularly strong effect on the decrease of DNA-histone interactions [95]. On the other hand, histone acetylation can also recruit factors involved in active transcription or lead to the displacement of repressive complexes which cannot bind acetylated nucleosomes [96].

Histone methylation at the lysine residues of H3 can have different consequences [97]: H3K9 di- and trimethylation are marks characteristic for constitutive heterochromatin while H3K4 mono-, di- or trimethylation is associated with active chromatin. H3K4 trimethylation is characteristically found in the promoter regions, at highest concentrations at transcriptional start sites [97-100] and also in so-called bivalent chromatin regions. Such regions, carrying a combination of active (H3K4me3) as well as inactive (H3K27me3) marks, do usually not contain actively transcribed genes but are held in a bistable state, already poised or prepared for later expression [101, 102]. Bivalent chromatin is often found in ES cells controlling the expression of transcription factors involved in developmental processes [103-105], but was recently also oberserved under other conditions [102]. Unless found in a bivalent chromatin state, H3K27 trimethylation marks polycomb repressor binding sites [106] and is thus a mark for facultative heterochromatin.

The modification status of histone tails has recently been directly implicated in the regulation of promiscuous gene expression as two laboratories found the transcriptional regulator Aire to be specifically binding to non-methylated H3K4 (H3K4me0) with its PHD1 domain [82, 83]. This implies that Aire may act on the gene expression of those promiscuously expressed genes which do not carry active chromatin marks such as H3K4me3 and would thus, without Aire-binding, not be expressed.

1.3.3 The role of chromatin structure and nuclear organization

It is an established concept today, that gene expression is controlled at two levels: On a smaller scale, the binding of transcription factors and gene-specific epigenetic modifications is important, as described in 1.3.1 and 1.3.2. On a larger scale, the chromatin organization of larger gene clusters or regions and their location as well as accessibility seem to play a significant role [107]. The analysis of nuclear structure and chromatin positioning is a rather new and dynamic field of research, thus many hypothesis concerning gene positioning and organization in the nucleus have been put forward recently. It is now considered a fact that genomes are non-randomly organized in the nucleus of higher eukaryotes while the pathways and effects of organization often remain largely unclear [108]. Light microscopic studies have confirmed the so-called chromosome territory – interchromatin department (CT-ITC) model which holds that each chromosome occupies its own distinct territory in the nucleus. Active genes are believed to be more at the surface of chromosome territories, in the interchromatin departments [84, 109, 110]. Chromosomes are also thought to intermingle, but the role and extent of intermingling is yet unclear [111-113].

Since Croft et al. found the radial position of genes to be non-random and to play a role in gene expression [114], many studies have followed this concept and often found radial positions of genes and chromosomes to correlate with expression status. Whether this link is a causal one is still unclear, as most studies primarily focused on the identification of correlations, functional studies being rare [115].

Concerning the positioning of chromosomes, gene-rich chromosomes and early-replicating DNA were found preferentially in the interior of the nucleus, gene-poor chromosomes (and late-replicating DNA) more often near the nuclear lamina [115 85, 116]. Similar studies have been made identifying so-called "ridges ("regions of increased gene expression", containing transcriptionally active chromatin and housekeeping genes) near the nuclear center and antiridges (containing heterochromatin and genes transcribed at low-level) more peripheral [113]. Lately, studies in human fibroblasts rather found large chromosomes located centrally and smaller ones located peripherally [117, 118].

Apart from these general correlations of activity with radial position, individual genes may not position according to these schemes and may possibly be found anywhere in the nucleus independent of their transcriptional state [108]. Others investigated positioning of genes with respect to heterochromatic regions[85] or with respect to nuclear speckles [108], but radial position still remains the most frequently used parameter which has been shown to correlate with activity.

A number of individual genes have been found to move upon differentiation and transcriptional activation in a more central position such as the genes for c.maf in T cells, IgH in committed B

lymphocytes, Cftr in adenocarcinoma cells and Mash1 in neuronal cells [119-122]. On the other side of the coin, a number of publications also present genes which do not change position upon activation or move to the periphery [123-125]. Along the same lines, active sites of RNA polymerase II transcription as well as heterochromatic regions are found throughout the nucleus, which speaks against a clear differential localization of active and inactive chromatin regions [108].
Two recent key experiments could show a direct causal link between nuclear position and transcriptional status: Chuang et al. proved that a gene can be relocated upon transcriptional activation by the example of a transgene which moved away from the nuclear lamina towards the interior after binding a strong transcriptional activator [126]. In 2008, Reddy et al. could show that artificial tethering of multiple endogeneous genes to the nuclear lamina by a lamin-associated protein leads to an approx. twofold reduction in gene expression of specific genes [127]. Thus, causal relations between position and transcriptional activity can be present for individual genes in both ways: Gene expression changes may lead to a change in position [126] and vice versa it is also possible that positional changes may have an impact on gene expression [127].
Such functional experiments will need to be expanded in the future in order to get a better understanding of this two-way regulation.

Another parameter which crucially influences transcriptional activity on a higher level of regulation is the packaging density of chromatin in a particular region. Open, decondensed chromatin has been known to be associated with transcriptional activity from early examples of lampbrush chromosomes with looping out of extremely decondensed, active loci which was confirmed by ultrastructural analysis [128]. In the 1980's, several experiments with chicken erythrocytes showed that chromatin from the β-globin locus sedimented more slowly than bulk chromatin in expressing cells, but not in non-expressing cells, indicating a less compact chromatin structure in the actively transcribed locus [116]. The question to what extent chromatin decompaction can be induced locally has been pursued by recent studies: Upon induction of Hsp70 transgenes, Hu et al. found a large scale 1.5-3-fold decompaction of chromatin in a previously highly condensed area. Even stronger decompaction on a smaller scale in confined areas surrounding active Polymerase II is likely to occur as to allow sufficient access to the DNA [129]. Previously, the same group had identified visible large-scale chromatin decondensation of heterochromatin through an apparently transcription-independent mechanism by targeting VP16, an acidic activator to an artificial lac-operator construct inserted in a heterochromatic region [130]. In parallel, Chambeyron et al. discovered significant decondensation in the HoxB locus in development upon induction of transcription [131, 132]. These studies lead to the conclusion that transcriptional activation of individual genes or gene clusters can lead to large-scale physical decompaction of chromatin. The reverse conclusion however cannot be inferred, i.e a gene which is located in decompacted chromatin is not necessarily expressed: When compact and open chromatin fibers were separated by sucrose sedimentation, open chromatin fibers correlated directly with regions of highest gene density but not with gene expression. Actively expressed genes could also be found in gene-poor regions, thus in compact chromatin [116].

Although still controversially discussed, radial positioning as well as chromatin compaction are two important aspects which describe higher order chromatin structure and may have a strong effect on transcriptional regulation of gene regions and clusters.
These aspects have not yet been analyzed in the context of pGE. Yet, with their widespread effects on transcription these important mechanisms could play a role in pGE.

1.4 Objective of this study

This project was undertaken in order to analyze different molecular aspects of promiscuous gene expression in the thymus in more detail. A particular focus was set on the implication of epigenetic mechanisms in pGE.
This work was started with the goal to test the model of terminal differentiation as described in [17] with particular focus on the aspect of progressive epigenetic opening of a promiscuously expressed gene cluster. This model holds that entire clusters of promiscuously expressed genes are opened epigenetically in the whole mTEC population either preceding pGE (in $mTEC^{low}$) or concomitantly with pGE (in $mTEC^{high}$). Once an open chromatin configuration is reached in the whole locus, transcriptional regulators will act upon individual genes within the cluster leading to a different transcriptional pattern in every cell.

Epigenetic mechanisms had previously been implicated in pGE but have not yet been directly addressed in detail. A role for epigenetic control was suggested by the following observations: a) Promiscuously expressed genes tend to localize in clusters, suggesting epigenetic co-regulation ; b) the imprinting status of Igf2 is lost in mTEC; c) expression of certain TRA including cancer germ cell antigens and a gene in the casein gene locus, Csnb, correlates with promoter hypomethylation [52] [17] (L. Tykocinski, unpublished results); d) the transcription factore Aire was shown to bind selectively to unmethylated H3K4.

For the assessment of epigenetic implications in pGE, the casein gene locus was chosen as a typical locus of clustered TRA. Two different types of epigenetic regulation were studied: First, histone tail modifications which play a significant role in regulating the accessibility of genetic loci for the transcription machinery were analyzed via chromatin immunoprecipitation (ChIP). Second, the radial position in the nucleus as well as the local compaction of a locus are two measures of higher-order chromatin structure implicated in the overall accessibility of larger genomic regions for transcription. These parameters were analyzed in the casein locus with fluorescence in situ-hybridization and spectral distance microscopy (Fish/SPDM).
Both methods, ChIP and FISH/SPDM were performed by other groups with much larger amounts of material primarily from cell lines. Therefore, both methods needed to be first established in the laboratory for the small numbers of ex-vivo isolated mTEC, which represent a rare and heterogeneous cell population from the thymus.

Additionally, other aspects of regulation of pGE were assessed such as co-expression patterns, mono-or biallelic expression patterns and concordance between protein and mRNA expression frequencies on a single cell level.
For all corresponding parts of the study, the main focus was to compare regulation patterns observed in the tissue with those in mTEC as to understand which specific mechanisms differentiate the regulation of pGE from regular tissue-specific gene expression.
Deciphering the mechanisms which cause the very particular gene expression patterns we observe in pGE will lead to a better understanding of tolerance induction in the thymus.

2. Materials and Methods

2.1 Materials

2.1.1 Chemicals

Table 2.1: Overview of chemicals

Product	**Supplier**
Acetic acid, glacial	Sigma
Agarose	Invitrogen
Citrate, sodium salt	Merck
Chelex-100	Biorad
DMEM culture medium	Gibco
EDTA	Sigma
Ethanol, absolute, p.a.	Riedel-de Haen
Ethidiumbromide (10 g/l)	Roth
F12 medium	Biochrom
Fetal calf serum (FCS)	Biochrom
Fixogum	Marabu
Formaldehyde	Merck
Formamide	Merck
Formamide, deionized	Merck
Glucose	Merck
Glycerol	Merck
HCl, 1M	J.T.Baker
HEPES	Invitrogen/Gibco
Igepal® CA-630	Sigma
Kaiser's Glycerolgelatine	Merck
KCl	Sigma
Linear polyacrylamide (LPA)	Ambion
MACS beads, CD45 microbeads	Miltenyi
2-Mercaptoethanol	Invitrogen/Gibco
Methanol	Merck
$MgCl_2$	Sigma
$MgCl_2$ solution for SC PCR	Applied Biosystems
NaCl	Fluka
Percoll™	Amersham
10 x Phosphate Buffered Saline (PBS)	Biochrom

PFA (Paraformaldehyde)	Merck
PMSF	Sigma
Poly-L-Lysine, 0.01%	Sigma
ProLong® Gold	Invitrogen
Protease Inhibitor Cocktail Tablets "Complete"	Roche
Salmon Sperm Protein A Agarose slurry	Upstate (Millipore)
SDS, 10% solution	Biorad
Sodium acetate	Sigma
Sodium deoxycholate	Sigma
TetraSpeck™ beads, 0,2 µm	Invitrogen
Tissue-Tek®	Sakura
Tris, Trizma® Base	Sigma
Triton-X-100	Merck
Trypan blue	Merck
Tryptone	Roth
Tween 20	Merck
Yeast Extract	Roth

2.1.2 Buffers and commercial solutions

Table 2.2: Overview of buffers and solutions

Buffer/solution	**Experiment used**	**Content or supplier**
LB medium	cloning	yeast extract 5 g/l; tryptone 10 g/l; NaCl 10 g/l
20x SSC	FISH	3 M NaCl and 300 mM trisodium citrate; pH 7.0
1x TAE	electrophoresis	40 mM Tris-acetate; 1 mM EDTA; pH 7.5-8.0
ChIP lysis buffer	ChIP	1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0, Roche Protein Inhibitors
ChIP dilution solution	ChIP	0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.0, 167 mM NaCl, +Roche Protein Inhibitors
TSEI	ChIP	0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.0, 150 mM NaCl
TSEII	ChIP	0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.0, 500 mM NaCl
TSEIII	ChIP	1 mM EDTA, 10 mM Tris-HCl, pH 8.0, 1% Igepal 630,1% sodium deoxycholate, 0.25 M LiCl
TE	ChIP	10 mM Tris (pH 8.0), 1 mM EDTA
Trypan blue	Cell counting	0.2% Trypanblau (w/v) 150 mM NaCl pH 7.0
RPMI 1640	Cell preparation	Gibco
Collagenase digest solution	TEC preparation	0.2 mg/ml Collagenase Type IV; 10 mM HEPES; 2% FCS (v/v) in RPMI 1640-medium, pH 7.2-7.4

Collagenase/Dispase solution	TEC preparation	0.2 mg/ml Collagenase Type IV; 0.2 mg/ml Dispase Grade I; 25 µg/ml DNase; 10 mM HEPES; 2% FCS (v/v) in RPMI 1640-medium, pH 7.2-7.4
MACS buffer	TEC preparation	0.5% BSA (v/v); 5 mM EDTA in PBS, pH 7.2-7.4
Trypsin digest solution	MEC preparations	0.25% Trypsin (v/v); 0.2% EDTA in PBS, pH 7.2-7-4
Percoll/RPMI	MEC preparation	12.6 ml Percoll; 16 ml PBS pH 6.5; 1.6 ml 10x RPMI
DMEM/F12	Brain preparation	50% DMEM medium, 50% F12 medium
0.5% Trypsin 10x in PBS	Brain preparation	Gibco
0.01% Poly-L-Lysine	FISH protocol	Sigma Aldrich

2.1.3 Enzymes, Proteins

Table 2.3: Overview of Enzymes and Proteins

Product	**Supplier**
AmpliTaq Gold	Applied Biosystems
BSA	Sigma
Calf Intestine Phosphatase (CIP)	Fermentas
CD45 microbeads	Miltenyi
Collagenase type IV	Cell Systems
Dispase grade 1	Cell Systems
DNA Polymerase	Roche
DNAse 1 for digest solution	ICN
DnaseI	Roche
Fetal calf serum (FCS)	Biochrom
MuLV Reverse Transcriptase	Applied Biosystems
Power Sybr Green Master Mix	Applied Biosystems
Proteinase K	Invitrogen
RedTaqTM DNA-Polymerase (1 U/µl) incl 10 x Puffer	Sigma
Restriction enzymes (BamH1, Sac1, Kpn1, Pst1, Hind III) including digestion buffers	Fermentas
RNase Block Ribonuclease Inhbitor	Stratagene
SuperaseIN	Ambion
Superscript II	Invitrogen
T4 Ligase	Fermentas
T4gp32 protein	UBS
Trypsin for cell culture	Gibco

2.1.3 Antibodies, dyes

Table 2.4: Overview of Antibodies and dyes

Antibody/Epitope	Clone	Species/Isotype	Conjugate	Reference/Supplier
CD45	30-F11	rat IgG2b, κ	PerCP Cychrome	Pharmingen
CD80	16-10A1	hamster IgG2, κ	PE	Pharmingen
CDR1	CDR1	rat IgG2b	Alexa 680	[133]
FcR (FcγIII/IIR)	2.4G2	rat IgG2b, κ	-	supernatant from P. Altevogt, DKFZ Heidelberg
Gad65/67	polyclonal	rabbit IgG	-	Sigma
Gad67	1G10.2	IgG2a	-	Millipore
gp40 (Ep-CAM)	G8.8	rat IgG	Alexa 647 biotin	[134]
Hoechst 33342	-	-	-	Sigma
Ly51	6C3	rat IgG2a, κ	FITC	Pharmingen
Streptavidin	-	-	PE-cy7	Pharmingen
Anti-mouse	polyclonal	goat IgG	cy3	Dianova (Jackson)
Anti-rabbit	polyclonal	goat IgG	cy3	Jackson
ChIP antibodies:				
H3	polyclonal	rabbit IgG	-	ab1791, Abcam
H3K4me3	polyclonal	rabbit IgG	-	ab8480, Abcam
H3K27me3	polyclonal	rabbit IgG	-	07-449, Upstate
Pan-acetyl H4	polyclonal	rabbit IgG	-	06-866, Upstate
HA-probe (Y-11)	polyclonal	rabbit IgG	-	sc-805, Santa Cruz

2.1.4 Primers and (Oligo-) Nucleotides

All primers were HPSF-purified and obtained from MWG-Biotech.

2.1.4.1 Primers for conventional PCR

Table 2.5: Overview of primers for conventional PCR

Primer Name	Primer sequence 5'-3'	Annealing Temp
actin beta forward	TGGAATCCTGTGGCATCCATGAAAC	58°C, 40 cycles
actin beta reverse	TAAAACGCAGCTCAGTAACAGTCCG	
eGFP – II for (for genotyping)	CTGCTGCCCGACAACCA	62°C, 30 cycles
eGPP – II rev (for genotyping)	CCATGTGATCGCGCTTCT	

2.1.4.2 Primers for realtime PCR

Gene expression analysis:

Table 2.6: Overview of primers for gene expression analysis with realtime PCR

Primer Name	Primer sequence 5'-3'	Concentration(nM)
UbiquitinC forward	AGCCCAGTGTTACCACCAAG	300 nM/ 900 nM
UbiquitinC reverse	ACCCAAGAACAAGCACAAGG	
Aire forward	GTACAGCCGCCTGCATAGC	300 nM/ 900 nM
Aire reverse	CCCTTTCCGGGACTGGTTTA	
Gad67 forward = Gad67 1 new	CCTTCTTCAGGCTCTCCCGT	50 nM/ 50 nM
Gad67 reverse = Gad67 3 new	GTAGGGCGCAGGTTGGTAGT	
eGFP forward = Gad67 1 new	CCTTCTTCAGGCTCTCCCGT	50 nM/ 50 nM
eGFP reverse = eGFP3 new2	CTGAACTTGTGGCCGTTTAC	

ChIP:

Table 2.7: Overview of primers for ChIP (for realtime PCR)

Primer Name and distance from transcription start	Primer sequence 5'-3'	Concentration(nM)
Bmp6 +178F	TCTTCGGGCTTCCTCTATCG	900 nM / 900 nM
Bmp6 +250R	GCACCGACAGGATCTCCTTTT	
Csna-53F (= Csna4947F)	TCAGTGAGTTTAAATAGTCAAGGAGCAA	900 nM / 900 nM
Csna+28R (= Csna5028R)	CCCAAGAGTAGGGAGAGGGAAT	
Csnb – 191F	TGCCTTGTTTAATGTACCCTAGAATTT	900 nM / 900 nM
Csnb -106R	TCAATTCCAAGAAGTCTACGTGATTAG	
Csng + 710F	TCAAGAGCCATGCTAGGTAAAGG	900 nM / 900 nM
Csng +804R	GGCTCTAAACACTCTGGCTATACCA	
Csnd -94F	CCACACAGTTGCCATATCAGATAGA	900 nM / 900 nM
Csnd +25R	TATGTAGAGGTTACCTTGCTTCCTAAGTT	
Csnk-1625F	GAATGTCCCATGCCAATAGCA	900 nM / 900 nM
Csnk-1552R	CCAGTCACTTCTACCACGACACA	
Gad67eGFP+1923F	GCTACCCCGACCACATGAAG	900 nM / 900 nM
Gad67eGFP+2036R	CGGCGCGGGTCTTGTA	
Gad67Promotor-545F	CGCTTTGCCTGTACAGCCATA	900 nM / 900 nM
Gad67Promotor-456R	GTTGTCCTGGATTGGCGATT	
Gad67+1637F	TGAACCGTAGAGACCCCAAGA	900 nM / 900 nM
Gad67+1709R	GCGTTCGAGGAGGTTGCA	
Hoxc10 +5F	CCTCCGCTGTAGTATTGCTCCTT	900 nM / 900 nM
Hoxc10 +86R	CGCGTACGAGTTCGGAGTTAC	
Muc10 -206F	GCAGAGTTCCAACCTGCTAGACA	900 nM / 900 nM
Muc10 -140R	GCCCTCACCTTGGCAATTC	
CD45 -393F (=Ptprc 4607F)	TTCAGAGCCTCGTACCAGCTTAG	900 nM / 900 nM
CD45 -319R (=Ptprc 4681R)	CAAGCTTCCACATGAGTATAACAACA	
Smr1 +9F	AATAACTGACCAGACCGCTTCTG	900 nM / 900 nM
Smr1 +80R	TCCCAGGAACCTACCTTGGA	
Sult1d1+74F	TGAGCTACCCAGGGACATCATC	900 nM / 900 nM

Sult1d1+58R	AATTCTCGGAGCTTTAAATGAGTTTT	900 nM / 900 nM
Sult1e1 -429F	TGTGCTCTTGTTCCATGTTCTTAACT	900 nM / 900 nM
Sult1e1-340R	GCTGATGTTACAAACAATAGCTTTGG	
Ugt2a3 +64F	GGCCGGCTGTGGATTCT	900 nM / 900 nM
Ugt2a3 +131R	TTAGATTCAGCCAGTGGCTCATAT	

SC PCR

Table 2.8: Overview of primers for SC PCR (for realtime PCR)

Primer Name	**Primer sequence 5'-3'**	**Primer type (ABC)**
Alt EpCAM 1	GAGGCGTTCACATCTCGATA	A
EpCAM 2	ACAAGACGACGTGGACATAG	B
Alt EpCAM 3	GGGTGCCTTTTCATCAACGT	C
Aire 1	GATGTGGACCTAAACCAGTC	A
Aire 2	AGCAACTCTGGCCTCAAAGA	B
Aire 3	CTGGATAAGGATCCCTTCCA	C
Gorasp2 1	TACAGGAAAACTCCCCAGGA	A
Gorasp2 2	GAAGGATCTGCTGAAAGCCA	B
Alt Gorasp2 3	GACGGTATCAGCTCCAATGA	C
Rik 1	CTTAACAAACCAGCAGTCCG	A
Rik 2 alt	GCAGGTGAAACAAGCAAGGA	B
Rik 3	TCTGCTCTCAGGAATTCTGC	C
Gad67 1 new	CCTTCTTCAGGCTCTCCCGT	A
Gad67 2 new	GCTGAACCGAGCCTGTTCCT	B
Gad67 3 new	GTAGGGCGCAGGTTGGTAGT	C
eGFP3 new2	CTGAACTTGTGGCCGTTTAC	C
Gad67 1 new and Gad67 2 new were also used as primers A and B in combination with eGFP3 new2 (Primer C).		

2.1.4.3 DNA, Nucleotides and Oligonucleotides

Table 2.9: Overview of DNA, Nucleotides and Oligonucleotides

Product	**Supplier**
10 mM dNTP	MBI Fermentas
2 mM dNTP	MBI Fermentas
Alexa 647 – OBEA –dCTP	Molecular Probes (Invitrogen)
Bac clone RP23-110B6 (in pBACe3.6)	RZPD
dATP, dTTP, dCTP, dGTP	GE Healthcare
dNTP mix 10 mM, 2.5 mM each	Applied Biosystems
Mouse COT-1 DNA	Invitrogen
Oligo (dT)7 Primer GCATTAGCGGCCGCGAAATTAATACGACTCACTATA GGGAGA$(T)_{21}$(AGC)	MWG Biotech, according to [135]
Oligo(dT)20 (500 µg/ml)	DKFZ Heidelberg
Oregon Green-dUTP	Molecular Probes (Invitrogen)

pUC19	Fermentas
Salmon Sperm DNA	Invitrogen

2.1.5 Commercial Kits

Table 2.10: Overview commercial kits

Product	**Supplier**
Alexa 647 Protein labelling kit	Invitrogen
DNeasy Blood & Tissue Kit Alexa 647 Protein labelling kit	Qiagen
EZ-Link Micro Sulfo-NHS-Biotinylation Kit	Pierce
High Pure RNA Isolation Kit	Roche
HiSpeed Plasmid Midi Kit	Qiagen
QIAEX II Gel Extraction Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
QIAquick PCR Purification Kit	Qiagen

2.1.6 Mice, cell lines, bacteria

C57BL/6 mice were bought from Charles River Wiga (female, 4 weeks old) and were used for organ preparations at the age of 6-8 weeks. For the isolation of mammary glands, C57BL/6 mice from inhouse breeding facilities were used directly after weaning or alternatively, mother mice were separated from their pups 2-5 days after birth.

Gad67/eGFP heterozygous mice were kindly provided by M. Götz from Munich and bread in inhouse facilities subsequently. Gad67/eGFP mice were used for organ preparations at the age of 6-12 weeks.

EL4 cells, a murine thymoma cell line, were cultured in RPMI 1640 supplied with 5% FCS in an incubator with 5% CO_2.

TOP10 E.coli strain was used for heat shock transfections.

2.1.7 Consumables

Table 2.11: Overview of Consumables

Product	**Supplier**
ABI PRISM 96-well Optical Reaction Plate und Optical Caps	AbGene
Cell-Strainer-Cap-Tubes (5 ml)	Becton Dickinson
Centrifuge tubes (15 ml and 50 ml)	TPP
Chromatography columns (for ChIP washing)	Biorad
Coverslips, all sizes	Lankenbrinck
8-tube strips, Ultra Clear Cap Strips	Abgene
Filter tips (10 µl und 1000 µl)	ART
Filter tips (20 µl und 200 µl)	Nerbeplus
Gauze (PA-60 Nybolt)	Eckert
Glass slides (Histobond)	Marienfeld Laborglas
MACS LS-columns	Miltenyi
MicroAmp Optical Adhesive Film	Applied Biosystems
MicroSpin™ G-50 columns	Amersham
Multi- pipets (0.2 ml, 2.5 ml, 5 ml)	Eppendorf
Pasteur pipets	WU-Mainz
PCR reaction tubes (200 µl)	Biozym
Petri dishes (all sizes)	TPP
Round bottom tubes, 13 ml	Becton Dickinson
Round bottom tubes, 25 ml	Rundbodenröhrchen
Safe-Lock reaction tubes (0.5-2ml)	Eppendorf
Sterile filters (0.2 µm and 0.45 µm)	Millipore
Syringes (10 and 50 ml)	Terumo
Thermo-Fast® 96 Detection Plate	Abgene

2.1.8 Equipment

Table 2.12: Overview of Equipment

Product	**Supplier**
Bacteria incubator	Memmert
Bacteria shaking incubator	Edmund Bühler GmbH
BD FACSAria™ Cell Sorter	Becton Dickinson
BD FACSVantageIITM Cell Sorter	Becton Dickinson
Bioruptor™	Diagenode
Centrifuge Rotatanta 460 R	Hettich
DNA Engine Dyad Peltier Thermal Cycler	MJ Research
Feinwaage CP224S	Sartorius

Gel electrophoresis chamber	AGS
GeneAmp®7300 Sequence Detector	Applied Biosystems
Incubator HeraCell 240	Kendro
LEICA CM 1900 Cryostat	Leica
Magnetic stirrer (MR 2000)	Heidolph
Nano drop® photometer	Thermo Scientific
Quadro MACS® separator	Miltenyi
Spinning disc confocal on Nikon TE2000 inverted microscope with Hamamatsu EM-CCD high sensitive CCD (black and white) camera	Nikon
Tabletop centrifuge Biofuge fresco	Heraeus
Tabletop microscope Axiostar Plus	Zeiss
Zeiss Axio Imager.Z1 fluorescence microscope	Zeiss

2.1.9 Software

Table 2.13: Overview of Software used

Software	**Company**
AxioVision 4.5	Zeiss
Amplify 3.1.4	freeware
BD FACSDiva	Becton Dickinson
ChromasLite	freeware
EditSeq 5.06	DNAStar
FlowJo	Tree Star
GeneAmp 7300 SDS Software	Applied Biosystems
LinRegPCR	LinRegPCR
MatLab	The MathWorks, Inc.
Microsoft Office	Microsoft
MegAlign	DNAStar
Primer ExpressTM 1.0 Applied Biosystems	Applied Biosystems
Primer ExpressTM 2.0 Applied Biosystems	Applied Biosystems
Ultraview ERS	Perkin Elmer

2.2 Methods

2.2.1 Distance measurements using fluorescence – in situ – hybridization (FISH)

The goal of this technique is to analyze the chromatin structure of a target region in an intact cell nucleus. In order to do so, two target regions close to each other are labeled in the nucleus with fluorescent probes. The distance between these probes as well as their distance to the nuclear rim is measured at high resolution with a fluorescent confocal microscope, followed by digital image processing.

High – resolution confocal laser scanning microscopy (CLSM) and labeling of genomic regions with FISH [136] are a method to analyze chromatin structure in an intact cell nucleus on the level of chromosomes and smaller genomic regions [137]. With this method, the resolution is limited to around 300 nm laterally and 700 nm axially [138], as preparations of cell nuclei, being usually highly heterogeneous, are not optimal optical objects and thus cannot be analyzed with the theoretical resolution limit of diffraction [139].

The optical resolution can be estimated as the full width at half maximum of the point spread function (PSF) of the optical system used. The PSF describes how a point object is imaged by an optical system, it includes several parameters which have an effect on the image produced, i.e. aberration by lenses, chromatic shifts, refractive index of the embedding medium and the sample. The PSF can be determined experimentally using small fluorescent objects e.g. beads. With the limits of resolution of conventional high resolution far field microscopy, measurements of sub-chromosomal regions in a nucleus would not be possible. But measurements of significantly smaller distances can be performed nevertheless, as long as the two objects differ in their spectral signatures (i.e. different absorption/emission spectra) with a method called Spectral Precision Distance Microscopy (SPDM) [140-143].

In SPDM, the maximum of intensity in the diffractive image of a point like object is expected to represent the geometrical localization of the object. For high quality objectives and negligible polarisation effects, the maximum of intensity should colocalize with the barycenter of the (diffraction limited) image. (The barycentre of an object takes into account both the size and the fluorescence intensity distribution of the object.) These bary-centers can be determined with digital image processing with an accuracy much better than the optimal resolution of modern microscopes. When fluorescent beads are used in CLSM, the position of the beads could be determined as exact <5 nm in lateral dimension and < 50 nm axial [144].

In theory, the principle of SPDM as shown in Figure 5 can be applied to any conventional high resolution far field microscopy method: When two small objects with the same spectral signatures are very close to each other, they cannot be resolved. For two small objects with different spectral signatures (i.e. two different fluorochromes) the diffraction images can be acquired independently of each other and their positions determined independently with high precision. In order to do this, careful calibration of chromatic shifts in the system is crucial because the accuracy of the distance between the two objects is given by the accuracy of the individual positions and the accuracy of the calibration of the microscope (microscope frame

calibration = chromatic shift). When chromatic shifts are measured using multispectral objects (i.e. TetraSpeck™ beads) under the same conditions as the FISH probes, the expected error of the calibration should be similar to the expected error of the positioning for each point in the coordinate system. The average error for measurements at different coordinate points will be slightly larger. It has been shown that measurements in 3D with particles of differing spectral signatures were possible with errors of 50-60 nm (between individual measurements) [144, 145], the minimal distance which can be resolved in SPDM in 3D is thus slightly larger than 50-60 nm. Fluorescently marked genes fulfil the requirement of being point-like objects in the nucleus as their dimensions are much smaller than the full width at half maximum of the corresponding PSF.

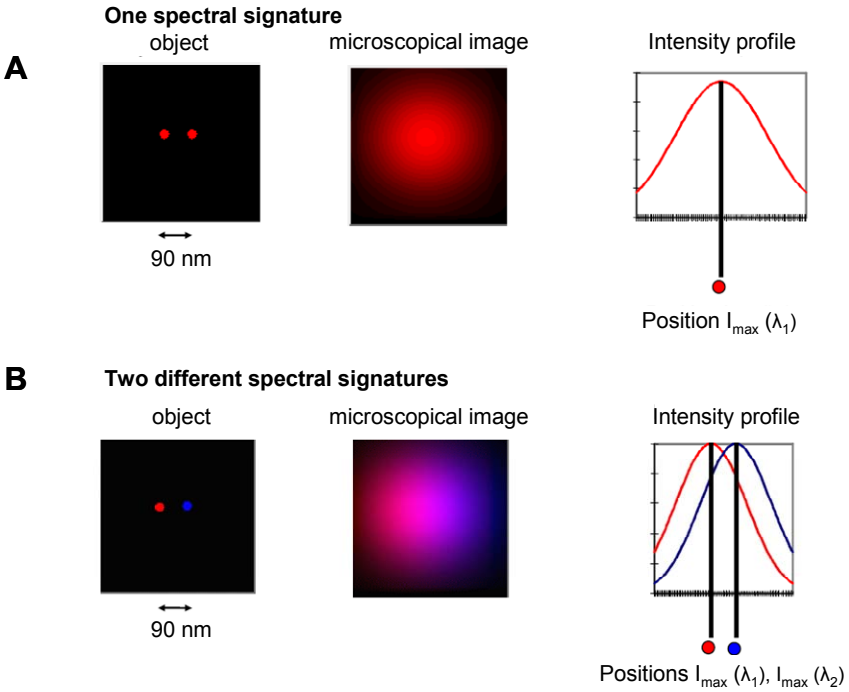


Figure 5
Principle of Spectral precision distance microscopy (SPDM). Two fluorescent signals are within 90 nm distance, a distance much smaller than the wavelength emitted. If both signals are marked with the same type of fluorophore (A), a point spread function (PSF) with a full width at half maximum of 250 nm will result in a microscopic image as shown. Only one wide red signal can be observed, also described by the intensity profile on the right. If the two objects are marked with different fluorophores (B), one can observe two different signals closer to each other than the resolution of the microscope. If both spectral signatures are detected independently by the microscope, it is possible to create two independent intensity profiles with clearly distinct bary-centers indicating the localization of the objects. This allows distance measurements under the optical resolution limit of the microscope. (given values for the figures: optical resolution = 250 nm; positioning error of I(λ1) and I(λ2) is 20 nm. Adapted from [146].

2.2.1.1 Construction of probes

In order to construct two probes of desired length in the target region, BAC clone "RP23-110B6", a 220 kB BAC clone encoding the genomic region around Csna/Csnb, was digested with different restriction enzymes to fragments which could be combined to make up the sequence target probes. The fragments were cloned into pUC19 and transfected into E.coli. The plasmids

expressing different fragments of the probes were prepared on a large scale, isolated, digested, fragments were purified via an Agarose gel and combined to make up the target probe DNA.

Digestion of BAC clone

20-30 µg BAC DNA was digested in a 100 µl volume for 4 h with 5 µl target enzyme in the adequate enzyme buffer at 37°C. The enzyme was inactivated according to manufacturer's specifications and the BAC clone together with appropriate markers loaded onto a 0.8 % agarose gel. Electrophoresis was done at 35 V over night (12-16 h). The desired fragments were excised with a clean scalpel, transferred to Eppendorf tubes and purified. Fragments larger than 10 kB were purified with *QIAEX II Gel Extraction Kit*, fragments smaller than 10 kB with *QIAquick Gel Extraction Kit*. Both kits were used according to manufacturer's specifications, elution was done in 20 µl.

Table 2.14: DNA fragments cloned for FISH probes

Fragment size (bp)	Digestion Enzyme	Fragment size (bp)	Digestion Enzyme
Probe 1:		*Probe 2:*	
9955	Hind III	14044	Kpn1
8018	Bam H1	11101	Kpn1
4727	Bam H1	8780	Pst1
4255	Sac1	8525	Pst1
3066	Sac1	3494	Pst1
1342	Sac1	2320	Pst1

Ligation

3-7 µl of eluted fragment and 1 µl of 1 µg/µl pUC19 vector (cut with restriction enzymes and dephosphorylated with CIP (Fermentas; according to manufacturer's specifications) were combined with 1 µl of T4 Ligase (Fermentas) and 1 µl of 10x Ligation buffer (Fermentas). Ligation was done at RT for 1-2 h.

Heat shock transfection

Complete 10 µl of ligation mix was combined with 200 µl TOP10 competent bacteria. For better efficiency with larger fragments (> 10 kB), competent bacteria were prepared fresh (with $CaCl_2$ method, see "Current Protocols in Molecular Biology", 2007, Wiley VCH), for all other fragments, bacteria stored at -80°C were used. Bacteria/plasmid mixture was incubated at 42°C for 2 min to allow the bacteria to take up the plasmid, then, 1 ml of 37°C warm LB medium was added and bacteria were incubated at 37°C with shaking for 1 h. Bacteria were pelleted carefully, resuspended in 200 µl of LB, plated onto agar plates with 100 µg/ml ampicillin and left at 37°C over night.

Preparation of target bacteria stocks

The following day, clones were picked and incubated with 1.5 ml LB medium at 37°C, in a shaking incubator for 16 h. Plasmids were purified with *QIAprep Spin Miniprep Kit* and a control digest was performed (2 µl of enzyme (Fermentas), 2 µl of 10x reaction buffer, 5 µl of miniprep-DNA, 11 µl of H_2O, 2 h at 37°C), analysed on a 1% agarose gel, checked for the proper fragment size and sequenced in house. From the correct bacterial clones, glycerol stocks were prepared: 10

ml of bacterial culture was pelleted and resuspended in 1 ml LB medium with 50% glycerol, frozen and stored at -80°C.

Preparation of DNA for FISH probes

Bacteria containing the target fragments were grown to high density in 200 ml LB medium and Plasmid DNA was prepared using *HiSpeed Plasmid Midi Kit* (Qiagen). Plasmids were digested in a preparative digest in 50 – 100 µl volume, depending on the amount of DNA digested. Digests were carried out as before in adequate reaction buffer (Fermentas), using 1/10 of volume Restriction enzyme. Samples were run on an agarose gel, fragments excised from the gel, purified with *QIAEX II Gel Extraction Kit* for fragments larger, *QIAquick Gel Extraction Kit* for fragments smaller than 10 kB. DNA concentration was measured. All fragments of each FISH probe were combined, taking into account the proper ratios between the fragments depending on the fragment size. A solution of 0.1 mg/ml DNA was obtained for each FISH probe.

2.2.1.2 Preparation and labelling of Probes for FISH

Probe DNA generated as in 2.2.1.1 was fluorenscently labeled using *Nick Translation*. In Nick Translation, low concentration of DNAse I introduces single strand breaks into the probe DNA, "nicks". DNA Polymerase exonuclease and polymerase activity leads to synthesis of a new DNA strand and thereby fluorescent nucleotides from solution are integrated into the probe. Successful nick translation yield fluorescently labeled probes of approx. 200 bp in size. As the two probes were too close to each other to be labeled with the same color, probes were labeled with two fluorochromes which have completely non-overlapping excitation and emission spectra. In this study, Oregon Green-dUTP (green) and Alexa 647 – OBEA –dCTP (dark red) (Molecular Probes, Invitrogen) were used.

Table 2.15: Components of Nick translation cocktail

Nick translation cocktail	
1 µg probe DNA (0.1 µg/µl), Probe 1 or Probe 2	10 µl
0.5 mM each dNTP mix (dATP, dCTP, dGTP for Probe 1; dATP, dTTP, dGTP for Probe 2)	5 µl
1 mM fluorochrome (Oregon Green-dUTP for Probe 1; Alexa 647 – OBEA –dCTP for Probe 2)	1 µl
10x Nick translation buffer (0.5M Tris-HCl pH 8; 50 mM $MgCl_2$; 0.5 mg/ml BSA)	5 µl
0.05 M β-Mercaptoethanol	10 µl
DNA polymerase I	1 µl
DNAse I (1:2500 diluted)	3 µl
H_2O	15 µl
total	50 µl

The Nick translation cocktail was incubated for 2 h at 15°C, then put on ice. 5 µl were analyzed on an agarose gel for the correct probe size.

15 µl of Probe 1 and 2 each were combined with 2 µg of Salmon Sperm DNA and 12 µg of mouse COT-1 DNA. 0.1x volume Sodium Acetate (3M) and 2.5x volume of 100% -20°C Ethanol were added and DNA was precipitated over night at -20°C. The next day it was centrifuged, washed with 70% Ethanol, dried for 5 min at 42°C to remove residual ethanol and

dissolved in 4 µl of deionized formamide for 4-6 h at 42°C (until the pellet was dissolved). 4 µl of 30% dextrane sulfate were added and the probes stored at -20°C for hybridization.

2.2.1.3 Cell fixation

The aim of fixation is on the one hand the optimal preservation of the nuclear structure with minimal loss of chromosomal DNA, while on the other hand, accessibility of the probe to the nucleus has to be guaranteed. All cells in this study were fixed with PFA which induces crosslinking of intracellular material and ensures optimal preservation of structure. Methanol/acetic acid which leads to precipitation of protein from the cell was only used for chromosome spreading in metaphase chromosome preparation.

Protocol for PFA fixation of cells:

Slides were coated with 0.01 mg/ml Poly-L-Lysine, washed 2x in water and dried. Target cells ($5x10^5$ in 300 µl 50% RPMI/ 50% FCS) were dropped onto the coated slides and left at 37°C for 45-60 min to settle down and adhere. The slides with adherent cells were washed 1 min in 0.3x PBS and fixed 10 min in 0.3% PBS/ 4% PFA. After 3x 5 min washing in 1x PBS, the nuclear membrane was permeabilized for 20 min in PBS/0.5% Triton-X-100 and slides were washed in 1x PBS 3x 3 min. The samples were immersed in PBS/20% Glycerol as a cryopreservative and left to equilibrate for 1.5 h at RT or over night at 4°C. Samples were subjected to 4 freeze/thaw cycles in liquid nitrogen, then washed 3x 5 min in PBS and incubated for 7.5 min in 0.1 N HCl.
After washing 3 x 5 min in 2x SSC, slides were equilibrated in 50% Formamide/2x SSC pH7 over night and stored in this solution at 4°C for up to three months.

2.2.1.4 Metaphase preparation

EL4 cells were grown up to confluency in 1640 RPMI/5% FCS, then treated with 0.1 mg/ml colcemid for 1h in order to arrest the dividing cells in metaphase and increase the percentage of metaphase chromosomes in the preparation.
Cells were pelleted and treated with hypotonic 75 mM KCl solution for 10 min at 37°C, pelleted carefully for 10 min at 200x g and resuspended drop by drop under constant shaking in 5 ml ice cold (-20°C) methanol-acetic acid solution (3:1). Methanol/acetic acid solution was added up to 12 ml, cells were pelleted again and the resuspension repeated 2-3 x until the pellet became translucent. The cell solution in methanol-acetic acid can be stored at -20°C up to 3 months.
15 µl of the cell suspension (after adequate cell concentration had been checked on a test slide) were dropped onto wet slides, left to dry, washed 5 min in PBs, incubated in 0.01 N HCl at 37°C for 1 min, washed 3x 3 min in PBS and subjected to an increasing ethanol series to dehydrate the cells (70%, 90%, 100% Ethanol for 3 min each). Slides were either dried and denatured directly for hybridization or stored up to one week in 70% Ethanol at 4°C.

2.2.1.5 Fluorescence-in-situ-hybridisation (FISH)

For FISH, both the DNA in the cell nucleus as well as the probe DNA need to be denatured, combined and renatured together. During the renaturation step, single stranded probe DNA can bind to single strands of the genomic DNA.

Metaphase preparations were denatured for 5 min in 75°C 70% Formamide/2x SSC pH 7, dehydrated by an increasing ethanol series at -20°C (70%, 90%, 100% ethanol, -20°C for 3 min each) and dried. PFA-fixed samples were denatured for 3 min in 72°C 70% Formamide/2x SSC pH 7 and hybridized directly. During the entire preparation of PFA-fixed nuclei it is crucial to keep samples wet at all times as drying will lead to a collapse of the 3D architecture of the nucleus.
Probes were heated for 7 min at 78°C and partially renatured for 20 min at 37°C. Then, probes were dropped onto 18x18 mm cover slips, combined with the denatured cell nuclei on slides, sealed with rubber cement (Fixogum) and left to renature for 2-3 nights in a moist chamber at 37°C. After renaturing, cover slips were removed and excess probe washed off: 3x 5 min in 4x SSC/0.5% Tween20 at 42°C followed by 3x 5 min stringent wash in 1x SSC at 60 °C. Slides were dipped again in 4x SSC/0.5% Tween20, equilibrated in PBS for 5 min and stained with HOECHST 33342 (1:4000 of 4 mg/ml stock) for 2-3 min. Slides were washed 2x 3m in in PBS and embedded in 15 µl ProlonGold (Invitrogen) embedding medium and 22x22 mm coverslips. Alternatively, calibration objects were added to the slides (as described in 2.2.1.6) before embedding. Prolon Gold was left to harden over 2 nights at 4°C before images were taken.

2.2.1.6 Measurement of chromatic shifts

As mentioned in 2.2.1, correction of systematic imaging errors, in particular correction of chromatic shifts, is crucial for the successful application of SPDM measurements. In order to correct for the chromatic shift, point-like multispectral objects are needed which can be measured under the same condition as the biological sample. In this study, 0.2 µm Tetraspeck™ (Invitrogen) beads fluorescent in four channels (UV, green, red and dark red) were used as calibration objects. As the only specimen on the slides fluorescent in the red channel, signals from beads can be clearly distinguished from probe signals.
After staining with HOECHST 33342 and PBS washes, Tetraspeck™ beads, diluted 1:300 and sonicated previously for 5 min, were incubated for 5 min on the slides. Slides were dipped in PBS and embedded as described in 2.2.1.5. This leads to adhesion of beads directly to the sample, beads can be acquired simultaneously with probe signals under almost identical conditions which leads to an optimal correction of the chromatic shift.

2.2.1.7 Image Acquisition

Z-stack images were aquired at the Nikon Imaging Center, Heidelberg, with a PerkinElmer Ultra-View spinning disc confocal on Nikon TE2000 inverted microscope with Laser lines 405, 440, 488, 514, 568, 640 and a Hamamatsu EM-CCD high sensitive CCD (black and white) camera, no binning was used..
For all acquisitions the 100x oil objective was used. Only the center quarter of the field of view was acquired as the chromatic shift was found to be smallest in this region (see 3.4.2). Images were acquired and saved using PerkinElmer UltraView software. Whenever possible, only cells in close proximity to a fluorescent bead were acquired.

2.2.1.8 Image processing

The image processing software used in this study is a custom made program by D. Baddeley on the Matlab platform as it was used in [147].

Briefly, it performs the following actions:

- z-projection of image data
- localization of cell nuclei via channels indicated as "nuclear counterstain" by the user
- filtering through a bandpass filter to suppress noise and background (leaving point-like objects only) according to a threshold defined by the user
- fit of the local maxima according to a multipoint model
- calculation of the bary center of intensity in voxels, conversion into μm
- localization of bary centers with respect to counterstained objects and calculation of distance from the rim of the counterstain

The localisation of all points found in μm is then used to calculate the distance between two points in Microsoft Excel.

2.2.1.9 Distance measurements

Each point analyzed was routinely checked for the quality of its fit. As the genomic distance between two points is small compared to the cell nucleus, there was no ambiguity concerning the identification of probe pairs belonging together on the same chromosome. For two points (x_{green}; y_{green}; z_{green}) and ($x_{d\text{-}red}$; $y_{d\text{-}red}$; $z_{d\text{-}red}$) the distance between the two points is determined as follows:

$$d=\sqrt{(x_{green} - x_{d\text{-}red} - \Delta c_x)^2+(y_{green} - y_{d\text{-}red} - \Delta c_y)^2+(z_{green} - z_{d\text{-}red} - \Delta c_z)^2}$$

(Δc_x; Δc_y; Δc_z) represents the chromatic shift, so the distance measured between the green and the dark red point coming from the Tetraspeck™ bead.

2.2.1.10 Localization of probes within the nucleus

It is also possible to use SPDM measurements together with a cellular counterstain (here: Hoechst 33342) to localize a probe within a stained nucleus. For these measurements, no correction of chromatic shifts is necessary as the chromatic shifts are small compared to the size of the cell nucleus. The digital image processing programme calculates directly the absolute distance of a probe from the rim of an object in μm or the normalized distance on a scale from 0=at the nuclear rim to 1=in the center of the respective nucleus.

2.2.2. Agarose gel electrophoresis

1 – 1.8 % horizontal agarose gels were used to separate DNA fragments coming from restriction enzyme digests, PCR products or sheared genomic DNA by size. Adequate amounts of agarose were boiled up in 100 ml of 1x TAE buffer and 3 μl ethidium bromide solution (10 mg/ml) was added. Once the gel had polymerized, DNA samples were loaded and electrophoresis was performed at 80-95 V for 30 – 60 min (for digested DNA from BAC clones, electrophoresis was done over night (14 -16 h) at 30 V in 0.8% Agarose due to the large size of the fragments).

Separation of fragments was visualized via exposure to UV light for 200 – 600 ms and acquisition of the image with a digital camera.

2.2.3 RNA isolation

Complete RNA from purified cell population in the range of $2x10^4$ – 10^6 cells was isolated with the *High Pure RNA Isolation Kit* (Roche) according to the manufacturer's instructions. RNA was eluted in 50 µl of Elution buffer and precipitated with 5 µg LPA and 125 µl EtOH over night. RNA was resuspended in 10 – 20 µl of RNase free water and stored at -70°C.

2.2.4 RT-PCR

cDNA synthesis was performed in 0.5 ml reaction tubes in a PCR machine (thermal cycler). For preparation of cDNA from 0.5 – 2 x 10^5 cells, RNA was precipitated, dissolved in 16 µl water of which 8 µl were used for RT PCR. RNA was re-digested with DNAse I for 30 min at 37°C followed by heat inactivation of the enzyme. Subsequently, RT-PCR was performed as described in the manual for SuperScript II ™ RT enzyme (Invitrogen). Oligo $(dT)_{20}$ was used instead of Oligo $(dT)_{12\text{-}18}$. For higher cell numbers (up to 10^6), total RNA concentration was measured using the *NanoDrop* photometer (Thermo Scientific), up to 5 µg total RNA were used for RT-PCR, all reaction volumes in the protocol were doubled in this case.
For RNA samples from cell numbers lower than 5 x 10^4 cells, an alternative protocol was used:
100 ng (200 ng/µl) (dT) – T7 Primer[135] were heated with 4.5 µl RNA for 5 min at 70°C in a PCR machine. 4.5 µl ice-cold RT mix (2 µl 5x First Strand Buffer, 0.5 µl 10 mM dNTP, 0.5 µl T4gp32 (8 mg/ml, USB), 0.5 µl SuperaseIN, 1 µl H_2O) were added. After 2 min incubation at 50°C, 0.5 µl SuperScript II ™ RT enzyme (Invitrogen) was added. The reaction was incubated for 60 min at 50 °C and mixed carefully with a pipette every 15 – 20 min. Enzymes were heat inactivated for 15 min at 65°C.
cDNA resulting from all protocols was finally purified on gel chromatography columns (*MicroSpin™ G50* columns) according to manufacturer's specifications in order to remove excess nucleotides and oligonucleotides.

Success of cDNA synthesis was controlled via PCR for β-Actin. As β-Actin primers are intron – spanning, PCR of β-Actin cDNA results in a product of 358 bp while PCR of the β-Actin gene results in a longer product. Successfully transcribed cDNA thus showed a strong PCR product at 358 bp and no larger bands on an agarose gel.

2.2.5 Isolation of genomic DNA from murine tails

For genotyping of mice, 2-5 mm of mouse tails were cut and stored in 1.5 ml Eppendorf tubes. DNA was isolated using *DNeasy Blood & Tissue Kit* (Qiagen) according to manufacturer's instructions. DNA was eluted in 50 – 200 µl of buffer (dependent on the tissue sizes) and 1-2 µl were used for genotyping PCR.

2.2.6 "Conventional" and Quantitatve PCR (qPCR)

"Conventional" PCR was used for genotyping of mice or as β-Actin control PCR for cDNA synthesis. For a 25 µl reaction volume, 2.5 µl each of foreward Primer (2.5 µM), reverse primer (2.5 µM), dNTP (2 mM) and 10x reaction buffer were combined with 1-2µl of cDNA/genomic DNA and 1 µl of RedTaq® DNA-Polymerase and filled up to 25 µl with H_2O.
The following parameters were used for the thermocycler:

- 3 min denaturing at 94 °C
- 1 min denaturing at 94 °C
- 1 min annealing at 54–66 °C 30-40 cycles
- 2 min elongation at 72 °C
- 3 min elongation at 72 °C

Annealing temperature and cycle number was adjusted accordingly for every primer pair.
Comparative quantitative gene expression analysis as well as ChIP analysis was done with realtime PCR using the *GeneAmp® 7300 Sequence Detector*. Optimal primer concentrations for expression analysis were determined by titration for each primer individually. For templates from ChIP experiments it proved to be best to use each primer pair at 900 nM concentration in order to get optimal amplification results.
Each reaction volume of 25 µl contained:

- Primer concentration: 50–900 nM
- 1 x *Power SYBR® Green PCR Master Mix* (Applied Biosystems)
- Template DNA

All reactions were performed in triplicates using the following parameters:

- Activation of Hot GoldStar Polymerase 10 min at 95°C
- Denaturing 15s at 95°C
- Annealing and elongation 1 min at 60°C
- Repeat denaturing and annealing/elongation for 40 cycles

A melting curve analysis of PCR products was performed and compared with adequate positive controls in order to ensure specificity of the PCR amplification.
Relative quantification of gene expression was performed using the ΔΔCt–method with Ubiquitin as a housekeeping gene for normalization. For ChIP experiments, quantification was done relative to an IP with anti-H3 antibody. (ΔΔCt–method according to ABI User Bulletin #2; can be obtained at www.appliedbiosystems.com).

2.2.7 Single-cell PCR (SC PCR)

SC PCR allows to analyze the transcriptome of a single cell for a number of genes of interest (up to 20). Primer design, reverse transcription, 1st PCR amplification and real-time quantitative PCR were performed essentially as described in [148] using DNA Engine Dyad (MJ Research) and 7300 Real Time PCR System (Applied Biosystems) machines. The cycle number for the first PCR reaction was lowered from 15 to 10, since we observed an improved correlation between input cDNA and resulting threshold cycle (Ct) values.

For SC PCR, individual cells were sorted into 96-well plates and cell lysis, RT PCR and following amplification were all performed in the same tube. Sensitivity of the SC PCR was high, as a first multiplex PCR was performed with outer primer pairs of all genes of interest (primers A and C) for 10 cycles. In a second amplification round, nested realtime PCRs were performed with the inner primer pairs (primers B and C) individually, using the first PCR as a template (Figure 6). Second round amplification was done for 50 cycles. This gave very high sensitivity and allowed to detect with adequate primers even the two copies of genomic DNA in an individual cell.

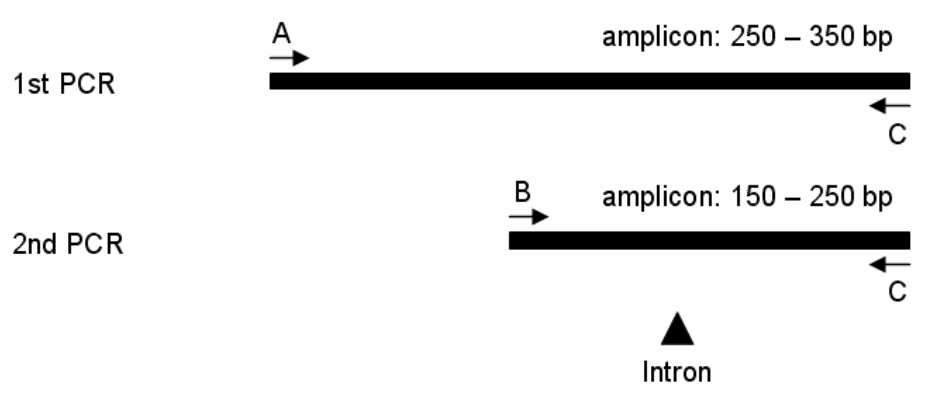


Figure 6

Schematic representation of the two amplification rounds performed in SC PCR. 10 cycles of preamplification are performed in multiplex PCR with an outer primer pair, second step amplification is performed as realtime PCR with a single inner primer pair.

Primers for SC PCR were designed with the Amplify 3 program to be 20mers with a preferable GC content of 50%. All primers were spanning exon-intron boundaries. Generation of primer dimers and potential crosspriming was tested *in silico*, primers were routinely tested by BLAST search for specificity. Primers were diluted from 1 mM stocks in TE buffer or 100 µM stocks in H_2O to their final concentration. No solution used in SC PCR passed more than three freeze/thaw cycles.

Handling of all chemicals and enzymes was done on a separate bench with completely separate equipment. On a second bench, RNA and cDNA were handled.

mTEC were sorted at the BD FACSVantageII (BD Biosciences) using an Automated Cell Deposition Unit (*ACDU*) in 96-well plates in 5 µl PBS. Plates were snap-frozen in liquid N_2 and stored at -80°C. Routinely, empty wells were included in the plates to check the specificity of the reaction.

Each primer pair (primer pairs A-C and B-C from all genes of interest) was tested with the same template and qPCR efficiency was assessed using LinRegPCR program ("efficiency test") (Figure 7). In order to test the primer competition in the multiplex PCR, 1st PCR was performed on the same template with either a) multiplex PCR, so a mix of all primer pairs A-C or b) individual primer pairs A-C alone. In a second qPCR, templates from a) multiplex PCR and b) individual primer pair PCR were analyzed with each primer pair B-C. Ct values and PCR efficiencies were compared between individual and multiplex 1st PCR to assess the degree of primer competition ("competition test") (Figure 7).

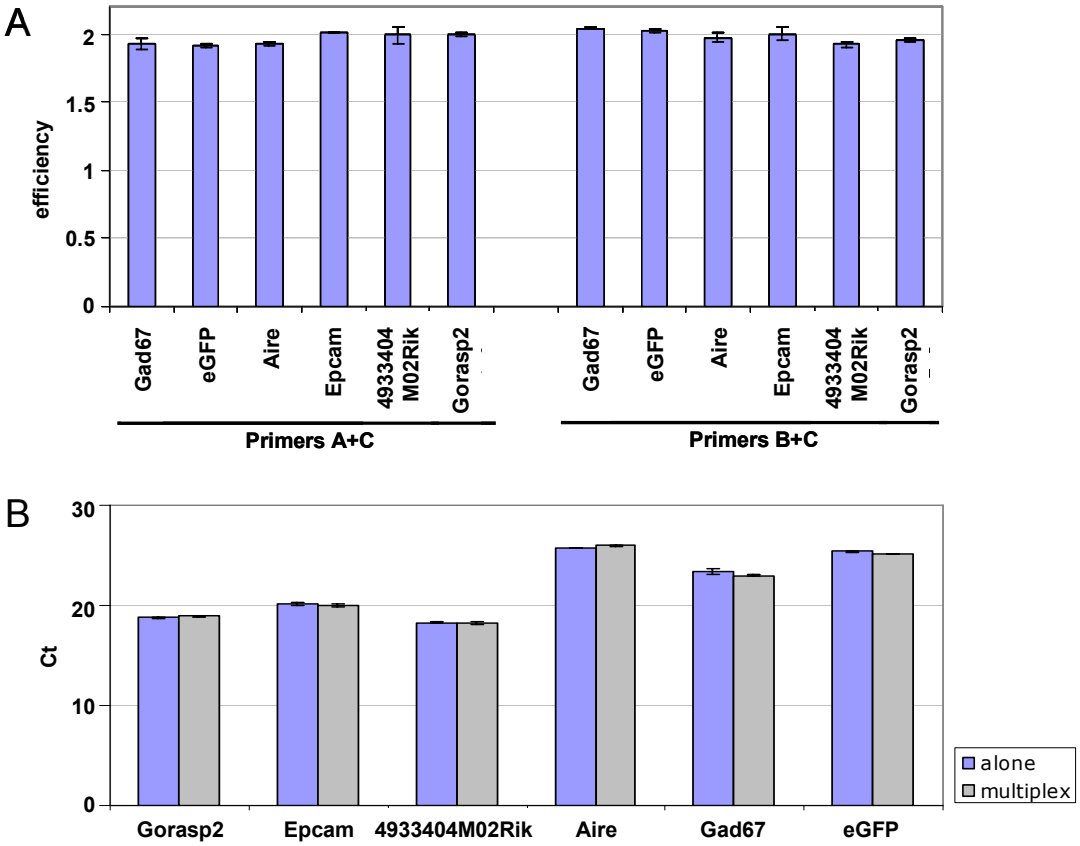


Figure 7

Efficiency and competition test of SC PCR. Efficiencies between the primer pairs tested individually did not vary significantly (A). Primer pairs A+C and B+C were tested in sequential PCR reactions, either only containing one primer pair alone in the first PCR (alone) or containing a mix of all primer pairs A+C (multiplex). There is no significant competition between the primer pairs as both "alone" and "multiplex" reactions achieved the same Ct values in the second amplification round (B).

Cells were lysed in the plate for 2 min at 65°C in the PCR machine. 10 µl of reverse transcription mix was added and reverse transcription performed in the PCR machine. 70µl master mix for the first PCR was added and the PCR was carried out in the PCR machine. From the resulting 85 µl of template, 4 µl per well were transferred to a new plate for each qPCR with specific inner primers (primer pair B-C). Routinely, melting curve analysis was performed and melting curves were compared to positive controls. Plates were frozen at -20°C after qPCR and samples with ambiguous melting curves were picked, reamplified for 15 cycles with "conventional" PCR,

analyzed on an agarose gel, purified with QIAquick PCR Purification Kit and sequenced in house.

Expression frequencies and co-expression patterns were tested for significance with the Chi-Square Test. When the Chi-Square test was not applicable due to low cell counts, Fisher's exact test was used. Co-expression pattern of Gad67/eGFP and Aire were tested for significance with Jonckheere-Terpstra Test.

Table 2.16: Reaction mixes and PCR programs for SC PCR

1x reverse transcription mix	
H_2O	2.34 µl
PCR buffer II (10x)	1.5 µl
$MgCl_2$	2 µl
dNTPs (10mM each)	1.5 µl
Oligos "C" (12.2 µM each)	0.16 µl x 6
RNase Block	1 µl
MuLV	0.7 µl
total	10 µl

PCR program for reverse transcription	
37°C	60 min
95°C	10 min
4°C	∞

1x Master Mix 1st PCR	
H_2O	46.3 µl
PCR buffer II (10x)	8.5 µl
$MgCl_2$	7 µl
dNTPs (2.5 mM each)	7 µl
Oligos (25 µM each of "A" and "C")	0.05 µl x 2 x 6
Taq	0.6 µl
total	70 µl

PCR program for 1st PCR		
95°C	10 min	
94°C	45 sec	10 cycles
60°C	1 min	
72°C	1 min 30 sec	
72°C	10 min	

1x Master mix for 2nd PCR (qPCR)	
H2O	7.52 µl
Primer "B" (25 µM)	0.24 µl
Primer "C" (25 µM)	0.24 µl
Power SYBRGreen Mix	12 µl
Total	20 µl

PCR program for 2nd PCR (qPCR)		
95°C	10 min	
95°C	30 sec	50 cycles
60°C	30 sec	
72°C	45 sec	
Melting curve analysis		

2.2.8 Chromatin-Immunoprecipitation

In order to analyze and compare the chromatin state at different promoter regions, the cell populations of interest were purified via cell sorting (for mTEC populations or brain cells) or enriched (for MEC and thymocytes). 10^5 cells per IP were used for all ChIP experiments unless stated otherwise. For ChIP with material from Gad67-eGFP mice, $3.5x10^4$ cells were used per IP. Cells were resuspended in 0.5 ml of medium or PBS (2-3% final FCS concentration) and fixed for 10 min at RT with 1% formaldehyde followed by blocking with final concentration of 0.125M Glycine for 5 min at RT. Fixed cells were washed twice with PBS in the presence of 0.1 mM PMSF, cells populations were lysed in 200 µl each (in 1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0, including Roche Protein Inhibitor Cocktail) on ice. Lysates were sonicated in parallel for 20 min (30 s ON/30 s OFF, setting "high") with the *Bioruptor* (Diagenode) in order to obtain an optimal length of chromatin fragments (100-500 bp) (Figure 8).

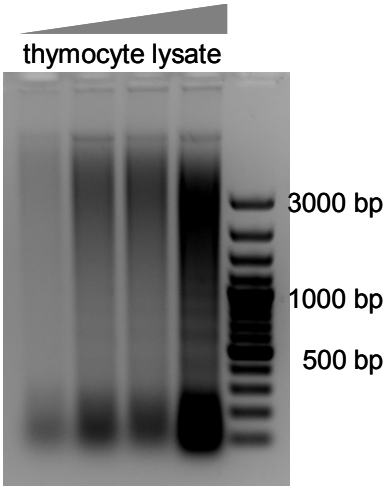


Figure 8
Chromatin shearing. Thymocyte lysate was sonicated with the Bioruptor™ for 20 min 30s ON/30s OFF on setting "high". Increasing concentrations were analyzed on an agarose gel. Optimal fragement size is between 100 and 500 bp.

Cellular debris was precipitated for 5 min at 13k rpm in a tabletop centrifuge. Chromatin solutions were diluted to 1 ml per antibody used for ChIP (usually 4-6 ml) with ChIP dilution solution (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.0, 167 mM NaCl, Roche Protein Inhibitors, PMSF). 1 ml of chromatin solution was combined with 2-3 µl of the respective antibodies for IP and incubated over night at 4°C with constant agitation.
The next day, 50 µl of ProteinA-Agarose slurry was added and bound to the Antibodies within 3h at 4°C.
Samples with beads were transferred to chromatography columns (Biorad). Beads bound to the antibody-chromatin complex were washed with different buffers (2x 1 ml TSEI, 2x 1 ml TSEII, 2x 1 ml TSE III, 3-4x 1.3 ml TE) and then transferred into Eppendorf tubes. The supernatant was removed and the bead pellet was incubated 10 min 95°C with 100 µl 10% Chelex-100 in order to remove the protein – DNA crosslinks introduced by fixation. Proteins were digested with 2 µl ProteinaseK (Invitrogen) for 30 min at 55°C followed by heat inactivation (10 min

95°C). Supernatant (80 µl) was transferred to a new tube, the bead pellet was washed again with 100 µl water and the supernatants were combined.
The 180 µl samples were analyzed by qPCR using 12 µl sample per triplicate.
qPCR was performed in the promoter regions (defined as 1000 bp upstream of the transcription start) of genes of interest with 40 – 45 cycles. All results were normalized to the results from ChIP with anti-H3 antibody. A negative control (anti-HA antibody) was always included in order to verify the specificity of the ChIP.

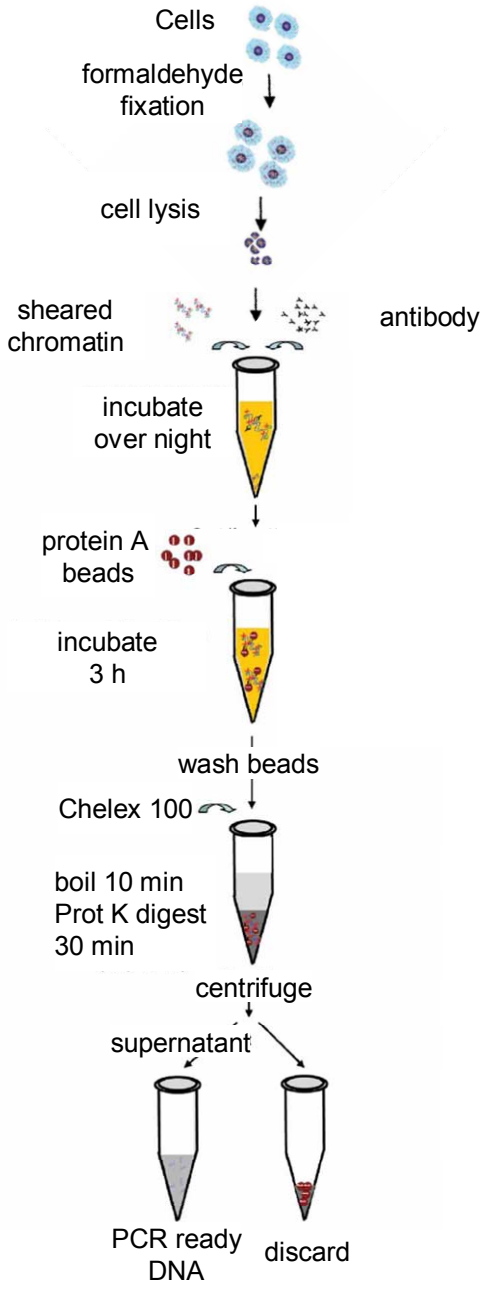


Figure 9
Chromatin Immunoprecipitation procedure. Adapted from Nelson et al. [149].

2.2.9 Counting of live cells

Cells were mixed to an adequate dilution with trypan blue and counted in a Neubauer counting chamber with a tabletop light microscope. Dead cells appear in blue and could be excluded, live cells appear translucent and were counted.

2.2.10 Antibody labeling

G8.8 (Ep-CAM) antibody was labeled and purified according to the manufacturer's instructions with *Alexa 647 Protein labelling kit* (Invitrogen) (1 mg antibody) or *EZ-Link Micro Sulfo-NHS-Biotinylation Kit* (Pierce) (0.4 mg antibody).

2.2.11 Isolation of specific cell populations from mice

Unless stated differently, RPMI-1640 Medium supplemented with 2 – 4 % FCS was used for all cell preparations.

2.2.11.1 Thymocytes

Mice were killed with CO_2 and thymi were removed from the thorax, cleaned from fat and connective tissue and cut into small pieces in few ml medium. Tissue pieces were stirred in a round-bottom tube at low speed in 10-15 ml of medium for 10-15 min at RT in order to release thymocytes. The supernatant was decanted and filtered through 70 μm gauze filters to remove residual fat and connective tissue pieces. The remaining tissue pieces were further processed for TEC isolation, the filtered supernatant now contains a almost pure thymocytes with very few contaminating MΦ and DC. Thymocytes were washed in medium and counted.

2.2.11.2 Thymic epithelial cell (TEC) populations

In order to obtain a single cell suspension of TEC from the remaining tissue pieces from 2.2.11.1, the tissue was subjected to multiple rounds of enzymatic digestion. Remaining tissue from up to 10 thymi was pooled per round bottom tube and digested 1x with Collagenase solution for 10 min at 37°C (which will release MΦ and DC from the tissue) followed by 5 – 6 rounds of digestion with Collagenase / Dispase solution for 20 min each at 37°C. After each round of digestion, the supernatant was decanted, filtered through 70 μm gauze and washed in medium. Tissue pieces in the filter were returned to the digestion solution. During incubation times, tissue was agitated regularly with a Pasteur pipette.
Fractions from Collagenase / Dispase digest 2 – 6 were pooled, washed and counted. Digestion with trypsin was avoided as the mTEC maturation marker CD80 is sensitive to trypsin [150].
In order to enrich for TEC, $CD45^+$ cells were depleted via magnetic cell sorting (*MACS*). Cells were resuspended in *MACS* buffer at $2x10^8$/ml and incubated for 15 min at 4°C with CD45-microbeads at 1:40 dilution. The cell solution was diluted to 10^8/ml and up to $4x\ 10^8$ cells were added step by step to a *LS* MACS column which had been prepared according to manufacturer's specifications. Columns were washed 2x with 3 ml MACS buffer, the flow-through containing the TEC-enriched fraction was collected, washed with medium and counted.

Cells (approx.1 – 3 x 10^7 per 20 mice, depending on the degree of enrichment) were blocked with 500 µl of anti-Fc-receptor supernatant (2.4G2) for 15 min on ice. Cells were then washed and stained with an antibody cocktail for 15 min on ice in 3% FCS/PBS, the antibodies vary according to the populations which were to be isolated downstream. Cells were washed, filtered through 35 µm filters and stained with PI directly prior to sorting. Cell populations were sorted on the BD FACSAria (BD Biosciences) on purity mode (Figure 10). (Only sorting for SC PCR was done at the BD FACSVantageII (BD Biosciences) using an Automated Cell Deposition Unit (*ACDU*).

Table 2.17: Overview of different antibody combinations used for cell sorting

TEC populations sorted	**Cell sorter**	**Antibodies used**	**Dilution**
$mTEC^{high}$ and $mTEC^{low}$	BD FacsAria	Ly-51 Fitc	1:50
		CD45 PerCP	1:50
		G8.8 Alexa 647	1:500
		CD80 PE	1:100
		PI (1mg/ml)	1:5000
$mTEC^{high}$ and $mTEC^{low}$ for FISH	BD FacsAria	Ly-51 Fitc	1:50
		CD45 PerCP	1:50
		G8.8 bio – SAV PEcy7	1:100
		CD80 PE	1:100
		PI (1mg/ml)	1:5000
$eGFP^+$ and $eGFP^-$ $mTEC^{high}$ $mTEC^{low}$ from Gad67/eGFP mice for SC PCR	BD FACSVantage II	CDR1 Alexa 680	1:50
		CD45 cychrome	1:100
		G8.8 Alexa 647	1:100
		CD80 PE	1:100
		PI (1mg/ml)	1:1000
$eGFP^+$ and $eGFP^-$ $mTEC^{high}$ $mTEC^{low}$ from Gad67/eGFP mice	BD FacsAria	CDR1 Alexa 680	1:50
		CD45 PerCP	1:50
		G8.8 Alexa 647	1:500
		CD80 PE	1:100
		PI (1mg/ml)	1:5000

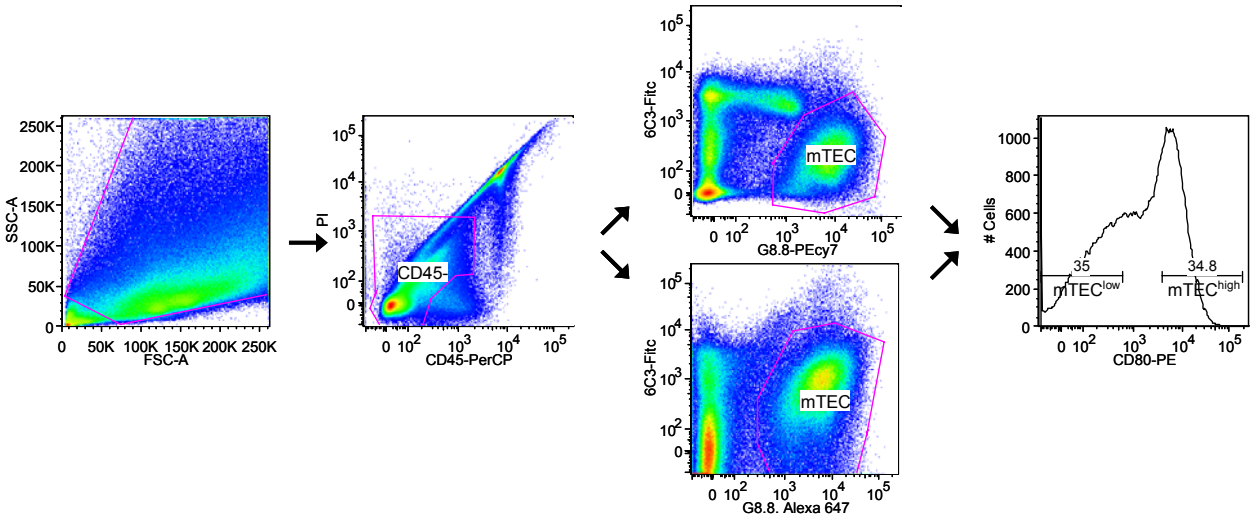


Figure 10
Sort gates for the isolation of mTEC^{high} and mTEC^{low} using G8.8^{-}PEcy7 or G8.8-Alexa 647 alternatively. mTEC were gated as CD45^{-}PI^{-}EpCAM(G8.8)$^{+}$ cells and subdivided into the mature CD80^{high} and the immature CD80^{low} subset.

2.2.11.3 Mammary gland epithelial cells (MEC)

For isolation of MEC, mice were killed on the same day of weaning (age of cubs: 3-4 weeks) or 2-7 days after birth directly after separation from the cubs (when young cubs were used for other experiments). All mammary glands were taken out, cleaned from connective tissue and cut into small pieces. In order to obtain a single cell solution, the tissue pulp was digested 1x with collagenase/dispase solution (15 min, 37°C) and 5-6 x with trypsin solution (20 min, 37°C). After each round of digestion, the digestion medium was removed from the bottom of the flask (as fatty mammary gland tissue floats on top), filtered through 70 µm gauze and washed in medium. Tissue pieces in the filter were returned to the digestion solution. During incubation times, the tissue was agitated regularly with a Pasteur pipet. When to much cell death occured and large amounts of DNA were released (solution turns slimy), the digestion was stopped.

Cells from all fractions were pooled, filtered and enriched via a biphasic Percoll gradient: MEC were resuspended in the more dense Percoll/RPMI-1640 mixture (δ=1.06) and pure RPMI-1640 (δ=1.00) was carefully layered on top. Upon centrifugation (30 min at 3500 x g without brake), MEC floated up and could be collected at the phase boundary. This leads to a ~50% enriched MEC population (6-12 x 10^6 per mouse).

For some applications, MEC were further purified via cell sorting: Cells were blocked with 500 µl of anti-Fc-receptor supernatant (2.4G2) for 15 min on ice, then washed and stained with G8.8 – Alexa 647 antibody (1:500) for 15 min on ice. Cells were filtered through 60 µm gauze and stained with PI 1:5000 directly prior to sorting on purity mode at the FACS Aria (BD Biosciences) (Figure 11). As MEC isolation itself is a very harsh treatment, many cells were found to be dead upon sorting or died due to additional treatment.

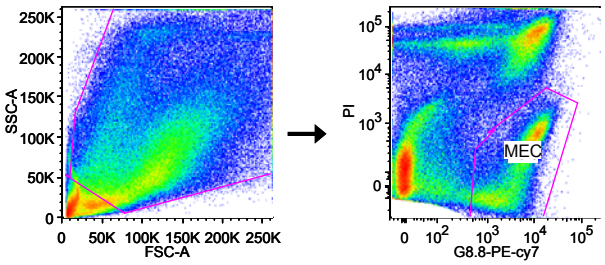


Figure 11
Sort gates for the isolation of pure MEC. MEC were gated as $PI^-EpCAM(G8.8)^+$ cells.

2.2.11.4. Neurons

Mice were killed with CO_2 and cut open in the neck region to remove the complete brain. The brain was placed in PBS with 0.45 % glucose on ice, the regions of interest (frontal cortex, olfactory bulbs and cerebellum) were isolated and cut into small pieces with a scalpel. Brain pieces were transferred to a 15 ml tube, supernatant was removed, tissue resuspended in 10 ml 0.05% trypsin solution and incubated for 15 min at 37°C. After the digest, the tissue was washed 2-3 times in medium (DMEM/F12 50% each). Finally, tissue pieces were resuspended in 1 ml of medium and homogenized first with a 1 ml tip, then with a Pasteur pipette followed by a drawn out Pasteur pipette. Cells were passed through a 70 µm gauze filter and a 35 µm cell strainer, PI (1mg/ml) 1:5000 was added and $eGPF^+$ live cells were sorted at the BD FACSAria (Figure 12).

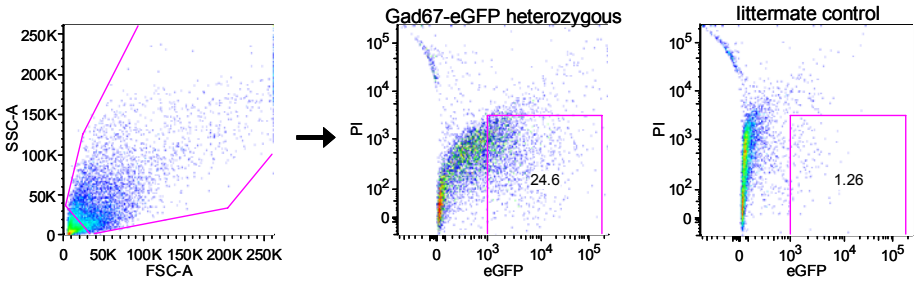


Figure 12
Sort gates for the isolation of $eGFP^+$ brain cells. PI^- $eGFP^+$ cells were sorted from mouse brain (olfactory bulbs, frontal cortex and cerebellum only).

2.2.12 Immunohistochemistry

2.2.12.1 Organ preparation

For Gad67/GFP mice, mice were killed with CO_2 and perfused with 30 ml 4% PFA in PBS after rinsing the blood circuit with PBS. Target organs were removed, cleaned from connective tissue and post-fixed in 4% PFA at 4°C over night on a rotating platform.

Organs were immersed in *Tissue-Tek*® on a small plastic frame and dipped into a beaker with isopentene which was immersed in liquid nitrogen. Frozen embedded tissue was taken out and stored at -80°C or directly taken for cryosectioning of 5 µm sections at a LEICA CM 1900 Cryoastat. Sections were dried on *Histobond*® slides for 30 min or over night.

Wt mice were killed with CO_2 , target organs removed, cleaned and directly processed in *Tissue-Tek*® as described above without further fixation.

2.2.12.2 Cytospins

Target cell populations were resuspended in PBS and 1-3x 10^4 cells in 10 µl were dropped onto *Histobond*® slides and dried over night.

2.2.12.3 Immunohistochemical staining

Unfixed tissue or cytospins were fixed in 4% PFA in PBS for 10 min at RT, tissues which had already been fixed were processed directly.
Sample was then washed 2x 5 min with PBS at RT and blocked with blocking solution (2% BSA, 0.5% Triton-X-100 in PBS) for 30 min at RT. The primary antibody was diluted in blocking solution, centrifuged (13,000 rpm, 5 min 4°C) to avoid precipitates and added to the sample either over night at 4°C or for 30 min 37°C and 30 min RT. Slides were washed 3x for 10 min in PBS, the secondary antibody was diluted in blocking solution, centrifuged (13,000 rpm, 5 min 4°C) to avoid precipitates and added to the sample for 1-2h at RT. Samples were washed again 3x for 10 min in PBS. Cytospins stained with Hoechst 33342 for 2 min and washed 2x 3 min in PBS. All samples were embedded with warm Kaiser's Glycerol Gelatine and covered with coverslips. Images were taken at a Zeiss Axio Imager Z.1 fluorescence microscope.

3. Results

3.1 Chromatin structure is modified at the level of individual genes, not the whole locus

pGE in mTEC affects a vast number of genes which are distributed over the entire genome. Genes which are promiscuously expressed are often regulated differently compared to their corresponding tissue [60] (Tykocinski, unpublished results). To date, the transcriptional regulator Aire is the only cis-acting factor known to control pGE. As the control excerted by tissue-specific cis-acting factors is dispensable in the thymus and pGE was at the same time found to encompass a plethora of genes expressed at low level, other mechanisms were suggested to play a major role in the control of pGE. Such a type of gene expression which is characterized to be a) low-level, b) noisy, i.e. with high background activity [73], c) to show mainly stochastic expression patterns on the level of gene expression in an individual cell, but at the same time d) to encompass large parts of the genome at the population level [52, 60, 76], was thought to be controlled by a higher-order mechanism. Individual factors could hardly be imaged to account for these characteristics alone [17, 74]. Epigenetic mechanisms have been suggested to control pGE for the following reasons: a) Promiscuously expressed genes tend to localize in clusters, b) the imprinting status of Igf2 is lost in mTEC, c) expression of certain TRA including cancer germ cell antigens correlates with promoter hypomethylation [17]. The two ways which are most commonly know as mechanisms of epigenetic regulation of gene expression are DNA CpG methylation and the modifications of histone tails (as described in 1.3). The latter are analyzed in this study in mTEC via chromatin immunoprecipitation (ChIP) with particular focus on the mouse casein gene locus and the mouse Gad67 locus.

3.1.1 The casein gene locus as a model locus for promiscuous gene expression

ChIP experiments were first performed on the mouse casein gene cluster which serves as a model locus for pGE for the following reasons: It is a large cluster of 1.2 MB on murine chromosome 5 in which 12 genes are expressed contiguously in the mTEC^{high} population [52]. The contiguous expression in such a large cluster is consistent with potential epigenetic regulation of this locus. Within the contiguously expressed region, the locus encompasses the following gene families: Casein genes normally expressed in the lactating mammary gland, flanked upstream by the sulfo-transferase and UDP glycosyl-transferase families normally expressed in the kidney and downstream by a family of salivary gland genes.

Genes from all families are expressed in mTEC^{high} and are classified as typical promiscuously expressed antigens, as they are highly tissue-specific as well as expressed in a temporally restricted fashion (i.e. casein genes are only expressed during late pregnancy and lactation) [52]. Within the casein locus, both Aire-dependent and Aire-independent genes are mixed, allowing for the

simultaneous analysis of both of these groups: Csna, Csng and Csnd are Aire-dependent, while Csnb and Csnk are Aire-independent [52]. Single cell expression analysis has already been performed in our group, defining expression frequencies as well as expression patterns for every individual casein gene on the single cell level [60].

The casein locus is thus suitable as a model locus for epigenetic analysis, as it encompasses many typical promiscuously expressed genes with known expression patterns and shows highly clustered co-expression at the mTEC^{high} population level.

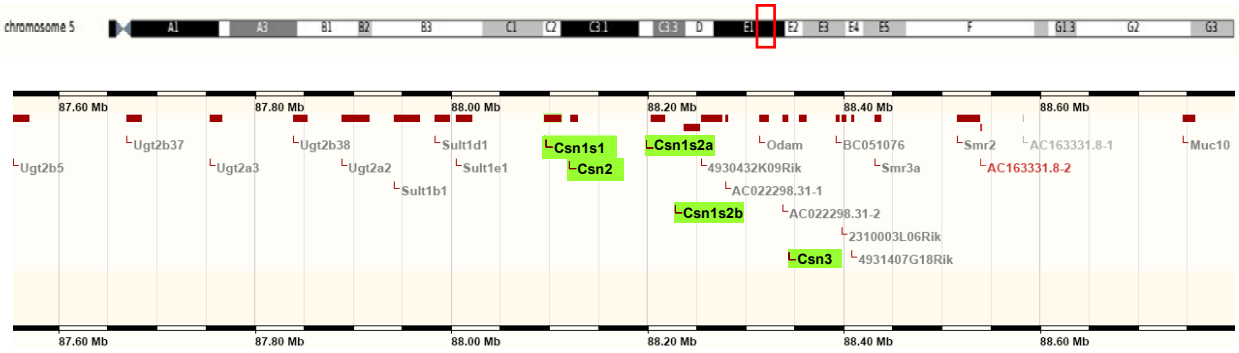


Figure 13

The mouse casein locus on chromosome 5. Casein genes Csna-Csnk are marked in green. They are flanked upstream by sulfo-transferase and UDP glycosyl-transferase gene families and downstream by salivary gland genes. Information on the gene locus was obtained at www.ensembl.org.

3.1.2 Optimization of the ChIP protocol for low numbers of *ex vivo* sorted cells

Chromatin immunoprecipitation (ChIP), a method to analyze the modifications of histone tails in a specific region of interest, is routinely performed with 10^8 – 10^6 tissue culture cells per IP [101, 151, 152]. This would exclude its application to the analysis of ex-vivo sorted TEC, which are available only in low numbers (approx. 3-4 x10^4 of mTEC^{high} and mTEC^{low} per mouse thymus). Recently, different modifications have been introduced to the ChIP protocol [153, 154]which made it feasible to downscale significantly the number of cells needed. ONeill et al. performed Carrier ChIP (CChIP) with as little as 100 cells per IP [153], but as quantitation of precipitated chromatin was performed with radioactive PCR, this method proved to be impractical to be implemented in the laboratory. Nelson et al. developed a method for so-called Fast-ChIP [149], which proved to be also more efficient for low cell numbers. This protocol was used after further optimization (see 2.2.8) in this study. Using promoter acetylation in the Csnb promoter in MEC and in the CD45 promoter in thymocytes as two positive control signals, as these two genes are known to be strongly expressed in the respective cell types, the method was shown to be applicable with a minimal number of 10^3 cells per IP (Figure 14) for both MEC and thymocytes. The background signal from precipitation with unspecific (anti-HA) antibody significantly increased with respect to the acetylation signal below 2x10^4 cells per IP. Therefore, between 10^5 and 2x10^4 cells per IP were used in the following experiments in order to obtain more robust signals. mTEC were too valuable to perform such pre-tests, therefore the specificity of ChIP was assumed to be similar. Nevertheless it is possible that ChIP with mTEC is less sensitive than with MEC or thymocytes,

as expression levels in mTEC are usually lower and mTEC promoters could possibly carry fewer marks at the histone tails.

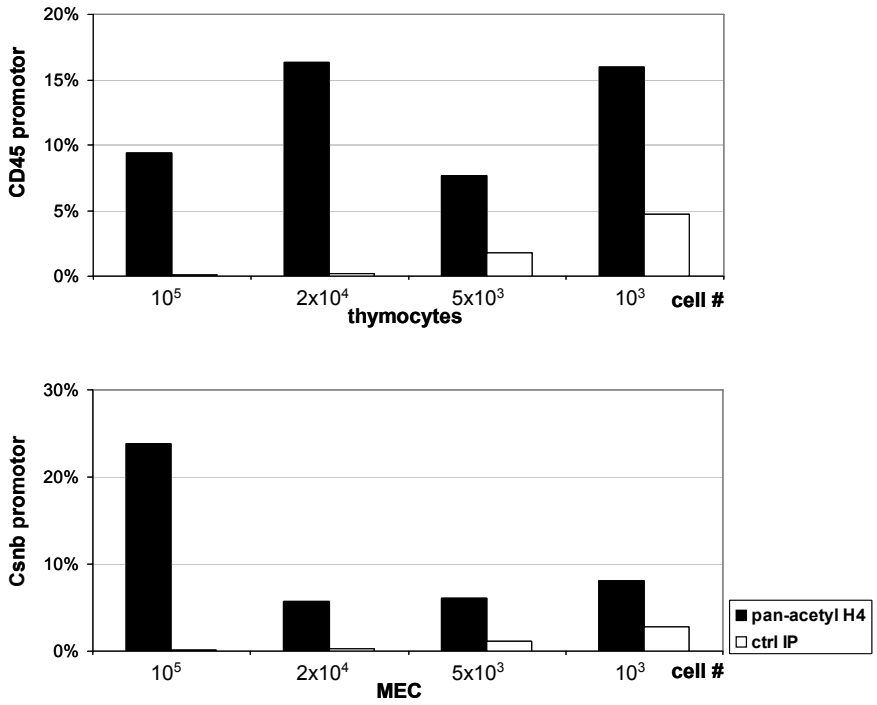


Figure 14

ChIP is functional down to 10^3 cells/IP. Chromatin IPs were performed with different numbers of thymocytes and MEC (50% enriched) per IP. Csnb and CD45 promoter regions were analyzed with qPCR. All samples were normalized to H3 levels. Experiment is representative of two independent experiments.

3.1.3 Active chromatin marks are present only at the Casein beta promoter

As the sensitivity of ChIP analysis had been assessed and found to be sufficient for the analysis of sorted mTEC populations, ChIP analysis was performed focusing on genes in the casein gene locus in different mTEC and control cell populations.

Histone tails can undergo a large variety of modifications including acetylation, methylation, and phosphorylation [155] (see 1.3.2). In this study, three characteristic modifications for an active or inactive chromatin state were analyzed. Histone H4 lysine acetylation is a mark primarily indicative of active chromatin (analogous to H3 lysine acetylation), broadly distributed at the promoter and over the 5' end of actively transcribed genes [100]. Histone 3 lysine 4 trimethylation (H3K4me3) is a more specific, less frequent mark found directly at the transcription start site of actively transcribed genes [97, 99, 100]. The H3K4me3 mark can also be found, together with H3K27me3, at so-called bivalent chromatin domains.

It had been suggested by our group that the expression of whole gene clusters transcribed in mTEC is regulated by epigenetic opening of the entire gene cluster either preceding gene expression (opening in $mTEC^{low}$) or concomitantly with gene expression (opening in $mTEC^{high}$). This was hypothesized to be followed by transcription factors activating individual genes from the cluster [17]. In order to test this hypothesis, specific histone tail modifications in $mTEC^{high}$ and $mTEC^{low}$ were analyzed in the casein locus.

ChIP was performed with 10^5 cells per IP of *ex vivo* sorted mTEC^{high} and mTEC^{low}. *Ex vivo* isolated MEC (50% enriched) were used as a positive control as casein genes are highly expressed in MEC, thymocytes were used as a negative control. IPs were performed with anti-pan-acetyl H4 and anti-H3K4me3 antibodies as characteristic active marks, anti-H3K27me3 as a characteristic mark for facultative heterochromatin, anti-HA as negative control and anti-H3 for normalization of signals. As nucleosomes are not positioned equally along the DNA [100 156], all signals for histone tail modifications need to be normalized to the nucleosome density in the region of interest. Without normalization, regions which are partially depleted of nucleosomes may give biased results, underestimating the percentage of modified histone tails.

All marks were analyzed at the transcriptional start regions of selected genes in the casein locus with realtime PCR. As chromatin fragments produced for ChIP analysis are usually 100-500 bp long, ChIP analysis always gives the average concentration of a particular mark in a region, but cannot identify the position of the exact nucleosome carrying the mark[157].

H4 acetylation and H3K4 trimethylation were both clearly above background in the positive controls, which are all casein genes in MEC and the CD45 gene in thymocytes. These genes are known to be strongly expressed in a high percentage of the respective cell populations and should thus show high levels of active histone marks. Note that percentages for H4 acetylation were generally higher, as the mark is on average more frequent than H3K4 trimethylation. In mTEC^{low}, no distinct H4 acetylation or H3K4 trimethylation could be detected. None of the genes analyzed is expressed in mTEC^{low}, thus there was no evidence for chromatin opening of the casein locus already at an immature state, preceding gene expression (Figure 15).

In mTEC^{high}, only Csnb carried active chromatin marks, while all other genes only showed background levels for both H4 – acetylation and H3K4 trimethylation. It can be concluded that chromatin is not opened locus-wide in the casein locus in the entire mTEC^{high} population as it had been suggested (see 1.2.2).

Csnb, in distinction to all other promiscuously expressed genes known so far, is expressed in over 80% of mTEC^{high}, as detected by SC PCR while all other genes in the casein locus analyzed by SC PCR only show expression frequencies between 2% and 15 % in the mTEC^{high} subset [60]. We found active chromatin marks only locally at the Csnb transcription start, a phenomenon for which two potential explanations are possible: Csnb, as it is already exceptional in expression frequency, could be regulated by different mechanisms than other promiscuously expressed genes with more "typical" expression frequencies. Thus it could be the only gene in the casein locus regulated by epigenetic mechanisms. On the other hand it is possible that modifications at genes, which are only expressed in a very low percentage of the cell population analyzed, fall below the detection threshold of the ChIP method. This could apply to all genes analyzed in the locus but Csnb. In this case, possibly every gene expressed in a particular single mTEC^{high} could be regulated epigenetically (as it has been observed for Csnb) and carry active chromatin marks at its promoter in all cells expressing the gene. Such modifications would not be visible in an assay like ChIP which averages signals over the whole cell population analyzed.

The H3K27me3 mark could not be detected to a significant degree on any gene in the casein locus in mTEC. In this case, the Hoxc10 gene, a characteristic PcG target [106], was used as a positive control. In relation to Hoxc10, which was highly trimethylated at H3K27, all genes in the casein locus were negative for this mark in all cell populations analyzed. This implies that the casein locus is characterized neither by a state of facultative heterochromatin nor by a bivalent chromatin state (carrying H3K4me3 and H3K27me3 simultaneously) in any of the populations analyzed.

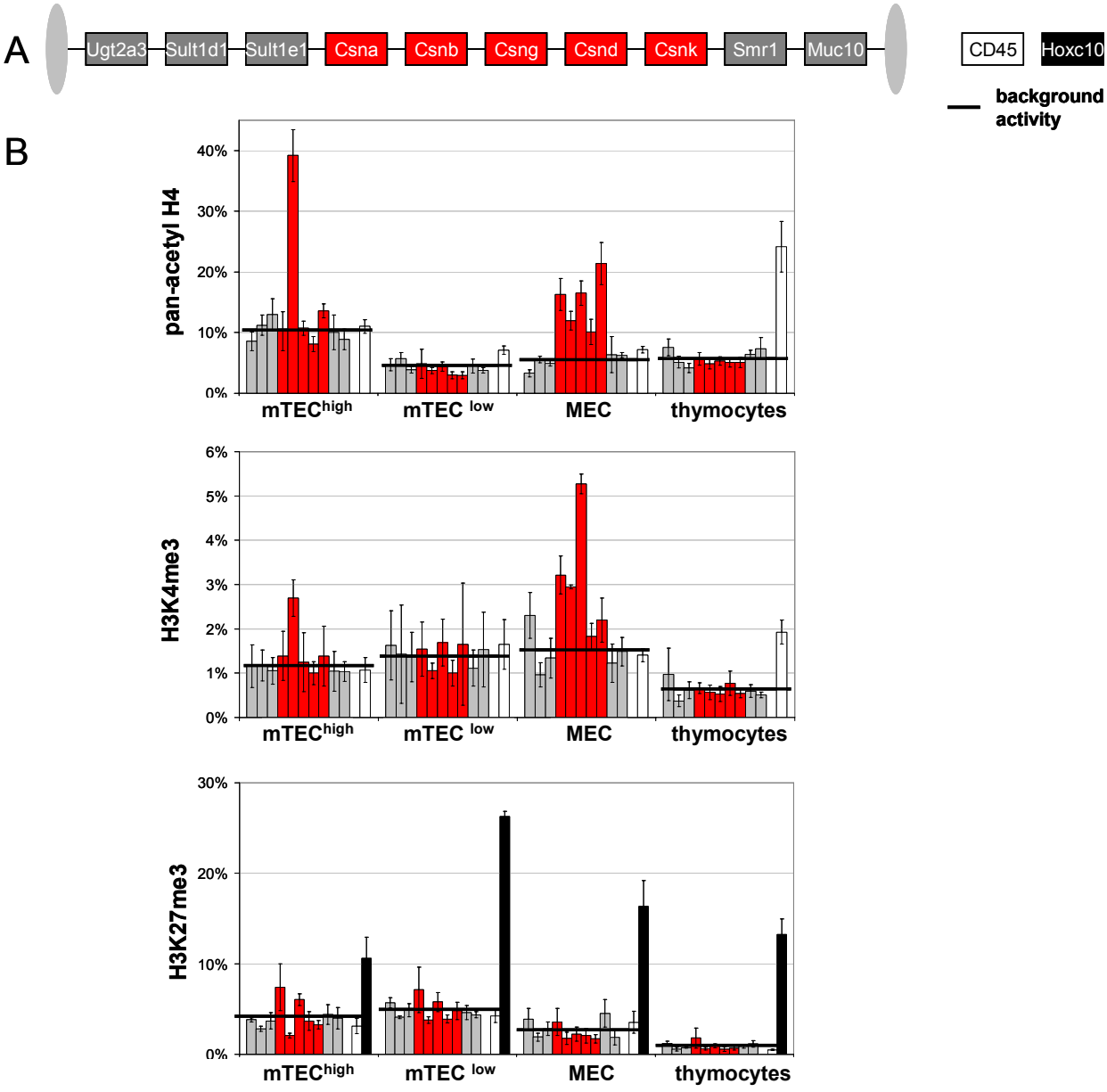


Figure 15

Only genes expressed at a high frequency show active marks in their promoter regions. ChIP was performed with freshly prepared mature and immature mTEC. MEC (50% enriched) and thymocytes were used as positive and negative controls (10^5 cells per IP). Samples were analyzed with qPCR focusing on different promoter regions from the casein gene locus as well as control genes for thymocytes and H3K27 methylation (A). Diagrams show the frequency of a particular histone modification mark at the promoter region, normalized to H3. Threshhold values were calculated as the mean modification level of all genes which are not expressed to a significant degree in the target cell population. In mature mTEC, only Csnb, the only gene expressed at high

frequency, carries marks for active chromatin to a significant degree while in MEC, all casein genes carry active marks (B). All data in Figure 2 are the mean of three to four independent experiments, errors represent the SEM.

3.1.4 Detection threshold of ChIP is above the frequency of most promiscuously expressed genes in the casein locus

To clarify whether histone modifications in genes which are less frequently expressed than Csnb are possibly under the detection threshold of the analysis, the detection limit for ChIP had to be assessed. This was done by mixing MEC (50% enriched) and thymocytes in various ratios, thus creating a situation where only a limited number of cells in the sample expressed certain genes. ChIP was performed with pan-acetyl H4 antibody and normalized to H3 antibody. Csnb and CD45 promoter regions were analyzed to determine the minimal number of MEC expressing Csnb needed in the sample to allow for detection of promoter acetylation (and vice versa for thymocytes expressing CD45). The detection threshold was defined here as the minimal frequency at which a gene must be expressed in the cell population in order to still allow detection of active chromatin marks.

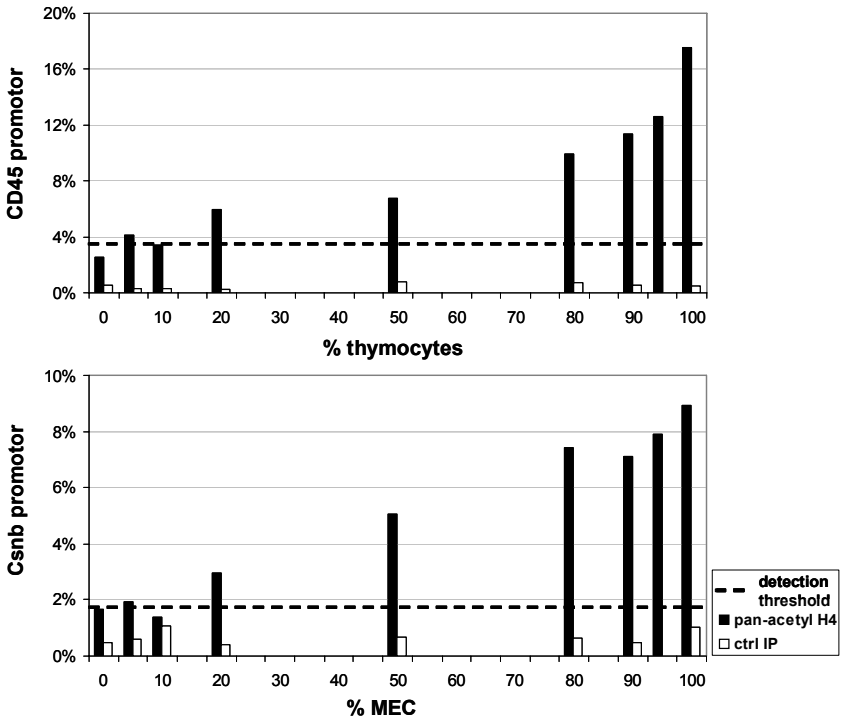


Figure 16
ChIP detection threshold. Chromatin IPs (20.000 cells/IP) were performed using thymocytes and MEC (50% enriched) mixed in different ratios. Csnb and CD45 promotor regions were analyzed with qPCR. All samples were normalized to H3 levels. Experiment is representative of two independent experiments.

H4 acetylation at the CD45 promoter could be distinguished from background levels starting at 20% thymocyte content; H4 acetylation at the Csnb promoter was similarly detectable at 20% MEC content (Figure 16). Note that as MEC were only purified to ~50%, 20% MEC content corresponded to a true threshhold of ~10% pure MEC for detection of H4 acetylation. The ChIP detection threshold for promoter acetylation of genes strongly expressed in MEC and

thymocytes was therefore determined to be 10-20%. As the genes analyzed in Figure 16 are expressed to high levels in the respective cells and thus probably carry high concentrations of modificated histone tails in the regions analyzed [100], these thresholds may easily be higher for low-level promiscuous gene expression. Detection thresholds for the H3K4me3 mark were not analyzed, but may be, as H3K4me3 is generally a less frequent mark, even higher. It is thus a reasonable assumption that modifications at the promoters of genes only expressed at low frequencies (up to 20%), such as the genes from the casein locus in mTEC, do carry active chromatin marks when expressed. This, however, cannot be detected by ChIP. In order to verify this hypothesis and observe the predicted histone modifications at other genes than Csnb it is necessary to enrich for mTEC expressing a single antigen, so that the expression frequency is above the detection threshold, and perform ChIP with this subpopulation,.

3.1.5 Enrichment of mTEC expressing a single antigen: Chromatin marks in Gad67/eGFP mice

As ChIP analysis on the whole mTEC^{high} and mTEC^{low} populations was performed in the casein gene locus, it would be preferable to enrich for mTEC expressing a particular gene from this locus. Unfortunately there are no highly specific antibodies against murine caseins available to date. Staining with a commercial antibody against Csnk has been tried, but proved to be difficult due to insufficient specificity of the antibody in FACS analysis and on cytospins. Suitable antibodies for membrane-bound promiscuously expressed antigens were not available. Cell sorting for intracellularly stained antigens generally posed – apart from the availability of specific antibodies – multiple problems for downstream applications: Firstly, fixation of cells is essential to allow for permeabilization of the membranes and entry of the antibody into the cytoplasm without destroying the cells. But fixed cells tend to loose large percentages of their RNA, thus RNA isolation and control of enrichment of target RNA (i.e. Csnk RNA) in the sorted population is virtually impossible. On the other hand, it is essential as to determine whether enrichment is sufficient for ChIP analysis. Secondly, if cells are fixed, fixation needs to be carried out in a way which allows for downstream ChIP analysis [158, 159]. Lastly, sorting mTEC subpopulations decreases the number of cells which can potentially be obtained by a factor of 10-50, depending on the expression frequency of the antigen. As many cells are lost during fixation and permeabilization procedures, cell numbers obtained can be very low (~1-2 x10^4 cells from 20 or 30 mice). Several antibodies against promiscuously expressed antigens have been tried, both from the casein locus and from other loci, but without success due to the limitations mentioned.

A simpler way to enrich for mTEC expressing a single antigen is to use reporter mice. Reporter mice for casein genes or other genes in the casein locus were unfortunately not available, therefore we switched the gene locus of interest and used Gad67/eGFP reporter mice [160] which express eGFP under the Gad67 promoter as a *knock in* into the first translated exon of Gad67, disrupting the endogenous Gad67 gene. Gad67 is a typical tissue restricted antigen ectopically expressed in mTEC and upregulated in mTEC^{high} in an Aire-independent fashion [161]. Glutamate decarboxylase (GAD), which exists in two isoforms of 65 kD and 67 kD, decarboxylates

glutamate to GABA, a neurotransmitter, and is primarily expressed in GABA-nergic neurons in the brain. Also being expressed in the pancreatic islets, Gad67 is a prominent auto-antigen in autoimmune diabetes [162, 163]. Heterozygous Gad67/eGFP mice show no apparent phenotype; for this study, only heterozygous mice on the C57BL/6 background were used.

eGFP$^+$ and eGFP$^-$ mTEC^{high} were sorted to high purity (Figure 17) and used to perform ChIP experiments. Due to limited material, 5x10^4 cells per IP of each sorted population, including thymocytes as negative control were used to ChIP with pan-acetyl H4 antibody, H3K4me3 antibody, H3 and HA antibodies. As a positive control, eGFP$^+$ cells were sorted from Gad67/eGFP mouse brain.

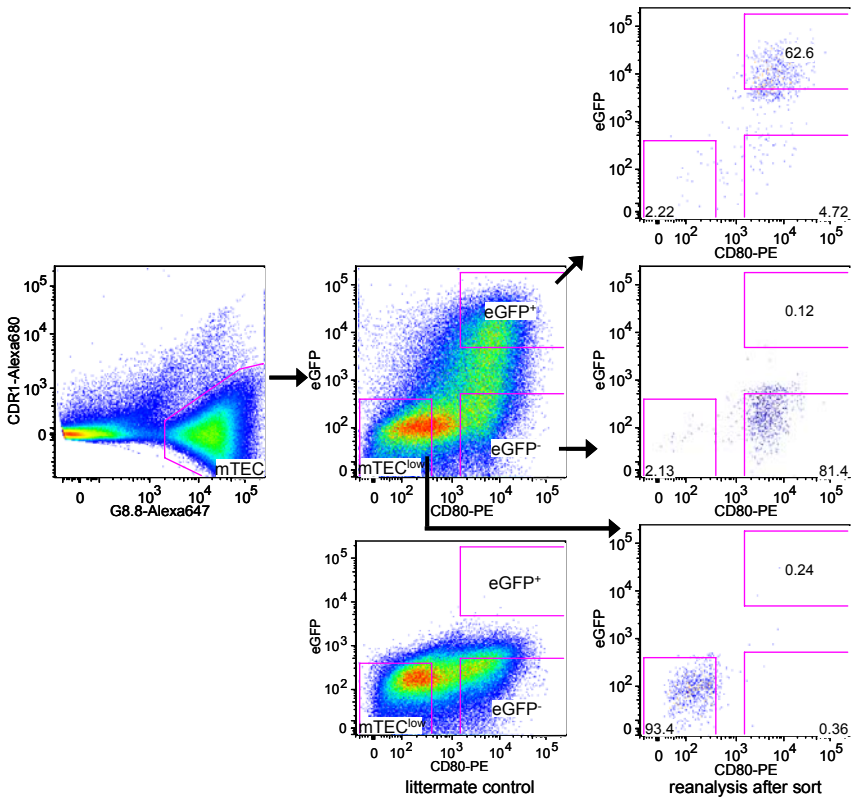


Figure 17

Sort gates for the isolation of eGFP$^+$/eGFP$^-$ mTEC^{high} from Gad67/eGFP mice. Cells were sorted as CD45$^-$PI$^-$ CDR1$^-$EpCAM(G8.8)$^+$CD80$^+$ eGFP$^+$ or eGFP$^-$ cells; mTEC^{low} as CD45$^-$PI$^-$CDR1$^-$EpCAM(G8.8)$^+$CD80$^-$ eGFP$^-$ cells. Littermate controls were used to determine the eGFP background fluorescence. Sorted populations were reanalyzed and shown to be of high purity.

Using heterozygous mice made it possible to distinguish between chromatin marks on the two different alleles by using allele-specific primers and a common primer for the promoter region which is the same on both alleles (Figure 18 A).

The ChIP shows significant promoter acetylation at the Csnb gene for both, eGFP$^+$ and eGFP$^-$ populations of mTEC^{high}, as expected, but does not show any significant acetylation in the Gad67/eGFP locus; neither in the common promoter region, nor for any of the allele-specific primers (Figure 18 B). eGFP$^+$ brain cells, where Gad67 and eGFP are both expressed at a high level [160], carry H4 acetylation in the Gad67/eGFP region as well as strong H3K4 trimethylation. eGFP, Gad67 and the common promoter region are also found to be H3K4 trimethylated in eGFP$^+$ mTEC^{high}, at a similar level as the Csnb promoter. The eGFP gene carries much higher

levels of H3K4me3 in eGFP$^+$ compared to eGFP$^-$ mTEC^{high}, in correlation with eGFP expression. However, in eGFP$^-$mTEC^{high} H3K4 trimethylation at the Gad67 gene and the common promoter region is found at similar levels as in eGFP$^+$mTEC^{high}. No modifications can be observed for any of the genes analyzed in mTEC^{low}.

Altogether, the Gad67 gene region is regulated by two epigenetic marks, H4 acetylation and H3K4 trimethylation in the tissue, (i.e. in brain cells), but only by H3K4 trimethylation in mTEC. The GFP gene carries more active marks in the eGFP$^+$ population where it is expressed to a higher level, while Gad67 and the common promoter region carry similar degrees of modification in both eGFP$^+$ and eGFP$^-$ mTEC^{high}.

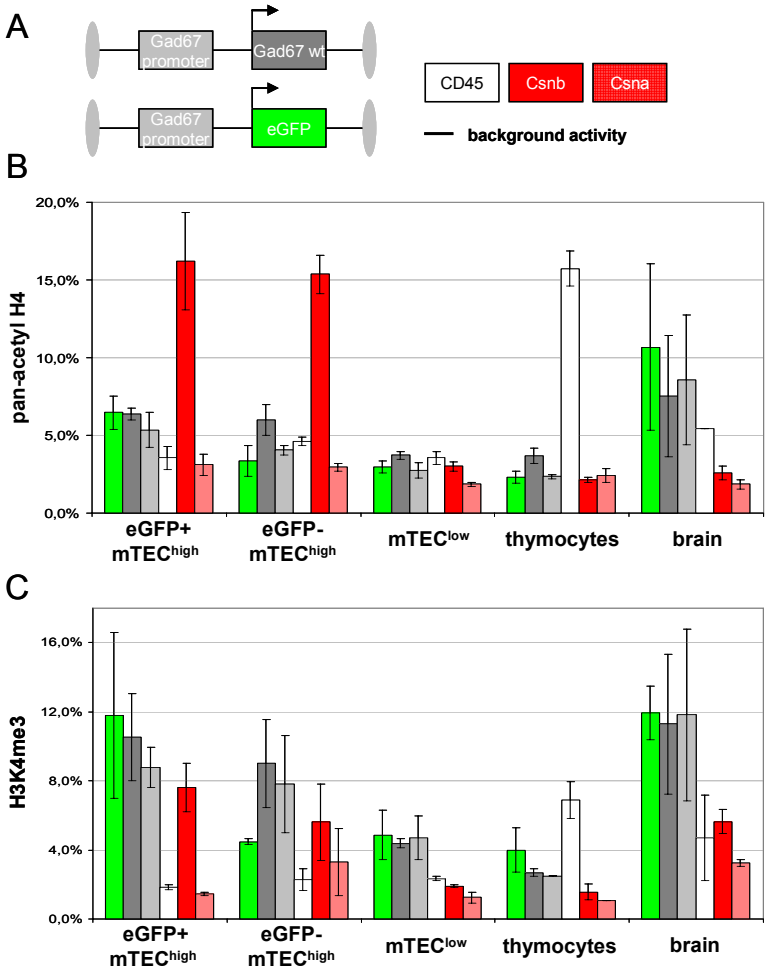


Figure 18

Gad67/eGFP region only carries H3K4 trimethylation but no H4 acetylation. ChIP was performed with freshly prepared thymocytes and TEC populations from heterozygous Gad67/eGFP mice. The common Gad67 promoter region and both alleles (Gad67 and eGFP) as well as three genes as internal controls for ChIP were analyzed independently with qPCR (A). Both mTEC^{high} populations show clear promoter acetylation of Csnb, but no significant modification in the Gad67/eGFP while Gad67/eGFP acetylation is present in brain (B). H3K4 trimethylation of eGFP is high in brain and increased in eGFP$^+$ over eGFP$^-$ mTEC. Gad 67 is H3K4 trimethylated at similar levels in both mTEC^{high} populations and in the brain. (C). All data in Figure 3 is representative of two or three independent experiments.

Some genes promiscuously expressed in mTEC were found to be expressed monoallelically [76]. This may also be the case for Gad67 and eGFP, therefore, Gad67 expression may be independent of eGFP expression. However, no significant differences are observed in active

chromatin marks between the Gad67 and the GFP allele in eGFP$^+$ mTEC^{high} which would suggest differential regulation between the two alleles.
In comparison with the casein locus one can conclude that eGFP expression, similar to Csnb expression is epigenetically regulated at the level of the individual gene in mTEC^{high}. But the type of regulation differs between Csnb and eGFP/Gad67: Both genes show different distributions of active chromatin marks, Csnb being marked with both acetyl H4 and H3K4me3 but Gad67/eGFP only with H3K4me3 in mTEC. H3K4 trimethylation in mTEC is dependent on gene expression for eGFP but independent of gene expression for Gad67. Csnb is similarly marked as in the tissue (MEC), but Gad67/eGFP is differently marked as it lacks H4 acetylation in mTEC but not in brain.

3.2 Expression patterns of a single promiscuously expressed antigen – lessons from the Gad67 locus using Gad67-eGFP knock-in mice

Most of the information generated on expression patterns and mechanisms of pGE stems from the casein locus [52, 60], therefore it is crucial to verify whether these findings also apply to a different genomic locus. Knowledge of the expression patterns of pGE in mTEC from a second genetic locus allows comparison between the two loci and provides a broader view of the regulation of pGE, especially since the analysis of heterozygous Gad67/eGFP reporter mice additionally allows for differentiation between mono- and biallelic expression patterns for Gad67. In the following chapter, pGE in the Gad67 locus will be analyzed in depth using gene expression analysis population wide and at the single cell level.

3.2.1 Co-enrichment for Gad67 expression in eGFP$^+$ mTEC^{high} on the population level

Gene expression in the Gad67 locus was first analyzed in eGFP$^+$ mTEC^{high}, eGFP$^-$ mTEC^{high} and mTEC^{low} cell populations sorted to high purity from Gad67/eGFP mice. eGFP was highly over-expressed in eGFP$^+$ vs. eGFP$^-$ populations as could be expected from the different eGFP protein levels in these populations. mTEC^{low} showed no eGFP expression at all. Gad67 was slightly co-enriched with eGFP in the eGFP$^+$ fraction which points towards biallelic, or at least partially biallelic expression (Figure 19). (Compared with the 60-fold upregulation of eGFP mRNA however, co-enrichment was low.) For more in-depth analysis, it was crucial to perform SC PCR which will give insights into the expression patterns in individual cells. Both biallelic (as in the tissue) and monoallelic expression patterns have been observed for different promiscuously expressed genes in mTEC [76]. We applied SC PCR to adress the question of mono- or biallelic expression in the Gad67 locus.

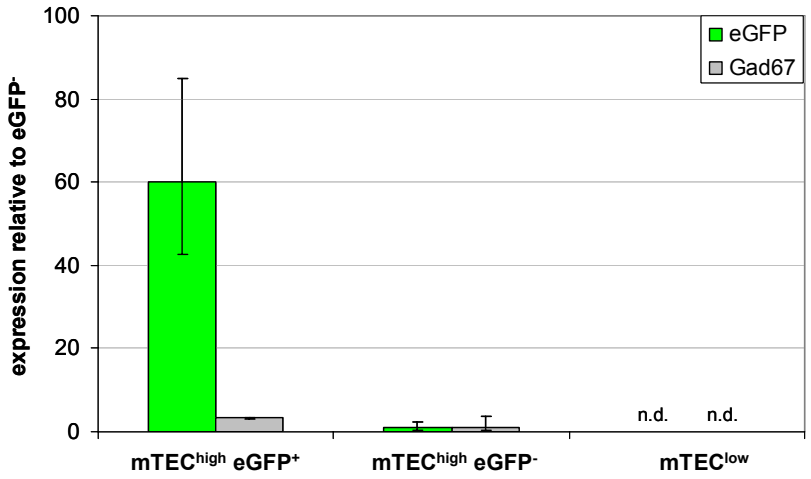


Figure 19
Slight co-enrichment of Gad67 expression in mTEC^{high} eGFP^{+} population. Expression analysis of sorted mTEC target populations from Gad67/eGFP mice. Expression analysis was performed using SC PCR primers A and C for Gad67 and eGFP expression and normalized to eGFP^{-} mTEC^{high}.

3.2.2 SC PCR analysis of the Gad67 locus shows decoupling between eGFP protein and mRNA expression

The Gad67 locus does not contain a particularly large gene cluster like the casein locus, but maps to a generally gene-rich region on murine chromosome 2. Neighboring genes – upstream the transcription factor Sp5 and the Riken clone 4933404M02Rik and downstream the Golgi reassembly-stacking protein 2 (Gorasp 2) – are not tissue specific and do not belong to a common gene family.

For more detailed analysis of gene expression patterns in this region, which has already been analyzed by ChIP, SC PCR analysis was performed. Gad67 and eGFP expression was examined, in addition genes directly upstream and downstream of Gad67 (4933404M02Rik and Gorasp2) were studied to check for potential co-enrichment of neighboring genes. Such "spreading" of actively transcribed regions hints towards epigenetic regulation[164] and had already been found in other cases: Carcinoembryonic antigen (CEA) positive sorted human mTEC co-express neighboring genes CEACAM6 and CEACAM3, similar co-enrichment of neighboring genes is found for Sglt1 in human mTEC [80]. SC PCR was furthermore performed for EpCAM, as a housekeeping gene, and Aire, as the single transcriptional regulator controlling pGE. With the SC PCR method, the expression of six genes, Gad67, eGFP, 4933404M02Rik, Gorasp2, EpCAM and Aire was analyzed in single cells from sorted eGFP^{+} mTEC^{high}, eGFP^{-} mTEC^{high} and mTEC^{low} populations from Gad67/eGFP mice.

276 eGFP^{+} and 275 eGFP^{-} cells were analyzed and as expected, almost all (97% vs. 92%) expressed the TEC marker EpCAM, proving the high sensitivity of the SC PCR method. Aire was also expressed at a high frequency (63% vs 72%) (Figure 20 B). These numbers correlated well with previously published data [60] and percentages of mTEC^{high} previously found positive for Aire protein in FACS analysis [65].

Surprisingly, eGFP expression was only detected in 33% of cells sorted positive for eGFP protein ($eGFP^{+}mTEC^{high}$) (Figure 20 B). Why protein and mRNA expression frequencies correlated so poorly for eGFP while they correlated well for EpCAM and Aire is unclear. On the other hand, in the population sorted as negative for eGFP protein ($eGFP^{-}mTEC^{high}$), still 10% of cells were found to express eGFP mRNA (Figure 20 B). Note that eGFP primers were highly specific as out of 36 $mTEC^{low}$ found positive for EpCAM, no cell was found positive for eGFP (Figure 20 B).

Gad67 is overall much less frequent than eGFP expression but strongly co-enriched in the $eGFP^{+}mTEC^{high}$ population (15% expression vs. only 6% in $eGFP^{-}$ $mTEC^{high}$) (Figure 20 B). The co-enrichment speaks for at least partial biallelic expression which will be analyzed more in detail in 3.2.3.

As Gad67 and eGFP expression are both driven by the same promoter, one would expect them to be expressed at the same frequency. However, eGFP expression frequency is about two-fold higher than Gad67 frequency in both $eGFP^{+}$ and $eGFP^{-}mTEC^{high}$ populations. Neither of the directly neighboring genes of Gad67, 4933404M02Rik and Gorasp2, are differentially expressed between $eGPF^{+}$ and $eGFP^{-}$ $mTEC^{high}$ populations ($p > 0.05$) (Figure 20 B). Thus, active transcription in Gad67/eGFP apparently does not have an effect on the expression of its directly neighboring genes. Gorasp2 is expressed in most cells (81% vs. 76%) while 4933404M02Rik is relatively rare and only expressed in 5-6 % of genes as it would be typical for a promiscuously expressed antigen in mTEC. As there is no information available on the expression patterns of 4933404M02Rik gene across different tissues, it is unclear whether it represents a typical tissue-specific promiscuously expressed antigen.

Altogether, protein and mRNA frequencies correlate well for "conventionally" expressed genes like EpCAM and Aire, while eGFP which is expressed as a TRA in the target cell populations, shows strong divergence between protein frequency and mRNA frequency. Gad67 is surprisingly expressed at lower levels than eGFP, but is co-enriched in the $eGFP^{+}mTEC^{high}$ population which suggests at least partially biallelic expression.

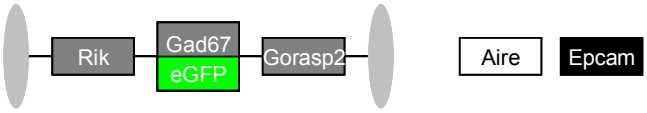
3.2.3 Strong preference for biallelic expression at the Gad67 locus

Villasenor et al.[76] have shown different promiscuously expressed antigens to be expressed both biallelically (as in the tissue) or monoallelically, thus differently regulated than in the tissue. Gad67 was found to be co-enriched with eGFP expression, thus suggesting at least partially biallelic expression. To resolve this question, SC expression data was examined for co-expression patterns.

Expression of Gad67 and eGFP is highly interdependent and shows a strong bias towards biallelic expression both in the $eGPF^{+}$ and the $eGFP^{-}$ cell populations (see Figure 20). A majority of cells expressing Gad67 co-express eGFP (73%) in the $eGFP^{+}$ sorted population, but nevertheless, both genes can also be expressed in a monoallelic fashion. Biallelic expression is thus favored, but not exclusive (Figure 20).

Co-expression patterns may of course also be influenced by different dynamics of RNA levels for the two genes. Such variances could bias the observed cell numbers co-expressing both genes as functional RNA from one allele may still be present in the cell after active transcription has already been turned off.

A



B

	mTEC^{high} eGFP^{+}	mTEC^{high} eGFP^{-}	P value	mTEC^{low}
Total cell #	276	275		36
Epcam	97%	92%	$2.5*10^{-3}$	100%
Aire	63%	74%	$6.3*10^{-3}$	6%
Rik	6%	5%	0.46	0%
eGFP	33%	10%	$<10^{-4}$	0%
Gad67	14%	5%	$4*10^{-4}$	0%
Gorasp2	81%	76%	0.14	25%
GFP/Gad67 coexpression:				
None	63%	87%	$<10^{-4}$ for eGFP^{+}	100%
eGFP only	22%	8%		0
Gad67 only	4%	3%	$1.7*10^{-3}$ for eGFP^{-}	0
eGFP and Gad67	11%	2%		0

C

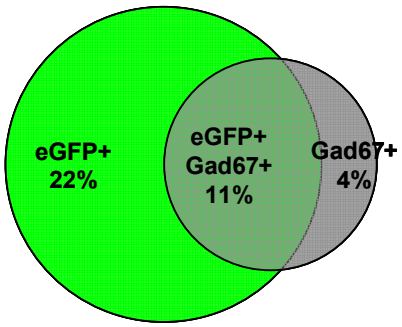


Figure 20

SC PCR analysis of the Gad67/eGFP locus in heterozygous Gad67/eGFP mice reveals low protein-mRNA correlation. Gad67/eGFP, two adjacent genes and Aire and Epcam as mTEC markers were analyzed in eGFP^{+} and eGFP^{-} mTEC^{high} with SC PCR (A). Sorting was done as shown in Figure 17. Only a partial correlation between eGFP protein and eGFP mRNA expression was observed. Gad67 was co-enriched with eGFP in the eGFP^{+} population (B). Co-expression analysis showed a strong bias towards biallelic expression in eGFP^{+}mTEC^{high}, although monoallelic expressors still existed (B and C).

3.2.4 Gad67/eGFP co-expressors are enriched in Aire$^+$ cells

Aire$^+$ mTEC^{high} have been described to be the most mature, terminally differentiated mTEC population which expressed most and the largest variation of promiscuously expressed antigens [65]. Therefore, we asked the question in which way Aire expression correlates with mono- or biallelic expression in the Gad67/eGFP locus on the single cell level. Control of Gad67 expression is independent of the transcriptional regulator Aire [75, 161]. Mature Aire$^+$ mTEC express both Aire-dependent and Aire-independent antigens [17, 60].

Transcriptional activity in Gad67/eGFP was separated into three groups: "no gene expressed", which stands for lack of gene expression in Gad67 and eGFP; "either Gad67 or eGFP expressed", equal to monoallelic expressors and finally "both Gad67 and eGFP expressed", equal to biallelic expressors. The percentage of Aire$^+$ cells among these three groups was analyzed and shown to increase significantly from low to high transcriptional activity ($p<<10^{-4}$, Jonckheere-Terpstra Test) (Figure 21). Cells expressing neither of the two genes were relatively depleted in the frequency of Aire expression: Only 63% of the group "no gene expressed" expressed Aire, compared to an overall frequency of 69%. The group of biallelic expressors was highly enriched in cells co-expressing Aire (91% Aire$^+$) (Figure 21), while the monoallelic expressors showed an intermediate percentage of Aire$^+$ cells.

Therefore, similarly to other Aire-independent genes, Gad67 and eGPF were also expressed preferentially in Aire$^+$ cells. The particular features of Aire$^+$mTEC thus seem not only to allow for the expression of Aire-dependent genes but also seem to be favorable for the expression of Aire-independent genes.

	% Aire+
all mTEC^{high}	69%
0 genes (neither Gad nor GFP)	63%
1 gene (Gad or GFP)	85%
2 genes (Gad and GFP)	91%

Figure 21

Gad67/eGFP are preferentially expressed in Aire$^+$ cells. Frequency of Aire expression was increased in eGFP$^+$ and/or Gad67$^+$ cells in spite of Gad67 being an Aire-independent antigen. Biallelic expressors showed a higher frequency or Aire expression than monoallelic expressors or cells expressing neither Gad67 nor eGFP ($p<<10^{-4}$, Jonckheere-Terpstra Test).

3.2.5 Gad67 and eGPF are co-expressed at the protein level in the brain

Expression patterns for Gad67 and eGFP mRNA were established in mTEC^{high} populations using SC PCR. As eGFP RNA expression was found to diverge strongly from the eGFP protein levels seen in FACS, eGFP and Gad67 co-expression was also examined on the protein level, in order to see a more complete picture of protein and mRNA dynamics in mTEC^{high} from both alleles in this particular locus.

Immunohistochemical staining of thymus and brain cryosections was performed with an anti-Gad67/65 antibody in heterozygous Gad67/eGFP mice. Crossreactivity with Gad65 of the antibody used for detection should not cause any problems as Gad65 expression in mTEC was found to be significantly lower than Gad67 expression[50]. No suitable specific anti-Gad67 antibody was available for staining.

On the protein level, eGFP$^+$ neurons strongly co-expressing Gad67 were found in the brain, especially in the cerebellum (Figure 22). Similar colocalization has already been observed in other brain regions of Gad67/eGFP mice [160]. Neither single eGFP$^+$ nor single Gad67$^+$ cells could be detected, thus expression in the brain must be biallelic.

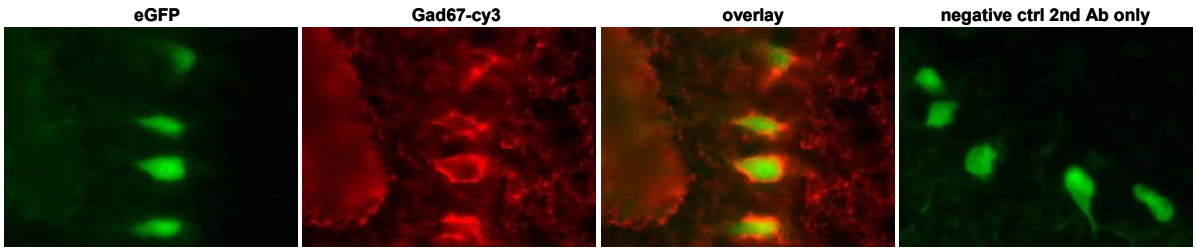


Figure 22
Staining of eGFP protein and Gad67 in murine cerebellum. Images of murine cerebellum from sagittal 5 µm brain sections.

eGFP staining in the thymus, although easily detectable in FACS analysis of enriched mTEC populations (Figure 17), was confounded by strong autofluorescence of phagocytosing macrophages in the green channel in histology, making rare eGFP$^+$ mTEC hard to identify. As Gad67 levels are probably much lower in the thymus than in the brain and Gad67$^+$ mTEC will be very rare on whole thymic sections, thus difficult to find, significant Gad67 staining could not be detected. Fixation of the brain was found to be critical as optimal preservation of the antigen for high-quality staining could only be obtained with PFA perfusion of mice and postfixation in PFA of the respective organs over night. Direct PFA perfusion could not be performed for mTEC which were isolated from the tissue, sorted and put on cytospins as the enrichment protocol for mTEC requires unfixed tissue. Thus, eGFP signals were found to be much lower on these cytospins, possibly even due to eGFP bleaching after FACS sorting of cells. The less optimal fixation method (PFA fixation of cells directly prior to staining) did not allow for efficient staining with the Gad65/67 antibody.

3.3 Nuclear positioning and DNA compaction of the casein locus during mTEC maturation

Measurement of local epigenetic modifications such as DNA methylation or histone tail modifications, as it was done in this study, shows only part of the possible epigenetic regulation mechanisms. The position of a locus in a particular environment in the nucleus and the overall compaction of chromatin around a target locus are other epigenetic parameters which can influence gene accessibility and gene expression. Preparation of a locus for active transcription can correlate with decompaction of chromatin to make it more accessible for the transcription machinery [129, 132]and with relocalization of genes in the nucleus towards the nuclear interior [119, 120, 122, 165, 166]. Nevertheless, a link between radial position and transcriptional activity has not consistently been found [108]. Concerning the compaction of chromatin, inactive chromatin was found to be tighly packed while actively transcribed regions tend to be more decondensed and thus more easily accessible.

Whether parameters describing the nuclear positioning and compaction of DNA also play a role in pGE will be further examined here, using fluorescence in situ-hybridization (FISH) and spectral precision distance microscopy (SPDM). Once again, the casein locus will be examined as a model locus for pGE.

Chromatin in the target area was marked with fluorescent probes and both parameters, a) localization of the probes within the nucleus, compared to a nuclear counterstain, and b) decompaction of the locus, i.e. distance between the two probes were measured with high resolution (below the resolution limit of conventional microscopes) using SPDM.

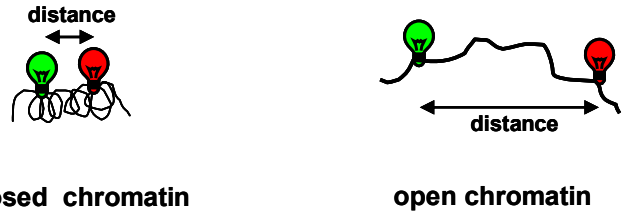


Figure 23
Localization of probes in open and closed chromatin conformation. Upon activation of transcription in a particular locus, the locus is expected to decompact, resulting in better accessibility for the transcription machinery. This may result in a larger distance of two fluorescent probes, located upstream and downstream of the genetic region of interest.

It was to be analyzed whether chromatin structure in the casein locus changes in mTEC compared to casein non-expressing cells (thymocytes) and MEC, which express high levels of caseins. As casein genes are only promiscuously expressed in mTEC^{high}, it is also of interest whether the changes in chromatin structure can only be observed in mTEC^{high} or whether the chromatin in mTEC^{low} is potentially already prepared (or "primed") for later expression.

3.3.1 Two probes are localized upstream and downstream of Csna/Csnb to measure chromatin decompaction

Two probes within the casein gene locus were chosen for distance measurements: The probes were 30 kb (Probe 1) and 46 kB (Probe 2) in size with 65 kb distance between them which encompasses the Csna and Csnb gene. The true distance between the bary centers of the two probes was 103 kB (Figure 24). All other casein genes were located downstream of Probe 2. As Csnb is expressed in most mTEC^{high}, it is an optimal target gene as it is sufficient to sort complete mTEC^{high} to obtain a high enrichment of Csnb expressors. Probe sizes were chosen to be as small as possible but still large enough to be labeled with a sufficient number of fluorochromes for detection. Genomic distances were chosen to give small geometric distances which are still well above the resolution equivalent of multispectral SPDM measurements [137]. A larger distance between the probes would increase the risk of measuring not the true length (and compaction) of the locus but twisted/coiled chromatin strands resulting in a higher apparent compaction. The probes were labeled with two different colors, Oregon Green and Alexa-Fluor-647, cells were counterstained with HOECHST 33342 for nuclear localization measurements.

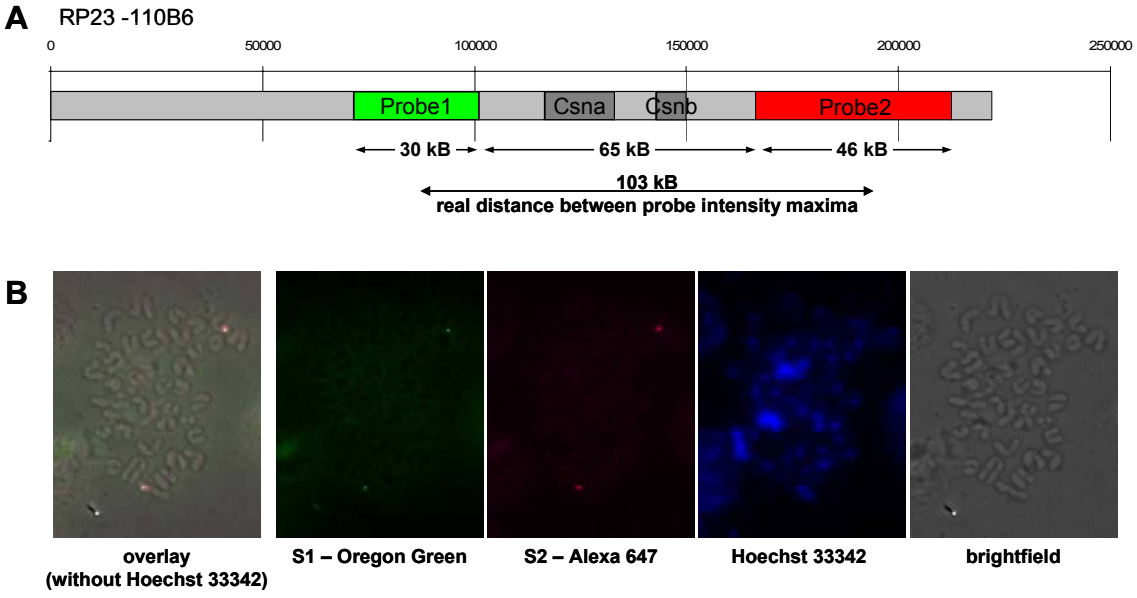


Figure 24
Probes are localized upstream and downstream of Csna/Csnb on mouse chromosome 5. 30 kB Probe 1 and 46 kB Probe 2 are constructed from 220kB BAC clone (RP23 -110B6) (A). Probe 1 is labeled in green with Oregon green fluorophore (Invitrogen), Probe 2 is labeled in red with Alexa –Fluor 647 fluorophore (Invitrogen). Probes are hybridized on metaphase spreads of EL4 murine lymphoma cell line. Note that the two probes colocalize very closely on the same chromosome, below the resolution limit of the microscope (B).

3.3.2 Both decompaction of the locus and nuclear localization can be measured with the same probes

With the same set of images, both the distance between Probe 1 and Probe 2, a measurement for decompaction of the locus, as well as the absolute and normalized distance of Probe 1 to the rim of the Hoechst 33342 counterstain can be acquired. All four channels are acquired sequentially as

to avoid spectral overlap between signals. Each probe gives two signals within each nucleus. The Probe 1 – Probe 2 signal pairs which belong to the same chromosome and which must be used for distance measurements can be easily identified without the need for extra chromosomal counterstains: As the distances between the two probes are below the resolution of the microscope, the pairs look as if they would overlap (Figure 25 A, B). Tetraspeck™ beads present on the samples for calibration can be distinguished from probe signals by their fluorescence in the red channel (excitation at 568 nm) and are mostly found to be much brighter and larger than FISH probes (Figure 25 A, B).

In order to obtain the 3D localization of the probes within the cell in μm, the raw images (shown in Figure 25 A, B) were processed with image analysis software in Matlab®. The software identifies and maps point-like objects and counterstain as shown in Figure 25 C. The 3D intensity profiles of the points identified are then fitted to a multipoint model which will lead to lower background fluorescence and better localization of the signal (Figure 25 D) All fits the software performed were controlled visually and only points which had been fitted sufficiently well (compare Figure 25) were taken into the calculation, as a poor fit will lead to an error-prone point localization.

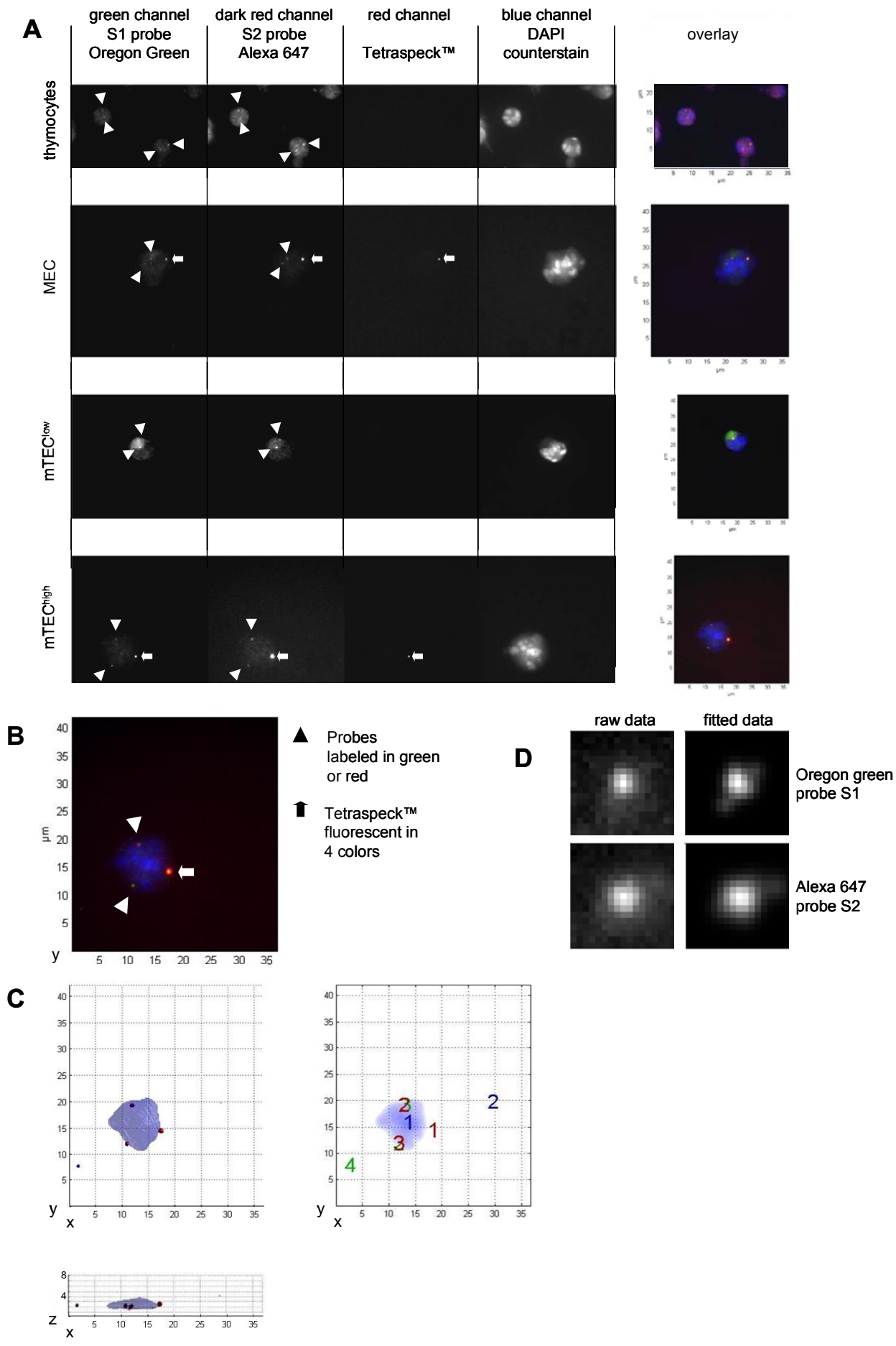


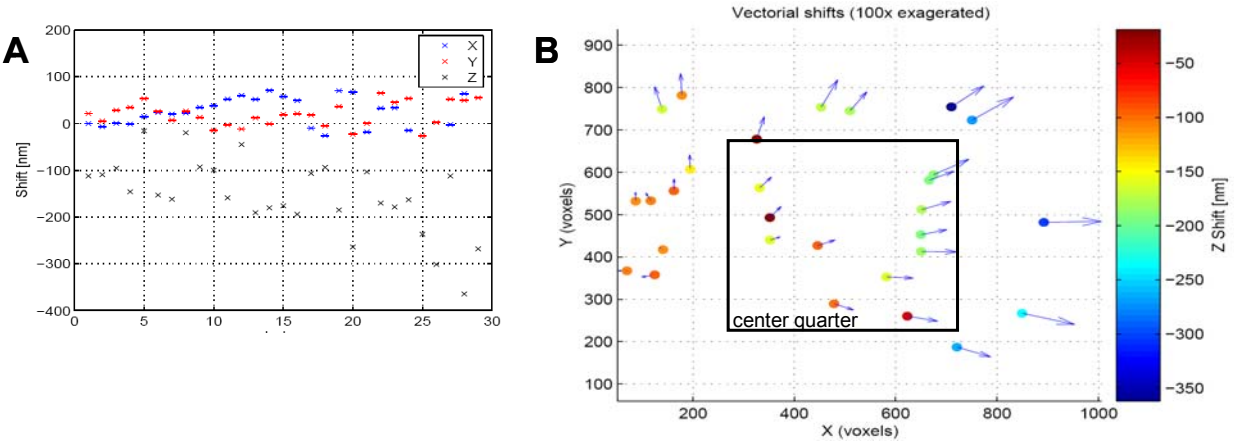
Figure 25 Image acquistion and data processing of FISH probes

Figure 25

Image acquistion and data processing of FISH probes. Compiled Z-stacks for each individual channel are shown for all four cell types analyzed. Each probe gives two signals inside the nucleus (see dark red and green channel), additional signals stem from beads fluorescent in four colors which can be identified by their additional fluorescence in the red channel (568 nm excitation wavelength) (A and B) Note that the two probes colocalize very closely and can hardly be distinguished by eye as the distance between them is below the microscopic resolution. Points and counterstain are identified in an image processing software in 3D, points and cells are numbered and xyz coordinates as well as localization to the nuclear rim determined for each point (C). Finally, distance between the points (i.e. green 3 – red 2 and green 2 – red 3) is calculated and corrected with the chromatic shift determined with the help of the bead (C). Example for original data fitted to a multipoint model (D).

3.3.3 Chromatic shifts between Oregon Green-Alexa Fluor 647 are small when directly measured in parallel with probes

For accurate distance measurements between probes with different spectral signatures, as they are used here, careful correction of chromatic shifts is necessary. Chromatic shifts occur due to imperfect microscope calibration and lead to differences in localization of a signal depending on the wavelength its fluorophore emits. The location of an object appears to change between different wavelengths. Chromatic shifts can be as large as several hundred nanometers, especially along the Z axis, thus their correction is particularly important for the measurement of small distances. In order to visualize the variation of chromatic shifts over the entire field of view, images of Tetraspeck™ beads which are fluorescent in four different colors were acquired and the chromatic shifts between the green channel (488 nm laser) and the dark red channel (647 nm laser) were calculated and visualized using Matlab software. The chromatic shifts vary considerably over the entire field of view of the microscope but are generally smaller and less variable in the center quarter of the field of view (Figure 26 A, B). Cells were therefore always placed in the center quarter before images were acquired. As the chromatic shifts within this center quarter still vary, it is crucial to correct the measurements with chromatic shifts which were determined under the same conditions and at the same location as the signal. Therefore, Tetraspeck™ beads were placed on every sample and chromatic shifts for correction were measured from beads on the same sample as close to the object of interest as possible. These precautions assured that: a) Chromatic shifts were overall smaller during the measurements in the center quarter than the ones acquired over the entire field of view (Figure 26 C). They tend to be more variable as the refractive index of the sample near cells is more heterogeneous than when beads are measured alone. b) The shift affecting the calibration object could be assumed to be very similar to the shift affecting the probe of interest, as both objects were close to each other. While the correction of chromatic shifts was essential for distance measurements, all chromatic shifts were small in relation to the size of the entire nucleus, hence no correction of chromatic shifts was necessary for localization measurements.



C

	Beads only, acquired over entire field of view		Beads directly on sample, acquired in the center quarter	
	mean (nm)	St. Dev. (nm)	mean (nm)	St. Dev. (nm)
Δx	33,9	36,96	17,70	37,54
Δy	19,28	25,08	-48,65	53,13
Δz	-155,29	79,56	-38,20	106,94

Figure 26

Chromatic shifts between green (488 nm excitation) and dark red (647 nm excitation) were overall smaller in the center quarter of the field of view. Fluorescence of Tetraspeck™ beads was acquired and localization differences between the different fluorescent channels were mapped via a program written for Matlab software as described in 2.2.1.8. All chromatic shifts are measured in the entire field of view (A) and are depicted according to their localization in the field of view (B) are depicted. Note that the shifts are overall lower in the center quarter of the field of view. Average shifts measured over the entire field of view were thus larger than shifts measured from Tetraspeck™ beads which were put on the samples together with target cells and exclusively measured in the center quarter of the field of view (C).

3.3.4 Chromatin in the Casein locus is more compact in mTEC^{low}

The distances between the Probe 1 and 2 were measured for all four cell types: thymocytes and mTEC^{low}, which do not express Csnb, MEC, expressing high levels of Csnb and mTEC^{high} which express lower levels of Csnb per cell. An average of 35 probe pairs was acquired per cell type and the distances corrected by their chromatic shifts and averaged in 3D and in 2D (table 3.1). The analysis of 2D distances is recommended when larger distances are measured and the cells analyzed are very flat. In such a case the z-axis could contributes only to a lesser extent to the measurement. For all four cell types, the 2D distances were significantly smaller than 3D distances, thus there is a crucial contribution of the z axis to all distances. Therefore the 3D distances are used for analysis – cells were found sufficiently round after fixation and during image acquisition considering the distance between the two probes which is relatively small (several 100 nms).

Thymocytes, MEC and mTEC^{high} were found to have highly similar distances (ca. 360 nm, p~1) between the two probes and thus to have similarly compact chromatin in the casein locus. In mTEC^{low} on the other hand, a smaller distance (237 nm) was measured which differs significantly from all the other samples (table 3.1). Pairwise comparisons with Tukey's HSD [167] gave p-values between 0.02 and 0.05 for the three sample pairs (mTEC^{high}-mTEC^{low}, thymocytes-mTEC^{low}, MEC-mTEC^{low}). As distances in mTEC^{high}, MEC and thymocytes are to nearly 100% (p~1)

similar, one can consider a p value between 0.02 and 0.05 for the comparison with $mTEC^{low}$ as significant.

		thymocytes	MEC	$mTEC^{high}$	$mTEC^{low}$
# of distances analyzed		44	36	35	25
mean (nm)	3D	361	353	359	237
Σ (nm)		193	154	186	140
error of the mean (nm)		29	26	31	28
mean (nm)	2D	247	160	226	87
Σ (nm)		182	149	180	113
error of the mean (nm)		27	25	30	23

Table 3.1
Chromatin compaction in the casein locus. Distances between Probe 1 and Probe 2 in the casein gene locus measured in 2D and 3D, depicted for the four cell populations: thymocytes, MEC, $mTEC^{high}$ and $mTEC^{low}$.

Apart from the analysis of the mean distance between the two points, which is a measure of the average chromatin compaction in the target locus, it is also important to check whether multiple different states of chromatin, consequently multiple subpopulations are present within any of the samples. This would be even more likely for mTEC populations which tend to be highly heterogeneous with respect to gene expression. The examination of dot plots showed for none of the target cell populations multiple maxima in the distribution of distances. Therefore, there is no indication for different subpopulations within any of the target populations (Figure 27 A), thus only one type of chromatin compaction status seems to be present in each cell type.
The variance of the distance datasets also did not vary between the four target cell populations (ANOVA F-test), thus chromatin structure was equally heterogeneous in all four populations. The variance of a data set can be detected as the slope in a plot of cumulative frequencies (Figure 27 B), where a steep slope is an indicator for less variable data distribution and a flat slope indicates high variation, accordingly. The cumulative frequency plot in this case shows similar slopes for all four cell populations, but significantly smaller distances between the two probes for $mTEC^{low}$ (Figure 27 B).

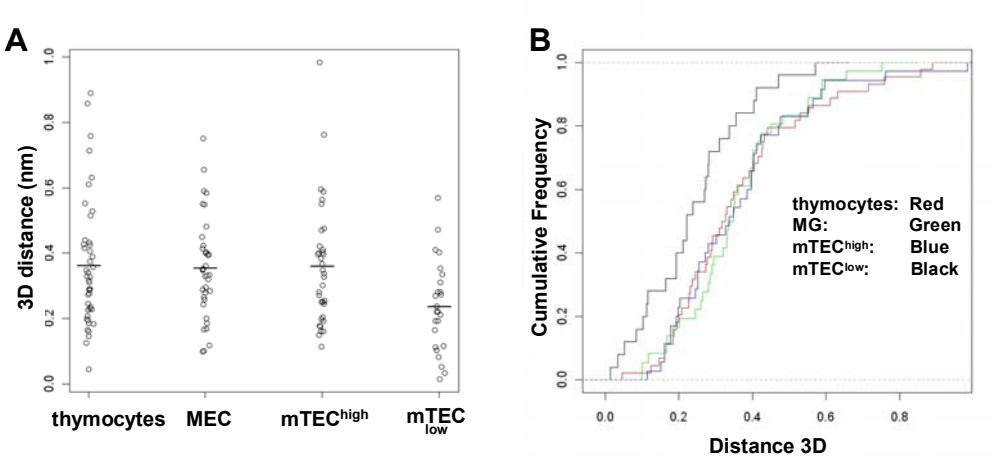


Figure 27

Distribution of distances does not show any local maxima or differences in heterogeneity between the four cell populations. Distances were plotted for all four cell populations and were found not to have multiple maxima, indicative of the presence of different subpopulations (A). Cumulative frequency plots showed similar slopes of the curves for all four cell populations, indicating similar degrees of heterogeneity for data from all four cell populations (ANOVA F-test) (B).

The distance between Probe 1 and Probe 2 in this experiment is a direct measure for chromatin compaction in the region between the two bary centers of the fluorescent probe signals. The true bary center of the probe fluorescence depends on the distribution and intensity of fluorochrome within each probe and thus cannot be determined here. Instead, the center of the maximal fluorescence intensity is used to measure the the exact geometric localization of each probe. The smaller the geometric distance between the two probes, which are set at a genomic distance of 103 kB (between the two centers of the probes) (Figure 24), the more compact the chromatin in the locus must be, assuming that all chromatin is arranged along the linear axis between each two target points. Of course the longer the distance between the two probes, the more coiling and folding of chromatin will occur, therefore the calculation of the chromatin compaction factor is only an approximation which works optimally for very small genomic distances.

The chromatin compaction factor is a set measure for comparing the geometric distances / extension of chromatin (in nms) with genomic distances (in kB). The geometric distance of 340 nm for 1 kB genomic distance corresponds to a chromatin compaction factor of 1, synonymous with maximally extended, decondensed DNA. The straight 30 nm chromatin fiber already has a compaction factor of 40 with 8.5 nm per 1 kB [137].

The casein locus between the two probes, around the Csnb gene, is found to be moderately compact with compaction factors around 100, thus 2.5 fold higher than an elongated 30 nm fiber. This speaks for the presence of higher order chromatin structure in the locus which allows for higher compaction factors. Higher order structure does not exclude the presence of shorter, highly decondensed stretches of chromatin, i.e. in promoter regions or at active transcription start sites.

The casein locus in mTEC^{low} is significantly more compact with a compaction factor of about 150 (Figure 28).

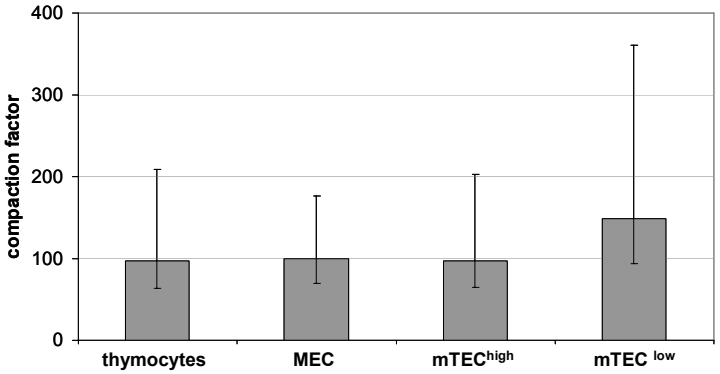


Figure 28
Chromatin compaction is equal for thymocytes, MEC and mTEC^{high} but higher for mTEC^{low}. mTEC^{low} show significantly higher chromatin compaction factors than the other three cell types analyzed. A compaction factor of 40 corresponds to the packing of a typical 30 nm chromatin fiber.

Altogether, chromatin compaction is estimated to be similar in thymocytes, mTEC^{high} and MEC, while only the latter two express genes from the casein locus. Chromatin in the casein locus of mTEC^{low} was found to be significantly more compact.
Why chromatin is more condensed in mTEC^{low} than in thymocytes is unclear, as both cell types lack casein gene expression. Thymocytes possibly differ biologically in cellular and chromatin structure from the other cell types analyzed as these are all of epithelial origin. Therefore, thymocytes may not be an optimal cell type for a negative control in this experiment. Another epithelial cell type with more similar cell structure also lacking casein expression may have been more appropriate.

3.4.5 The casein locus is localized quasi-randomly in the nucleus in all cell populations analyzed

Apart from the level of chromatin compaction, nuclear localization can be an indicator of transcriptional activity of a gene locus. Inactive gene loci have often been found nearer to the nuclear membrane while actively transcribed loci were often found closer to the nuclear center [108, 119, 120, 122, 165, 166]. Probe 1 in Oregon Green, as a measure for localization of the casein locus, was mapped in the nucleus with respect to Hoechst 33342 counterstain in 15 – 30 cell nuclei per cell population analyzed. Localization data was acquired in 2D and 3D and normalized to the radius of the nucleus. Localization is indicated as radial distance from the nuclear rim, thus a localization of zero corresponds to a probe at the nuclear rim while a probe in the center of the nucleus will give a localization signal of one. Normalization is essential as nuclear sizes between different cell types may vary and distort the measurement. The localization was again performed in 2D and in 3D, while always the shortest distance to the rim is given. The bias towards smaller distances in 3D is a clear indicator for flatter or ellipsoid cellular – and nuclear – shape, which was present for mTEC and MEC. In such nuclear shapes the distance in z to the nuclear rim is often relatively small compared to the xy-distances. This leads to an overestimation of the z-axis and confounds the measurements of nuclear localization. Therefore it is preferable to use the 2D localization for all cases where nuclei are not perfectly round. The chemical treatment of nuclei during FISH analysis may introduce additional bias into the measurement: Different preparation and fixation

techniques will almost certainly lead to subtle differences in the optical characteristics. While methanol/acetic acid (MAA) fixation is known to result in changes of the morphology of the interphase cell nucleus, it remains the standard technique for the preparation of metaphase chromosomes. MAA fixation results in an increase of nuclear diameter and a flattening of the nucleus compared with PFA fixed nuclei which have a much better preserved nuclear structure. Treatment with PFA also stabilizes against heat distortion of chromosomal structure during the hybridization steps [147]. As all cell types were subjected to the same PFA treatment and were controlled for proper 3D structure after image acquisition, subtle changes induced by the fixation method should not have an influence on the results. Distances or localization within MAA-fixed metaphase chromosome spreads were not measured.

The cellular 3D structure plays a larger role for localization data than it does for distance measurements, as the nuclear shape mainly influences measurements which are based on nuclear counterstain. Therefore, using 2D localization values may be more accurate and account better for the entire spectrum of localization possibilities (table 3.2).

Localization of Probe 1 in 2D as a measure for localization of the casein locus was not found to differ between any of the cell populations analyzed. For all populations, a localization of approx. 0.4, thus in an intermediate position almost half way to the center of the nucleus, is detected. The standard deviation of the data is very high indicating a high variance in the localization of the probe (table 3.2).

		thymocytes	**MEC**	**mTEC^{high}**	**mTEC^{low}**
# of probes localized		15	29	29	15
mean (normalized)	3D	0.32	0.29	0.34	0.26
Σ		0.21	0.20	0.19	0.14
error of the mean		0.5	0.04	0.04	0.04
mean (normalized)	2D	0.40	0.43	0.40	0.43
Σ		0.25	0.25	0.22	0.27
error of the mean		0.06	0.05	0.04	0.07

Table 3.2

Nuclear localization of the casein locus. Localization of Probe 1 with respect to Hoechst 33342 nuclear counterstain is measured and, normalized to the absolute size of the nucleus. A normalized localization of zero corresponds to the nuclear rim, a localization of one to the center of the nucleus. Localization is measured in 2D and 3D, depicted for the four cell populations: thymocytes, MEC, mTEC^{high} and mTEC^{low}.

Similarly to the data for chromatin compaction, two other parameters of the localization data sets are analyzed: a) The distribution of the different localizations measured, to check whether multiple maxima, thus multiple different states of the cell (or of the casein locus) can be detected (Figure 29 A) and b) the variance of the data acquired as to compare the heterogeneity of localization between the different cell populations (Figure 29 B).

No significant differences for any of these parameters can be observed between the four cell populations, neither in 2D nor in 3D analysis. The casein probe was found to be distributed quasi randomly throughout the nucleus with localizations occupying almost the entire radial spectrum

(Figure 29 A). The flat slope of all cumulative frequency curves is another measure which exemplifies the high variance of the localization data for all four cell populations (Figure 29).
Overall, the casein locus does not exhibit a preferential location within the nucleus in any of the cell populations analyzed. It does also not change its location upon activation as similar values were found for MEC and mTEC^{high} (expressing caseins to different degrees) compared to thymocytes and mTEC^{low} (not expressing caseins).
Other than for the compaction of chromatin, which was found to be significantly increased in mTEC^{low} alone, radial distribution in the nucleus does not appear to have an effect on the transcriptional state in the casein gene locus.

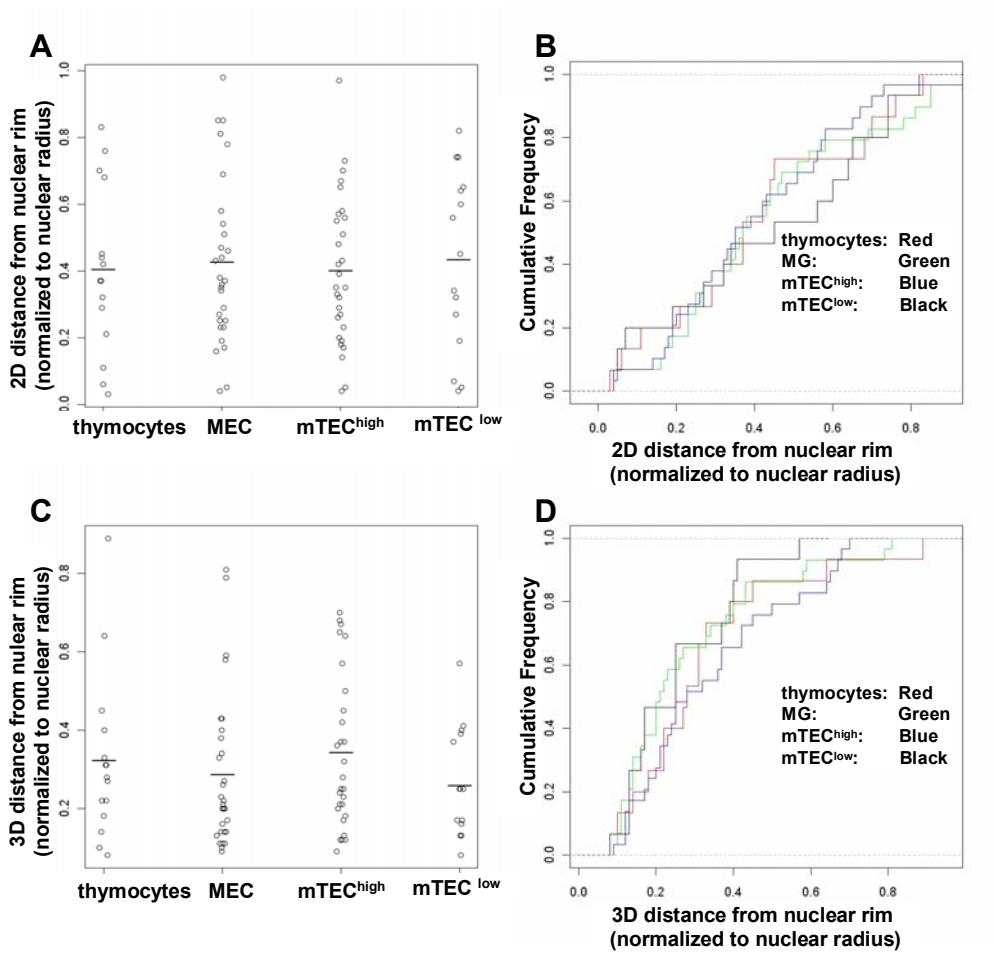


Figure 29

The casein locus is localized quasi randomly in the nucleus for all cell populations analyzed. Localization of Probe 1 with respect to Hoechst 33342 nuclear counterstain is measured and normalized to the absolute size of the nucleus. A normalized localization of zero corresponds to the nuclear rim, a localization of one to the center of the nucleus. Probe 1 is localized quasi homogenously throughout the entire nucleus in 2D (A) and more focused towards the rim in 3D (C). Localization does not differ between any of the cell populations. Cumulative frequency plots showed similar slopes of the curves for all four cell populations in 2D and 3D, indicating similar degrees of heterogeneity for data from all four cell populations (ANOVA F-test) (B, D).

4. Discussion

Promiscuous gene expression encompasses an ever increasing number of genes from virtually all tissues of the body. The group of Aire-dependent genes alone was recently found to comprise several thousand, making the impact of Aire on transcription 5-8 fold larger than that of other classical transcription factors [74]. Due to the number of affected genes it is unlikely that Aire will act primarily by binding to specific motifs in the DNA sequence. These motifs would need to be overly abundant, hence other aspects of regulation must be involved. Aire has been implicated in different processes of the regulation of pGE as diverse as transcriptional elongation, splicing, nuclear organization and recognition of epigenetic marks [74, 82, 83, 168, 169]. Significant interindividual differences have been found for the expression of Aire-dependent genes not only in human [58], but also in in-bred mice of the same strain [73]. About two thirds of differences in Aire-dependent promiscuous gene expression are thought to be of genetic origin, while one third is believed to be caused by stochastic variations. Either stochasticity in the activity of Aire or epigenetic effects were suggested to account for the intricate patterns of pGE [73], however it is not known to date which exact mechanism regulates the promiscuous expression of both Aire-dependent and Aire-independent TRA in mTEC.

Genes expressed promiscuously in mTEC were shown to be highly clustered in the genome, even more so the subpopulation of Aire-dependent promiscuously expressed genes [51, 52, 59]. The idea that genes with similar expression patterns would be arranged in clusters on the genome and coordinately regulated had been proposed previously [164, 170]. Such genes were believed to be co-regulated by changes in chromatin structure across large genomic regions [170]. For large gene clusters, for example the Hox gene cluster, such mechanisms of regulation seem to exist [171]. This concept was extended to clusters of differentially regulated genes (Figure 30). Clusters of genes differentially regulated in different tissues are co-expressed in mTEC at the population level, thus the expression of such clusters was hypothesized to be coordinately regulated employing epigenetic mechanisms [17] (Figure 30). This hypothesis could not yet be confirmed. Genes from one cluster are rarely co-expressed at the single cell level [60, 62], but entire loci could still be "opened" epigenetically and poised for transcription.
In this study, we focused on the implication of different epigenetic mechanisms in the regulation of pGE, which will be discussed in the following chapter.

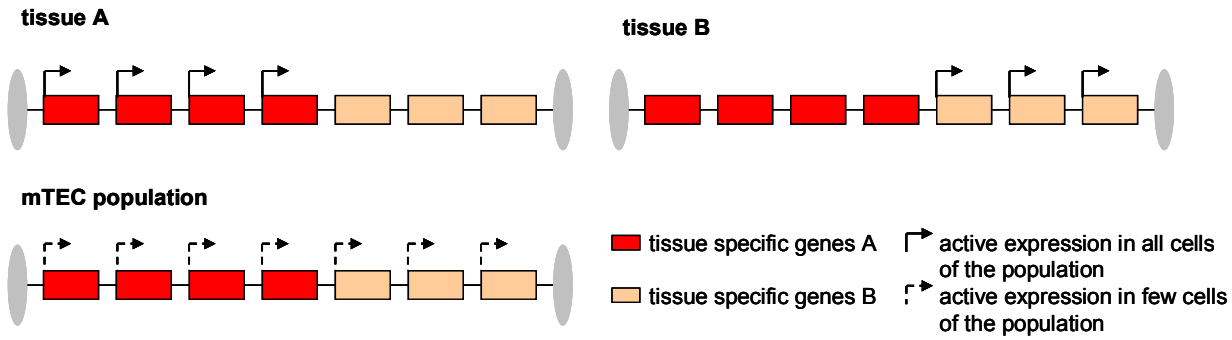


Figure 30
Clusters of tissue specific genes are co-expressed in the mTEC population. Different genes from a cluster of tissue restricted genes, isolated by boundary elements, are expressed in tissue A or tissue B. In the mTEC^{high} population, the entire cluster is co-expressed at the population level. This oberservation has lead to the hypothesis that the expression could be coordinately regulated. In single mTEC, individual genes from the cluster or combinations of genes will be expressed. These add up to full co-expression of the cluster on the population level.

While entire gene clusters were found to be co-expressed at the level of the mTEC population, gene expression at the single cell has been found to be "noisy" and probabilistic, meaning that a) gene expression is highly variable between cells, between individual thymic lobes and between genetically identical mice [73], b) genes on different alleles, within a certain cluster as well as across different clusters are expressed in a stochastic fashion [60, 76]. Aire is necessary on the single cell level for the expression of Aire-dependent genes, but not sufficient. Which other mechanisms are decisive for the expression of a particular antigen in an individual cell remains obscure.
Stochastic gene expression, as we observe it in pGE, has been proposed to be caused by slow transitions from inaccessible to open promoter regions. These transitions may be slow due to the high energy needed to transform an epigenetically closed into an accessible locus. Thus the probabilistic aspects of pGE on the single cell level may stem from intermediate states of transcriptional activation where chromatin is not completely and properly remodelled [76, 172, 173]. Epigenetic processes have also been specifically linked to pGE in a number of other ways (see 1.2.2).
Genome wide studies of epigenetic marks in mTEC have proven to be difficult: General epigenetic changes in the mTEC^{high} population as such would be detectable, but identification of the additional effects caused by 100-200 promiscuously expressed genes in a single mTEC^{high} has been precluded by a "background" of about 5000 ubiquitously and lineage-specifically expressed genes: In a study of genome-wide promoter DNA –methylation in the human system we could not find any significant differences in the overall promoter methylation state between mTEC^{low} and mTEC^{high}, or between mTEC and thymocytes. But when individual loci were examined, significant differences could be detected between TEC and thymocytes (L. Tykocinski, unpublished results).
In the study presented here we focused on the analysis of epigenetic marks at distinct, promiscuously expressed genetic loci. Particular focus was set on the Casein gene locus, where we analyzed higher-order chromatin structure by two parameters, namely chromatin compaction

and radial positioning of the locus in the nucleus. Furthermore we characterized the distribution of histone modifications at the promoter regions of target genes in the casein locus. This data was compared with the analysis of histone modifications in the Gad67 locus.

4.1 PGE is regulated by multiple epigenetic processes

4.1.1 Epigenetic opening occurs at the level of individual genes, not at the level of entire clusters

In order to assess the effect of histone modifications on gene expression in the casein locus in mTEC, we performed ChIP analysis and quantified the extent of H4 acetylation, H3K4 trimethylation and H3K27 trimethylation in the promoter regions of a representative group of genes in the mouse casein locus in $mTEC^{high}$ and $mTEC^{low}$. We did not find histone marks characteristic for open chromatin structure in the immature $mTEC^{low}$, where the casein genes are not expressed yet (Figure 15). This finding speaks against epigenetic opening of the locus prior to gene expression via the modification of histones. In $mTEC^{high}$ we found only the Csnb promoter to carry significant marks of acetyl H4 and trimethyl H3K4 while all other genes were unmodified. Previously, Csnb was found to be the only gene expressed to high frequency (over 80% of $mTEC^{high}$ express Csnb [60]). Thus we can conclude that Csnb assumes a special position within the casein gene locus, it is also the only gene for which both H4 acetylation and H3K4 trimethylation, two marks for active chromatin, could be detected (Figure 15). All other genes in the locus which are expressed at a much lower frequency, were not modified above background. In MEC, all casein genes are highly expressed and accordingly carry high levels of both histone acetylation and H3K4 trimethylation while genes specific for other tissues are not modified. Although the MEC population was enriched to only 50% purity, active chromatin modifications are clearly detectable (Figure 15) as the expression frequency is above the ChIP detection threshold. In the tissue we do thus find a clear cut correlation between expression and active histone modifications.

The lack of activating histone modifications at the promoters of genes expressed at a low frequency in $mTEC^{high}$ allows two possible conclusions: Either, the histone modifications analyzed do not play a role for pGE of these particular genes in the casein locus. Regulation would thus differ from the tissue, the mammary gland, where all casein genes have both H4 acetylated and H3K4 trimethylated promoter regions. We could show that the ChIP experiment has a detection threshold of up to 20% (Figure 16), thus a minimal frequency of modifications of 20% in the population analyzed is required for detection. All genes in the locus but Csnb are expressed below this threshold, therefore it is equally possible that each gene carries active chromatin marks only in those cells which express the gene, but these marks cannot be detected.

Population-wide histone modifications of all genes in the locus, leading to epigenetic opening of the entire cluster independent of the expression state on the single-cell level, were not detected in mTEC^{high}. Such modifications can be detected by the ChIP experiment, as shown by the example of Csnb.

As the detection threshold of ChIP will be problematic for the identification of histone modifications at any promiscuously expressed antigen, we chose the model of Gad67/eGFP *knock in* mice which permits the enrichment of mTEC expressing a particular antigen, namely Gad67, via the eGFP fluorescence.

The Gad67 gene was shown to be regulated differently by histone modifications than Csnb. eGFP$^+$ and eGFP$^-$ mTEC^{high} were sorted and histone modifications analyzed together with control populations (mTEC^{low}, thymocytes, eGFP$^+$ brain cells). Only eGFP$^+$ mTEC^{high} were H3K4 trimethylated in the eGFP gene but both eGFP$^+$ and eGFP$^-$ mTEC^{high} from Gad67/eGFP mice exhibited significant H3K4 trimethylation in the promoter region of Gad67. This implies that expression of eGFP is regulated by H3K4 trimethylation at the individual gene level, just as Csnb. Differences in this histone mark were clearly detectable despite the relatively poor enrichment of eGFP expression frequency in eGFP$^+$ vs. eGFP$^-$ mTEC^{high} (33% vs. 10% (Figure 21B)). The threefold enrichment in H3K4 trimethylation correlates well with the threefold increased expression frequency of eGFP in the eGFP$^+$mTEC^{high} population.

In eGFP$^+$ neurons, expression of eGFP and Gad67 was regulated differently from mTEC, namely by both H3K4 trimethylation and H4 acetylation (Figure 18). Similar degrees of modification at the Gad67 and the eGFP gene suggest identical regulation mechanisms in both alleles and thus biallelic expression. Co-expression of Gad67 and eGFP protein in neurons of Gad67/eGFP mice has been shown previously [160].

It is surprising that the same active mark at the Gad67 gene was equally detected in both eGPF$^+$ and eGFP$^-$ mTEC^{high}, suggesting an opening in the entire mTEC^{high} population, independently of transcription. The fact that Gad67 expression is not very frequent (14% in eGFP$^+$ and 5% in eGFP$^-$ mTEC^{high} (Figure 21B)) – its expression frequency generally lies below the threshold identified for the ChIP experiment – underlines the possibility of population-wide, transcription-independent epigenetic opening.. While epigenetic modification of eGFP was regulated according to transcriptional activity, the epigenetic state of Gad67 was found to be largely independent of active transcription. One can speculate that the locus may still be closed in mTEC^{low}, but, in contrast to the Casein locus and to the eGFP gene, opened population-wide in mTEC^{high}, independently of gene expression. Why these two genes, which are controlled by the same promoter region, differ in histone modifications is unclear.

In order to fully understand the discrepancies between the two loci in the regulation of pGE by histone modifications, one would need to analyze additional gene loci.

It has been proposed that Aire-dependent genes would be unmethylated at H3K4, because the Aire protein binds specifically to unmethylated H3K4 with its PHD1 domain [82, 83]. Highly active genes such as housekeeping genes typically carry the H3K4 trimethyl-mark at their promoters and would thus be actively transcribed. Tissue-restricted genes were shown to lack H3K4 trimethylation in embryonic stem cells and only acquire sufficient amounts of trimethyl H3K4 in cells which are committed to their differentiated lineages. In mTEC, such genes could equally

lack trimethylated H3K4 or only carry very low amounts. Such genes would require additional binding of Aire to support recruitment of the transcription machinery to the promoter in the absence of trimethyl H3K4 [82, 83, 174]. Our data is compatible with this concept because both Csnb and Gad67 are Aire-independent genes and both carry significant levels of H3K4 trimethylation at their promoters. In order to verify the hypothesis of Org et al., enrichment for mTEC expressing an Aire-dependent antigen such as Insulin or Csna followed by the analysis of histone modifications would be required. Unfortunately, no suitable antibodies or reporter mice were available during this study to perform such experiments.

Since the start of this project, the field of histone modifications has changed in some aspects. To date, opinions differ with regard to the significance of H3K4 trimethylation. Apart from marking active genes, H3K4me3 may also identify genes poised for transcription [102] or general promoter regions [98, 175]. This may be the case at the promoter of Gad67/eGFP and the Gad67 gene where H3K4 trimethylation can be found equally at $eGFP^+$ and $eGFP^-$ $mTEC^{high}$ while in $eGFP^-$ $mTEC^{high}$ the frequency of Gad67 or eGFP expressors is significantly lowered. In this context, H3K4 trimethylation would mark both actively transcribed promoter regions and regions prepared for expression.

Studies by L. Tykocinski on the DNA methylation at distinct promoter regions in the mouse casein gene locus and Gad67 locus detected strong differences in the regulation of DNA methylation between the two loci:
While the promoter regions of all casein genes were found to be demethylated in the mouse mammary gland, only the Csnb gene showed strong demethylation in $mTEC^{high}$. All other casein genes had relatively high DNA methylation levels similar to those in thymocytes. Strikingly, demethylation was observed for the Csnb promoter not only in $mTEC^{high}$, but already to a similar degree in $mTEC^{low}$. Thus, the Csnb promoter is already demethylated in immature $mTEC^{low}$, preceding gene expression. In the Gad67 locus, the wt Gad67 gene, the eGFP/Gad67 gene and the common promoter region were shown to be strongly demethylated in all cell types analyzed. (L. Tykocinski, unpublished results). This observation speaks in favor of a largely open chromatin structure in the Gad67 locus, other than in the casein locus. One can speculate that in the Gad67 locus, high levels of H4 acetylation are not necessary to allow for sufficient access to the promoter region for the transcription machinery in mTEC. In the brain, expression levels of Gad67 are much higher, which may explain the requirement for additional H4 acetylation. The degree of H4 acetylation was previously shown to be directly correlated with promoter activity with higher degrees of H4 acetylation at promoters of genes transcribed at a higher rate[100].
Gad67 is found only 150 kB downstream of the gene Myosin III b (Myo3b) which is strongly and ubiquitously expressed. As the presence of housekeeping genes in a locus has been proposed to induce a largely open chromatin structure which may also effect near genomic regions or entire loci [164], patterns we observe in the Gad67 locus may also be influenced by the vicinity of Myo3b. In this case, as Myo3b is ubiquitously expressed, the signal for epigenetic opening would be cell type independent. This could provide an explanation for the wide demethylation observed in the Gad67 locus and the lack of H4 acetylation in $mTEC^{high}$ (Figure 18).

The histone modification patterns in the two loci, Gad67 and the casein locus, still require further analysis in the future. Within the scope of this study we can come to the following conclusions: Epigenetic opening of the entire casein gene cluster by activating histone marks on the population level of mTEC^{high}, as it had been proposed [17], does not occur. Epigenetic marks are rather specifically regulated and applied at the level of the individual gene, as we have shown for Csnb. Epigenetic opening preceding gene expression, already in mTEC^{low}, has been detected by DNA methylation studies: The Csnb promoter was found to be demethylated at CpGs in both mTEC^{low} and mTEC^{high} by L. Tykocinski (unpublished data). With respect to the histone modifications analyzed in this study, opening preceding gene expression (in mTEC^{low}) was not observed, neither at the Casein gene locus, nor at the Gad67 gene. The Gad67/eGFP locus is regulated differently than the casein locus: Gad67 probably carries active chromatin marks in mTEC^{high} which do not yet – or not anymore – express the gene. It is likely that H3K4 trimethylation at Gad67 has a function involved in priming/poising of the gene rather than directly regulating the expression state.

It was shown that not only opening of the promoter and initiation of transcription, but also the elongation step may be of importance for the regulation of pGE. One study showed that Aire promotes transcriptional regulation by recruiting the positive transcription elongation factor b (P-TEFb) complex to RNA polymerase II at target gene promoters [168]. In some loci, active histone modification marks thus may be a prerequisite for transcription, but final activation of transcription may be induced by other parameters.

Villasenor et al. could show that some genes use multiple and alternative transcriptional start sites in pGE, different from those in the target tissue. If this is also the case for the casein genes and Gad67, we may not be looking at the right transcriptional start sites used in mTEC. The analysis of histone modifications at these alternative start sites will provide additional insight in the future [76].

We found specific histone modification patterns which differed between the two loci, Gad67 and the Casein locus. These two examples made it clear that there is variation also at the level of gene regulation by epigenetic marks between different promiscuously expressed loci.

4.1.2 Changes in chromatin structure upon differentiation from mTEC^{low} to mTEC^{high}

Regulation of chromatin structure has been hypothesized to be involved in pGE. Aire has been implicated in the regulation of nuclear organization by a) tethering genes to nuclear matrix and by b) localizing in the nuclear punctate dots which colocalize with nuclear speckles [17, 169, 176]. To date, correlations between promiscuous gene expression and nuclear organization have not been studied in detail. We analyzed the radial position and chromatin compaction of the casein locus, in particular of Csnb, in mTEC, as these two parameters had previously been shown to be related to gene activity [119, 120, 122, 129, 121].

Within any of the four cell populations analyzed, namely mTEC^{high} and mTEC^{low} as well as thymocytes and MEC as controls, we found no preferential radial position for the Csnb gene. In all four cell types, the position of Csnb varied evenly over the entire nuclear volume in the 2D

analysis. The 3D analysis showed a bias towards smaller distances from the nuclear rim (Figure 29). This again was similar in all four cell populations and can be explained by an overall ellipsoid shape of the nuclei. In ellipsoid nuclei, the distance to the rim is very similar to the distance in the z – axis which is smaller the flatter the cells are.
The radial position of the Csnb gene in the nucleus thus seems to be independent of its expression state.
The radial position of individual genes in the nucleus has been shown to be important for transcriptional activity: Many studies have found correlations between radial position and transcriptional activity and some studies also proved a functional coherence [126, 127 130-132]. However, not every gene is affected to the same extent by radial positioning and some genes have been identified to date which localize in the nucleus independently of their expression state [113, 177].

Genes whose activity does not correlate with radial position in the nucleus are not necessarily positioned randomly: In the chromatin territory – interchromatin compartment (CT-ITC) model, a widely accepted model for the organization of nuclear architecture, genetic activity is thought to correlate with the position of a gene with respect to the chromosome where it is located, rather than with respect to the entire nucleus: The CT-ITC model was first proposed by C. and T. Cremer in the late 1990s. It holds that each chromosome occupies its own distinct territory (CT) in the nucleus with larger/gene-rich chromosomes located more to the center and smaller/gene-poor chromosomes located more to the nucler periphery [178, 179]. The interchromatin territories (ICT) are thought to contain factors necessary for chromatin remodelling, transcription and splicing. Genes which are active loop out of their CT into the ITC where they obtain access to the transcription machinery [84 109, 110 180]. According to this model, active genes in the ITC can be found at different radial positions in the nucleus. It would thus be interesting to see with a chromosomal counterstain for mouse chromosome 5 whether the Csnb gene is consistently located on the surface or outside of its chromosome territory in those cell populations where it is actively transcribed.

The compaction of chromatin, although it does not represent an unsurmountable barrier for the trancription machinery, has in many cases been reported to inversely correlate with genetic activity [181]. Interestingly, when mTEC stained with the TO-PRO®-3 dye, a DNA dye, were analyzed under a microscope in ultrathin cryosections, they showed remarkably few DNA-dense, heterochromatic regions, in some cells, none were present at all (Figure 31, A. Pombo, unpublished data). This implies that chromatin in mTEC must overall be less densely packed, in some cases even completely lacking highly condensed heterochromatic regions. An overall more open chromatin structure could facilitate accessibility to transcriptional start sites and allow for genome-wide promiscuous gene expression. Which mTEC subpopulations exhibit larger and smaller amounts of heterochromatin and how this is implicated in the regulation of pGE on a cellular level has yet to be analyzed.

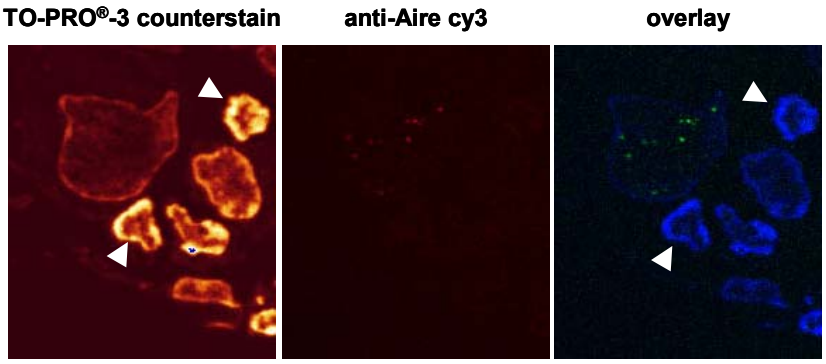


Figure 31
Aire$^+$ mTEC have very few heterochromatic regions. Ultrathin cryosections were stained with anti-Aire antibody and TO-PRO®-3 DNA dye as counterstain. Heterochromatic regions in Aire$^+$ mTEC were fewer or completely absent in comparison to thymocytes (white arrows). (A. Pombo, unpublished data).

Compaction of chromatin can vary between different domains and may show opposite trends in different regions. Opposing mechanisms in different genomic regions are possible as the structure of different chromatin domains may be regulated by different sets of proteins [147]. Therefore it is crucial to study chromatin compaction not only at a global level which only allows determination of an average measure of compaction, but primarily in individual domains. In this study we measured chromatin compaction in a 103 kB stretch in the casein gene locus which encompassed the Csna and Csnb gene. In three of the four cell populations analyzed, namely mTEC^{high}, MEC and thymocytes we found similar chromatin compaction while chromatin in mTEC^{low} was found to be 1.5 fold more compact. We concluded that the Csnb gene locus decompacts significantly upon maturation from mTEC^{low}, which do not express Csnb, to mTEC^{high}, which express Csnb to high frequency. The final level of chromatin compaction of the 103 kB stretch analyzed was found similar in mTEC^{high} and MEC, while MEC express high levels of Csnb and should have an open chromatin structure. (Figure 27). When comparing the compaction factors found in MEC and mTEC^{high} to the the 30 nm fibre which typically has a compaction factor of 40, it is obivous that a higher level of chromatin structure must still be present in the casein locus in the two cell populations (Figure 28). An additional degree of compaction must be introduced in the chromatin in mTEC^{low}, which is resolved upon maturation to mTEC^{high}.

The comparison of chromatin compaction shows a decompaction from mTEC^{low} to mTEC^{high} and is still valid when extended to other epithelial cells, i.e. MEC, which have a similarly open chromatin structure as mTEC^{high}. However, the comparison does not hold when extended to completely different cell types such as thymocytes which show a chromatin compaction similar to mTEC^{high} and MEC but do not express any casein genes. Thymocytes potentially have a completely different nuclear structure than epithelial cells (Figure 31), which complicates the comparison of chromatin compaction between completely different cell types.

All four cell populations analyzed, mTEC^{high}, mTEC^{low}, MEC and thymocytes, showed a similarly high variation for the distances measured. This extent of variability is frequently detected in measurements concerning the nuclear structure: Distance measurements often have large standard variations and colocalization studies rarely show complete colocalization or complete

separation of target loci, but rather subtle variations [137, 182]. This variability appears to be inherent to measurements of nuclear structure. We can hypothesize from these high variations that a certain nuclear structure can favor or impede certain expression patterns but is not compulsory for a particular expression state.

As the chromatin compaction factor represents an average chromatin density over a certain region, an overall more compact chromatin structure over the entire 103 kB (or an even larger region) does not exclude the possibility that small, localized areas of DNA exhibit a much more open conformation [129]. Such a conformatin would allow for very open chromatin for example at the promoter region of specific casein genes, allowing access for the transcription machinery in MEC or mTEC^{high}, while other regions would be much more condensed. Chromatin loops which emerge from more condensed areas have been found previously to be a crucial factor for proper gene regulation [181]. One equally has to keep in mind that the measurement of chromatin compactions factors by definition assumes the DNA to be compacted on a linear stretch between the two end points measured. This is of course not a very likely conformation, as looping of the DNA occurs constantly. But chromatin compaction factors can provide a rough measure of the overall chromatin state in a particular region, they are all the more exact the smaller the region of chromatin analyzed, as there is less room for looping.
From the data presented here we can conclude that chromatin decompaction at the Csnb gene region takes place upon differentiation from mTEC^{low} to mTEC^{high}.

4.1.3 A model of epigenetic changes in the mouse casein gene locus

In this study, we have analyzed the status of histone modifications, radial position and chromatin compaction in the mouse casein locus. Studies by L. Tykocinski on the DNA methylation at distinct promoter regions in the mouse casein gene locus complement these results and confirm the exceptional position of Csnb:
1) Only the Csnb promoter was found to be demethylated in mTEC^{high} while all other genes in the locus were methylated at CpGs in the promoter regions. Strikingly, CpG demethylation at Csnb was already detected in mTEC^{low}, preceding pGE (L. Tykocinski, unpublished). 2) Chromatin in the region of the Csnb gene was found to be less compact in mTEC^{high} compared to mTEC^{low}. 3) Similarly, histone modifications characteristic for open chromatin, namely acetyl H4 and trimethyl H3K4, were detected only in the Csnb gene of mTEC^{high}, but not in mTEC^{low}.
These three aspects allow us to postulate a hypothetical chronological order for the epigenetic opening of Csnb in promiscuous gene expression (Figure 32):

1) Already in mTEC^{low}, the DNA at the Csnb promoter becomes demethylated. As DNA methylation is a very stable and long-lived mark, there are two options for this process: Either, the promoter region is passively demethylated concomitantly with cell division and incompletely remethylated after division [90], or alternative mechanisms such as deamination, base and nucleotide-excision repair lead to an active demethylation of the region. An enzyme directly

containing DNA demethylase activity has not been identified to date [89] [183]. Why the methylation state of the Csnb promoter already changes preceding gene expression is not clear. Maybe Csnb serves as an entry point into pGE in the entire casein locus. This would explain why the locus is prepared so early on in differentiation and why Csnb plays such a special role with regard to epigenetic modifications, expression frequency and level.

2) Upon differentiation from mTEC^{low} to mTEC^{high}, chromatin at the Csnb locus decompacts:
Decompaction and potential reorganization of genes in the nucleus underlies certain constraints: Upon cell division, chromatin was rarely observed to move over distances larger than 1-2 μm [184]. During interphase, positions of subchromosomal areas are fixed within a radius of 0.5-1 μm, rarely individual loci have been observed to travel further [85] [126]. Reorganization and reshuffling of genes in the nucleus upon differentiation is thus spatially limited. Additionally, locus decompaction, especially chromatin remodelling has been shown to be energy-consuming and potentially slow [173]. Different enzymes involved in chromatin remodelling and in the modification of histone tails supposedly act in concert to decompact chromatin in the casein locus. Frequently a combinatorial pattern of many modifications is needed to assemble a certain chromatin state [185].

3) After or in parallel with the establishment of an open, less compact chromatin state in mTEC^{high}, active chromatin marks (H4 acetylation and H3K4 trimethylation) are introduced at the histone tails in the Csnb promoter region. This leads to the recruitment of additional factors and the transcription machinery.
Changes in histone modifications such as histone lysine acetylation or methylation have been shown to change as fast as within minutes [92]. With the data present it is not possible to temporally resolve whether histone modification are introduced prior, at the same time or after chromatin decompaction as both changes occur on such different time scales.

Neither histone modifications nor DNA CpG methylation have been shown to act on the entire casein cluster, both were found to act exclusively on the Csnb gene within the detection limits of the experiments. Thus it is unclear whether the other promiscuously expressed genes in the locus are just stochastically expressed at a low level as a consequence of transcriptional spreading mechanism from a highly active Csnb gene. It had been suggested previously that a strongly expressed gene could cause ectopic expression of neighbouring genes which would be tolerated if it was not harmful to the cell [164] [186]. Preliminary data indicates that during mTEC ontogeny, ectopic expression in the casein locus starts at the casein beta gene while other casein genes follow a few days later (J. Derbinski, unpublished).
Alternatively, each gene could undergo epigenetic opening on the single cell level prior to expression but only specifically in those individual cells which eventually express the gene. As most genes are expressed at low frequency, such a mechanism would be below the detection limit of the ChIP experiment (see Figure 16).

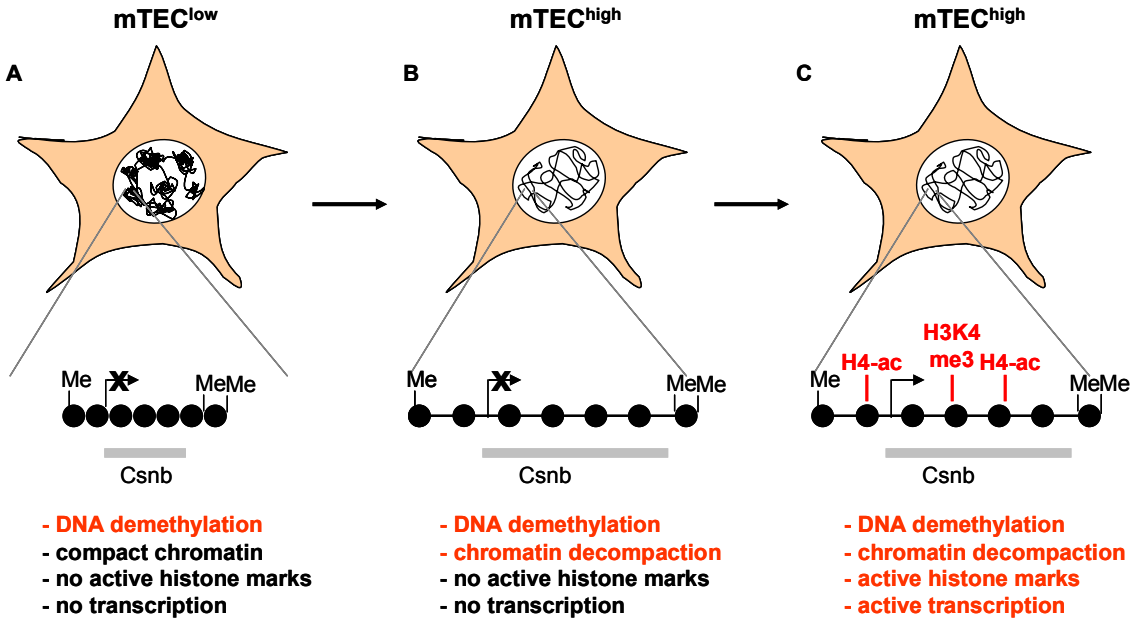


Figure 32

Model of a chronological order of epigenetic changes at the mouse Csnb gene. The Csnb gene is demethylated already in mTEC^{low}, which do not express any casein genes (A). Upon maturation from mTEC^{low} to mTEC^{high}, chromatin decompacts in the region of the Csnb gene (B) followed or accompanied by the introduction of active histone modification marks, acetyl H4 and trimethyl H3K4, at the Csnb promoter (C). Csnb is expressed in over 80% of mTEC^{high} [60].

4.2 pGE: Fluctuating, stochastic and increasingly complex

Aside from the involvement of epigenetic regulation mechanisms, a number of other issues have been been raised in this study regarding the intricate expression patterns observed in pGE. Particularly at the single cell level, more information was needed in order to understand the stochasticity, allelic expression patterns and complexity. What leads to a specific expression pattern in an individual mTEC? Are these patterns stable or do they change over time?

Two studies have been published concerning the single cell expression pattern in mTEC: One study focused on the casein locus, showing pGE within the borders of a particular locus to be probabilistic, the other on a number of promiscuously expressed genes from different clusters, which were equally shown to be expressed in a stochastic manner [60, 76]. In this study, we have additionally performed single cell expression analysis in the murine Gad67 locus for two reasons: First, the model of Gad67/eGFP mice allows for enrichment of a subpopulation of mTEC expressing a particular gene – eGFP – and for a detailed analysis of mono- or biallelic expression patterns of an Aire-independent gene which has not been done previously. Second, single cell expression patterns could be directly linked to the pattern of histone modifications identified in this locus.

4.2.1 pGE is largely stochastic and increasingly complex in Aire$^+$mTEC

With the help of single cell expression analysis we found a number of characteristics in the pattern of the promiscuously expressed genes Gad67-wt (named here "Gad67") and transgenic eGFP-Gad67 (named here "eGFP") which are atypical for "conventional" gene expression.
In the heterozygous Gad67/eGFP mice used in this study, one allele remains wt Gad67 while eGFP is inserted in the other allele directly at the transcriptional start of Gad67 [160]. Gad67 and eGFP are thus controlled by the same promoter region, however they differ twofold in their expression frequency (see Figure 20). Both in eGFP$^+$mTEC^{high} and eGFP$^-$mTEC^{high}, eGFP expression is approx. two times more frequent than Gad67 expression (33% eGFP vs. 14% Gad67 in eGFP$^+$mTEC^{high} and 10% eGFP vs 5% Gad67 in eGFP$^-$mTEC^{high}, respectively) (Figure 20). As both genes are placed in the same genetic context, differences in promoter activation or transcription initiation should not play a role. Differences in RNA stability or efficiency of RNA processing could lead to the accumulation of eGFP RNA in a larger percentage of cells. Deadenylation-dependent mRNA decay which is a principal mRNA degradation pathway is often regulated by the presence of A+U rich elements near the 3' UTR which destabilize or cytosine-rich elements which stabilize the mRNA against deadenylation [187]. None of such elements were found in the eGFP or Gad67 mRNAs, but other mRNA degradation pathways may still play a role. As the eGFP mRNA only contains one exon while the Gad67 mRNA is spliced from 18 exons, it is possible that mRNA processing, in particularly mRNA splicing, accounts for the different frequencies. If the Gad67 mRNA is spliced less efficiently or more slowly, it will accumulate in fewer cells although the promoter on the Gad67 allele is as active as the promoter on the eGFP allele. Similar promoter activity of Gad67 and eGFP may also account for the similar degrees of histone modifications observed between Gad67 and eGFP regions in eGFP$^+$ mTEC^{high} in ChIP (Figure 18).
Gad67/eGFP are expressed in a partially biallelic fashion. Co-expression of Gad67 and eGFP is highly significant ($p<<10^{-4}$) thus biallelic expression is favored, but monoallelic expressors still exist. Aire-dependent genes, which were previously analyzed by Villasenor et al., were shown to be expressed monoallelically[76]. Monoallelic expression implies that every allele is regulated individually thus monoallelically expressed genes exhibit increased probabilistic behaviour in their expression patterns compared to the Aire-independent gene Gad67.
pGE can be described as a largely stochastic, but not completely random process: The expression pattern of Gad67 is partially biallelic and significant co-expression could be shown for two genes in human mTEC, MUC1 and CEA. They are located on different chromosomes but co-expressed to a high degree at the protein and mRNA level in individual human mTEC [80] (S. Pinto, unpublished). We can infer that pGE must be controlled by multiple, seemingly opposing mechanisms: On the one hand, mechanisms control co-regulation of certain alleles and certain genes. On the other hand, potentially different mechanisms create largely stochastic expression profiles. How both of these aspects are compatible is still unclear.

Cells which co-expressed Gad67 and eGFP were enriched in Aire expression over cells which expressed only one of the two genes. The cell population expressing neither of the two was shown to be relatively depleted in Aire expression (Figure 20). Aire expression thus favors the

biallelic expression of Gad67 and eGFP. Gad67 is Aire-independent, thus Aire protein expression should not be required for Gad67/eGFP expression. The finding that the presence of Aire nevertheless correlates with Gad67/eGFP expression shows that the particular differentiation state marked by Aire influences pGE of Aire-independent antigens.
An overall increased complexity of pGE has been observed in Aire$^+$mTEC due to the expression of Aire-dependent genes. Increased expression frequency of Aire-independent genes in Aire$^+$mTEC further adds to the complexity [60]. Aire has been found to be not only implicated in transcriptional regulation of pGE but also in the differentiation program of mTEC [75]. How this program affects gene expression at the molecular level, making it more variable and increasingly complex, is unclear. With increasing complexity, more genes are expressed per single cell and even rare individual cells can be identified which co-express all genes from a gene cluster (i.e. the casein cluster [60]) (Figure 33). The increasing complexity during the lifetime of an mTEC could be explained by a mechanism which may be started early on in mTEC differentiation and would act in a cumulative manner once set in motion, leading to a maximally complex gene expression pattern in terminally differentiated Aire$^+$mTEC..

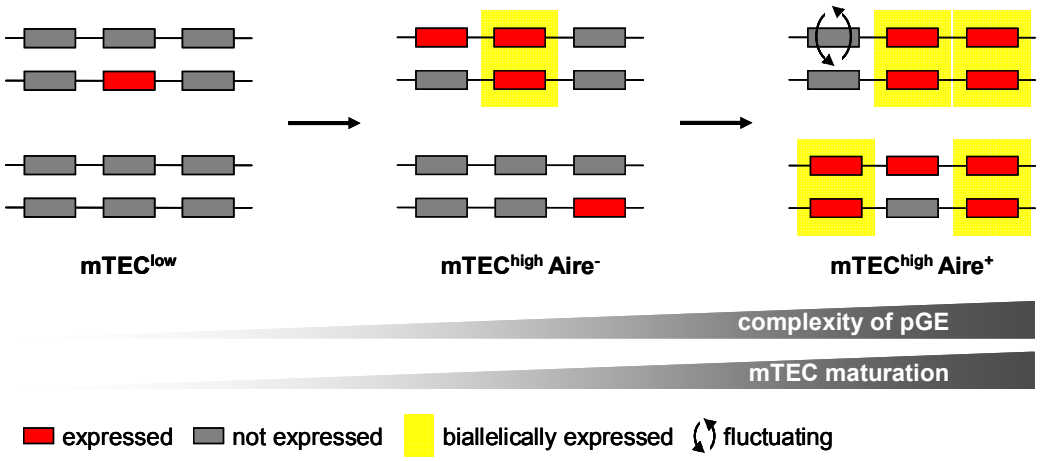


Figure 33
Model of increasing complexity of pGE. Individual mTEC express an increasing number of genes concomitantly with maturation, potentially leading to co-expression of an entire gene locus. At the same time, pGE expression patterns are fluctuating, thus individual genes may be turned on and off over time. Biallelic expression of some genes is favored in terminally differentiated Aire$^+$ mTEC^{high}.

4.2.2 Discrepancies between protein and mRNA expression speak for a fluctutating repertoire

The mRNA expression frequencies found both for eGFP$^+$mTEC^{high} and eGFP$^-$mTEC^{high} differ significantly from the eGFP protein levels determined by FACS analysis: While the gates on eGFP$^+$ and eGPP$^-$ mTEC^{high} were set stringently (Figure 17), only 33% of eGFP$^+$ but 10% of eGFP$^-$mTEC^{high} expressed eGFP mRNA (Figure 20). Protein-mRNA discrepancies seem to be an effect specifically observed in promiscuously expressed genes. Protein-and mRNA levels for both EpCAM and Aire are largely consistent: mTEC^{high} were sorted as EpCAM$^+$ and EpCAM

mRNA was expressed to nearly 100% (Figure 20). Similarly, Aire protein expression was previously found to be 56% [65] which is in a similar range as the average frequency of Aire mRNA of 68% found in this study (Figure 20). Which factors can explain these discrepancies only found for promiscuously expressed antigens?

10% of eGFP mRNA expressing cells were detected in the $eGFP^-$ population. The eGFP primers were shown to be highly specific (no $eGFP^+$ cells were found in the $mTEC^{low}$ population) (Figure 20), thus we can exclude non-specific amplification. A possible explanation could be delayed translation and expression of eGFP protein due to limited access to the translation machinery or slow mRNA processing prior to translation. This could lead to the detection of mRNA prior to eGFP protein.

On the other hand, only 33% of sorted $eGFP^+mTEC^{high}$ were found to express eGFP mRNA.

eGFP is known to be a particularly stable protein which may introduce a certain bias into the measurement, caused by the artificial eGFP *knock in* situation: The half life of eGFP protein was measured in different studies to be between 17 and 26 h [188, 189 190], the half life of eGFP mRNA was determined as a) over 6 h and b) significantly longer than GFP mRNA half life which had been measured to be 1h elsewhere [191, 192]. While the eGFP protein is particularly stable, the half-life of eGFP mRNA is well within the average range measured for different mRNA lifetimes.

As we could only detect about one third of the cells to express eGFP in the cell population sorted as $eGFP^+$, the remaining two thirds do express the protein but must have turned off the mRNA. An exemplary calculation will illustrate how the half-lifes of eGFP protein and mRNA may affect the results of SC PCR. We will assume eGFP protein half life to be 24h, mRNA half life to be 8h. Two thirds of the sorted cell population were found to only express eGFP protein but not eGFP mRNA (population 3 in Figure 34). The remaining one third of the sorted cell population still has detectable quantities of eGFP mRNA in the nucleus, but only a certain subpopulation still carries out active transcription of eGFP (population 1 in Figure 34), the remaining cells have turned off active expression while eGFP mRNA (with a half-life of 8h) is still present in the cell (population 2 in Figure 34). The population actively transcribing eGFP must thus be significantly smaller than the 33% of the sorted population, where eGFP mRNA was detected (population 1+2, Figure 34). The other much over 66% of cells have turned eGFP expression off recently, probably within the last 32 h (according to the half-lifes assumed for the model calculation, 24h protein half-life + 8h mRNA half life) (see Figure 34).

Gad67 protein has a half life of only 4 h (in human) and a roughly estimated mRNA half life of 9 h (in mouse) [193, 194]. For such a gene, mRNA and protein expression frequencies would correlate much better in a similar scenario, we would find fewer cells expressing Gad67 protein without expressing Gad67 mRNA due to the shorter protein half-life (see Figure 34). Unfortunately we cannot measure the expression of Gad67 protein in parallel in FACS analysis.

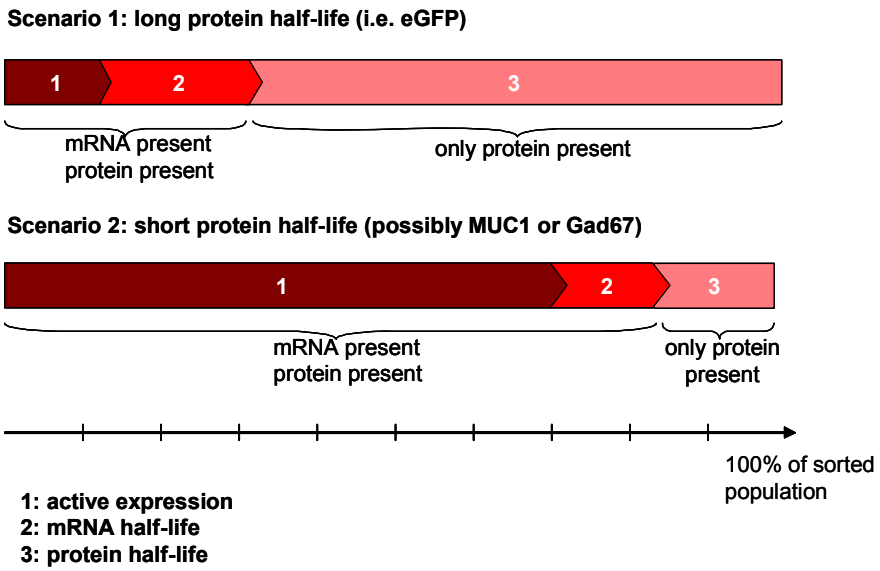


Figure 34: Protein and mRNA half-lifes: Implications for SC PCR measurements. When a cell population is sorted by the expression of a particular protein and is subsequently analyzed for mRNA expression, only stably expressed genes will show perfect correlation between protein and mRNA frequencies. For fluctuating genes, the sorted cell population encompasses three pools: 1) Cells actively synthesizing mRNA and translating protein. 2) Cells having turned off expression, but mRNA is still present in the cell and protein is being translated. The size of this population depends on the mRNA half-life. 3) Cells having turned off expression, mRNA is decayed but protein is still present in the cell. The size of this population depends on the protein lifetime. Populations 1, 2 and 3 are positive for the protein but only populations 1 and 2 are positive for the respective mRNA. Protein and mRNA frequencies thus diverge most when the protein half-life is long.

The most important point is: The pattern we observed for eGFP protein and mRNA expression can only be explained by the fact that mRNA expression had been turned off in a significant percentage of the population.This implies that the gene pool expressed in pGE is not only the sum of different expression patterns of individual cells but is also fluctuating over time in a single cell, adding to the complexity of pGE (Figure 33).

The hypothesis of fluctuation in the repertoire of pGE in an individual cell had been proposed previously [76], but not shown to date. Observations in support of fluctuation were made previously in our group: Human mTEC with maximal expression of the surface antigen SGLT1 ($SGLT1^{++}$) showed a decrease in SGLT1 mRNA levels compared to $SGLT1^{+}$ mTEC. This implies that at a maximal protein concentration, the mRNA expression is apparently already turned off [80].

However, this discordance between mRNA and protein is not found for every promiscuously expressed gene: When $MUC1^{+}$ human mTEC were analyzed in single cell PCR, they expressed MUC1 mRNA to a high frequency (approx. 85%, S. Pinto, unpublished data). Maybe, MUC1 protein half life is overall very short, in this case, fluctuating gene expression would not lead to a discordance between protein and mRNA frequencies (Figure 34). On the other hand it is possible that some promiscuously expressed genes fluctuate in their expression, such as eGFP (and thus probably Gad67) in mouse and SGLT1 in human. Others may be consistently expressed in

particular cells, such as for example MUC1 in human and Csnb in mouse. Csnb is expressed in such a high percentage of mTEC^{high} that it leaves almost no room for fluctuating expression.

4.2.3 Can tolerance induction benefit from fluctuating pGE?

Can these two features, partial stochasticity in expression and fluctuating expression patterns of pGE contribute to the efficiency of tolerance induction?

Stochastic expression will have two benefits for tolerance induction: First, a largely stochastic expression process ensures the expression of most promiscuously expressed antigens over the entire mTEC population, thus avoiding holes in the repertoire. Second, the fact that a single mTEC only expresses a small slice of the multitude of promiscuously expressed antigens allows these antigens to attain a sufficient presentation by the limited number of MHC molecules on the surface. The attempt to represent the full repertoire in a single cell would overload the transcription machinery and most antigens could not be presented on surface-MHC or only at insufficient levels.

Fluctuating expression patterns within individual mTEC have previously been discussed in the literature [74, 76] but were mostly considered to be of low relevance for tolerance induction as terminally differentiated mTEC have a limited lifetime of just a few days [65].

Within the lifetime of terminally differentiated mTEC, fluctuating expression may serve as an additional mechanism to increase antigen diversity within the thymic microenvironment. The efficiency of this increase in diversity depends on two factors: The time a promiscuously expressed antigen needs to reach presentation at the mTEC surface and the time and efficiency with which the antigen can be crosspresented on neighboring DC. As the turnover of the MHC-peptide complex is estimated to be around 20h (B. Kyewski, unpublished data), fluctuating antigens expressed early in the lifetime of a postmitotic mTEC may still be presented while TRA expressed later will not reach the surface before the cell undergoes apoptosis. Additionally, DC take up antigens from mTEC via different transfer routes. Antigens expressed late in mTEC can thus be taken up by DC from apoptotic bodies. However, other crosspresentation pathways such as the transfer of membrane complexes between mTEC and DC are more likely to occur[195]. Transfer from mTEC to DC has been confirmed as a mechanism which amplifies the number of cells capable of presenting a certain antigen [21, 195]. Such transfer mechanisms would allow for regular capture of promiscuously expressed antigens from mTEC and presentation on DC. Crosspresentation on DC would overall increase the number of TRA presented in the microenvironment and even include TRA expressed late in the lifetime of an mTEC. Therefore, despite the relatively slow turnover of peptide-MHC complexes, all TRA can possibly be presented to developing thymocytes in the situation of a fluctuating repertoire in mTEC.

Could antigen diversity in the thymic medulla be a limiting factor for "complete" tolerance induction?

Medullary thymocytes have been proposed to extensively scan APCs during the approximately four days they spend in the medulla [196]. Calculations have estimated the number of mTEC a developing thymocyte could scan to 1000-5000 per day. Scanning the entire mTEC^{high} population

of about 30,000 cells by all $15x10^6$ thymocytes in a single thymus was then calculated to take approx. four days [17, 21]. According to this calculation, a developing thymocyte would be able to get a "complete" view of the antigen repertoire without the need for increased antigen variability through a fluctuating pGE repertoire.

However, the calculation is based on a fast and sterically unhindered scanning process. This is not the case for mice in the phase of maximal thymic output at the age of two weeks. Thymi of such young mice contain about 600 isolated medullary islands of only oligoclonal origin which merge later on to form a single coherent medulla [197]. An individual thymocyte only scans one medullary island and may thus only see a limited antigen pool. Fluctuating pGE would increase the antigen diversity within a medullary island and thus broaden the scope of tolerance induction. This is particularly important as the early postnatal period has been identified to be crucial in the induction of tolerance to TRA[198].

Also in adult mice scanning may not be sterically unhindered: A new study has assessed the scanning speed and paths of medullary thymocytes in *ex vivo* thymic explants by two-photon microscopy. Such measurements had previously been precluded by the depth of the thymic medulla within the thymic tissue [199]. The speed of medullary thymocytes was shown to be high (>10 μm/min) with frequent and transient contacts with DC, similar to previous measurements of lymph node T cells [200] Surprisingly, thymocytes do not roam freely via a random walk in the medulla, as it had been shown for cortical thymocytes [201], but behave as if they had been „leashed". Movement of medullary thymocytes is random, but only within distinct confinement zones of 55 μm diameter and even more restrained in the presence of negatively selecting ligand. Within the observation times of approx. 20 minutes, thymocytes rarely left their confinement zones [199]. It is unclear how many of such zones a thymocyte scans during its four day residence time in the medulla as the measurement time was relatively short. However, we can conclude that thymocyte movement is likely to be spatially restricted compared to a random-walk type movement. This leads to a reduction of total number of different APCs scanned compared to previous calculations. LeBorgne et al. estimate a thymocyte to scan about 5 APCs per hour, so 500 APC (including mTEC and DC) in 4 days, a number which would probably only be sufficient for tolerance against ubiquitous antigens but not for rare TRA only expressed by up to 3% of the mTEC population [199].

When thymocyte scanning is spatially restricted, a local increase in antigen diversity due to a fluctuating pGE repertoire could increase the diversity of antigens scanned per thymocyte. Note that only if a particular thymocyte remains in a certain microenvironment long enough for local expression and antigen presentation to change, antigen fluctuation will have an effect on the diversity of the repertoire scanned. A fluctuation frequency of one expression change per 1-2 days (estimated from more than 66% of $eGFP^+mTEC^{high}$ having turned off eGFP within 32 h, see 4.2.2) would allow for multiple expression patterns during the lifetime of a terminally differentiated mTEC and possibly also for the (cross-) presentation of multiple repertoires to a thymocyte in a certain confinement zone.

This increase in antigen diversity may be crucial to reach the critical threshold of diversity required to obtain a tolerant T cell repertoire.

4.3 Concluding remarks and future perspectives

Since the discovery of promiscuous gene expression in thymic epithelial cells, the pattern and regulatory mechanisms of pGE as well as their implications for central tolerance have been studied in great detail. Yet, the exact mechanism of regulation of pGE remains poorly understood.

In this study, we have added new aspects to the regulation of pGE:

1. A three-step model to accommodate the involvement of different epigenetic mechanisms in the induction of the Csnb gene. Epigenetic opening occurs at the level of the individual gene, opening of the entire casein locus on a population-wide level can be excluded.
2. pGE can be described as a partially stochastic and fluctuating process. The regular fluctuation of expression patterns may enhance central tolerance induction both in neonates and in adult animals.

The identification of these additional characteristics of pGE still do not allow us to devise a model which can integrate:

1. The apparent stochasticity of gene expression in conjunction with biallelic and inter-chromosomal co-expression patterns.
2. The epigenetic mechanisms correlating with pGE with the observed co-expression patterns and expression frequencies at the single cell level.

Future experiments should be designed to prove the functional relevance of the correlations reported here, both in the case of Aire-dependent and Aire-independent antigens.

Further characterization of the expression patterns and regulation mechanisms of pGE in the future will allow to identify to which extent pGE recruits combined elements from conventional gene regulation and to which extent it uses completely unique pathways.

This will hopefully lead to a better understanding of the onset of autoimmune reactions and potentially contribute to the development of novel strategies to improve the prevention and cure of such diseases.

References

1. Miller, J.F. Immunological function of the thymus. *Lancet* **2**, 748-749 (1961).
2. Boyd, R.L. *et al.* The thymic microenvironment. *Immunol Today* **14**, 445-459 (1993).
3. von Boehmer, H., Teh, H.S. & Kisielow, P. The thymus selects the useful, neglects the useless and destroys the harmful. *Immunol Today* **10**, 57-61 (1989).
4. Paul, W.E. Fundamental Immunology, 5th ed. *Lippincott/William&Wilkins*, 259-301 (2003).
5. Jameson, S.C., Hogquist, K.A. & Bevan, M.J. Positive selection of thymocytes. *Annu Rev Immunol* **13**, 93-126 (1995).
6. Ueno, T. *et al.* CCR7 signals are essential for cortex-medulla migration of developing thymocytes. *J Exp Med* **200**, 493-505 (2004).
7. Kurobe, H. *et al.* CCR7-dependent cortex-to-medulla migration of positively selected thymocytes is essential for establishing central tolerance. *Immunity* **24**, 165-177 (2006).
8. Petrie, H.T. Cell migration and the control of post-natal T-cell lymphopoiesis in the thymus. *Nat Rev Immunol* **3**, 859-866 (2003).
9. Merkenschlager, M. Tracing interactions of thymocytes with individual stromal cell partners. *Eur J Immunol* **26**, 892-896 (1996).
10. Merkenschlager, M., Benoist, C. & Mathis, D. Evidence for a single-niche model of positive selection. *Proc Natl Acad Sci U S A* **91**, 11694-11698 (1994).
11. McCaughtry, T.M., Baldwin, T.A., Wilken, M.S. & Hogquist, K.A. Clonal deletion of thymocytes can occur in the cortex with no involvement of the medulla. *J Exp Med* **205**, 2575-2584 (2008).
12. Kisielow, P. & von Boehmer, H. Kinetics of negative and positive selection in the thymus. *Adv Exp Med Biol* **292**, 31-42 (1991).
13. Huseby, E.S. *et al.* How the T cell repertoire becomes peptide and MHC specific. *Cell* **122**, 247-260 (2005).
14. Hogquist, K.A. Signal strength in thymic selection and lineage commitment. *Curr Opin Immunol* **13**, 225-231 (2001).
15. Hogquist, K.A. *et al.* T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17-27 (1994).
16. Hogquist, K.A. & Moran, A.E. Treg cells meet their limit. *Nat Immunol* **10**, 565-566 (2009).
17. Kyewski, B. & Klein, L. A central role for central tolerance. *Annu Rev Immunol* **24**, 571-606 (2006).
18. Lo, D. & Sprent, J. Identity of cells that imprint H-2-restricted T-cell specificity in the thymus. *Nature* **319**, 672-675 (1986).
19. Marrack, P. *et al.* The effect of thymus environment on T cell development and tolerance. *Cell* **53**, 627-634 (1988).
20. Matzinger, P. & Guerder, S. Does T-cell tolerance require a dedicated antigen-presenting cell? *Nature* **338**, 74-76 (1989).
21. Gallegos, A.M. & Bevan, M.J. Central tolerance to tissue-specific antigens mediated by direct and indirect antigen presentation. *J Exp Med* **200**, 1039-1049 (2004).
22. Fontenot, J.D. *et al.* Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity* **22**, 329-341 (2005).
23. Sakaguchi, S. *et al.* T cell-mediated maintenance of natural self-tolerance: its breakdown as a possible cause of various autoimmune diseases. *J Autoimmun* **9**, 211-220 (1996).
24. Maloy, K.J. & Powrie, F. Regulatory T cells in the control of immune pathology. *Nat Immunol* **2**, 816-822 (2001).
25. Pacholczyk, R. *et al.* Nonself-antigens are the cognate specificities of Foxp3+ regulatory T cells. *Immunity* **27**, 493-504 (2007).
26. Pacholczyk, R., Ignatowicz, H., Kraj, P. & Ignatowicz, L. Origin and T cell receptor diversity of Foxp3+CD4+CD25+ T cells. *Immunity* **25**, 249-259 (2006).
27. Spence, P.J. & Green, E.A. Foxp3+ regulatory T cells promiscuously accept thymic signals critical for their development. *Proc Natl Acad Sci U S A* **105**, 973-978 (2008).
28. Fontenot, J.D., Dooley, J.L., Farr, A.G. & Rudensky, A.Y. Developmental regulation of Foxp3 expression during ontogeny. *J Exp Med* **202**, 901-906 (2005).
29. Bensinger, S.J., Bandeira, A., Jordan, M.S., Caton, A.J. & Laufer, T.M. Major histocompatibility complex class II-positive cortical epithelium mediates the selection of CD4(+)25(+) immunoregulatory T cells. *J Exp Med* **194**, 427-438 (2001).

30. Watanabe, N. *et al.* Hassall's corpuscles instruct dendritic cells to induce CD4+CD25+ regulatory T cells in human thymus. *Nature* **436**, 1181-1185 (2005).
31. Aschenbrenner, K. *et al.* Selection of Foxp3+ regulatory T cells specific for self antigen expressed and presented by Aire+ medullary thymic epithelial cells. *Nat Immunol* **8**, 351-358 (2007).
32. Wirnsberger, G., Mair, F. & Klein, L. Regulatory T cell differentiation of thymocytes does not require a dedicated antigen-presenting cell but is under T cell-intrinsic developmental control. *Proc Natl Acad Sci U S A* **106**, 10278-10283 (2009).
33. Linsk, R., Gottesman, M. & Pernis, B. Are tissues a patch quilt of ectopic gene expression? *Science* **246**, 261 (1989).
34. Heath, W.R. *et al.* Autoimmune diabetes as a consequence of locally produced interleukin-2. *Nature* **359**, 547-549 (1992).
35. von Herrath, M.G., Dockter, J. & Oldstone, M.B. How virus induces a rapid or slow onset insulin-dependent diabetes mellitus in a transgenic model. *Immunity* **1**, 231-242 (1994).
36. Jolicoeur, C., Hanahan, D. & Smith, K.M. T-cell tolerance toward a transgenic beta-cell antigen and transcription of endogenous pancreatic genes in thymus. *Proc Natl Acad Sci U S A* **91**, 6707-6711 (1994).
37. Pribyl, T.M., Campagnoni, C., Kampf, K., Handley, V.W. & Campagnoni, A.T. The major myelin protein genes are expressed in the human thymus. *J Neurosci Res* **45**, 812-819 (1996).
38. Wakkach, A. *et al.* Expression of acetylcholine receptor genes in human thymic epithelial cells: implications for myasthenia gravis. *J Immunol* **157**, 3752-3760 (1996).
39. Egwuagu, C.E., Charukamnoetkanok, P. & Gery, I. Thymic expression of autoantigens correlates with resistance to autoimmune disease. *J Immunol* **159**, 3109-3112 (1997).
40. Kojima, K., Reindl, M., Lassmann, H., Wekerle, H. & Linington, C. The thymus and self-tolerance: co-existence of encephalitogenic S100 beta-specific T cells and their nominal autoantigen in the normal adult rat thymus. *Int Immunol* **9**, 897-904 (1997).
41. Pugliese, A. *et al.* The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. *Nat Genet* **15**, 293-297 (1997).
42. Hanahan, D. Peripheral-antigen-expressing cells in thymic medulla: factors in self-tolerance and autoimmunity. *Curr Opin Immunol* **10**, 656-662 (1998).
43. Heath, V.L., Moore, N.C., Parnell, S.M. & Mason, D.W. Intrathymic expression of genes involved in organ specific autoimmune disease. *J Autoimmun* **11**, 309-318 (1998).
44. Klein, L., Klein, T., Ruther, U. & Kyewski, B. CD4 T cell tolerance to human C-reactive protein, an inducible serum protein, is mediated by medullary thymic epithelium. *J Exp Med* **188**, 5-16 (1998).
45. Sospedra, M. *et al.* 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 (1998).
46. Mallet, V. *et al.* HLA-G in the human thymus: a subpopulation of medullary epithelial but not CD83(+) dendritic cells expresses HLA-G as a membrane-bound and soluble protein. *Int Immunol* **11**, 889-898 (1999).
47. Klein, L., Klugmann, M., Nave, K.A., Tuohy, V.K. & Kyewski, B. Shaping of the autoreactive T-cell repertoire by a splice variant of self protein expressed in thymic epithelial cells. *Nat Med* **6**, 56-61 (2000).
48. Diez, J. *et al.* Differential splicing of the IA-2 mRNA in pancreas and lymphoid organs as a permissive genetic mechanism for autoimmunity against the IA-2 type 1 diabetes autoantigen. *Diabetes* **50**, 895-900 (2001).
49. Bruno, R. *et al.* Multiple sclerosis candidate autoantigens except myelin oligodendrocyte glycoprotein are transcribed in human thymus. *Eur J Immunol* **32**, 2737-2747 (2002).
50. Derbinski, J., Schulte, A., Kyewski, B. & Klein, L. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nat Immunol* **2**, 1032-1039 (2001).
51. Gotter, J., Brors, B., Hergenhahn, M. & Kyewski, B. 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 (2004).
52. Derbinski, J. *et al.* Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. *J Exp Med* **202**, 33-45 (2005).

53. Smith, K.M., Olson, D.C., Hirose, R. & Hanahan, D. Pancreatic gene expression in rare cells of thymic medulla: evidence for functional contribution to T cell tolerance. *Int Immunol* **9**, 1355-1365 (1997).
54. Klein, L., Roettinger, B. & Kyewski, B. Sampling of complementing self-antigen pools by thymic stromal cells maximizes the scope of central T cell tolerance. *Eur J Immunol* **31**, 2476-2486 (2001).
55. Avichezer, D. *et al.* An immunologically privileged retinal antigen elicits tolerance: major role for central selection mechanisms. *J Exp Med* **198**, 1665-1676 (2003).
56. Chentoufi, A.A., Palumbo, M. & Polychronakos, C. Proinsulin expression by Hassall's corpuscles in the mouse thymus. *Diabetes* **53**, 354-359 (2004).
57. Cloosen, S. *et al.* Expression of tumor-associated differentiation antigens, MUC1 glycoforms and CEA, in human thymic epithelial cells: implications for self-tolerance and tumor therapy. *Cancer Res* **67**, 3919-3926 (2007).
58. Taubert, R., Schwendemann, J. & Kyewski, B. Highly variable expression of tissue-restricted self-antigens in human thymus: implications for self-tolerance and autoimmunity. *Eur J Immunol* **37**, 838-848 (2007).
59. Johnnidis, J.B. *et al.* Chromosomal clustering of genes controlled by the aire transcription factor. *Proc Natl Acad Sci U S A* **102**, 7233-7238 (2005).
60. Derbinski, J., Pinto, S., Rosch, S., Hexel, K. & Kyewski, B. Promiscuous gene expression patterns in single medullary thymic epithelial cells argue for a stochastic mechanism. *Proc Natl Acad Sci U S A* **105**, 657-662 (2008).
61. Gillard, G.O. & Farr, A.G. Contrasting models of promiscuous gene expression by thymic epithelium. *J Exp Med* **202**, 15-19 (2005).
62. Gillard, G.O. & Farr, A.G. Features of medullary thymic epithelium implicate postnatal development in maintaining epithelial heterogeneity and tissue-restricted antigen expression. *J Immunol* **176**, 5815-5824 (2006).
63. Gray, D.H. *et al.* Developmental kinetics, turnover, and stimulatory capacity of thymic epithelial cells. *Blood* **108**, 3777-3785 (2006).
64. Gabler, J., Arnold, J. & Kyewski, B. Promiscuous gene expression and the developmental dynamics of medullary thymic epithelial cells. *Eur J Immunol* **37**, 3363-3372 (2007).
65. Gray, D., Abramson, J., Benoist, C. & Mathis, D. Proliferative arrest and rapid turnover of thymic epithelial cells expressing Aire. *J Exp Med* **204**, 2521-2528 (2007).
66. Rossi, S.W. *et al.* RANK signals from CD4(+)3(-) inducer cells regulate development of Aire-expressing epithelial cells in the thymic medulla. *J Exp Med* **204**, 1267-1272 (2007).
67. Tykocinski, L.O., Sinemus, A. & Kyewski, B. The thymus medulla slowly yields its secrets. *Ann N Y Acad Sci* **1143**, 105-122 (2008).
68. Akiyama, T. *et al.* The tumor necrosis factor family receptors RANK and CD40 cooperatively establish the thymic medullary microenvironment and self-tolerance. *Immunity* **29**, 423-437 (2008).
69. Ferguson, B.J., Cooke, A., Peterson, P. & Rich, T. Death in the AIRE. *Trends Immunol* **29**, 306-312 (2008).
70. Ferguson, B.J. *et al.* AIRE's CARD revealed, a new structure for central tolerance provokes transcriptional plasticity. *J Biol Chem* **283**, 1723-1731 (2008).
71. Anderson, M.S. *et al.* Projection of an immunological self shadow within the thymus by the aire protein. *Science* **298**, 1395-1401 (2002).
72. Halonen, M. *et al.* Subcellular location and expression pattern of autoimmune regulator (Aire), the mouse orthologue for human gene defective in autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED). *J Histochem Cytochem* **49**, 197-208 (2001).
73. Venanzi, E.S., Melamed, R., Mathis, D. & Benoist, C. The variable immunological self: genetic variation and nongenetic noise in Aire-regulated transcription. *Proc Natl Acad Sci U S A* **105**, 15860-15865 (2008).
74. Mathis, D. & Benoist, C. Aire. *Annu Rev Immunol* **27**, 287-312 (2009).
75. Yano, M. *et al.* Aire controls the differentiation program of thymic epithelial cells in the medulla for the establishment of self-tolerance. *J Exp Med* **205**, 2827-2838 (2008).
76. Villasenor, J., Besse, W., Benoist, C. & Mathis, D. Ectopic expression of peripheral-tissue antigens in the thymic epithelium: probabilistic, monoallelic, misinitiated. *Proc Natl Acad Sci U S A* **105**, 15854-15859 (2008).

77. McKnight, R.A. *et al.* An Ets site in the whey acidic protein gene promoter mediates transcriptional activation in the mammary gland of pregnant mice but is dispensable during lactation. *Mol Endocrinol* **9**, 717-724 (1995).
78. Zhou, J. *et al.* Elf5 is essential for early embryogenesis and mammary gland development during pregnancy and lactation. *EMBO J* **24**, 635-644 (2005).
79. Harris, J. *et al.* Socs2 and elf5 mediate prolactin-induced mammary gland development. *Mol Endocrinol* **20**, 1177-1187 (2006).
80. Arnold, J. Regulation ektopischer Genexpression im Thymus-Microenvironment: Vom Medulla Kompartiment zur Einzelzellebene. *Faculty for Biology, Ruprecht-Karls-Universität Heidelberg* (2006).
81. Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome. *Nat Rev Genet* **2**, 21-32 (2001).
82. Org, T. *et al.* The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. *EMBO Rep* **9**, 370-376 (2008).
83. Koh, A.S. *et al.* Aire employs a histone-binding module to mediate immunological tolerance, linking chromatin regulation with organ-specific autoimmunity. *Proc Natl Acad Sci U S A* **105**, 15878-15883 (2008).
84. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2**, 292-301 (2001).
85. Lanctot, C., Cheutin, T., Cremer, M., Cavalli, G. & Cremer, T. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet* **8**, 104-115 (2007).
86. Bestor, T.H. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J* **11**, 2611-2617 (1992).
87. Okano, M., Bell, D.W., Haber, D.A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).
88. Barreto, G. *et al.* Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* **445**, 671-675 (2007).
89. Ma, D.K., Guo, J.U., Ming, G.L. & Song, H. DNA excision repair proteins and Gadd45 as molecular players for active DNA demethylation. *Cell Cycle* **8**, 1526-1531 (2009).
90. Brandeis, M. *et al.* Sp1 elements protect a CpG island from de novo methylation. *Nature* **371**, 435-438 (1994).
91. Bird, A.P. & Wolffe, A.P. Methylation-induced repression--belts, braces, and chromatin. *Cell* **99**, 451-454 (1999).
92. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
93. Roth, S.Y., Denu, J.M. & Allis, C.D. Histone acetyltransferases. *Annu Rev Biochem* **70**, 81-120 (2001).
94. Marmorstein, R. Structure of histone deacetylases: insights into substrate recognition and catalysis. *Structure* **9**, 1127-1133 (2001).
95. Robinson, P.J. *et al.* 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. *J Mol Biol* **381**, 816-825 (2008).
96. Kuo, M.H. & Allis, C.D. Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays* **20**, 615-626 (1998).
97. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837 (2007).
98. Heintzman, N.D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-318 (2007).
99. Bernstein, B.E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169-181 (2005).
100. Pokholok, D.K. *et al.* Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517-527 (2005).
101. Bernstein, B.E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326 (2006).
102. Araki, Y. *et al.* Genome-wide analysis of histone methylation reveals chromatin state-based regulation of gene transcription and function of memory CD8+ T cells. *Immunity* **30**, 912-925 (2009).
103. Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* **8**, 532-538 (2006).

104. Pan, G. *et al.* Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* **1**, 299-312 (2007).
105. Zhao, X.D. *et al.* Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1**, 286-298 (2007).
106. Boyer, L.A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-353 (2006).
107. van Driel, R., Fransz, P.F. & Verschure, P.J. The eukaryotic genome: a system regulated at different hierarchical levels. *J Cell Sci* **116**, 4067-4075 (2003).
108. Takizawa, T., Meaburn, K.J. & Misteli, T. The meaning of gene positioning. *Cell* **135**, 9-13 (2008).
109. Cremer, T. & Cremer, C. Rise, fall and resurrection of chromosome territories: a historical perspective. Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present. *Eur J Histochem* **50**, 223-272 (2006).
110. Cremer, T. & Cremer, C. Rise, fall and resurrection of chromosome territories: a historical perspective. Part I. The rise of chromosome territories. *Eur J Histochem* **50**, 161-176 (2006).
111. Albiez, H. *et al.* Chromatin domains and the interchromatin compartment form structurally defined and functionally interacting nuclear networks. *Chromosome Res* **14**, 707-733 (2006).
112. Branco, M.R. & Pombo, A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol* **4**, e138 (2006).
113. Goetze, S. *et al.* The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Mol Cell Biol* **27**, 4475-4487 (2007).
114. Croft, J.A. *et al.* Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol* **145**, 1119-1131 (1999).
115. Mateos-Langerak, J. *et al.* Nuclear architecture: Is it important for genome function and can we prove it? *J Cell Biochem* **102**, 1067-1075 (2007).
116. Gilbert, N. *et al.* Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* **118**, 555-566 (2004).
117. Bolzer, A. *et al.* Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* **3**, e157 (2005).
118. Neusser, M., Schubel, V., Koch, A., Cremer, T. & Muller, S. Evolutionarily conserved, cell type and species-specific higher order chromatin arrangements in interphase nuclei of primates. *Chromosoma* **116**, 307-320 (2007).
119. Hewitt, S.L., High, F.A., Reiner, S.L., Fisher, A.G. & Merkenschlager, M. Nuclear repositioning marks the selective exclusion of lineage-inappropriate transcription factor loci during T helper cell differentiation. *Eur J Immunol* **34**, 3604-3613 (2004).
120. Kosak, S.T. *et al.* Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* **296**, 158-162 (2002).
121. Zink, D. *et al.* Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. *J Cell Biol* **166**, 815-825 (2004).
122. Williams, R.R. *et al.* Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. *J Cell Sci* **119**, 132-140 (2006).
123. Gartenberg, M.R., Neumann, F.R., Laroche, T., Blaszczyk, M. & Gasser, S.M. Sir-mediated repression can occur independently of chromosomal and subnuclear contexts. *Cell* **119**, 955-967 (2004).
124. Janicki, S.M. *et al.* From silencing to gene expression: real-time analysis in single cells. *Cell* **116**, 683-698 (2004).
125. Nielsen, J.A., Hudson, L.D. & Armstrong, R.C. Nuclear organization in differentiating oligodendrocytes. *J Cell Sci* **115**, 4071-4079 (2002).
126. Chuang, C.H. *et al.* Long-range directional movement of an interphase chromosome site. *Curr Biol* **16**, 825-831 (2006).
127. Reddy, K.L., Zullo, J.M., Bertolino, E. & Singh, H. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* **452**, 243-247 (2008).
128. Ericsson, C., Mehlin, H., Bjorkroth, B., Lamb, M.M. & Daneholt, B. The ultrastructure of upstream and downstream regions of an active Balbiani ring gene. *Cell* **56**, 631-639 (1989).
129. Hu, Y., Kireev, I., Plutz, M., Ashourian, N. & Belmont, A.S. Large-scale chromatin structure of inducible genes: transcription on a condensed, linear template. *J Cell Biol* **185**, 87-100 (2009).

130. Carpenter, A.E. & Belmont, A.S. Direct visualization of transcription factor-induced chromatin remodeling and cofactor recruitment in vivo. *Methods Enzymol* **375**, 366-381 (2004).
131. Chambeyron, S., Da Silva, N.R., Lawson, K.A. & Bickmore, W.A. Nuclear re-organisation of the Hoxb complex during mouse embryonic development. *Development* **132**, 2215-2223 (2005).
132. Chambeyron, S. & Bickmore, W.A. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev* **18**, 1119-1130 (2004).
133. Rouse, R.V., Bolin, L.M., Bender, J.R. & Kyewski, B.A. Monoclonal antibodies reactive with subsets of mouse and human thymic epithelial cells. *J Histochem Cytochem* **36**, 1511-1517 (1988).
134. Farr, A., Nelson, A., Truex, J. & Hosier, S. Epithelial heterogeneity in the murine thymus: a cell surface glycoprotein expressed by subcapsular and medullary epithelium. *J Histochem Cytochem* **39**, 645-653 (1991).
135. Baugh, L.R., Hill, A.A., Brown, E.L. & Hunter, C.P. Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res* **29**, E29 (2001).
136. Lichter, P., Boyle, A.L., Cremer, T. & Ward, D.C. Analysis of genes and chromosomes by nonisotopic in situ hybridization. *Genet Anal Tech Appl* **8**, 24-35 (1991).
137. Rauch, J. *et al.*, Vol. 4164. (eds. K. Karsten, J.T. Hans & S. Herbert) 1-9 (SPIE, 2000).
138. Edelmann, P., Esa, A., Hausmann, M. & Cremer, C. Confocal laser-scanning fluorescence microscopy:In situ determination of the confocal pointspread function and the chromatic shifts in intact cell nuclei. *Optik* **110**, 194-198 (1999).
139. Pawley, J. *Handbook of biological confocal microscopy*. (Plenum Press, New York; 1995).
140. Burns, D.H., Callis, J.B., Christian, G.D. & Davidson, E.R. Strategies for attaining superresolution using spectroscopic data as constraints. *Appl Opt* **24**, 154 (1985).
141. Schmidt, M., Nagorni, M. & Hell, S.W. Subresolution axial distance measurements in far-field fluorescence microscopy with precision of 1 nanometer. *Review of Scientific Instruments* **71**, 2742-2745 (2000).
142. Lacoste, T.D. *et al.* Ultrahigh-resolution multicolor colocalization of single fluorescent probes. *Proc Natl Acad Sci U S A* **97**, 9461-9466 (2000).
143. van Oijen, A.M., Köhler, J., Schmidt, J., Müller, M. & Brakenhoff, G.J. Far-field fluorescence microscopy beyond the diffraction limit. *J. Opt. Soc. Am. A* **16**, 909-915 (1999).
144. Bornfleth, Sätzler, Eils & Cremer High-precision distance measurements and volume-conserving segmentation of objects near and below the resolution limit in three-dimensional confocal fluorescence microscopy. *Journal of Microscopy* **189**, 118-136 (1998).
145. Esa, A. *et al.* Three-dimensional spectral precision distance microscopy of chromatin nanostructures after triple-colour DNA labelling: a study of the BCR region on chromosome 22 and the Philadelphia chromosome. *J Microsc* **199**, 96-105 (2000).
146. Cremer, C. *et al.* Spektrale Präzisionsdistanzmikroskopie in der Genomforschung. *Z. Med. Physik* **9**, 14-20 (1999).
147. Batram, C., Baddeley, D., Kreth, G. & Cremer, C. High precision size measurement of centromere 8 and the 8q24/c-myc gene region in metaphase and interphase human fibroblasts indicate differential condensation. *Journal of Structural Biology* **164**, 293-303 (2008).
148. Peixoto, A., Monteiro, M., Rocha, B. & Veiga-Fernandes, H. Quantification of multiple gene expression in individual cells. *Genome Res* **14**, 1938-1947 (2004).
149. Nelson, J.D., Denisenko, O. & Bomsztyk, K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat Protoc* **1**, 179-185 (2006).
150. Gray, D.H., Chidgey, A.P. & Boyd, R.L. Analysis of thymic stromal cell populations using flow cytometry. *J Immunol Methods* **260**, 15-28 (2002).
151. Wirbelauer, C., Bell, O. & Schubeler, D. Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. *Genes Dev* **19**, 1761-1766 (2005).
152. Foster, S.L., Hargreaves, D.C. & Medzhitov, R. Gene-specific control of inflammation by TLR-induced chromatin modifications. *Nature* **447**, 972-978 (2007).
153. O'Neill, L.P., VerMilyea, M.D. & Turner, B.M. Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. *Nat Genet* **38**, 835-841 (2006).
154. Nelson, J.D., Denisenko, O., Sova, P. & Bomsztyk, K. Fast chromatin immunoprecipitation assay. *Nucleic Acids Res* **34**, e2 (2006).

155. Bernstein, B.E., Meissner, A. & Lander, E.S. The mammalian epigenome. *Cell* **128**, 669-681 (2007).
156. Segal, E. *et al.* A genomic code for nucleosome positioning. *Nature* **442**, 772-778 (2006).
157. Liu, C.L. *et al.* Single-nucleosome mapping of histone modifications in S. cerevisiae. *PLoS Biol* **3**, e328 (2005).
158. Esser, C., Gottlinger, C., Kremer, J., Hundeiker, C. & Radbruch, A. Isolation of full-size mRNA from ethanol-fixed cells after cellular immunofluorescence staining and fluorescence-activated cell sorting (FACS). *Cytometry* **21**, 382-386 (1995).
159. Diez, C., Bertsch, G. & Simm, A. Isolation of full-size mRNA from cells sorted by flow cytometry. *J Biochem Biophys Methods* **40**, 69-80 (1999).
160. Tamamaki, N. *et al.* Green fluorescent protein expression and colocalization with calretinin, parvalbumin, and somatostatin in the GAD67-GFP knock-in mouse. *J Comp Neurol* **467**, 60-79 (2003).
161. Mathis, D. & Benoist, C. A decade of AIRE. *Nat Rev Immunol* **7**, 645-650 (2007).
162. Baekkeskov, S. *et al.* Identification of the 64K autoantigen in insulin-dependent diabetes as the GABA-synthesizing enzyme glutamic acid decarboxylase. *Nature* **347**, 151-156 (1990).
163. Kaufman, D.L. *et al.* Autoimmunity to two forms of glutamate decarboxylase in insulin-dependent diabetes mellitus. *J Clin Invest* **89**, 283-292 (1992).
164. Spellman, P.T. & Rubin, G.M. Evidence for large domains of similarly expressed genes in the Drosophila genome. *J Biol* **1**, 5 (2002).
165. Roix, J.J., McQueen, P.G., Munson, P.J., Parada, L.A. & Misteli, T. Spatial proximity of translocation-prone gene loci in human lymphomas. *Nat Genet* **34**, 287-291 (2003).
166. Ragoczy, T., Bender, M.A., Telling, A., Byron, R. & Groudine, M. 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 (2006).
167. Faraway, J.J. *Linear Models with R, Texts in Statistical Science.* (hapman & Hall/CRC., Boca Raton; 2005).
168. Oven, I. *et al.* AIRE recruits P-TEFb for transcriptional elongation of target genes in medullary thymic epithelial cells. *Mol Cell Biol* **27**, 8815-8823 (2007).
169. Tao, Y. *et al.* AIRE recruits multiple transcriptional components to specific genomic regions through tethering to nuclear matrix. *Mol Immunol* **43**, 335-345 (2006).
170. Dillon, N. Gene regulation and large-scale chromatin organization in the nucleus. *Chromosome Res* **14**, 117-126 (2006).
171. Sproul, D., Gilbert, N. & Bickmore, W.A. The role of chromatin structure in regulating the expression of clustered genes. *Nat Rev Genet* **6**, 775-781 (2005).
172. Kaern, M., Elston, T.C., Blake, W.J. & Collins, J.J. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* **6**, 451-464 (2005).
173. Paldi, A. Stochastic gene expression during cell differentiation: order from disorder? *Cell Mol Life Sci* **60**, 1775-1778 (2003).
174. Peterson, P., Org, T. & Rebane, A. Transcriptional regulation by AIRE: molecular mechanisms of central tolerance. *Nat Rev Immunol* **8**, 948-957 (2008).
175. Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. & Young, R.A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77-88 (2007).
176. Su, M.A. *et al.* Mechanisms of an autoimmunity syndrome in mice caused by a dominant mutation in Aire. *J Clin Invest* **118**, 1712-1726 (2008).
177. Misteli, T. Concepts in nuclear architecture. *Bioessays* **27**, 477-487 (2005).
178. Tanabe, H. *et al.* Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci U S A* **99**, 4424-4429 (2002).
179. Cremer, M. *et al.* Inheritance of gene density-related higher order chromatin arrangements in normal and tumor cell nuclei. *J Cell Biol* **162**, 809-820 (2003).
180. Heard, E. & Bickmore, W. The ins and outs of gene regulation and chromosome territory organisation. *Curr Opin Cell Biol* **19**, 311-316 (2007).
181. Misteli, T. Beyond the sequence: cellular organization of genome function. *Cell* **128**, 787-800 (2007).
182. Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R. & Flavell, R.A. Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637-645 (2005).
183. Niehrs, C. Active DNA demethylation and DNA repair. *Differentiation* **77**, 1-11 (2009).

184. Soutoglou, E. & Misteli, T. Mobility and immobility of chromatin in transcription and genome stability. *Curr Opin Genet Dev* **17**, 435-442 (2007).
185. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**, 897-903 (2008).
186. Cajiao, I., Zhang, A., Yoo, E.J., Cooke, N.E. & Liebhaber, S.A. Bystander gene activation by a locus control region. *EMBO J* **23**, 3854-3863 (2004).
187. Wilusz, C.J., Wormington, M. & Peltz, S.W. The cap-to-tail guide to mRNA turnover. *Nat Rev Mol Cell Biol* **2**, 237-246 (2001).
188. Ward, C.M. & Stern, P.L. The human cytomegalovirus immediate-early promoter is transcriptionally active in undifferentiated mouse embryonic stem cells. *Stem Cells* **20**, 472-475 (2002).
189. Ruan, H. *et al.* Killing of brain tumor cells by hypoxia-responsive element mediated expression of BAX. *Neoplasia* **1**, 431-437 (1999).
190. Holtkamp, S. *et al.* Modification of antigen-encoding RNA increases stability, translational efficacy, and T-cell stimulatory capacity of dendritic cells. *Blood* **108**, 4009-4017 (2006).
191. Cybulsky, A.V. *et al.* The 3'-untranslated region of the Ste20-like kinase SLK regulates SLK expression. *Am J Physiol Renal Physiol* **292**, F845-852 (2007).
192. Al-Ahmadi, W., Al-Haj, L., Al-Mohanna, F.A., Silverman, R.H. & Khabar, K.S. RNase L downmodulation of the RNA-binding protein, HuR, and cellular growth. *Oncogene* **28**, 1782-1791 (2009).
193. Pinal, C.S. & Tobin, A.J. Uniqueness and redundancy in GABA production. *Perspect Dev Neurobiol* **5**, 109-118 (1998).
194. Sharova, L.V. *et al.* Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res* **16**, 45-58 (2009).
195. Koble, C. & Kyewski, B. The thymic medulla: a unique microenvironment for intercellular self-antigen transfer. *J Exp Med* **206**, 1505-1513 (2009).
196. Scollay, R. & Godfrey, D.I. Thymic emigration: conveyor belts or lucky dips? *Immunol Today* **16**, 268-273; discussion 273-264 (1995).
197. Rodewald, H.R., Paul, S., Haller, C., Bluethmann, H. & Blum, C. Thymus medulla consisting of epithelial islets each derived from a single progenitor. *Nature* **414**, 763-768 (2001).
198. Guerau-de-Arellano, M., Martinic, M., Benoist, C. & Mathis, D. Neonatal tolerance revisited: a perinatal window for Aire control of autoimmunity. *J Exp Med* **206**, 1245-1252 (2009).
199. Le Borgne, M. *et al.* The impact of negative selection on thymocyte migration in the medulla. *Nat Immunol* **10**, 823-830 (2009).
200. Bousso, P. & Robey, E.A. Dynamic behavior of T cells and thymocytes in lymphoid organs as revealed by two-photon microscopy. *Immunity* **21**, 349-355 (2004).
201. Witt, C.M., Raychaudhuri, S., Schaefer, B., Chakraborty, A.K. & Robey, E.A. Directed migration of positively selected thymocytes visualized in real time. *PLoS Biol* **3**, e160 (2005).

Acknowledgements

I would like to express my sincere gratitude to the following persons who have supported and helped me during the course of my PhD:

First of all, Prof. Dr. Bruno Kyewski for the great supervision and for always being willing to discuss, help and motivate.

Prof. Dr. Günter Hämmerling for willingly agreeing to be the referee for this thesis.

Dr. Lars Tykocinski for helping me with epigenetics, working together and constantly being open to discuss my results and technical problems.

Dr. Jens Derbinski for introducing me to a number of methods and for your constant helpfulness and many discussions.

Dr. Steffen Schmidt and Klaus Hexel for the many hours you spent in FACS-sorting as well as helping to solve problems when I was using the sorter on my own.

Prof. Dr. Dr. Christoph Cremer for the collaboration with his group.

Yanina Weiland for teaching me the FISH and SPDM methods, for always finding time for discussion and helping me getting the method to work.

Dr. David Baddeley for allowing me to use his Matlab algorithms for the analysis of SPDM data and for teaching me to use the program.

Thanks also to the other members of the Cremer Group for helping me with the physics, microscopy and discussing my data.

Dr. Steven Johnson from the EMBL, Heidelberg, for sharing his ChIP protocol.

Prof. Dr. Magalena Götze for kindly providing the Gad67/eGFP mice.

Prof. Dr. Annette Kopp-Schneider for the statistic evaluation of the SPDM and single cell expression data.

Esmail Reszavandy and Stefanie Egle for their help with mouse preparations and a number of other experiments.

The animal keepers at the DKFZ animal isolator station, in particular Ute Riesterer, for taking excellent care of the mice.

Last but not the least all members of the Department of Developmental Immunology for all the help, the great work atmosphere and all the good moments in (and out of) the lab! Thank you Janna, Jochen, Jana, Christian, Jens, Lars, Sheena, Martina, Chloé, Paula, Martin, Janine!

This note of thanks will not be complete without a special mention of the constant support and encouragement from my parents, and, in particular, from Pascal. Thank you.