Medullary Epithelial Cells of the Human Thymus Express a Highly Diverse Selection of Tissue-specific Genes Colocalized in Chromosomal Clusters

Jörn Gotter,1 Benedikt Brors,2 Manfred Hergenhahn,3 and Bruno Kyewski1

1Tumor Immunology Program, 2Intelligent Bioinformatics Systems, and 3Genetic Alterations in Carcinogenesis, German Cancer Research Center, D-69120 Heidelberg, Germany

Abstract

Promiscuous expression of tissue-specific self-antigens in the thymus imposes T cell tolerance and protects from autoimmune diseases, as shown in animal studies. Analysis of promiscuous gene expression in purified stromal cells of the human thymus at the single and global gene level documents the species conservation of this phenomenon. Medullary thymic epithelial cells over-express a highly diverse set of genes (>400) including many tissue-specific antigens, disease-associated autoantigens, and cancer-germline genes. Although there are no apparent structural or functional commonalities among these genes and their products, they cluster along chromosomes. These findings have implications for human autoimmune diseases, immuno-therapy of tumors, and the understanding of the nature of this unorthodox regulation of gene expression.

Key words: self-tolerance • autoimmunity • tumor antigens • epigenetics • gene array

Introduction

Self/nonself discrimination is an essential property of adaptive immunity. The T cell repertoire acquires a state of self-tolerance specific for each individual by imprinting different fates on developing immature T cells depending on their propensity to recognize self. The extent of self-tolerance is necessarily dictated by the diversity of self-antigens accessible to the nascent repertoire within the thymus, when lymphocytes are most sensitive to tolerance induction. Although this diversity of self-antigen display may be high due to distinct APC subsets in the thymus, each presenting its unique sets of self peptides (1), it has been held inconceivable that the immunological “self” can be comprehensively represented in the thymus. This reasoning is particularly true for “tissue-specific” self-antigens, which play a prominent role in most common autoimmune diseases and by definition should be restricted in their expression to lineage-specific cells of the respective organ. Various mechanisms collectively known as peripheral tolerance have been described which offer an explanation for this conundrum (2).

More recently, however, the relative contributions of central versus peripheral tolerance have been reevaluated based on accruing evidence that tissue-specific antigens are expressed within the thymus and displayed there for repertoire selection. This physiological expression of tissue-specific antigens by thymic epithelial cells (TECs), in particular medullary TECs (mTECs), has been termed promiscuous as distinguished from ectopic (3, 4). Promiscuous gene expression is a cell-autonomous property of mTECs and is maintained during the entire period of thymic T cell output (3). By inclusion of numerous self-antigens expressed by different parenchymal organs, this mechanism widens the scope of central tolerance in a manner not anticipated by original concepts of self-tolerance (5).

Although promiscuous gene expression by mTECs and its role in self-tolerance in mice are by now undisputed, it is still unclear whether this tolerance mechanism has been conserved across species barriers during evolution including humans. This issue is of particular interest in view of the recent report showing that the rare human Autoimmune Polyglandular Syndrome (APS) (6), a recessive autosomally inherited monogenic disease, is modeled by mice deficient in the autoimmune regulator (Aire) gene (7, 8). These mutant mice display a partial defect in promiscuous gene expression.
Materials and Methods

After stepwise digestion with collagenase-dispase (3 10 min, 4 37 •C), low density cells were incubated with mAbs anti–CD45 (sense and antisense, respectively): AIRE, 5'CTGGCTCCCAACACTGCCAACGTC-3', MART-1, tyrosinase), 33 (PLP), 32 (insulin, GAD67, thyroglobulin, collagen II, 5'-TCCTTTGGGTTTGCAACGGATTGT-3'. Cells were washed with FITC-conjugated mAb HLA-DR (BD Biosciences), PE-conjugated mAb anti-CD11c (BD Biosciences), and epithelial cell adhesion molecule (EpCAM; clone HEA125)-biotin/sav-CyChrome. Cells were sorted with a FACSVantage Plus cell sorter, and EpCAM+ epithelial cells were excluded by appropriate gating. TECs were isolated from the third trypsin fraction. Cells were enriched in Percoll (p 1.07 g/cm3, 1,700 g, 10 min, 4°C), and myogenic cells were depleted from the low density fraction with anti–CD45 Dynabeads (cells/beads = 1:3). After staining with mAbs biotinylated EpCam/sav-PE and cortical dendritic reticulum antigen 2 (CDR2)-Alexa488 (Alexa Fluor 488 Protein Labeling kit; Molecular Probes) (19), cells were sorted, excluding remaining myogenic (CD45+) cells (stained with mAb anti–CD45-CyChrome; Immunotech). Thymocytes were sorted as CD3+ cells (mAb anti–CD3–FITC; BD Biosciences) from the medium fraction. Dead cells were always excluded with propidium iodide (1 µg/ml). The average cell yield per thymus aliquot (3–5 cm3, equivalent to 15 mouse thymi) was 6 106 mTECs, 106 cTECs, and 6 105 DCs.

Immunohistochemistry. Human thymic tissue pieces were embedded in Tissue-Tek (Sakura) and snap frozen in liquid nitrogen. Cryosections (4 µm) were fixed in ice-cold acetone for 20 min, air dried, and soaked in PBS supplemented with 0.01% Tween 20. To reduce endogenous peroxidase activity, sections were incubated with H2O2 in 0.1 M sodium acetate (0.3%) for 20 min, and unspecific binding sites were blocked by incubation with 5% mouse serum in PBS/Tween for 20 min. Specific staining was performed at room temperature in PBS/Tween for 45 min with the biotinylated mAbs CDR2 (cortex) and HEA125 (medulla), respectively. After a 45-min incubation with peroxidase-conjugated streptavidin (Jackson ImmunoResearch Laboratories), sections were developed with a substrate buffer containing 5 mM 3-amino–9-ethylcarbazol and 0.15% H2O2 in 0.1 M sodium acetate.

RNA Preparation and cDNA Synthesis. RNA was isolated from single-cell suspensions with the High Pure RNA Isolation kit (Roche) and eluted in 50 µl H2O. 48 µl of this RNA solution or 1 µg total RNA from commercially available control tissues (Stratagene) were treated with 6 U DNase I (Invitrogen) and reverse transcribed into cDNA with Oligon(dT)12 Primer and Superscript II Reverse Transcriptase (Invitrogen); this was followed by RNase H digestion (Promega).

RT-PCR Analysis. PCRs were performed in a final volume of 25 µl with 1 U REDTaq DNA Polymerase (Sigma–Aldrich). Final concentrations of the PCR mix were: 250 nM for each primer, 200 µM dNTP (MBI Fermentas), 10 nM Tris·HCl (pH 8.3), 50 mM KCl, 1.1 mM MgCl2, and 0.01% gelatin. The amplification was performed in a Thermacycler (PTC-100; MJ Research) under standard conditions: a single denaturation step at 94°C for 3 min followed by either 35 (GAD65, IA-2, TPO, H/K-ATPase (MOG), RetSag, IRBP, MAGE-A1, -A3, -A4, NY-ESO-1, MART-1, tyrosinase), 33 (PLP), 32 (insulin, GAD67, thyroglobulin, DC-LAMP), 30 (WHN, AIRE, collagen II), or 21 cycles (for GAPDH normalization) of 94°C for 1 min, 54–62°C for 1 min and 72°C for 2 min, followed by a final extension step of 72°C for 10 min. The following oligonucleotide pairs were used (sense and antisense, respectively): AIRE, 5’-TGCCAAGATGATCACTGTC-3’ and 5’-TTCCAGAGTAGAAGGTCGGGC-3’; collagen II, 5’-CTGGCCTCCCAACTGTGCAAAGTC-3’ and 5’-TCTCTTGGGTTTGGACACGATTGT-3’; DC-LAMP, 5’-GCCAGATGCCAGTCAATAGA-3’ and 5’-GAATATCCTC-

Downloaded from jem.rupress.org on July 9, 2017

156 Promiscuous Gene Expression in Human Thymic Epithelium

Materials and Methods

Tissues. Samples of thymic tissue from five patients (014: male, 8 mo; 015: male, 18 mo; 017: male, 5 mo; 018: female, 5 mo; and 019: female, 2 mo) were obtained in the course of corrective cardiac surgery at the Department of Cardiac Surgery, Medical School, University of Heidelberg. This study has been approved by the Ethics Committee of the University of Heidelberg.

Isolation of Human Thymic Cells. Pieces of thymic tissue were minced into very small fragments and stirred in RPMI-1640 medium on ice (2 10 min) to remove the majority of thymocytes. After stepwise digestion with collagenase-dispase (3 20 min, 37°C) and trypsin–EDTA (3 15 min, 37°C), DCs were purified according to a protocol described previously (18). Briefly, rosettes from the pooled collagenase-dispase fractions were dissociated for 5 min by incubation with EDTA/PBS (25 mM, 4°C), low density cells were incubated with mAbs anti-CD3 (OKT3; American Type Culture Collection) and anti-CD19 (HD37; DKFZ–Heidelberg), and remaining thymocytes and contaminating B cells were depleted with anti–pan-IgG Dynabeads (Dynal). Cells were stained with FITC-conjugated mAb HLA-DR (BD Biosciences), PE-conjugated mAb anti-CD11c (BD Biosciences), and epithelial cell adhesion molecule (EpCAM; clone HEA125)-biotin/sav-CyChrome. Cells were sorted with a FACSVantage Plus cell sorter, and EpCAM+ epithelial cells were excluded by appropriate gating. TECs were isolated from the third trypsin fraction. Cells were enriched in Percoll (p 1.07 g/cm3, 1,700 g, 10 min, 4°C), and myogenic cells were depleted from the low density fraction with anti–CD45 Dynabeads (cells/beads = 1:3). After staining with mAbs biotinylated EpCam/sav-PE and cortical dendritic reticulum antigen 2 (CDR2)-Alexa488 (Alexa Fluor 488 Protein Labeling kit; Molecular Probes) (19), cells were sorted, excluding remaining myogenic (CD45+) cells (stained with mAb anti–CD45-CyChrome; Immunotech). Thymocytes were sorted as CD3+ cells (mAb anti–CD3–FITC; BD Biosciences) from the medium fraction. Dead cells were always excluded with propidium iodide (1 µg/ml). The average cell yield per thymus aliquot (3–5 cm3, equivalent to 15 mouse thymi) was 6 106 mTECs, 106 cTECs, and 6 105 DCs.
Promiscuous Gene Expression in Medullary Epithelial Cells Is Conserved between Mice and Man. We first assessed the expression patterns, we regarded these populations as sufficient to pursue gene expression analysis. Based on these expression patterns, we regarded these populations as sufficient to pursue gene expression analysis.

**Microarray Analysis.** Total RNA from thymic samples 017 and 018 (4.5 μl, corresponding to ~6 × 10^6 mTECs, 10^3 cortical TECs (cTECs), and 6 × 10^5 DCs) was preamplified and biotinylated in two independent experiments by two rounds of cDNA synthesis and in vitro transcription as described previously (16, 17, 20). Biotinylated aRNA was hybridized to Gene Chip Arrays (Human Genome U95Av2; Affymetrix) and sequentually stained in a GeneChip Fluidics Station 400 (Affymetrix) with streptavidin-PE (Molecular Probes), biotinylated antistreptavidin mAb (Linaris), and again with streptavidin-PE. Arrays were scanned with a Gene Array Scanner (Hewlett Packard) and evaluated using the software Microarray Suite 5.0 (Affymetrix) and Excel 97 (Microsoft). To identify genes overexpressed in the different thymic cell types, comparison analyses between arrays from each dataset were performed. We defined overexpression of a probe set as reliable when the following criteria were met in both experiments: present or marginally present, increased or marginally increased, and signal log ratio low ≥1 compared with the reference subset.

Statistical Analysis. To test for statistical significance of clustering, we used randomized lists of genes from the same chip for comparison according to the method described by Roy et al. (21) Briefly, we produced random lists of genes of the same length as the number of genes overexpressed in mTECs. We calculated the number of pairs of genes for each of the lists that were located on the same chromosome within a distance of 35, 50, 80, 120, 200, 300, 500, 1,000, 2,000, 3,000, or 5,000 kb. The numbers obtained from the list of genes overexpressed in mTECs were compared with those obtained from 10,000 random lists. P-values were calculated from the empirical distribution (22). To test the number of clusters of a given size, we calculated the number of genes that were located within a sliding window of 10 consecutive genes, taking the highest local maximum as cluster size. The frequency of clusters of a given size was determined for both the list of overexpressed genes and for 1,000 random lists.

**Online Supplemental Material.** Fig. S1 shows RT-PCR results for the expression of selected genes in thymic stromal cells. Table S1 lists all genes overexpressed in mTECs versus cTECs, Table S2 lists all genes overexpressed in cTECs versus mTECs, and Table S3 lists genes overexpressed in mTECs and which presumably are involved in gene regulation. Supplemental material is available at http://www.jem.org/cgi/content/full/jem.20031677/DC1.

**Results**

**Purification of Stromal Cells of the Human Thymus.** A protocol previously established for the isolation of thymic stromal cells in mice has been adapted to purify the corresponding cell types from the human thymus (3). A combination of stepwise enzymatic digestion, density centrifugation, magnetic cell depletion, and multicolor sorting yielded pure populations of mature thymic DCs, cTECs, and mTECs. The differential coexpression of EpCAM and CDR2, a cTEC-specific antigen (19), allowed separation of the latter two cell subsets (Fig. 1 a). Although CDR2 is specifically expressed on cTECs and rare nonepithelial cells of the medulla, EpCAM is expressed both on mTECs and at 10-fold lower levels on cTECs rather than DC contamination. Based on these expression patterns, we regarded these populations as sufficient to pursue gene expression analysis. The cell yield per tissue volume was comparable between mouse and human. The purity of these populations was verified by PCR expression analysis of indicator genes, i.e., FOXL1 was expressed in mTECs and cTECs but not in DCs, AIRE in mTECs, and weakly in DCs and DC-LAMP in DCs and surprisingly also in mTECs (Fig. 1 b). We presume that expression of DC-LAMP in mTECs either reflects promiscuous expression or a physiological feature of mTECs rather than DC contamination. Based on these expression patterns, we regarded these populations as sufficiently pure to pursue gene expression analysis.
were analyzed by RT-PCR (Table I). All antigens were reproducibly expressed in mTECs with the exception of GAD65 and the H/K-ATPase β chain, which were below detection in all thymic cells analyzed. Messenger RNAs for PLP and Golli-MBP showed a broader expression pattern beyond mTECs, a finding already reported for the mouse thymus (3, 4). The unequivocal detection of MOG mRNA, the intrathymic expression of which had been controversial (14), documents the gain in sensitivity when purified cells compared with whole thymus or laser-captured medullary areas were analyzed.

Expression of a second class of peripheral autoantigens, so called tumor-associated antigens, has been claimed to be absent from the thymus and the peripheral immune system and consequently thought to be exempted from self-tolerance (23, 24). We tested the expression pattern of four members of the cancer-germ line group (MAGE–A1, -A3, -A4, and NY-ESO-1) and two members of the group of melanoma differentiation antigens (MART-1 and tyrosinase), some of which had been selected for current clinical trials (25). Cancer-germ line antigens are expressed in male germ cells in uterus and various tumors and melanoma differentiation antigens in melanocytes and melanomas (26). All of these genes were detectable in purified mTECs, albeit with interindividual differences in their expression levels (Fig. 2). Whereas MAGE–A1 was only detectable in one out of five thymi, MART–1 was present in all five samples. It is also notable that the expression levels were much lower than in the corresponding peripheral tissues. Notwithstanding these qualifications, the data show that tumor-associated antigens previously thought to be secluded from the immune system are also expressed by mTECs and

Figure 1. Isolation of distinct stromal cells of the human thymus. (a) Sort regions selected for the isolation of thymic stromal cell subsets; the coexpression profile for CDR2 and EpCAM of low density, CD45 negative trypsin-digested cells is shown. Cortical TECs are defined as CDR2, EpCAM, and mTECs as CDR2, EpCAM. The corresponding histological staining patterns of both Abs are shown. Thymic DCs are defined as CD11c, HLA-DR, EpCAM cells of low density. (b) Expression of a selected panel of marker genes was assessed in purified cTECs, mTECs, and DCs of five thymi by RT-PCR. The cell type–specific expression patterns document the purity of the respective subsets. The expression of DC-LAMP in mTECs probably reflects its promiscuous expression rather than contamination by DCs. The amount of input cDNA was normalized according to signals obtained for GAPDH in a titration experiment (not depicted).
thus presumably displayed to developing T cells. The selectivity of this expression in mTECs underpins the unique role this cell type plays in promiscuous gene expression. The high score with which “arbitrarily” selected genes were found to be expressed in mTECs and the diversity of these promiscuously expressed genes in mice and man implies that this gene pool is complex and comprehensive. To more precisely delineate this gene pool, we defined the transcriptome of these cell types by microarray analysis.

Promiscuously Expressed Genes Are Highly Diverse. We performed a comparison of the gene expression profiles of cTECs, mTECs, DCs, and mature thymocytes. RNA was isolated from independent replicates of each subset, and amplified cDNA were prepared by two rounds of cDNA synthesis and in vitro transcription steps and hybridized to GeneChip® oligonucleotide arrays (Affymetrix U95Av2) containing 12,500 probe sets. Scanned arrays were analyzed with Affymetrix software Microarray Suite 5.0 to identify cell type-specific gene expression profiles and assess quantitative difference in gene expression among these subsets. Calculation of the fold change in gene expression was based on the Lower Confidence Bound as a measure of enrichment of gene expression. Genes for which the signal log ratio low was equal to a difference of at least two-fold and which at the same time were designated increased or marginally increased were considered (see Materials and Methods). Comparisons were performed among the different cell subsets within each experiment, and only genes fulfilling these criteria in both comparisons were considered (Fig. 3, a–c).

The mutual comparisons of mTECs with cTECs, DCs, and thymocytes revealed in each case a much higher number of genes being expressed in mTECs compared with the reference population (Fig. 3, a compared with c). The number of differentially expressed genes correlates with the ontogenetic kinship of these cell lineages (Fig. 3 d). We regard the comparison between mTECs and cTECs as most

Table I. Cell Type-specific Expression Pattern of Selected Self-antigens Purportedly Associated with Common Autoimmune Diseases

<table>
<thead>
<tr>
<th>Organ</th>
<th>Autoantigen</th>
<th>cTECs</th>
<th>mTECs</th>
<th>DCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pancreas</td>
<td>Insulin*</td>
<td>------</td>
<td>++++</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>GAD67*</td>
<td>------</td>
<td>++++</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>GAD65*</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>IA-2a</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Thyroid</td>
<td>Thyroglobulin</td>
<td>------</td>
<td>++++</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>TPO*</td>
<td>------</td>
<td>++++</td>
<td>------</td>
</tr>
<tr>
<td>Stomach</td>
<td>H/K-ATPase α</td>
<td>------</td>
<td>++++</td>
<td>++++</td>
</tr>
<tr>
<td></td>
<td>H/K-ATPase β</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Brain</td>
<td>Golli-MBP</td>
<td>++++</td>
<td>++++</td>
<td>++++</td>
</tr>
<tr>
<td></td>
<td>PLP/DM20</td>
<td>++++</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>MOG</td>
<td>------</td>
<td>++++</td>
<td>------</td>
</tr>
<tr>
<td>Eye</td>
<td>RetSAg*</td>
<td>------</td>
<td>++++</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>IRBP*</td>
<td>------</td>
<td>++++</td>
<td>------</td>
</tr>
<tr>
<td>Cartilage</td>
<td>Collagen II</td>
<td>++++</td>
<td>++o++</td>
<td>---</td>
</tr>
</tbody>
</table>

Expression was analyzed by RT-PCR applying 30–35 cycles. Quantitative differences in signal strength were not considered (see also Fig. S1). +, present; −, absent; Ø, not determined.

*aThe TEC-specific expression pattern of antigens has been confirmed in six additional thymi. In addition, mRNA expression of albumin, α-fetoprotein, C-reactive protein, and erythropoietin in mTECs has been found in six thymi (not depicted).
informative, given the close relationship between these two cell types. It is worth noting that 70% of the signals in this case differed by more than fourfold (SLR $\geq 2$). The pool of genes overexpressed in mTECs comprised 443 genes (Table S1, available at http://www.jem.org/cgi/content/full/jem.20031677/DC1), and the pool overexpressed in cTECs comprised 162 genes (Table S2, available at http://www.jem.org/cgi/content/full/jem.20031677/DC1). The ratio between mTECs and cTECs pool as a measure of differential gene expression between these two cell types was remarkably invariant between man and mouse (2.7 versus 2.6; Derbinski, J., personal communication). However, this ratio is probably higher for two reasons. First, the pool of genes overexpressed in mTECs is underestimated, since many genes detected by RT-PCR did not fulfill the criteria of our array comparison. This was not unprecedented since the array analysis is known to be less sensitive than RT-PCR analysis. Note that mTEC-specific genes (i.e., absent in cTECs) were predominantly expressed at low levels, in concordance with the low signal strength of promiscuously expressed genes in mice (8) (Fig. 3 a). Second, the cTEC-specific pool is overestimated, since it comprises many genes specific for T cells due to contamination of this subset by thymic nurse cells.

The genes overexpressed in mTECs fall into at least two categories: those which constitute the differentiation program of this epithelial cell lineage and those which are regarded as tissue specific, i.e., expressed in a promiscuous manner. We based the gene assignment on data reported by Su et al. (27). In this study, gene expression in >40 human tissues and cell lines had been assessed by gene array analysis using the Affymetrix U95A arrays. This assignment, however, remains preliminary, since different reports on tissue specificity of gene transcription vary widely, depending on the methods applied. Notwithstanding this qualification, the expression of at least 80 out of 443 genes in this pool can be qualified as tissue restricted based on their reported expression pattern and/or functional designation (27). Most tissues including mammary gland, liver, muscle, kidney, pancreas, placenta, thyroid, salivary glands, testis, and prostate are represented as shown for 80 selected genes (Table II).

Given the recent demonstration that Aire, a putative regulator of gene transcription, partly controls promiscuous gene expression in murine mTECs (8), we searched for gene products which are known or suspected to control gene expression at the genetic or epigenetic level. 25 genes overexpressed in mTECs submit to these criteria, yet none of these factors appeared to be specific for mTECs. Noteworthy, AIRE showed the highest enrichment in mTECs (Table S3, available at http://www.jem.org/cgi/content/full/jem.20031677/DC1).

Although global gene analysis in mTECs documents the diversity of promiscuously expressed genes, it does not reveal commonalities with regard to structure, function, or gene ontology, which may have selected this set of genes. Alternatively, promiscuous gene expression may target certain chromosomal domains via epigenetic control mechanism. Hence, we analyzed the chromosomal and subchromosomal location of this gene pool.

**Promiscuously Expressed Genes Cluster along the Chromosome.** The chromosomal position of 415 of the 443 genes overexpressed in mTECs had been mapped. The relative distribution of these genes on the various chromosomes showed no marked under- or overrepresentation for particular chromosomes compared with the distribution of all mapped genes of the array (Fig. 4 a). Since the two datasets
are from a male and a female thymus, we did not evaluate the Y chromosome. Notably, we did not find preferential X chromosomal locations as reported for genes expressed in spermatogonia, which partially overlap with genes expressed in mTECs (28, 29). Next, we analyzed whether promiscuously expressed genes are clustered along chromosomes. Recently, several studies in different species showed that genes coexpressed in a particular tissue are aligned in clusters (21, 30). We specifically calculated the probability with which 2 genes of the 415 gene pool would occur within a DNA window of 200 kb. This number was compared with the corresponding probabilities of 400 randomly sampled genes present on the arrays probing 10,000 permutations. There was a highly significant clustering of overexpressed genes (Fig. 4b). This difference also holds when windows ranging between 35 and 5,000 kb were tested (not depicted). In addition, the experimental and random frequencies of 2, 3, 4, and 5 genes being clustered within a window of 10 adjacent genes present on the arrays irrespective of genetic distance has been calculated. This analysis revealed a clear overrepresentation of triplets and quadruplets in the experimental gene set compared with 1,000 random draws of the same number of genes (Fig. 4c). An example of 10 genes clustered within 5 Mbp on chromosome 1, SELENBP1, selenium binding protein 1; S100A, calcium binding protein A; FLG, filaggrin; IVL, involucrin; SPRR1B, small proline rich protein 1B; NPR1, atrial natriuretic peptide receptor A; MUC1, mucin 1.

Discussion

This study documents that promiscuous gene expression by thymic epithelial cells is highly conserved between mouse and human. This correspondence at the expression level is most likely based on a common molecular regulation in both species. The conservation lends credit to anal-
Table II. Expression of Tissue-specific Genes in Human TECs

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name/Function</th>
<th>Symbol</th>
<th>Name/Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammary gland</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSN1</td>
<td>Casein</td>
<td>SCGB2A2</td>
<td>Mammaglobin</td>
</tr>
<tr>
<td>CSN2</td>
<td>Casein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adipose tissue</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APM2</td>
<td>Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AADAC</td>
<td>Esterase</td>
<td>FGG</td>
<td>Fibrinogen</td>
</tr>
<tr>
<td>APOA1</td>
<td>Apolipoprotein</td>
<td>PKC2</td>
<td>PEP-carboxykinase</td>
</tr>
<tr>
<td>APOC3</td>
<td>Apolipoprotein</td>
<td>SERPINA5</td>
<td>Antitrypsin</td>
</tr>
<tr>
<td>DPYS</td>
<td>Dihydropyrimidinase</td>
<td>UGT2B10</td>
<td>UDP-glucuronosyl-transferase</td>
</tr>
<tr>
<td>FGB</td>
<td>Fibrinogen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muscle</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACTC</td>
<td>Alpha actin</td>
<td>MYH7</td>
<td>Myosin</td>
</tr>
<tr>
<td>AK1</td>
<td>Adenylate kinase</td>
<td>MYL1</td>
<td>Myosin</td>
</tr>
<tr>
<td>CASQ2</td>
<td>Calcium-binding protein</td>
<td>MYL2</td>
<td>Myosin</td>
</tr>
<tr>
<td>CHRNA1</td>
<td>Acetylcholine receptor</td>
<td>TNNC1</td>
<td>Troponin</td>
</tr>
<tr>
<td>CSRP3</td>
<td>Heart-LIM protein</td>
<td>TNNI1</td>
<td>Troponin</td>
</tr>
<tr>
<td>ENO3</td>
<td>Enolase</td>
<td>TNN1</td>
<td>Troponin</td>
</tr>
<tr>
<td>MB</td>
<td>Myoglobin</td>
<td>TTN</td>
<td>Myofilament</td>
</tr>
<tr>
<td>Nervous system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADAM23</td>
<td>Metalloproteinase</td>
<td>HAPiP</td>
<td>Vesicle transport</td>
</tr>
<tr>
<td>CADPS</td>
<td>Calcium-binding protein</td>
<td>KCNK1</td>
<td>Potassium channel</td>
</tr>
<tr>
<td>CALB1</td>
<td>Calbindin</td>
<td>NEF3</td>
<td>Neurofilament</td>
</tr>
<tr>
<td>CCK</td>
<td>Prohormone</td>
<td>PADI2</td>
<td>Deiminase</td>
</tr>
<tr>
<td>CHGB</td>
<td>Secretogranin</td>
<td>PCDH8</td>
<td>Protocadherin</td>
</tr>
<tr>
<td>CNR1</td>
<td>Receptor</td>
<td>PRSS12</td>
<td>Neurotrypsin</td>
</tr>
<tr>
<td>FGF1</td>
<td>Growth factor</td>
<td>SST</td>
<td>Somatostatin</td>
</tr>
<tr>
<td>GABBR1</td>
<td>Receptor</td>
<td>SV2B</td>
<td>Synaptic vesicle protein</td>
</tr>
<tr>
<td>GABRA5</td>
<td>GABA receptor</td>
<td>SYN-G1</td>
<td>Synaptogyrin</td>
</tr>
<tr>
<td>GAL</td>
<td>Galanin</td>
<td>SYT1</td>
<td>Synaptotagmin</td>
</tr>
<tr>
<td>GPR37</td>
<td>G protein-linked receptor</td>
<td>TAC1</td>
<td>Neuropeptide</td>
</tr>
<tr>
<td>Kidney</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABP1</td>
<td>Membrane glycoprotein</td>
<td>TM7SF1</td>
<td>Transmembrane protein</td>
</tr>
<tr>
<td>SPP1</td>
<td>Osteopontin</td>
<td>UGT1A4</td>
<td>UDP-glucuronosyl-transferase</td>
</tr>
<tr>
<td>Pancreas</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMY2B</td>
<td>Amylase</td>
<td>PRSS2</td>
<td>Trypsin II</td>
</tr>
<tr>
<td>E1A3A</td>
<td>Elastase</td>
<td>PRSS4</td>
<td>Trypsin III</td>
</tr>
<tr>
<td>GAD1</td>
<td>Glutamate-decarboxylase</td>
<td>RNASE1</td>
<td>RNase A1</td>
</tr>
<tr>
<td>GCC</td>
<td>Glucagon</td>
<td>SCGN</td>
<td>Secretagogin</td>
</tr>
<tr>
<td>LEFTB</td>
<td>L/R determination factor B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placenta</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGA</td>
<td>Glycoprotein hormone</td>
<td>PRG2</td>
<td>Proteoglycan</td>
</tr>
<tr>
<td>CLDN4</td>
<td>Tight junction protein</td>
<td>SDC1</td>
<td>Membrane protein</td>
</tr>
<tr>
<td>CSH2</td>
<td>Placental lactogen hormone</td>
<td>SEMA3B</td>
<td>Semaphorin</td>
</tr>
<tr>
<td>Prostate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACPP</td>
<td>Phosphatase</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ogies frequently drawn between experimental disease models and human autoimmune diseases. In the context of this study, this applies in particular to central tolerance as imparted by promiscuous gene expression and to Aire−/− mice and APS1 patients.

The isolation of purified thymic stromal cells of the human thymus has been a prerequisite to unambiguously and reproducibly detect expression of certain genes in the human thymus. Thus, MOG expression, in contrast to other myelin–specific antigens, had not been detected when whole thymus tissue or laser-captured medullary regions were analyzed (11, 14). Likewise, tumor-associated antigens including differentiation antigens (e.g., tyrosinase and MART-1) and cancer-germline antigens (i.e., the members of MAGE-A group or NY-ESO-1) had been considered absent from the thymus and the peripheral immune system for this matter (23, 24). In both cases, it had been inferred that spatial seclusion from the immune system precludes self-tolerance to these antigens. The expression of these antigens in mTECs, albeit at varying levels, is compatible with self-tolerance induction to these antigens. Given the gain in sensitivity by purifying mTECs, we conclude that promiscuous gene expression in the human thymus can only be reliably assessed in purified stromal cell subsets.

The expression analysis by RT-PCR documents a remarkable degree of species conservation between mouse and man. This does not only apply to the selection of genes expressed (as far as analyzed by RT-PCR, all genes transcribed in man are also expressed in mice except for Mage-a [3, 4]) but also to the cell type–specific expression pattern. All promiscuously expressed genes are found in mTECs, re-emphasizing their unique role. This has been confirmed by analysis of six additional thymi (unpublished data). However, expression was not always confined to mTECs; thus, PLP and golli-MBP were also expressed at comparable levels in cTECs, and this is true both for mouse and man (3, 4).

It is worth pointing out that GAD65 is apparently absent from human mTECs and thus likely to be absent from the human thymus. Absence of expression in the thymus and predominant expression in a peripheral tissue, as is the case for GAD65, is a particular liability for autoreactivity. GAD65 has been suspected to be the inciting or at least one of the initial autoantigens targeted in the early course of Diabetes mellitus type 1 (DM T1) (33–37), and the reason for this may be lack of central tolerance. In contrast, GAD67 shows the inverse expression pattern. This could explain the prevalence of specific autoantibodies to these two antigens in the course of DM T1, which is inversely correlated with their intrathyhmic expression levels (Fig. S1, available at http://www.jem.org/cgi/content/full/jem.20031677/DC1). A similar reasoning may apply to the two chains of the proton ATPase pump of the stomach. Although the α chain is clearly expressed in mTECs and DCs, the β chain is at best marginally expressed in mTECs. In the mouse model, the β chain but not the α chain is disease inducing and protective when ectopically expressed in the thymus (38). Thus, not only the absence of a particular exon, as exemplified by PLP (39), but also lack of intrathyhmic expression of an entire antigen may predispose to organ-specific autoimmunity. This supposition is all the more likely, since already moderate reductions in the levels of self-antigen expression (two- to fourfold) seem to affect the degree of self tolerance (40–43). A correlation between intrathyhmic expression levels and autoantibody prevalence is less apparent for insulin and IA-2. The wide range of insulin–specific autoantibody prevalence in DM T1 may be due to variations in intrathyhmic expression levels of insulin. Such variations have been documented for this self-antigen (40–42).

The set of 443 genes overexpressed in mTECs compared with cTECs delineates the extent and diversity of promiscuous gene expression in a first approximation. The size of this gene pool is comparable in mouse and man (555 versus 443 genes, respectively). It is certainly underestimated, since most genes detected by PCR have not been found by global gene analysis, despite representation on the chip. In addition, the U95Av2 array set represents only 1/3 of all human transcripts. Based on these considerations, we estimate the pool to contain at least a total of 1,200–3,000 genes or

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name/Function</th>
<th>Symbol</th>
<th>Name/Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory tract</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MUC4</td>
<td>Mucin</td>
<td>TSPAN-1</td>
<td>Transmembrane protein</td>
</tr>
<tr>
<td>SCG1B1A1</td>
<td>Uteroglobin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thyroid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLDN3</td>
<td>Tight junction protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salivary gland</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRB2</td>
<td>Secreted protein</td>
<td>SLPI</td>
<td>Protease inhibitor</td>
</tr>
<tr>
<td>Testis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSTM3</td>
<td>Glutathione transferase</td>
<td>SPAG6</td>
<td>Sperm antigen</td>
</tr>
</tbody>
</table>

Genes overexpressed in mTECs, as identified by array analysis, were evaluated for their function and/or tissue-specific expression based on the analysis by Su et al. (27). Note the functional diversity of the selected genes and their broad tissue distribution.
5–10% of all transcripts, of which only part are likely to be strictly restricted to a particular cell type. This estimate is well in accord with the number of genes being regulated by the Aire gene in mice, which, according to a recent report ranges between 200 and 1,200 genes (8). Of note, Aire only regulates a subset of promiscuously expressed genes.

The mTEC-overexpressed genes have no obvious structural or functional commonalities, they show no preference for any chromosome (Fig. 4 a), and they represent most if not all peripheral tissues (Table II). Despite this apparent arbitrariness, the pool includes genes, which are of particular physiological or pathophysiological relevance. Thus, human mTECs express self-antigens known or suspected to be targeted in common autoimmune diseases, like DM T1, Multiple Sclerosis, Hashimoto’s thyroiditis, Myasthenia gravis, autoimmune gastritis, retinitis, and rheumatoid arthritis (Table I). Most relevant to the role of AIRE in regulating promiscuous gene expression, we find many candidate antigens eliciting auto antibodies in APS-1 patients, e.g., thyroid peroxidase, thyroglobulin, P450 cytochrome subfamilies, and IA-2 expressed in mTECs (44, 45). Down-regulation or erasure of expression of these genes due to AIRE mutations may underlie the preference in organ pathology of APS-1 patients. The high species conservation between mouse and man strengthens a causal link between partial defunct promiscuous gene expression in mTECs and resultant multiorgan autoimmunity as the underlying pathophysiology of this rare human autoimmune syndrome.

The unequivocal and specific expression of tumor-associated antigens in human mTECs, in particular those of the cancer-germ line group, calls for a reappraisal of their immune-privileged status. Cancer-germ line antigens had so far only been found in male germ cells, a few cells of the uterus and tumor cells (25, 26). Since male germ cells lack MHC expression and the testis is sequestered from circulating lymphocytes, it had been assumed that these antigens are precluded from tolerance. Given the exquisite sensitivity of central tolerance, even the low and variable expression of these tumor antigens may result in some degree of self-tolerance (46, 47). Whether central tolerance contributes to the limited success of current clinical vaccination trials with peptides derived from these antigens remains conjectural (48).

Promiscuous expression of certain genes implies that developmental patterns of gene expression are overridden. Thus, self-antigens specifically induced or up-regulated during pregnancy (i.e., casein) are already found in the postnatal thymus. This “uncoupled” expression offers an explanation for a long-standing enigma, the induction of self-tolerance to antigens, which arise only in adulthood, i.e., during puberty and pregnancy. Likewise sex-specific gene expression is abolished. Testis-specific antigens, i.e., sperm-associated antigen 6, are expressed in female mTECs and placental antigens, i.e., placental lactogen hormone, in male mTECs. As a corollary, certain antigens are probably only expressed in mTECs and no other cell of the body.

Although promiscuously expressed genes are quite diverse, they still incompletely represent the vast array of tissue-specific self-antigens. How is self-tolerance then preserved to those antigens which are not expressed by mTEC-like GAD65, notwithstanding peripheral tolerance? As argued elsewhere, promiscuous gene expression and dominant tolerance are closely linked both experimentally and conceptually (4, 49, 50). Dominant tolerance, as opposed to recessive tolerance, offers a mechanism to compensate for incomplete representation of self in the thymus. Thus, in case of β cells of the pancreas, T regulator cells specific for GAD67 could also contain the autosensitization of naïve T effector cells specific for GAD65, when both antigens are copresented by the same DC, a mechanism termed bystander or linked suppression (51). Since thymic expression levels of self-antigens do, however, correlate with susceptibility to autoimmunity, the protection afforded by bystander suppression may be more easily breached than cognate suppression.

The structural and functional diversity of promiscuously expressed genes and their expression uncoupled from cell lineage specificity and temporal regulation poses a puzzle as to their molecular regulation. In this respect, the finding that the transcription (co) factor Aire regulates a host of promiscuously expressed genes provided an important step toward the molecular definition of this phenomenon (8). Moreover, the fact that Aire also regulates a truncated promotor (the rat insulin promotor) in a transgenic context across species barriers implies sequence-specific genetic control, at least as far as the Aire-dependent pool is concerned (52). DNA-encoded sequences, however, are only one aspect of gene regulation; genetic regulation itself is controlled by epigenetic mechanisms targeting genes, which share common chromosomal location. The highly significant clustering of overexpressed genes supports a role for epigenetic regulation in this case. Clustering was significant for triplets and quadruplets by statistical analysis. These findings are reminiscent of recent reports showing clustering of nonhomologous genes coexpressed in a particular tissue in C. elegans and D. melanogaster (21, 30). This clustering has been interpreted as the result of juxtaposition during evolution of genes involved in formation of a particular tissue, thus facilitating their coregulation. In contrast, promiscuously expressed genes obviously do not share such a function; hence, we interpret the clustering as the result of colocalization in chromatin domains which become accessible in mTECs irrespective of tissue-specific differentiation patterns akin to recently described “gene expression neighborhoods” (53). For this reason, we did not exclude homologous genes like those of the keratin type-I and -II group in our analysis, since individual members of these genes are expressed differentially in a tissue-specific manner and their coexpression in purified mTECs violates this regulation. However, clustering is not confined to homologous genes (Fig. 4 d). It will be important to define the actual size of such expression domains and their boundaries and to determine whether expression is contiguous within such domains as suggested by preliminary data (Derbinski, J., personal communication). We emphasize
that expression of nonhomologous genes of different ontol-
ogy in such domains points to epigenetic regulation but
does not prove it.

The species conservation of promiscuous gene expression
between mouse and man as documented in this study
strengthens the validity of analogies drawn between human
autoimmune diseases and corresponding experimental ani-
mal models. The extent of promiscuous gene expression by
human mTECs and the composition of this gene pool
potentially extends the scope of central tolerance to most pe-
ripheral tissues including self-constituents, which are only
encountered during adulthood, e.g., in the lactating mam-
mmary gland or placenta. Promiscuous gene expression is not
only sufficient for self-tolerance but likely has been essential
for survival of the species (54), given the high prevalence of
infertility associated with gonad-specific autoimmunity in
APS-1 patients (44) and in AIRE−/− mice (7, 8). The ident-
ification of mTECs as the responsible cell type and its pu-
rification will permit expression studies to be conducted in
a standardized fashion, which has not been possible to date.
This will be a prerequisite to establish a database on promis-
cuous gene expression in humans, thereby documenting in-
terindividual differences (11, 14), allelic polymorphisms (40,
41), and splice variants (55). The identification of the mo-

ercular mechanisms and genes responsible for the regulation
of promiscuous gene expression and its deviations will
helpfully aid our understanding of the complex genetic
control of human autoimmune disease.

We would like to express our gratitude to the Department of Car-
diac Surgery, University of Heidelberg (Director: Prof. S. Hagl)
for making human tissue available according to the guidelines of
the local Ethics Committee. We are indebted to Gerd Moldenhauer for
generously supplying antibodies (Deutsches Krebsforschungszent-
rum, Heidelberg, Germany), Marc Kenzelmann and Ralf Klären
for invaluable advice with RNA amplification, Klaus Hexel and
Gordon Barkowsky for relentless support with cell sorting and
Francis Brasseur (Ludwig Institute, Brussels, Belgium) and H.
Burkhardt (University of Erlangen, Erlangen, Germany) for pro-
viding primer sequences. We thank J. Derbinski and W. Schmid for
critical reading of the manuscript.

Submitted: 29 September 2003
Accepted: 21 November 2003

References

thymic stromal cells: a subtle division of labor. Curr. Opin.
within: keeping self-reactive T cells out in the periphery.
Promiscuous gene expression in medullary thymic epithelial
miscuous gene expression and central T-cell tolerance: more
5. Burnet, F.M. 1959. The Clonal Selection Theory of Ac-
from loss of function to autoimmunity. Genes Immun.
4:12–21.
7. Ramsey, C., 0. Winqvist, L. Puhakka, M. Halonen, A.
Moro, O. Kampe, P. Eskeln, M. Pelto-Huikkko, and L. Pel-
tonen. 2002. Aire deficient mice develop multiple features of
APECED phenotype and show altered immune response.
8. Anderson, M.S., E.S. Venanzi, L. Klein, Z. Chen, S.P.
Berzins, S.J. Turley, H. von Boehmer, R. Bronson, A. Dier-
ich, C. Benoist, and D. Mathis. 2002. Projection of an im-
munological self-shadow within the thymus by the aire pro-
9. Sakaguchi, S. 2000. Animal models of autoimmunity and
12:684–690.
10. Pribyl, T.M., C. Campagnoni, K. Kampf, V.W. Handley,
and A.T. Campagnoni. 1996. The major myelin protein
45:812–819.
11. Sospedra, M., X. Ferrer-Francesch, O. Domiguez, M. Juan,
M. Foz-Sala, and R. Pujol-Borrell. 1998. Transcription of a
broad range of self-antigens in human thymus suggests a role
for central mechanisms in tolerance toward peripheral anti-
Conti–Fine. 2001. Human thymuses express incomplete sets of
muscle acetylcholine receptor subunit transcripts that sel-
M. Redondo, J. Diez, G.S. Eisenbarth, D.D. Patel, and C.
Ricordi. 2001. Self-antigen-presenting cells expressing diabe-
tes-associated autoantigens exist in both thymus and periph-
Multiple sclerosis candidate autoantigens except myelin oli-
godendrocyte glycoprotein are transcribed in the human thym-
15. Pugliese, A., and J. Diez. 2002. Lymphoid organs contain di-
16. Ernst, T., M. Hergenhahn, M. Kenzelmann, C.D. Cohen,
M. Bonrouhi, A. Weninger, R. Klären, E.F. Grone, M.
expression are equally discriminatory markers for prostate
carcinoma: a gene expression analysis on total and microdis-
17. Kenzelmann, M., R. Klären, M. Hergenhahn, M. Bonrouhi,

nity amplification of nanogram total RNA amounts for gene
18. Vandenabeele, S., H. Hochrein, N. Mavaddat, K. Winkel,
and K. Shortman. 2001. Human thymus contains 2 distinct
1988. Monoclonal antibodies reactive with subsets of mouse
and human thymic epithelial cells. J. Histochem. Cytochem.
36: 1511–1517.
Quantitative analysis of mRNA amplification by in vitro


