

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

Jörn Gotter,<sup>1</sup> Benedikt Brors,<sup>2</sup> Manfred Hergenbahn,<sup>3</sup> and Bruno Kyewski<sup>1</sup>

<sup>1</sup>Tumor Immunology Program, <sup>2</sup>Intelligent Bioinformatics Systems, and <sup>3</sup>Genetic 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 overexpress 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

The online version of this article contains supplemental material.

Address correspondence to Bruno Kyewski, Div. of Cellular Immunology, Tumor Immunology Program, German Cancer Research Center, INF 280, D-69120 Heidelberg, Germany. Phone: 49-6221-423-734; Fax: 49-6221-423-702; email: b.kyewski@dkfz.de

*Abbreviations used in this paper:* AIRE, autoimmune regulator; APS, autoimmune polyglandular syndrome; CDR2, cortical dendritic reticulum antigen 2; cTEC, cortical thymic epithelial cell; DM T1, Diabetes mellitus type 1; EpCAM, epithelial cell adhesion molecule; MBP, myelin basic protein; MOG, myelin-oligodendrocyte-associated glycoprotein; mTEC, medullary thymic epithelial cell; PLP, proteolipid protein; TEC, thymic epithelial cell.

(8). This important finding offers a valuable opportunity to link a host of experimental data on tolerance mechanisms and their failure to the human situation, provided that promiscuous gene expression, its biological role, and molecular regulation are strictly conserved among mouse and humans (9). Several studies reported on the expression of selected autoantigens in human thymus, prominent examples being insulin, proteolipid protein (PLP), and myelin basic protein (MBP); yet, the identity of the cells expressing these tissue antigens remained ill defined and controversial (10–15). The identification and isolation of the responsible cell type in mice, however, has been a prerequisite to unambiguously confirm expression of certain tissue antigens of low abundance. Although almost all known or suspected autoantigens of common autoimmune diseases could be detected in murine mTECs by RT-PCR as far as analyzed, the full extent of expression of tissue-specific genes in the thymus remains to be defined. The array of promiscuously expressed genes being controlled by the transcriptional (co) factor Aire in mice implied that this gene pool extends beyond the collection of autoantigens known to be targeted in tissue-specific autoimmunity (8). A comprehensive expression analysis of promiscuously expressed genes may not only identify new candidate autoantigens but also reveal structural or functional commonalities among these genes and thus offer clues as to their molecular regulation.

Based on the isolation of pure thymic stromal cells of human thymus and an improved protocol for amplification of small RNA amounts (16, 17), we show that promiscuous gene expression is highly conserved between mouse and human. The array of promiscuously expressed genes in human mTECs is diverse with regard to gene ontology, tissue specificity, and chromosomal allocation. They encompass clinically relevant autoantigens and tumor antigens including cancer-germline antigens. The clustered location within the genome points to epigenetic mechanisms to control at least in part the expression of this gene pool.

## 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 ( $2 \times 10$  min) to remove the majority of thymocytes. After stepwise digestion with collagenase-dispase ( $3 \times 20$  min,  $37^\circ\text{C}$ ) and trypsin-EDTA ( $3 \times 15$  min,  $37^\circ\text{C}$ ), 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,  $37^\circ\text{C}$ ). After enrichment by a one-step Percoll gradient ( $\rho$  1.07 g/cm<sup>3</sup>, 1.700 g, 10 min,  $4^\circ\text{C}$ ), 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 (Dyna). 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<sup>pos</sup> epithelial cells were excluded by appropriate gating. TECs were isolated from the third trypsin fraction. Cells were enriched in Percoll ( $\rho$  1.07 g/cm<sup>3</sup>, 1,700 g, 10 min,  $4^\circ\text{C}$ ), and myelogenic 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 myelogenic (CD45<sup>+</sup>) cells (stained with mAb anti-CD45-CyChrome; Immunotech). Thymocytes were sorted as CD3<sup>high</sup> cells (mAb anti-CD3-FITC; BD Biosciences) from the medium fraction. Dead cells were always excluded with propidium iodide (1  $\mu\text{g}/\text{ml}$ ). The average cell yield per thymus aliquot (3–5 cm<sup>3</sup>, equivalent to 15 mouse thymi) was  $6 \times 10^5$  mTECs,  $10^6$  cTECs, and  $6 \times 10^5$  DCs.

**Immunohistochemistry.** Human thymic tissue pieces were embedded in Tissue-Tek (Sakura) and snap frozen in liquid nitrogen. Cryosections (4  $\mu\text{m}$ ) 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 H<sub>2</sub>O<sub>2</sub> 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% H<sub>2</sub>O<sub>2</sub> 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  $\mu\text{l}$  H<sub>2</sub>O. 48  $\mu\text{l}$  of this RNA solution or 1  $\mu\text{g}$  total RNA from commercially available control tissues (Stratagene) were treated with 6 U DNase I (Invitrogen) and reverse transcribed into cDNA with Oligo(dT)<sub>20</sub> 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  $\mu\text{l}$  with 1 U REDTaq DNA Polymerase (Sigma-Aldrich). Final concentrations of the PCR mix were: 250 nM for each primer, 200  $\mu\text{M}$  dNTP (MBI Fermentas), 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.1 mM MgCl<sub>2</sub>, and 0.01% gelatin. The amplification was performed in a Thermalcycler (PTC-100; MJ Research) under standard conditions: a single denaturation step at  $94^\circ\text{C}$  for 3 min followed by either 35 (GAD65, IA-2, TPO, H/K-ATPase  $\alpha$  and  $\beta$  chain Golli-MBP, myelin-oligodendrocyte-associated glycoprotein (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  $94^\circ\text{C}$  for 1 min,  $54$ – $62^\circ\text{C}$  for 1 min and  $72^\circ\text{C}$  for 2 min, followed by a final extension step of  $72^\circ\text{C}$  for 10 min. The following oligonucleotide pairs were used (sense and antisense, respectively): AIRE, 5'-TGCCAAGGATGACACTGC-3' and 5'-TTCCAGAGTAGAGAAGTGGGGC-3'; collagen II, 5'-CTGGCTCCCAACTGCCAACGTC-3' and 5'-TCCTTTGGGTTTGAACGGATTGT-3'; DC-LAMP, 5'-GCACGATGGCAGTCAAATGA-3' and 5'-GAAGTATCTC-

CGAGGTGAAA-3'; GAD65, 5'-TGCTCCAAAGTGGATGT-CAACTA-3' and 5'-ATGTTAGTATTTGCTGTTGATGTCA-3'; GAD67, 5'-ATGGCGTCTTCGACCCCATCTT-3' and 5'-AGCTGGTTGAAAAATCGAGGA-3'; GAPDH, 5'-AACAGCCTCAAGATCATCAGC-3' and 5'-CTGTTGCTGTAGCCAAATTCG-3'; Golli-MBP, 5'-AAACCACGCAG-GCAAACG-3' and 5'-AGGGTCTCTTCTGTGACG-3'; H/K-ATPase  $\alpha$  chain, 5'-GTCAACGAGCCCCTGGCTGC-3' and 5'-GTAGAGTTCCTGGTCCCACC-3'; H/K-ATPase  $\beta$  chain, 5'-GCAGGAGAAGAAGACGTGT-3' and 5'-GGATGTTGAGGAGCTTC-3'; IA-2, 5'-AGACAGGGCTCCAAATCTTGC-3' and 5'-GGCATGGTCATAGGGCAGAA-3'; insulin, 5'-CAACACCTGTGCGGCTCA-3' and 5'-TATTCATCTCTCTCGGTGTCAG-3'; IRBP, 5'-GCTGATAACTATGCCTCTGCCG-3' and 5'-CTTCTTCCAGATGTGCTCCACC-3'; MAGE-A1, 5'-CGGCCGAAGAACCTGACCCAG-3' and 5'-GCTGGAACCCCTCACTGGTTGCC-3'; MAGE-A3, 5'-TGGAGGACCAGAGGCCCCC-3' and 5'-GGACGATTATCAGGAGGCCTGC-3'; MAGE-A4, 5'-GAGCAGACAGGCCAACCG-3' and 5'-AAGACTCTGCGTCAGGC-3'; MART-1, 5'-ACTGCTCATCGGCTGTTG-3' and 5'-TTCAGCATGTCTCAGGTG-3'; MOG, 5'-CTCCTCTCCAAGTGTCTTC-3' and 5'-GTAGCTCTTCAAGGAATTGC-3'; NY-ESO-1, 5'-AGTTCTACCTCGCCATGCCT-3' and 5'-TCCTCCTCCAGCACAACAA-3'; PLP/DM20, 5'-ACTACAAGACCACATCTGC-3' and 5'-CCATACATTCTGGCATCAGC-3'; RetSag, 5'-GCCAATGTGGTTCTCTACTCG-3' and 5'-CGAGCAAACCTCTCAAAAATA-3'; thyroglobulin, 5'-CCCTGGCCTGACTTTGTACC-3'; and 5'-TCACTTGCTGTAGGTCTTAGAGCC-3'; TPO, 5'-TACAAGCATCCTGACAACATCG-3' and 5'-ATTCTCCACGCTCTCTGG-3'; tyrosinase, 5'-TTGGCAGATTGTCTGTAGCC-3' and 5'-AGGCATTGTGCATGCTGCTT-3'; and WHN (FOXP1), 5'-TTCCTTACTTCAAGACAGCAC-3' and 5'-GGTTCTTGCAGGAATGG-3'. All primers were synthesized in the oligonucleotide synthesis facility of the German Cancer Research Center and were, wherever possible, designed to span at least one intron. Reaction products were separated on a 1.2% agarose gel in TAE (40 mM Tris-acetate, 1 mM EDTA) containing ethidium bromide and revealed with the Lumi-Imager F1 Workstation (Roche). LumiAnalyst 3.0 software (Roche) was used to quantify PCR products for normalization to GAPDH expression before testing promiscuous gene expression.

**Microarray Analysis.** Total RNA from thymic samples 017 and 018 (4.5  $\mu$ l, corresponding to  $\sim 6 \times 10^4$  mTECs,  $10^4$  cortical TECs (cTECs), and  $6 \times 10^4$  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 sequentially 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 cells, 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  $\geq 1$  compared with the reference subset.

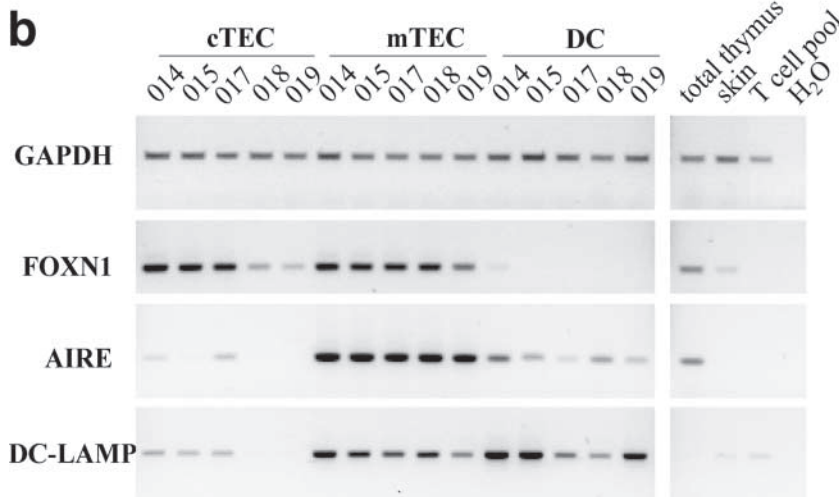
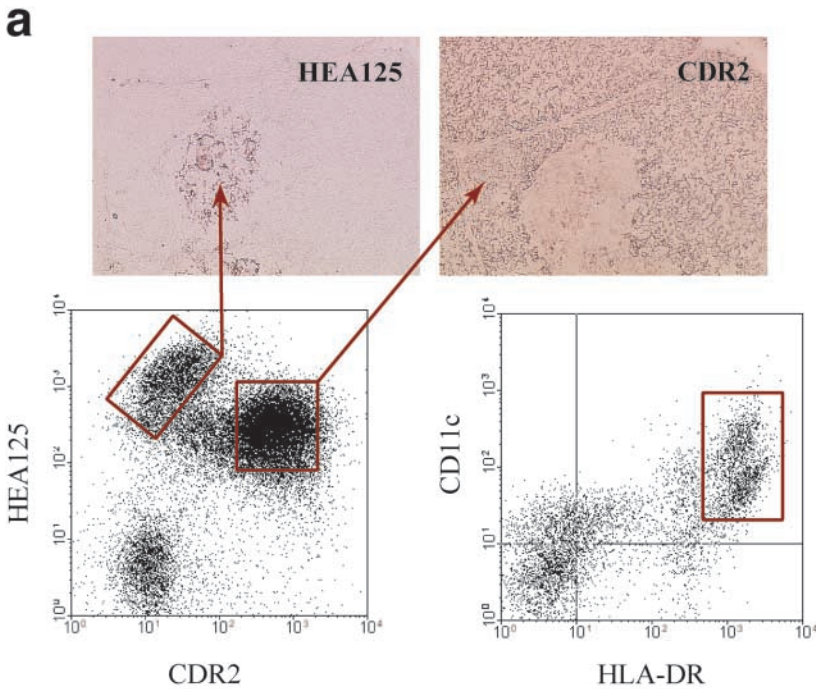
**Statistical Analysis.** To test for statistical significance of clustering, we used randomized lists of genes from the same chip for comparison according to the method described by Roy et al. (21). Briefly, we produced random lists of genes of the same length as the number of genes overexpressed in 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. This expression pattern thus yields cTECs, i.e., CD45<sup>-</sup>, CDR2<sup>hi</sup>, EpCAM<sup>int</sup>, and mTECs, i.e., CD45<sup>-</sup>, CDR2<sup>-</sup>, EpCAM<sup>hi</sup>. Thymic DCs were isolated as HLA-DR<sup>hi</sup>, CD11c<sup>+</sup>, EpCAM<sup>-</sup> cells according to a previously published protocol (18). 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., *FOXP1* 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.

**Promiscuous Gene Expression in Medullary Epithelial Cells Is Conserved between Mice and Man.** We first assessed the expression of organ-specific autoantigens of putative clinical relevance in these cell types isolated independently from five human thymi. Examples of self-antigens of the endocrine pancreas, thyroid, stomach, brain, cartilage, and eye



**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<sup>hi</sup>, EpCAM<sup>int</sup>, and mTECs as CDR2<sup>lo</sup>, EpCAM<sup>hi</sup>. The corresponding histological staining patterns of both Abs are shown. Thymic DCs are defined as CD11c<sup>int/hi</sup>, HLA-DR<sup>hi</sup>, EpCAM<sup>neg</sup> 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).

were analyzed by RT-PCR (Table I). All antigens were reproducibly expressed in mTECs with the exception of GAD65 and the H/K-ATPase  $\beta$  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

**Table 1.** Cell Type-specific Expression Pattern of Selected Self-antigens Purportedly Associated with Common Autoimmune Diseases

Organ	Autoantigen	Antigen-presenting cell		
		cTECs	mTECs	DCs
Pancreas	Insulin <sup>a</sup>	-----	+++++	-----
	GAD67 <sup>a</sup>	-----	+++++	-----
	GAD65 <sup>a</sup>	-----	-----	-----
	IA-2 <sup>a</sup>	-----	-++++	-----
Thyroid	Thyroglobulin	-----	+++++	-----
	TPO <sup>a</sup>	-----	+++++	-----
Stomach	H/K-ATPase $\alpha$	-----	+++++	+++++
	H/K-ATPase $\beta$	-----	-----	-----
Brain	Golli-MBP	+++++	+++++	+++++
	PLP/DM20	++++-	-++++	-----
	MOG	-----	+++++	-----
Eye	RetSAG <sup>a</sup>	--+-	+++++	+-----
	IRBP <sup>a</sup>	-----	+++++	-----
Cartilage	Collagen II	+++--	++ $\emptyset$ ++	- $\emptyset$ ---

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

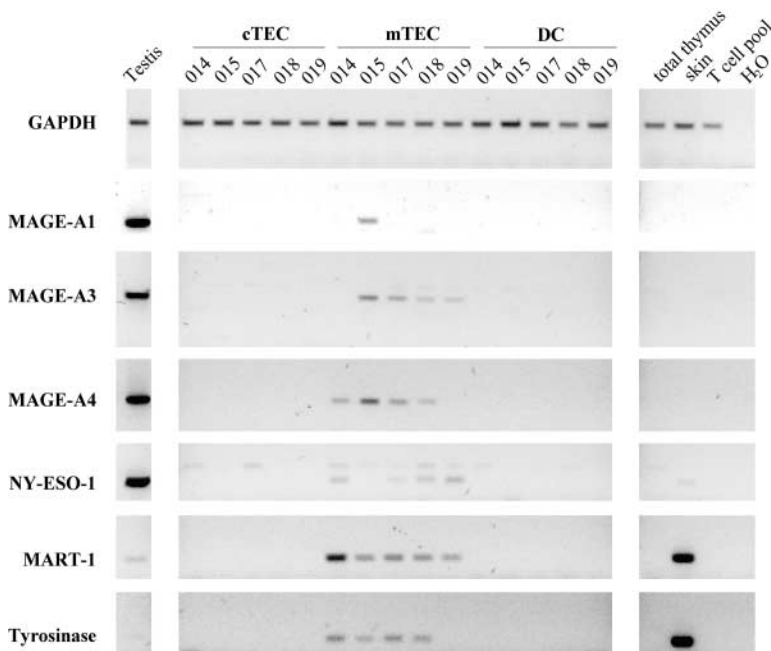
<sup>a</sup>The TEC-specific expression pattern of antigens has been confirmed in six additional thymi. In addition, mRNA expression of albumin,  $\alpha$ -fetoprotein, C-reactive protein, and erythropoietin in mTECs has been found in six thymi (not depicted).

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.

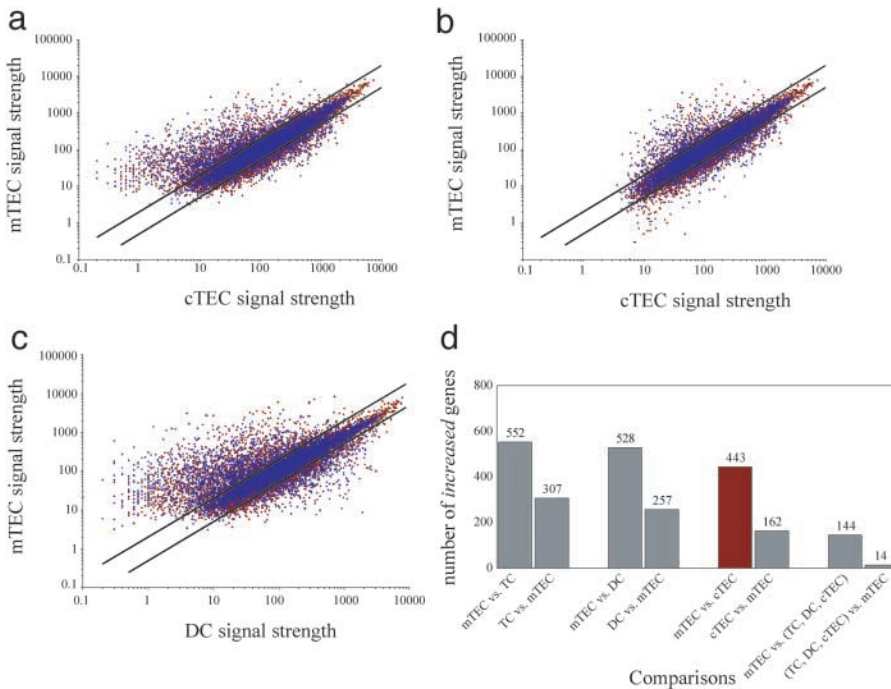
The high score with which “arbitrarily” selected genes were found to be expressed in mTECs and the diversity of these promiscuously expressed genes in mice and man implies that this gene pool is complex and comprehensive. To more precisely delineate this gene pool, we defined the transcriptome of these 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<sup>®</sup> 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  $\geq 1$  (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



**Figure 2.** Thymic expression of tumor-associated antigens. Expression of selected tumor-associated antigens in thymic stromal cells of five thymi was assessed by RT-PCR. Expression in each case was confined to mTECs. Note inter-individual variations and the difference in signal strength compared with control tissues. The amount of input cDNA was normalized according to signals obtained for GAPDH in a titration experiment (not depicted).



**Figure 3.** Comparative global gene expression in thymic stromal cells. The relative expression levels as detected by gene arrays are depicted as scatter plots of two experiments (blue and red). (a) Signal strength of genes present or marginally present in mTECs in both comparisons and present, marginally present, or absent in cTECs. (b) Signal strength of genes present or marginally present in cTECs in both comparisons and present, marginally present, or absent in mTECs. (c) Signal strength of genes present or marginally present in mTECs in both comparisons and present, marginally present, or absent in DCs. Note the higher proportion of genes specific for mTECs (absent in cTECs) and their low expression levels. (d) Quantitative analysis of relative gene expression in mutual comparisons between cTECs, mTECs, DCs, and mature thymocytes. Numbers above bars denote the frequency of genes overexpressed in each case. The gene pool overexpressed in mTECs versus cTECs (red bar) has been analyzed in more detail. For details of calculations, see Materials and Methods. TC, thymocytes.

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.

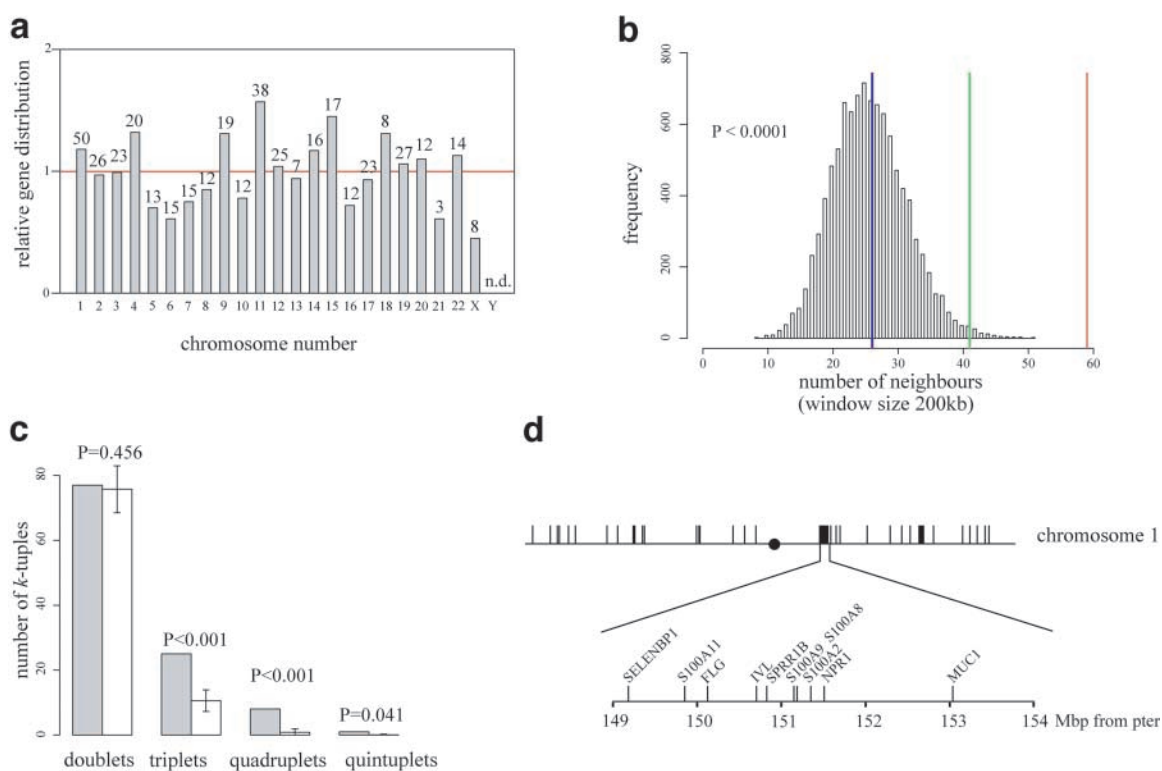
The genes overexpressed in mTECs fall into at least two categories: those which constitute the differentiation program of this epithelial cell lineage and those which are regarded as tissue specific, i.e., expressed in a promiscuous manner. We based the gene assignment on data reported by Su et al. (27). 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



**Figure 4.** Genomic distribution of mTECs overexpressed genes. (a) Chromosomal assignment of genes overexpressed in mTECs versus cTECs. Pooled data from a female and a male donor have been analyzed. The relative distribution of all transcripts present on the chip with known chromosomal assignment was normalized to 1 for each chromosome. The absolute number of genes overexpressed in mTECs per chromosome is indicated above the bars. The relative distribution of mTECs genes shows no particular chromosomal preferences. n.d., not determined. (b) Distribution of the number of neighbors in 10,000 random gene lists that are located on the same chromosome within a distance of 200 kb (calculated from transcription start). The median and the 95th percentile of the distribution are marked by blue and green lines, respectively. The actually observed number in the list of 415 genes overexpressed in mTECs is indicated by a red line. (c) Number of clusters of size 2, 3, 4, and 5 within a window of 10 consecutive genes. The gray bars refer to the number in the list of genes overexpressed in mTECs and the white bars to the number in 1,000 random gene lists. The SD among these 1,000 lists is indicated by the error bar. (d) Clustered expression of 10 promiscuously expressed genes in mTECs within 5 Mbp on chromosome 1. SELENBP1, selenium binding protein 1; S100 A, calcium binding protein A; FLAG, filaggrin; IVL, involucrin; SPRR1B, small proline rich protein 1B; NPR1, atrial natriuretic peptide receptor A; MUC1, mucin 1.

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. 4 b). 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. 4 c). An example of 10 genes clustered within 5 Mbp encompassing genes of different ontology, i.e., three members of the S100 gene family, MUC1 and SELENBP1, predominantly expressed in the liver (27) is shown in Fig. 4 d. We emphasize that we did not exclude homologous genes in this analysis, which may have arisen from gene duplication, since the expression pattern of individual members of such gene families may also reflect promiscuous gene expression. Thus, different type II keratin genes, including hair keratins were coexpressed in purified mTECs despite their differential regulation in epithelial cells of other tissues (31, 32).

## 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

Symbol	Name/Function	Symbol	Name/Function
Mammary gland			
<i>CSN1</i>	Casein	<i>SCGB2A2</i>	Mammaglobin
<i>CSN2</i>	Casein		
Adipose tissue			
<i>APM2</i>	Unknown		
Liver			
<i>AADAC</i>	Esterase	<i>FGG</i>	Fibrinogen
<i>APOA1</i>	Apolipoprotein	<i>PCK2</i>	PEP-carboxykinase
<i>APOC3</i>	Apolipoprotein	<i>SERPINA5</i>	Antitrypsin
<i>DPYS</i>	Dihydropyrimidinase	<i>UGT2B10</i>	UDP-glucuronosyl-transferase
<i>FGB</i>	Fibrinogen		
Muscle			
<i>ACTC</i>	Alpha actin	<i>MYH7</i>	Myosin
<i>AK1</i>	Adenylate kinase	<i>MYL1</i>	Myosin
<i>CASQ2</i>	Calcium-binding protein	<i>MYL2</i>	Myosin
<i>CHRNA1</i>	Acetylcholine receptor	<i>TNNC1</i>	Troponin
<i>CSRP3</i>	Heart-LIM protein	<i>TNNI1</i>	Troponin
<i>ENO3</i>	Enolase	<i>TNNT2</i>	Troponin
<i>MB</i>	Myoglobin	<i>TTN</i>	Myofibrin
Nervous system			
<i>ADAM23</i>	Metalloproteinase	<i>HAPIP</i>	Vesicle transport
<i>CADPS</i>	Calcium-binding protein	<i>KCNK1</i>	Potassium channel
<i>CALB1</i>	Calbindin	<i>NEF3</i>	Neurofilament
<i>CCK</i>	Prohormone	<i>PADI2</i>	Deiminase
<i>CHGB</i>	Secretogranin	<i>PCDH8</i>	Protocadherin
<i>CNR1</i>	Receptor	<i>PRSS12</i>	Neurotrypsin
<i>FGF1</i>	Growth factor	<i>SST</i>	Somatostatin
<i>GABBR1</i>	Receptor	<i>SV2B</i>	Synaptic vesicle protein
<i>GABRA5</i>	GABA receptor	<i>SYNGR1</i>	Synaptogyrin
<i>GAL</i>	Galanin	<i>SYT1</i>	Synaptotagmin
<i>GPR37</i>	G protein-linked receptor	<i>TAC1</i>	Neuropeptide
Kidney			
<i>ABP1</i>	Membrane glycoprotein	<i>TM7SF1</i>	Transmembrane protein
<i>SPP1</i>	Osteopontin	<i>UGT1A4</i>	UDP-glucuronosyl-transferase
Pancreas			
<i>AMY2B</i>	Amylase	<i>PRSS2</i>	Trypsin II
<i>ELA3A</i>	Elastase	<i>PRSS4</i>	Trypsin III
<i>GAD1</i>	Glutamate-decarboxylase	<i>RNASE1</i>	RNase A1
<i>GCG</i>	Glucagon	<i>SCGN</i>	Secretagogen
<i>LEFTB</i>	L/R determination factor B		
Placenta			
<i>CGA</i>	Glycoprotein hormone	<i>PRG2</i>	Proteoglycan
<i>CLDN4</i>	Tight junction protein	<i>SDC1</i>	Membrane protein
<i>CSH2</i>	Placental lactogen hormone	<i>SEMA3B</i>	Semaphorin
Prostate			
<i>ACPP</i>	Phosphatase		

(continues)



**Table II.** (continued)

Symbol	Name/Function	Symbol	Name/Function
Respiratory tract			
<i>MUC4</i>	Mucin	<i>TSPAN-1</i>	Transmembrane protein
<i>SCGB1A1</i>	Uteroglobin		
Thyroid			
<i>CLDN3</i>	Tight junction protein		
Salivary gland			
<i>PRB2</i>	Secreted protein	<i>SLPI</i>	Protease inhibitor
Testis			
<i>GSTM3</i>	Glutathione transferase	<i>SPAG6</i>	Sperm antigen

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.

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*<sup>-/-</sup> 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 intrathymic 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  $\alpha$  chain is clearly expressed in mTECs and DCs, the  $\beta$  chain is at best marginally expressed in mTECs. In the mouse model, the  $\beta$  chain but not the  $\alpha$  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 intrathymic 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 intrathymic 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 intrathymic 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 *U95A1v2* array set represents only  $\sim 1/3$  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

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 partially 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 tis-

sue-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  $\beta$  cells of the pancreas, T regulator cells specific for GAD67 could also contain the auto-sensitization of naive 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 promoter (the rat insulin promoter) 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 ontology 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 animal 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 peripheral tissues including self-constituents, which are only encountered during adulthood, e.g., in the lactating mammary 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*<sup>-/-</sup> mice (7, 8). The identification of mTECs as the responsible cell type and its purification 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 promiscuous gene expression in humans, thereby documenting interindividual differences (11, 14), allelic polymorphisms (40, 41), and splice variants (55). The identification of the molecular mechanisms and genes responsible for the regulation of promiscuous gene expression and its deviations will hopefully aid our understanding of the complex genetic control of human autoimmune disease.

We would like to express our gratitude to the Department of Cardiac 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 Krebsforschungszentrum, 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 Bresseur (Ludwig Institute, Brussels, Belgium) and H. Burkhardt (University of Erlangen, Erlangen, Germany) for providing primer sequences. We thank J. Derbinski and W. Schmid for critical reading of the manuscript.

Submitted: 29 September 2003

Accepted: 21 November 2003

## References

1. Klein, L., and B. Kyewski. 2000. Self-antigen presentation by thymic stromal cells: a subtle division of labor. *Curr. Opin. Immunol.* 12:179–186.
2. Walker, L.S., and A.K. Abul-Abbas. 2002. The enemy within: keeping self-reactive T cells at bay in the periphery. *Nat. Rev. Immunol.* 2:11–19.
3. Derbinski, J., A. Schulte, B. Kyewski, and L. Klein. 2001. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nat. Immunol.* 2:1032–1039.
4. Kyewski, B., J. Derbinski, J. Gotter, and L. Klein. 2002. Promiscuous gene expression and central T-cell tolerance: more than meets the eye. *Trends Immunol.* 23:364–371.
5. Burnet, F.M. 1959. The Clonal Selection Theory of Acquired Immunity. Cambridge University Press, London.
6. Pitkänen, J., and P. Peterson. 2003. Autoimmune regulator: from loss of function to autoimmunity. *Genes Immun.* 4:12–21.
7. Ramsey, C., O. Winqvist, L. Puhakka, M. Halonen, A. Moro, O. Kampe, P. Eskelin, M. Peltö-Huikko, and L. Peltonen. 2002. Aire deficient mice develop multiple features of APECED phenotype and show altered immune response. *Hum. Mol. Genet.* 11:397–409.
8. Anderson, M.S., E.S. Venanzi, L. Klein, Z. Chen, S.P. Berzins, S.J. Turley, H. von Boehmer, R. Bronson, A. Dierich, C. Benoist, and D. Mathis. 2002. Projection of an immunological self-shadow within the thymus by the aire protein. *Science.* 298:1395–1401.
9. Sakaguchi, S. 2000. Animal models of autoimmunity and their relevance to human diseases. *Curr. Opin. Immunol.* 12:684–690.
10. Pribyl, T.M., C. Campagnoni, K. Kampf, V.W. Handley, and A.T. Campagnoni. 1996. The major myelin protein genes are expressed in the human thymus. *J. Neurosci. Res.* 45:812–819.
11. Sospedra, M., X. Ferrer-Francesch, O. Domínguez, 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 antigens. *J. Immunol.* 161:5918–5929.
12. Navaneetham, D., A.S. Penn, J.F. Howard, Jr., and B.M. Conti-Fine. 2001. Human thymuses express incomplete sets of muscle acetylcholine receptor subunit transcripts that seldom include the  $\delta$  subunit. *Muscle Nerve.* 24:203–210.
13. Pugliese, A., D. Brown, D. Garza, D. Murchison, M. Zeller, M. Redondo, J. Diez, G.S. Eisenbarth, D.D. Patel, and C. Ricordi. 2001. Self-antigen-presenting cells expressing diabetes-associated autoantigens exist in both thymus and peripheral lymphoid organs. *J. Clin. Invest.* 107:555–564.
14. Bruno, R., L. Sabater, M. Sospedra, X. Ferrer-Francesch, D. Escudero, E. Martinez-Caceres, and R. Pujol-Borrell. 2002. Multiple sclerosis candidate autoantigens except myelin oligodendrocyte glycoprotein are transcribed in the human thymus. *Eur. J. Immunol.* 32:2737–2747.
15. Pugliese, A., and J. Diez. 2002. Lymphoid organs contain diverse cells expressing self-molecules. *Nat. Immunol.* 3:335–336.
16. Ernst, T., M. Hergenbahn, M. Kenzelmann, C.D. Cohen, M. Bonrouhi, A. Weninger, R. Klären, E.F. Grone, M. Wiesel, C. Gudemann, et al. 2002. Decrease and gain of gene expression are equally discriminatory markers for prostate carcinoma: a gene expression analysis on total and microdissected prostate tissue. *Am. J. Pathol.* 160:2169–2180.
17. Kenzelmann, M., R. Klären, M. Hergenbahn, M. Bonrouhi, H.-J. Gröne, W. Schmid, and G. Schütz. 2004. High accuracy amplification of nanogram total RNA amounts for gene profiling. *Genomics.* In press.
18. Vandenaabeele, S., H. Hochrein, N. Mavaddat, K. Winkel, and K. Shortman. 2001. Human thymus contains 2 distinct dendritic cell populations. *Blood.* 97:1733–1741.
19. Rouse, R.V., L.M. Bolin, J.R. Bender, and B.A. Kyewski. 1988. Monoclonal antibodies reactive with subsets of mouse and human thymic epithelial cells. *J. Histochem. Cytochem.* 36:1511–1517.
20. Baugh, L.R., A.A. Hill, E.L. Brown, and C.P. Hunter. 2001. Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res.* 29:E29.

21. Roy, P.J., J.M. Stuart, J. Lund, and S.K. Kim. 2002. Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature*. 418:975–979.
22. Efron, B., and R. Tibshirani. 1993. An Introduction to the Bootstrap. Chapman & Hall, New York/London.
23. Boon, T., and L.J. Old. 1997. Cancer tumor antigens. *Curr. Opin. Immunol.* 9:681–683.
24. Rosenberg, S.A. 1999. A new era for cancer immunotherapy based on the genes that encode cancer antigens. *Immunity*. 10: 281–287.
25. Chomez, P., O. De Backer, M. Bertrand, E. De Plaen, T. Boon, and S. Lucas. 2001. An overview of the MAGE gene family with the identification of all human members of the family. *Cancer Res.* 61:5544–5551.
26. Renkvist, N., C. Castelli, P.F. Robbins, and G. Parmiani. 2001. A listing of human tumor antigens recognized by T cells. *Cancer Immunol. Immunother.* 50:3–15.
27. Su, A.I., M.P. Cooke, K.A. Ching, Y. Hakak, J.R. Walker, T. Wiltshire, A.P. Orth, R.G. Vega, L.M. Sapinoso, A. Moqrich, et al. 2002. Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci. USA.* 99:4465–4470.
28. Wang, P.J., J.R. McCarrey, F. Yang, and D.C. Page. 2001. An abundance of X-linked genes expressed in spermatogonia. *Nat. Genet.* 27:422–426.
29. Lorient, A., T. Boon, and C. DeSmet. 2003. Five new human cancer-germline genes identified among 12 genes expressed in spermatogonia. *Int. J. Cancer.* 105:371–376.
30. Boutanaev, A.M., A.I. Kalmykova, Y.Y. Shevelyov, and D.I. Nurminsky. 2002. Large clusters of co-expressed genes in the *Drosophila* genome. *Nature*. 420:666–669.
31. Heid, H.W., I. Moll, and W.W. Franke. 1988. Patterns of expression of trychoctytic and epithelial cytokeratins in mammalian tissue. II. Concomitant and mutually exclusive synthesis of trychoctytic and epithelial cytokeratins in diverse human and bovine tissues (hair follicle, nail bed, and matrix, lingua papilla, thymic reticulum). *Differentiation*. 37:215–230.
32. Langbein, L., M. Rogers, H. Winter, S. Praetzel, and J. Schweizer. 2001. The catalog of human hair keratins. *J. Biol. Chem.* 276:35123–35132.
33. Tisch, R., and H. McDevitt. 1996. Insulin-dependent diabetes mellitus. *Cell*. 85:291–297.
34. Yoon, J.W., R.S. Sherwin, H. Kwon, and H.S. Jun. 2000. Has GAD a central role in type 1 diabetes? *J. Autoimmun.* 15: 273–278.
35. Baekkeskov, S., J. Knaani, J.C. Jaume, and S. Kash. 2000. Does GAD have a unique role in triggering IDDM? *J. Autoimmun.* 15:279–286.
36. Viglietta, V., S.C. Kent, T. Orban, and D.A. Hafler. 2002. GAD65-reactive T cells are activated in patients with autoimmune type 1a diabetes. *J. Clin. Invest.* 109:895–903.
37. Lernmark, Å. 2002. Controlling the controls: GAD65 autoreactive T cells in type 1 diabetes. *J. Clin. Invest.* 109:869–870.
38. Toh, B.H., J.W. Sentry, and F. Alderuccio. 2000. The causative H<sup>+</sup>/K<sup>+</sup> ATPase antigen in the pathogenesis of autoimmune gastritis. *Immunol. Today*. 21:348–354.
39. Klein, L., M. Klugmann, K.A. Nave, V.K. Tuohy, and B. Kyewski. 2000. Shaping of the autoreactive T-cell repertoire by a splice variant of self protein expressed in thymic epithelial cells. *Nat. Med.* 6:56–61.
40. Pugliese, A., M. Zeller, A. Fernandez, Jr., L.J. Zalcberg, R.J. Bartlett, C. Ricordi, M. Pietropaolo, G.S. Eisenbarth, S.T. Bennett, and D.D. Patel. 1997. The insulin gene is transcribed in the human thymus and transcription levels correlate with allelic variation at the *INS VNTR-IDDM2* susceptibility locus for type 1 diabetes. *Nat. Genet.* 15:293–297.
41. Vafiadis, P., S.T. Bennett, J.A. Todd, J. Nadeau, R. Grabs, C.G. Goodyer, S. Wickramasinghe, E. Colle, and C. Polychronakos. 1997. Insulin expression in human thymus is modulated by *INS VNTR* alleles at the *IDDM2* locus. *Nat. Genet.* 15:289–292.
42. Chentoufi, A.A., and C. Polychronakos. 2002. Insulin expression levels in the thymus modulate insulin-specific autoreactive T-cell tolerance: the mechanism by which the *IDDM2* locus may predispose to diabetes. *Diabetes*. 51:1383–1390.
43. Miyamoto, K., S. Miyake, M. Schachner, and T. Yamamura. 2003. Heterozygous null mutation of myelin P0 protein enhances susceptibility to autoimmune neuritis targeting P0 peptide. *Eur. J. Immunol.* 33:656–665.
44. Vogel, A., C.P. Strassburg, P. Obermayer-Straub, G. Brabant, and M.P. Manns. 2002. The genetic background of autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy and its autoimmune disease components. *J. Mol. Med.* 80:201–211.
45. Perniola, R., A. Falorni, M.G. Clemente, F. Forini, E. Accogli, and G. Lobreglio. 2000. Organ-specific and non-organ-specific autoantibodies in children and young adults with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED). *Eur. J. Endocrinol.* 143:497–503.
46. Engelhard, V.H., T.N.J. Bullock, T.A. Colella, S.L. Sheasley, and D.W. Mullins. 2002. Antigens derived from melanocyte differentiation proteins: self-tolerance, autoimmunity, and use for cancer immunotherapy. *Immunol. Rev.* 188:136–146.
47. Lotz, C., E.A. Ferreirai, I. Drexler, S.A. Mutallibi, C. Huber, G. Sutter, and M. Theobald. 2004. Partial tyrosinase-specific self tolerance by HLA-A\*0201-restricted cytotoxic T lymphocytes in mice and man. *Int. J. Cancer.* 108:571–579.
48. Coulie, P.G. 2002. Cancer immunotherapy with MAGE antigens. *Suppl. Tumori*. 1:S63–S65.
49. Seddon, B., and D. Mason. 2000. The third function of the thymus. *Immunol. Today*. 21:95–99.
50. Apostoulu, I., A. Saukhan, L. Klein, and H. von Boehmer. 2002. Origin of regulatory T cells with known specificity for antigen. *Nat. Immunol.* 3:756–763.
51. Waldmann, H., and S. Cobbold. 2001. Regulating the immune response to transplants. A role for CD4<sup>+</sup> regulatory cells? *Immunity*. 14:399–406.
52. Liston, A., S. Lesage, J. Wilson, L. Peltonen, and C.C. Goodnow. 2003. Aire regulates negative selection of organ-specific T cells. *Nat. Immunol.* 4:350–354.
53. Oliver, B., M. Parisi, and D. Clark. 2002. Gene expression neighborhoods. *J. Biol.* 1:4.
54. Boehmer H. v., Aifantis, I., Gounari, F., Azogui, O., Haughn, L., Apostolou, I., Jaeckel, E., Grassi, F., and L. Klein. 2003. Thymic selection revisited: how essential is it? *Immunol. Rev.* 191:62–78.
55. Diez, J., Y. Park, M. Zeller, D. Brown, D. Garza, C. Ricordi, J. Hutton, G.S. Eisenbarth, and A. Pugliese. 2001. 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.