Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels

Jens Derbinski,[1] Jana Gäbler,[1] Benedikt Brors,[2] Sascha Tierling,[6] Sunitha Jonnakuty,[3] Manfred Hergenhahn,[4] Leena Peltonen,[5] Jörn Walter,[6] and Bruno Kyewski[1]

[1]Division of Developmental Immunology, [2]Division of Theoretical Bioinformatics, [3]Division of Molecular Biophysics, and [4]Division of Genetic Alterations in Carcinogenesis, German Cancer Research Center, D-69120 Heidelberg, Germany
[5]Department of Medical Genetics, University of Helsinki, 00029 HUS, Helsinki, Finland
[6]Department of Genetics, Saarland University, D-66041 Saarbrücken, Germany

The role of central tolerance induction has recently been revised after the discovery of promiscuous expression of tissue-restricted self-antigens in the thymus. The extent of tissue representation afforded by this mechanism and its cellular and molecular regulation are barely defined. Here we show that medullary thymic epithelial cells (mTECs) are specialized to express a highly diverse set of genes representing essentially all tissues of the body. Most, but not all, of these genes are induced in functionally mature CD80[hi] mTECs. Although the autoimmune regulator (Aire) is responsible for inducing a large portion of this gene pool, numerous tissue-restricted genes are also up-regulated in mature mTECs in the absence of Aire. Promiscuously expressed genes tend to colocalize in clusters in the genome. Analysis of a particular gene locus revealed expression of clustered genes to be contiguous within such a cluster and to encompass both Aire-dependent and –independent genes. A role for epigenetic regulation is furthermore implied by the selective loss of imprinting of the *insulin–like growth factor 2* gene in mTECs. Our data document a remarkable cellular and molecular specialization of the thymic stroma in order to mimic the transcriptome of multiple peripheral tissues and, thus, maximize the scope of central self-tolerance.

CORRESPONDENCE
Bruno Kyewski:
b.kyewski@dkfz.de

Abbreviations used: Aire, autoimmune regulator; ANOVA, analysis of variance; *Cdkn1c*, *cyclin-dependent kinase inhibitor 1C*; cTEC, cortical TEC; *H19*, *H19 fetal liver* mRNA; *Igf2*, *insulin-like growth factor 2*; LOI, loss of imprinting; mTEC, medullary TEC; SNuPE, single nucleotide primer extension; TEC, thymic epithelial cell; *Tlbp*; *testis lipid binding protein*; TRA, tissue-restricted antigen.

Self-tolerance is inextricably linked to immunity; only when both features of the immune system are balanced is the body's integrity safeguarded. Our perception of how self-tolerance of the plethora of self-antigens is initially imposed and maintained throughout life has recently changed. Two areas of research, initially pursued independently but which now converge, contributed to this development. First, the observation was made that a diverse array of tissue-restricted antigens (TRAs) is expressed in the thymus and displayed there for repertoire selection (1). Second, unambiguous experimental evidence emerged that dominant tolerance mechanisms, foremost CD4 regulatory T cells, are essential rather than supplementary to recessive tolerance modes such as deletion (2). These new insights, apart from their conceptual implications, also open new therapeutic possibilities, not least for the treatment of autoimmune diseases.

The notion that aberrant expression of TRAs (termed promiscuous gene expression) is

an inherent property of the thymic stroma has been established by studies reporting the transcription of genes coding for proteins that serve cell type–specific functions; e.g., *proteolipid protein* (*PLP*) in oligodendrocytes or interphotoreceptor retinoid binding protein in retinal cells (3, 4). Although promiscuous gene expression by thymic epithelial cells (TECs) in various species is by now undisputed, many aspects of this phenomenon remain to be explored. Thus, the complete scope of promiscuously expressed genes is unknown. An initial expression analysis focused on selected self-antigens including many prominent autoantigens (5). Two recent reports in mice and humans, however, indicate that this gene pool is broadly inclusive rather than selective (6, 7). A comprehensive analysis of promiscuously expressed genes will delineate the ultimate extent of tissue representation in the thymus and possibly offer new clues as to the underlying molecular regulation.

Particular features of promiscuous gene expression (i.e., being uncoupled from tissue or developmental regulation) appear unique among

The online version of this article contains supplemental material.

somatic cells, suggesting a novel mode of molecular regulation. An important initial clue as to this regulation has been the finding that the autoimmune regulator (Aire) controls the expression of numerous genes in murine medullary TECs (mTECs) with a predilection for TRAs (6). This important finding offers a cogent explanation for the pathophysiology of the monogenic human autoimmune polyglandular syndrome 1 (APS-1), which is caused by mutations in the *AIRE* gene (8). APS-1 patients suffer to various degrees

from failures of multiple endocrine organs and show heightened autoantibody titers to organ-specific self-antigens (9, 10), most of which are promiscuously expressed in human mTECs (7). Based on these particular features of APS-1, the functional properties of Aire as a transcriptional coregulator and its conspicuous overexpression in mTECs, we had previously proposed a role for Aire in controlling promiscuous gene expression (11). With Aire influencing intrathymic expression of numerous TRAs in a dose-dependent manner
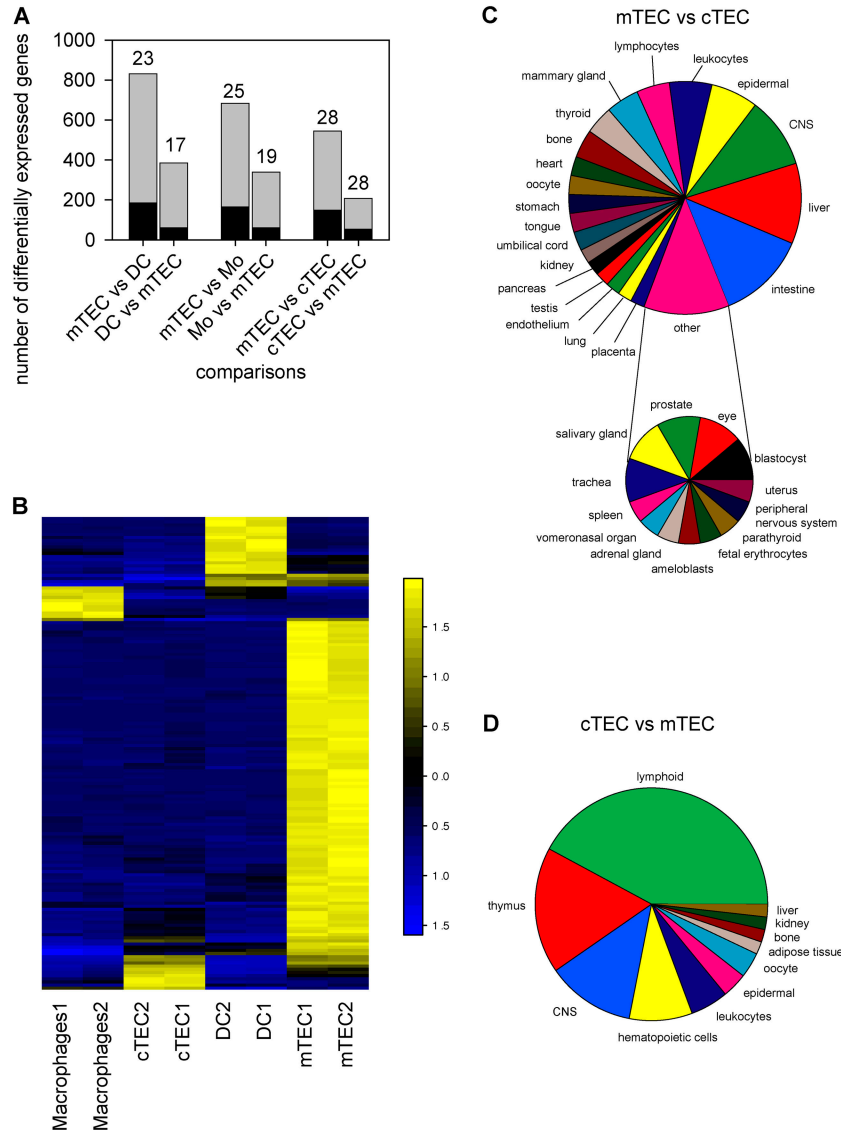

Figure 1. Global analysis of promiscuous gene expression in distinct thymic stromal cells. (A) Quantitative analysis of differentially expressed genes in mutual comparisons between mTECs and cTECs, DCs, and macrophages based on Affymetrix microarray analysis. The black bars indicate the fraction of TRAs contained in all genes overexpressed in the respective subsets and the relative percentages of TRAs are indicated above the bars. (B) ANOVA of all four stromal cell subsets showing the top 150 probe sets with adjusted p-values of <0.01. Note that mTECs display the highest proportion (106 out of 150) of differentially expressed genes. Yellow, up-regulated genes; blue, down-regulated genes; black, approximately the same gene expression as the mean for that gene across all samples. Numbers beside the color key represent mean ± SD of the respective gene. (C) Diverse tissue representation by genes overexpressed in mTECs vs. cTECs. (D) The reverse gene set shows limited tissue representation. Genes were assigned to tissues according to their predominant expression (Materials and methods). Mo, macrophages.

(12), it becomes apparent that the regulation of Aire itself will be an important determinant in self-tolerance control. The lymphotoxin β receptor has been recently identified as one upstream component of this molecular pathway (13).

Promiscuous gene expression, however, cannot solely be accounted for by the action of this molecule. The contribution of additional mechanisms is clearly documented by the fact that transcription levels of tissue-restricted antigens are dependent on Aire to various degrees, with some genes not being influenced by Aire at all (e.g., *CRP* or *GAD67*; reference 6). The complexity of the regulation of promiscuous gene expression is further exemplified by differences in cell type–specific expression patterns. The expression of certain TRAs (e.g., *GAD67*) is restricted to mTECs, whereas others (e.g., thyroglobulin) are found in both cortical TECs (cTECs) and mTECs at similar levels (14). To decipher the apparently complex cellular and molecular regulation of promiscuous gene expression, we analyzed gene expression in distinct thymic stromal cell lineages and subsets thereof at the level of global gene expression and defined genomic regions.

RESULTS

MTECs specialize in promiscuous gene expression

To define the scope of promiscuous gene expression and the antigenic representation of peripheral tissues in the thymus, we performed a large-scale analysis of gene expression in murine thymic stromal cells using Affymetrix chips. Different stromal cell types were purified by a combination of sequential enzymatic digestion, density gradient centrifugation, and multicolor sorting yielding pure populations of thymic DCs, macrophages, cTECs, and mTECs (5). The mutual comparisons of mTECs with cTECs, DCs, and macrophages revealed in each case a much higher number of genes being overexpressed in mTECs compared with the reference population (see Materials and methods; Fig. 1 A). This observation has been corroborated by an analysis of variance (ANOVA) of the gene expression pattern of these four stromal cell subsets. Clearly, the highest number of genes, which were differentially expressed among all four groups, was found in mTECs (Fig. 1 B).

Because promiscuous gene expression is obviously a particular property of mTECs, we regard the set of genes overexpressed in mTECs versus cTECs as most informative to delineate this gene pool. Given the close relationship between these two cell types (15), cell lineage–specific differences should be minimized, whereas aberrantly expressed genes should be included. The pool of genes overexpressed in mTECs versus cTECs is probably >545 genes (Table S1, available at http://www.jem.org/cgi/content/full/jem.20050471/DC1); when we compared the gene chip results with RT-PCR data of promiscuously expressed genes, we found that only ∼50% of genes analyzed by PCR could be reliably detected as present on arrays. Therefore, the total number of promiscuously expressed genes will be underestimated by at least a factor of two (unpublished data). This is probably because of the fact that the array analysis is less sensitive than RT-PCR and that promiscuously expressed genes are often expressed at low levels. To validate our criteria for the identification of overexpressed genes, we analyzed the expression of several mTEC-specific genes identified with the microarray analysis by real-time PCR. Certain "marker genes" for mTECs were reliably confirmed, and we therefore regarded the chosen criteria as valid. *Casein β* and the *testis lipid binding protein* (*Tlbp*) are detectable at equal levels in mTECs of both genders, including nonlactating female mice. (Fig. S1, available at http://www.jem.org/cgi/content/full/jem.20050471/DC1, and not depicted). Thus, promiscuous gene expression in mTECs overrides the normal tissue-, sex-, and development-dependent gene regulation of these two genes. *Casein β* is typically induced late during pregnancy in the mammary gland (16), and *Tlbp* is expressed in male germ cells (17).

Promiscuous expression has been operationally defined as the expression of genes that so far have not been known to be part of the physiological gene expression program of thymic stromal cells. To apply more stringent criteria, we determined the percentage of genes with restricted tissue expression, a definition relying on our present knowledge of cell type–specific gene expression programs. On account of published gene expression data, we categorized genes as tissue restricted if expressed in <5 out of 45 tissues tested. Approximately 28% of all genes overexpressed in mTECs (152 out of 545 genes) could be categorized as tissue restricted according to this approach (Fig. 1 C). One key finding is that most, if not all, tissues are represented by at least one or multiple genes in mTECs. In contrast, genes overexpressed in cTECs versus mTECs do not show such a bias. Although the relative percentage of TRAs in cTECs appeared similar at first sight (∼28%), most of these transcripts are lymphocyte specific and likely derived from "contamination" of the cTEC population with thymic nurse cells containing thymocytes (Fig. 1 D). This interpretation is supported by the finding that cTECs isolated from *Rag2*$^{-/-}$ mice lack expression of most of these lymphocyte-specific transcripts (unpublished data). In the same vein, genes overexpressed in DCs versus mTECs and macrophages versus mTECs (Fig. S2, available at http://www.jem.org/cgi/content/full/jem.20050471/DC1) only showed limited tissue diversity with the majority being restricted to hematopoietic cell lineages. The comparative analysis of global gene expression patterns among thymic stromal cells clearly singles out mTECs as a cell type specialized in expressing TRAs.

Promiscuous gene expression in mTECs is differentiation dependent

It has been unclear so far whether promiscuous gene expression in mTECs is tightly correlated with lineage commitment or whether it requires further differentiation steps within this lineage. mTECs are heterogeneous with regard to their phenotype, expressing varying levels of MHC class

II, CD80, or binding sites for the lectin UEA (18). It is presumed that an increase in these surface molecules denotes progressive maturation of mTECs either as a cell autonomous differentiation program or induced by mature thymocytes. Here, we chose to separate mTECs according to relative expression levels of the CD80 coreceptor and addressed the question of whether there is any correlation between promiscuous gene expression and induction of this costimulatory molecule. mTECs were sorted into low, intermediate, and high CD80-expressing cells (Fig. 2 A), and the expression of various promiscuously expressed genes was examined by quantitative PCR. 21 out of 22 genes tested showed a clear positive correlation with CD80 expression levels. The same was true for the transcriptional regulator Aire (Fig. 2 B and not depicted). This expression pattern of Aire fits previous observations that describe Aire expression in situ only in a subset of mTECs (13, 19). However, the acute phase protein *CRP* was already strongly expressed in the CD80^{lo} subset and showed no maturation-dependent induction. In keeping with this pattern, *CRP* has been shown to be independently regulated in Aire (6).

Given the concomitant induction of promiscuous gene expression and Aire in CD80^{hi} mTECs, we asked whether Aire directs all or only part of these up-regulated genes. To address this issue, we isolated CD80^{lo} and CD80^{hi} mTECs from Aire-deficient mice and initially analyzed a limited set of promiscuously expressed genes (Fig. 2 C). The expression of genes that are reportedly Aire dependent, such as *casein α* and *γ* and *insulin2*, were barely detectable in CD80^{lo} WT mTECs and both subsets of Aire-deficient mTECs (Fig. 2 C and see Fig. 5 B). In contrast, several genes, such as *casein β* and *κ*, *GAD67*, and *Tlbp*, showed a strong up-regulation concomitant with CD80 expression levels in *Aire$^{-/-}$* mice. Thus, additional transcriptional regulators apart from Aire become operative during the functional maturation of mTECs and drive expression of different promiscuously expressed genes. To delineate the size and diversity of this Aire-independent gene pool, we extended this analysis by comparing the gene expression profile of CD80^{hi} versus CD80^{lo} mTECs of WT and Aire-deficient mice with gene arrays (Fig. 3 A; Tables S2 and S3, available at http://www.jem.org/cgi/content/full/jem.20050471/DC1). This analysis confirmed that the WT CD80^{hi} mTEC subset expressed the highest number, as well as the highest relative percentage (33%), of TRAs among the subsets tested.

The CD80^{hi} mTEC subset of Aire-deficient mice showed a reduction in the total number of overexpressed genes by 25% (120 genes), and the relative percentage of TRAs decreased from 33 to 21% (74 out of 347 genes); nevertheless this pool still comprised >300 genes that were representative of many tissues (Fig. 3 B). In addition, the distributions of fold changes among the genes overexpressed in the CD80^{hi} versus CD80^{lo} subsets in *Aire$^{+/+}$* and *Aire$^{-/-}$* mice were very similar; i.e., the quantitative representation of self-antigens in both gene pools is similar (Fig. 3 C). The fold changes de-
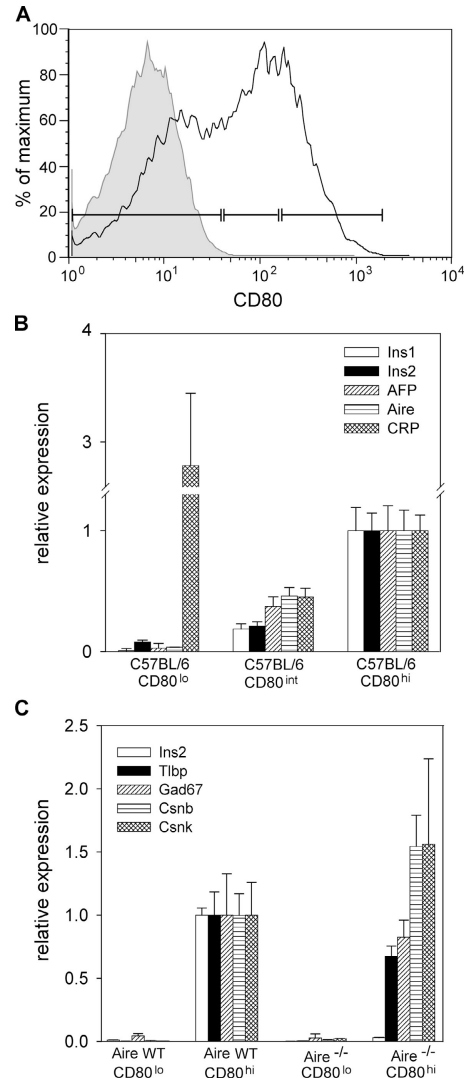

Figure 2. Promiscuous gene expression correlates with CD80 expression levels on mTECs. (A) Expression profile of CD80 expression on mTECs with indicated gates that were chosen for sorting CD80^{lo}, CD80^{int}, and CD80^{hi} mTEC subpopulations. The gray region corresponds with the isotype control staining. (B and C) The expression levels of different promiscuous transcripts were analyzed in the various mTEC subsets by real-time PCR normalized to the relative quantity of β-actin. Note the stepwise up-regulation of gene expression with increasing CD80 levels in WT mTECs, with the exception of *CRP* (B). Expression analysis of a selected panel of promiscuously expressed genes in the CD80^{lo} and CD80^{hi} mTEC subsets of WT and Aire–deficient mice. (C) Aire-independent up-regulation of expression of four of the analyzed genes in the CD80^{hi} subset. Error bars indicate the SD of triplicates of the same cDNA preparation. AFP, α-fetoprotein; Aire, autoimmune regulator; *CRP*, C-reactive protein; Csnb, *casein β*; Csnk, *casein κ*; *GAD67*, glutamic acid decarboxylase 67 kD; Ins1 or -2, insulin 1 or 2; Tlbp, testis lipid binding protein.

rived from the array analysis were validated for a selected set of genes by quantitative PCR. Strikingly, some genes were >100-fold up-regulated (Fig. 3 D).
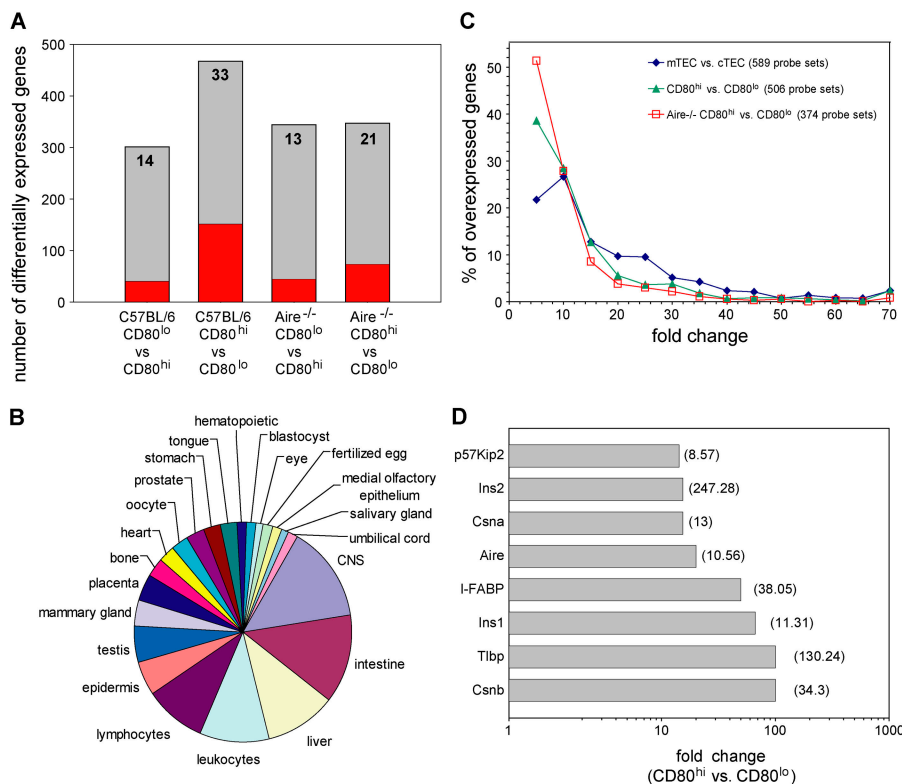
Figure 3. Induction of promiscuous gene expression in CD80[hi] mTECs shows distinct levels of control. (A) The number of differentially expressed genes in the various CD80 subsets indicated is shown by gray regions and the fraction of TRAs contained within these gene pools is shown by red regions. The corresponding relative percentages are given above the bars. (B) Diverse tissue representation by genes induced in the CD80[hi] vs. CD80[lo] subsets of Aire-deficient mTECs. (C) The curves represent the percentage of genes overexpressed in the different mTEC subsets (mTEC vs. cTEC, CD80[hi] vs. CD80[lo] of WT and Aire-deficient mice) at the indicated fold changes. Fold changes ≥70 were combined. (D) Verification of gene expression data derived from the microarray analysis by quantitative PCR. Values in parentheses indicate fold changes in the microarray analysis; values on the x axis denote fold changes of the quantitative PCR analysis that were normalized to β-actin expression.

Notably, Aire had hardly any effect, both in number and fraction of TRAs, on the set of genes that was down-regulated during mTEC maturation (Fig. 3 A). In addition, the fraction of TRAs is lower among genes that were down-regulated compared with those that were up-regulated in CD80[hi] mTECs.

Promiscuously expressed genes colocalize in chromosomal clusters

The structure, function, and physiological regulation of promiscuously expressed genes did not insinuate any commonalities, which would explain their coexpression and coregulation in mTECs. Preferential chromosomal localization has been reported for genes expressed in certain cell lineages. Thus, genes expressed in spermatogonia cluster on the X chromosome (20), and genes expressed in stem cells (a composite of embryonic, hematopoietic, and neuronal stem cells) were overrepresented on chromosome 17 (21). This, however, was not the case for the different gene pools expressed in thymic cell types. There was no marked under- or overrepresentation for particular chromosomes

compared with the distribution of all mapped genes of the array (not depicted). We analyzed whether the genes overexpressed in mTECs and subsets thereof localize to clusters on chromosomes, as has been recently reported, for the expression of tissue-specific or housekeeping genes in different species (22). Indeed, we found that mTEC-specific genes tend to colocalize in clusters comprising up to 16 genes. This clustering was highly significant when compared with random distributions of genes mapped to the same chromosomes (P < 0.001; Fig. 4 A). The same holds true for the array of genes up-regulated in CD80[hi] mTECs, though the mean and maximal number of clustered genes was reduced (Fig. 4 B). Interestingly, genes up-regulated in CD80[hi] mTECs in the absence of Aire still clustered, yet the number of clusters (from 40 to 24) and the number of genes per cluster were further reduced. Although differences in pool size may explain differences in the number of clusters of size 3 and 4, this does not apply to clusters larger than four genes because these clusters have not been observed at all by simulation of randomly sampled genes. Genes overexpressed in cTECs were only enriched in clus-
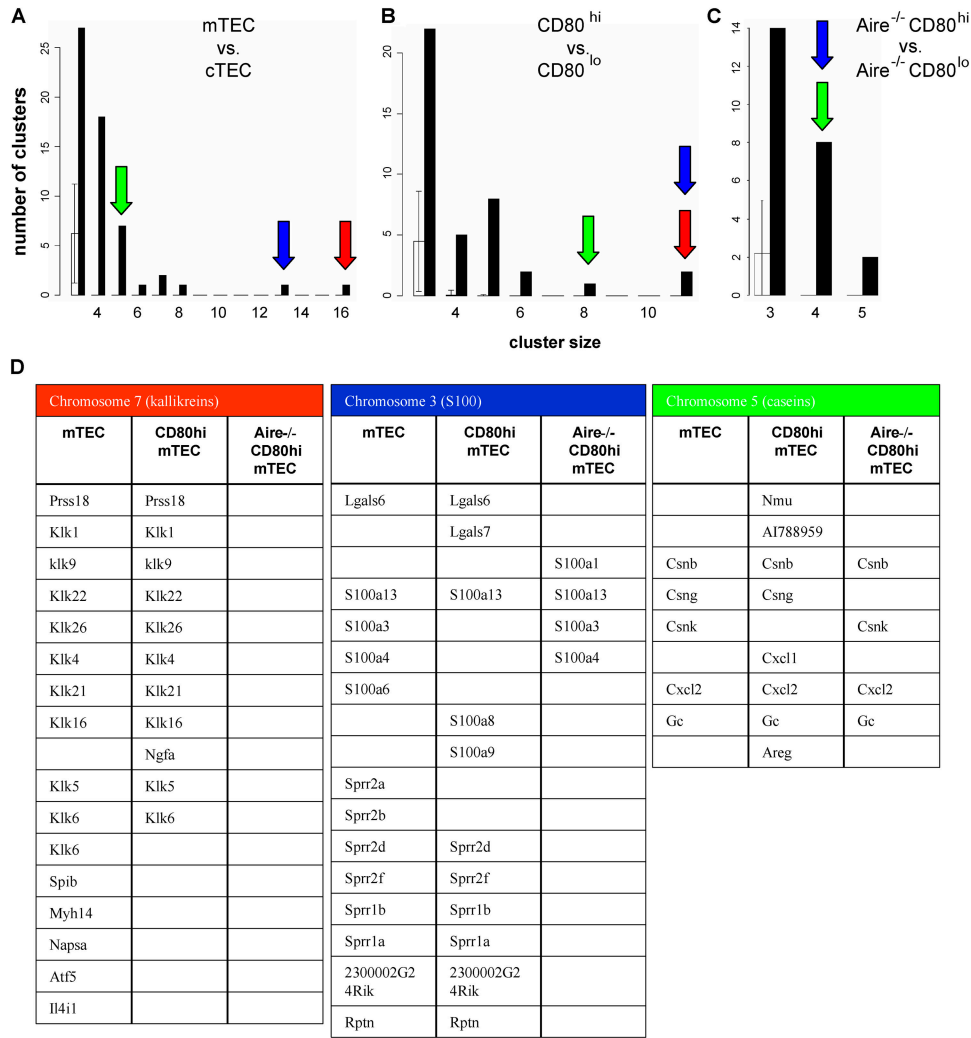
Figure 4. Chromosomal clustering of genes overexpressed in mTEC subsets. (A–C) The number of clusters of 3–16 genes recorded within a sliding window of 10 consecutive genes is shown for the different subsets (see Materials and methods). The black bars refer to the experimental values, and the white bars to the number of clusters observed in randomly generated gene lists. The error bars indicate the SD among the randomly generated gene lists. The identity and size of three clusters is indicated by arrows: red, kallikrein cluster on chromosome 7; blue, S100 cluster on chromosome 3; green, casein cluster on chromosome 5. Note the progressive reduction in number and size of clusters in the different gene pools. P < 0.001, except where indicated. Clusters of two genes were in no case significantly different from randomly generated gene lists. (D) Composition of the three different clusters as shown in A by arrows. The arrangement from top to bottom reflects the alignment from centromere to telomere on the respective chromosomes.

ters of three genes (not depicted), thus confirming that colocalization in larger clusters is specific for the mTEC gene sets. As a second quantitative measure of colocalization, the frequency of neighboring genes within windows of different size (50–5,000 kb) was determined (Fig. S3, available at http://www.jem.org/cgi/content/full/jem.20050471/DC1). This analysis confirmed the tendency of promiscuously expressed genes to colocalize; the degree of colocalization showed the same hierarchy as the cluster analysis shown in Fig. 3. The progressive reduction in the frequency of clustered genes was not necessarily caused by the entire loss of a given cluster, but also because of reduc-

tion of genes within a cluster (Fig. 4, A–C). Reduction in numbers of colocalized genes either affected contiguous stretches or scattered genes, as defined by the array analysis (Fig. 4 D). Notably, the gene pools with the highest degree of gene clustering also displayed the highest percentage of TRAs (compare Fig. 1 A, Fig. 3 A, and Fig. S3).

These data document clustering as a distinctive feature of promiscuously expressed genes that is conserved across different species (7). Moreover, they suggest that genes within a particular gene cluster are not subject to strict coregulation, but seem differentially regulated. To further corroborate this finding, we chose to analyze one cluster in detail.

Contiguous expression and coregulation of clustered genes

The gene clustering deduced from bioinformatic processing of the Affymetrix gene chip analysis is likely to be incomplete. These chips are estimated to cover about one third of all murine transcripts, and many transcripts will escape detection because of low expression. Therefore, we decided to address this directly by analyzing gene expression in a contiguous chromosomal region by RT-PCR in mTECs and their subsets. We chose the casein gene region on mouse chromosome 5 that we identified during the cluster analysis because expression of casein genes can be classified as bona fide promiscuous (Fig. S1). In addition to the casein gene family, this region of ∼1 Mb encodes genes shown to be expressed in the salivary glands, testis, epididymis, liver, kidney, and olfactory bulb and epithelium. Expression of the various genes was analyzed in mTECs and the respective tissues by semiquantitative PCR. Of the 14 genes analyzed, 11 were contiguously expressed in mTECs (Fig. 5 A). In contrast, expression of this locus was much more restricted in the various peripheral tissues. The expression pattern largely conformed to the prediction, but was often broader than that deduced from published data. It is obvious from even this incomplete analysis that expression of a given gene is rarely confined to one tissue. None of the control tissues, however, showed a similar transcriptional "read-through" of this genomic region as the mTECs, in which contiguous expression of genes purportedly specific for different tissues is observed in a region covering ∼800 kb. Interestingly, this read-through ends in the upstream region of the casein locus in mTECs. Although *Ugt2a1* is still transcribed, three different members of the UDP glycosyl-transferase family, located more distal to the casein region, are barely transcribed in mTECs, but clearly in liver and kidney. The nature of this transcriptional boundary in mTECs is currently unclear and a corresponding demarcation in the 3′ region has not yet been defined. Our finding of contiguous gene expression within this genomic region is necessarily based on the current status of gene mapping. We cannot preclude that, with future refinement of gene maps, genes that are not expressed in mTECs may be identified in this region.

To further characterize gene regulation in this local region we asked whether neighboring genes are coregulated en bloc or whether individual genes are subject to differential regulation. When we examined the expression of the core region of this cluster in CD80[hi] and CD80[lo] mTECs by real-time PCR, we found that all six adjacent genes were coinduced in CD80[hi] mTECs. Extension of this comparative analysis to *Aire*[−/−] mice, however, revealed that individual genes within this cluster differed in their dependency on Aire (Fig. 5 B). Aire-dependent genes were not necessarily grouped together, but dispersed among Aire-independent genes. These data document two levels of regulation: one targets clusters at large and requires mTEC maturation and the other targets individual genes and requires Aire and still unidentified transcriptional regulators.

Loss of imprinting (LOI) of *insulin–like growth factor 2* (*Igf2*) in mTECs

Colocalization of promiscuously expressed genes in regional clusters and their coregulation within such clusters is strongly suggestive of epigenetic gene regulation. Genes for which epigenetic regulation has been extensively documented are imprinted genes (23). Among the genes overexpressed in mTECs versus cTECs, we observed several imprinted genes, including the *pleiomorphic adenoma gene-like 1* (*Plagl1*), *cyclin-dependent kinase inhibitor 1C* (*Cdkn1c*), and the colocalized genes *Igf2* and *H19 fetal liver* mRNA (*H19*). Up-regulation of expression of these genes in mTECs could either be caused by enhanced expression of the imprinted active allele (either paternal or maternal) or by biallelic transcription after derepression of the silent allele. Mouse strains exhibiting single nucleotide polymorphisms on the distal arm of chromosome 7, where *H19*, *Igf2*, and *Cdkn1c* are located, allowed us to experimentally distinguish between these two mechanisms. Expression of *Igf2,* which was 17-fold overexpressed in mTECs, was monitored in crosses between C57BL/6 and SD7 mice in different organs and in mTECs by RT-PCR, followed by single nucleotide primer extension (SNuPE)/ HPLC analysis (Fig. 6). *Igf2* expression showed normal imprinting in the kidney and liver (i.e., only the paternal allele was expressed), but biallelic expression in the brain, as previously reported (24–26). Intriguingly, we also observed biallelic expression in mTECs; this finding was confirmed in the reciprocal crossing between SD7 and C57BL/6 mice. Thus, mTECs represent an additional somatic cell lineage in adult mice apart from certain cell types in the central nervous system in which imprinting of the *Igf2* gene is abolished. To determine whether this LOI extends to other imprinting loci as a reflection of more widespread epigenetic deregulation in mTECs, we analyzed the expression of the cell cycle inhibitor *Cdkn1c*, also known as *p57Kip2*, which also was 34-fold up-regulated in mTECs. *Cdkn1c* is encoded ∼800 kb telomeric on the same chromosome but, in contrast to *Igf2*, is paternally imprinted and controlled by a different imprinting center (23). Remarkably, imprinting of this gene was not abolished when both reciprocal crossings were analyzed. We only observed expression from the maternal allele both in control tissues and in mTECs.

DISCUSSION

The comparative analysis of gene expression in distinct thymic stromal cells at the global level, the level of chromosomal regions, and individual genes documents the complexity of promiscuous gene expression. Although promiscuous gene expression is a basal feature of TECs, mTECs clearly display the highest degree of promiscuous expression both with regard to number and diversity of expressed genes. Extrapolating from our gene array analysis, we estimate that 2,000–3,000 genes, or 5–10% of all known mouse genes, are turned on in mTECs, in addition to their cell lineage–specific expression program. The complexity of promiscuous gene expression in-
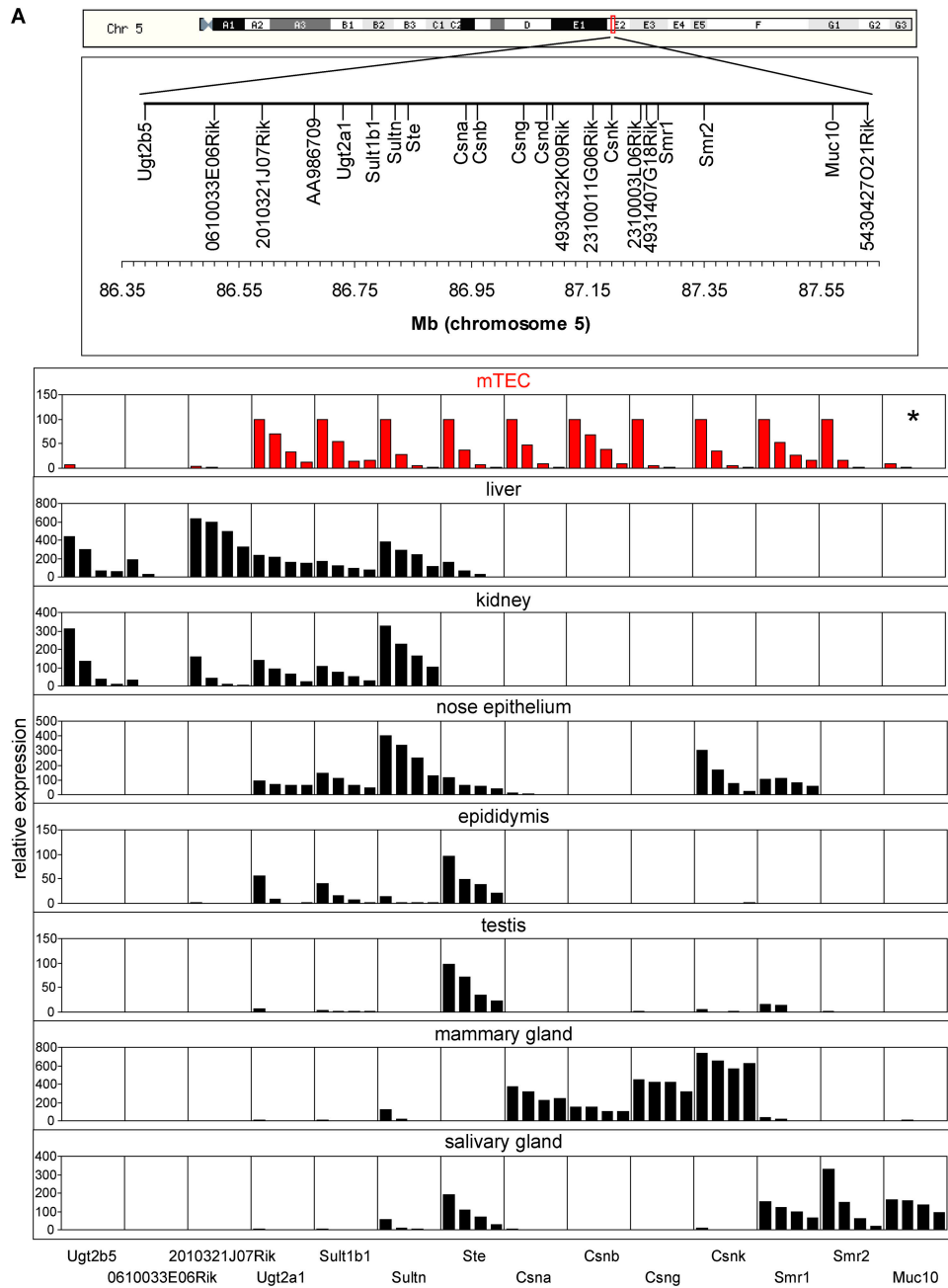
Figure 5.   Contiguous promiscuous gene transcription in the casein cluster. (A) Schematic representation of 1.2 Mb of chromosome 5, depicting the casein gene region flanked upstream by members of the sulfo-transferase and UDP glycosyl-transferase families and flanked downstream by salivary gland genes (top). The bottom panel shows the expression profile of the various genes of this region in mTECs, as analyzed by semiquantitative RT-PCR (fourfold serial dilutions). The same analysis was performed for the various tissues in which the various genes are specifically transcribed. Note that contiguous gene expression was only observed in mTECs. These expression patterns are representative of two independent experiments; discordant expression results were only observed for one gene (*).

creases in ascending order from cTECs, to immature mTECs, to mature CD80[hi] mTECs. These different gene pools are not complementary but additive, with the more complex one encompassing the less complex pool. Promiscuous gene expres-

sion is thus not only specified by commitment into the TEC lineages, but also by differentiation after lineage commitment.

Our observation that the bulk of promiscuously expressed genes are only turned on in CD80[hi] mTECs strongly supports

B
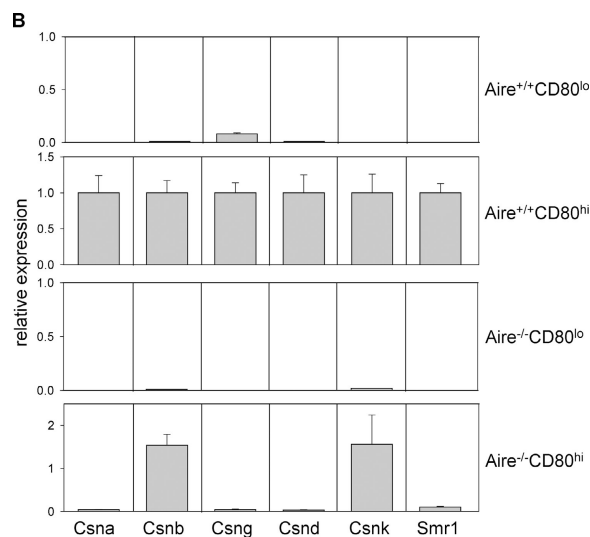
(B) Expression analysis of selected casein genes in the core region of this cluster by real-time PCR. Expression levels in CD80[lo] and CD80[hi] mTEC subsets of WT and Aire–deficient mice were examined. Although all six genes were coinduced in mature mTECs, they still differed in their dependency on Aire. Error bars indicate SD of triplicates of the same cDNA preparation.

Figure 6. Imprinting in mTECs (*Igf2* vs. *Cdkn1c* genes). Expression of *Igf2* and *Cdkn1c* was analyzed by RT-PCR amplification and SNuPE/HPLC in mTECs and control tissues from the F_1 generation of C57BL/6 × SD7 and SD7 × C57BL/6 crosses. Elution profiles of the SNuPE products are shown. The first peak corresponds to unextended primers and the second and third peak to products transcribed from the maternal or paternal allele, respectively, as indicated. *Igf2* is paternally expressed with the exception of the choroid plexus and leptomeninges. Note that biallelic expression also occurs in mTECs. In contrast, imprinting of *Cdkn1c* is maintained in all tissues tested including mTECs; i.e., the gene is maternally expressed. The analysis of genomic DNA (top right) indicates the position of both allele-specific PCR products. pat, paternal; mat, maternal.

the "terminal differentiation model." This model holds that promiscuous gene expression is enacted during mTEC differentiation/maturation. Terminally differentiated mTEC clones would express an assortment of TRAs of mixed tissue and germ layer derivation. Alternatively, it had been proposed that mTECs emulate the gene expression program of tissue-specific cell lineages. The medulla would thus represent a patch quilt of different tissues (the "mosaic model"; reference 27). According to this model, promiscuous expression of a given gene would follow rules of tissue-specific regulation. This prediction is, however, not supported by recent findings. Thus, insulin expression in mTECs and β cells of the pancreas is differently regulated in response to decreasing copy numbers of the insulin genes (28).

At least four pools of promiscuously expressed genes can be discerned: (a) genes that are expressed at similar levels in cTECs and mTECs and at much lower levels, if at all, in hematopoietic cells (e.g., *PLP* or thyroglobulin); (b) genes that are only expressed in mTECs, irrespective of the maturation stage (e.g., *CRP*); (c) genes that are strongly induced in CD80[hi] mTECs contingent on Aire (e.g., insulin or *casein α*); and (d) genes that are induced in CD80[hi] in the absence of Aire (e.g., *GAD67* or *casein β*). These different gene pools obviously differ in size, composition, and mode of regulation. To date, only one molecular component of this regulation has been identified, the transcriptional regulator Aire, which directs the maturation-dependent induction of a few hundred genes. Considering the role of promiscuous gene expression in tissue-specific tolerance, not only the number of expressed genes but also the balanced antigenic represen-
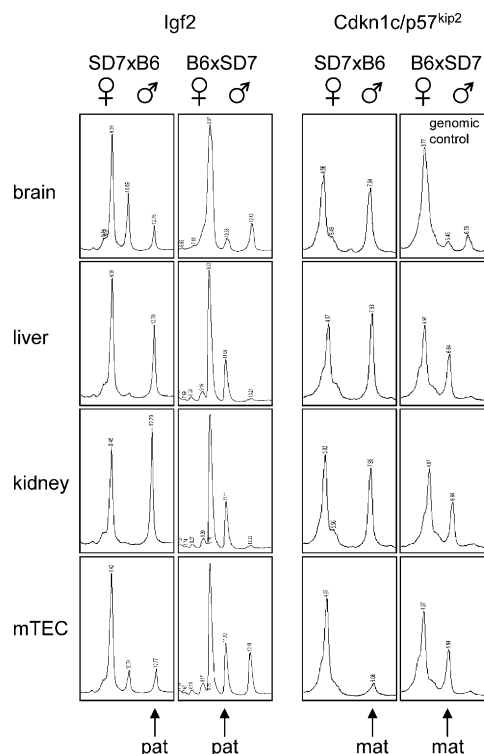
tation of diverse tissues matters. In this regard, the fraction of bona fide TRAs (i.e., genes expressed in less than five tissues, based on currently available data) is of particular interest. The percentage of TRAs among genes up-regulated in CD80[hi] mTECs of WT mice was 33% (i.e., 152 genes; Fig. 3). Interestingly, this enrichment of TRAs was even higher when the CD80[hi] subset was examined between WT and $Aire^{-/-}$ mice; i.e., 45% of Aire-dependent genes were categorized as TRAs (unpublished data). This remarkable feature of Aire to preferentially target genes of restricted tissue expression (6) awaits a molecular explanation.

Despite the important quantitative and qualitative contribution of Aire in directing promiscuous gene expression in mTECs and, thus, protecting from autoimmunity, a sizable number of genes are still expressed in the absence of Aire. These genes still represent diverse tissues and are composed of up to 21% TRAs (Fig. 3). This "residual" promiscuous gene expression could explain the relative mild autoimmune

phenotype of $Aire^{-/-}$ mice. Promiscuous gene expression independent of Aire falls into three categories as represented by pools (a), (b), and (d). Possible genetic and/or epigenetic mechanisms directing expression of these pools remain to be identified. An important inference from the terminal differentiation model of promiscuous gene expression (1) is that arrest of differentiation at the CD80[lo] stage will result in the absence of both gene pools (c) and (d), and this should result in a more severe autoimmune phenotype than that of Aire knock–out mice.

Clustering is one distinctive feature of promiscuously expressed genes that otherwise do not show obvious commonalities. Chromosomal gene clustering has been described in different species and has been interpreted to be the result of juxtaposition in order to facilitate their coregulation (22). In contrast, we suggest that the observed clustering of promiscuously expressed genes in mTECs is a result of accessibility of chromosomal regions to transcription irrespective of tissue-specific differentiation patterns. The selection of genes within clusters would thus be determined by the genetic history of the species rather than immunological selection criteria. Such clusters encompass up to 1 Mb, as exemplified with the casein region, but may be larger because the boundaries have not been defined. The coordinated induction of the core region of the casein cluster in CD80[hi] mTECs speaks for long-range regulatory effects (opening of domains), whereas the differential dependency of adjacent casein genes on Aire shows gene-specific regulation. *Caseins β* and *κ* are Aire independent, and the remaining genes are Aire dependent. Interestingly, this pattern does not concur with the gene ontology of the casein family members; although *casein β* belongs to the group of calcium-sensitive genes (*caseins α, β, γ,* and *δ*), which have originated from a common ancestral gene, the *casein κ* gene is not related to this group (16).

One interpretation of these findings is as follows. During mTEC maturation, alterations in the accessibility of scattered, local regions in the genome would allow DNA–binding complexes to differentially control transcription of genes within such open domains. Aire may be part of such complexes. The position-independent control of several transgenes directed by tissue-restricted promoters by Aire suggests that promoter-specific sequences directly or indirectly specify the activity of Aire (29). The control of promiscuous gene expression by Aire would thus be contingent on specific conditions in mTECs. This interpretation concurs with a recent report that shows that Aire, when overexpressed in different human monocytic cells, targets a different set of genes than in mTECs (30).

It is currently unclear whether mTECs are unique among all somatic cells in expressing such a diversity of TRAs. Yet, a limited comparison showed that none of the seven different tissues of mixed cellular composition, including epithelial cells, showed a similar read-through of the casein locus as purified mTECs. In addition, when comparing global gene expression between four thymic stromal cells and unseparated liver tissue (unpublished data), mTECs clearly expressed the largest set of genes not shared by the other cell types. Promiscuous gene expression has also been reported for multi- and oligopotent stem cells (31, 32). Although the biological role in this context is most certainly different, the molecular regulation may share common features.

A striking observation of our study is that the four imprinted genes *Igf2, Cdkn1c, Plagl1,* and *H19* are overexpressed in mTECs. Because imprinted gene expression is controlled by DNA methylation and chromatin alterations (23), this suggests that changes in such epigenetic marks may play an important role in promiscuous gene expression. This view is supported by the additional observation that several promiscuously expressed genes in mTECs are also found to be overexpressed in epigenetically modified, hypomethylated mouse fibroblasts (33). However, the simple explanation that global DNA methylation changes might be the trigger for LOI and overexpression in mTECs seems to be rather unlikely. First, several other imprinted genes known to be controlled by DNA-methylation are not overexpressed in mTECs. Second, *Cdkn1c* and *H19* overexpression is not accompanied by LOI (unpublished data). Third, the level of overexpression by far exceeds the expected twofold increase as a consequence of LOI.

We can envisage two scenarios as possible explanations for the overexpression of imprinted genes and other genes whose expression was shown to be methylation sensitive. Epigenetic silencing may simply be overridden (or masked) by other expression mechanisms, or locally induced epigenetic changes affect only selected genes but do not abrogate general epigenetic marks like imprints. A striking observation in this context is the LOI and overexpression of the *Igf2* gene in mTECs and the simultaneous overexpression of *H19*. Biallelic expression of *Igf2* has been also reported in other somatic cells; i.e., the choroid plexus and the leptomeninges (24, 25). Here the *H19* gene is expressed monoallelically, with the paternal allele being silent. This uncoupling of *Igf2* expression from imprinting at the *H19* locus has been shown to involve the centrally conserved domain enhancer between *Igf2* and *H19* (26). So far, the mechanisms driving biallelic *Igf2* overexpression in mTECs remain unclear.

LOI alone would not explain the 17-fold up-regulation of *Igf2* transcription. Intriguingly, *Igf2* is not only biallelically expressed in mTECs, but is also Aire dependent. Aire was shown to direct promiscuous expression of *Ins2* (6). Because *Ins2* is located in direct proximity to *Igf2*, it is conceivable that Aire may also affect the neighboring *Igf2* gene completely independent of the imprinting control by the *H19–Igf2* region.

Induction of promiscuous gene expression in CD80[hi] mTECs correlates with concomitant up-regulation of different sets of genes involved in antigen processing and presentation, including MHC class II, H-2M, CD80, and several cathepsins (E, H, K, S, and Z). The induction of these two gene expression programs probably occurs independently because up-regulation of MHC class II, CD80, H-2M, and cathepsin S was also observed in $Aire^{-/-}$ mice (Table S4,

available at http://www.jem.org/cgi/content/full/jem. 20050471/DC1). The acquisition of professional antigen presentation competence parallel to the induction of the complete complement of promiscuously expressed genes enables mature mTECs to present a host of self-peptides at sufficient epitope density. In conjunction with the display of appropriate coreceptors, mTECs are thus able to autonomously tolerize tissue-reactive T cells, possibly both via deletion and induction of T regulatory cells (1).

The maturation sequences of mTECs and DCs into competent APCs share certain common features. Both cell types, despite their different origins, express CD80/86 and the immuno-proteasome (34) and are able to activate naive T cells in vitro (Koble, C., personal communication). In contrast, cTECs, which share a common precursor with mTECs (15), do not express CD80 or the immunoproteasome and lack complete competence to activate naive T cells (35). The common function of mTECs and DCs in tolerance induction in the thymic medulla thus overrides their different lineage derivation. A further aspect of DC maturation is the up-regulation of chemokines, which attract naive T cells and, thus, facilitate the encounter between rare antigen-specific T cells and antigen-laden DCs (36). MTECs also up-regulate an array of chemokines (Table S5, available at http://www.jem.org/cgi/content/full/jem.20050471/DC1), and this may serve the same purpose, namely to attract highly mobile thymocytes to those mTECs that display the complete repertoire of self-determinants.

CD80^{hi} mTECs also up-regulate genes that characterize terminally differentiated keratinocytes; e.g., *claudin-4* and *-7*, *keratin 10*, and the epidermal differentiation complex (37, 38). This gene complex serves to provide the barrier activity of stratified squamous epithelia. Interestingly, human mTECs also build up a barrier activity when forming Hassall's corpuscles, which presumably are formed by terminally differentiated mTECs (37). Whether this is of any physiological significance or is a byproduct of the differentiation program is not clear.

In conclusion, our data document a remarkable cellular and molecular specialization of the thymic stroma, which is highly conserved between mice and humans (reference 7; unpublished data). This serves to comprehensively mimic the transcriptome of peripheral tissues and, thus, maximize the scope of central self-tolerance. Understanding these processes in more detail will be of considerable biological interest and may also help to unravel the complex genetic regulation of organ-specific autoimmune diseases.

MATERIALS AND METHODS

Animals. C57BL/6 mice were obtained from Charles River Laboratories. *Aire*$^{-/-}$ mice were genotyped as previously described (39). These mice were of a mixed genetic background. For analyses of allele-specific gene expression, *Mus musculus domesticus* (C57BL/6) and *domesticus* mice harboring a *Mus spretus* allele on distal chromosome 7 (SD7) were mated and various tissues were dissected from the F$_1$ generation. All mice were kept under specific pathogen-free conditions at the animal facilities of the German Cancer Research Center.

Isolation of murine thymic stromal cells. Thymic stromal cells were purified as described previously (5) with the following staining modifications. Thymic rosettes were stained with anti-CD11c–PE (HL3; BD Biosciences) and anti-F4/80–FITC (CI:A3-1; Serotec). The TEC-enriched fraction was stained with either anti-CDR1–Alexa488 or anti-Ly51–FITC (6C3; BD Biosciences), anti–Ep-CAM–Cy5 (G8.8), and anti-CD45–PE (30-F11; BD Biosciences). After staining, cells were resuspended in FACS buffer containing 1 μg/ml propidium iodide to exclude dead cells. For subdivision of mTECs, the following combinations were used: anti-CD80–PE (16-10A1) or anti-KLH–PE (Ha 4/8; isotype control), anti-CD45–PerCP (30-F11), anti-Ly51–FITC (6C3; all were obtained from BD Biosciences), and anti–Ep-CAM–Cy5 (G8.8). FcR blocking with the anti-FcR mAb 2.4G2 preceded all stainings. Cell sorting was performed with a cell sorter (FACSVantage Plus; Becton Dickinson).

RNA preparation and cDNA synthesis. Whole tissue RNA was isolated using DNaseI digestions on-column with an Ultra-Turrax T25 (IKA) and the RNeasy Mini Kit (QIAGEN) and from single-cell suspensions with the High Pure RNA Isolation Kit (Roche). Total RNA (4 μg of tissue-extracted RNA or an equivalent of $4 \times 10^4 - 1 \times 10^6$ single cells) was reverse transcribed into cDNA with Oligo(dT)$_{20}$ Primer and Superscript II Reverse Transcriptase (Invitrogen), followed by RNase H digestion (Promega).

RT-PCR analysis. PCRs were performed as previously described (5). All primer pairs were synthesized by the oligonucleotide synthesis facility of the German Cancer Research Center and, when possible, were designed to span at least one intron. PCR products were revealed with the Lumi-Imager F1 Workstation (Roche) and bands were quantified with LumiAnalyst 3.0 software (Roche). For semiquantitative PCR, the different cDNA preparations were normalized to β-actin expression before testing expression of the gene of interest.

Quantitative PCR. Real-time PCR reactions were performed in a final volume of 25 μl with optimal concentrations of the forward and reverse primers (50–900 nM) using the qPCR Core Kit for SybrGreen I (Eurogentec) containing Hot GoldStar polymerase and uracil-DNA glycosylase. Probes were used with a concentration of 200 nM in combination with the qPCR Core Kit (Eurogentec). Reactions were run on a sequence detection system (GeneAmp 5700; Applied Biosystems) in triplicates, and expression values were normalized to β-actin expression using the comparative C$_T$ method. Primers were purchased from MWG and, when possible, were designed to span at least one intron. Probes were purchased from Eurogentec.

SNuPE analysis. *Igf2* and p57kip2 were RT-PCR–amplified from RNA isolated from organs and thymic stromal cells of the F$_1$ generation derived from crossings between SD7 and C57BL/6 mice using AmpliTaq DNA polymerase (Applied Biosystems) and the following primer pairs (sense and antisense, respectively): *Igf2*, 5′-GGCCCCGGAGAGACTCTGTGC-3′ and 5′-TGGGGGTGGGTAAGGAGAAACCT-3′; and p57kip2, 5′-TTCAGATCTGACCTCAGACCC-3′ and 5′-AGTTCTCTTGCGCT-TGGC-3′. PCR products were separated in agarose gels; bands of an appropriate size were excised and purified using the QIAquick Gel Extraction Kit (QIAGEN).

For the SNuPE reaction, 20–130 ng/μl of gel-purified RT-PCR products were used as templates. SNuPE primers were placed immediately adjacent to the polymorphic sites and had the following sequences: *Igf2*, 5′-TCAGTGAATCAAATTA-3′; and *p57kip2*, 5′-CTGTTCCTCGC-CGTCC-3′. Before performing the SNuPE reaction on *p57kip2*, PCR products had to be digested with 0.2 U *FOKI* to avoid secondary structures. The primers were extended in 20-μl reactions using the following conditions: 3.6 μM SNuPE primer, 0.05 mM ddNTPs, and 0.15 U Thermo-Sequenase (GE Healthcare) in reaction buffer supplied by the manufacturer. After denaturation for 2 min at 96°C, 50 cycles (15 s at 96°C, 30 s at 37°C, and 2 min at 52.5°C) were performed. Extension products were separated on an IP-HPLC system (WAVE DNA Fragment Analysis System; Transgenomics).

Microarray analysis. Microarray analysis using MGU74Av2 chips (Affymetrix, Inc.) was performed as previously described (7, 40). Gene expression data have been deposited in the GEO database (available under accession no. GSE2585) at http://www.ncbi.nlm.nih.gov/projects/geo/. Reagents were provided by T. Wintermantel and D. Engblom (German Cancer Research Center, Heidelberg, Germany).

Identification of tissue-restricted genes. Gene expression data from the public database at http://symatlas.gnf.org (41) were taken as a starting point for the identification of tissue-restricted genes among the total number of genes overexpressed in the different thymic stromal cell populations. This database contains expression assignments for many different tissues and cell types derived by gene array analysis using Affymetrix U74A or custom-made GNF1M arrays. In combination with data from the mouse genome informatics database (http://www.informatics.jax.org), Swissprot, and the literature, genes were assigned to tissues of their predominant expression when applicable. Genes with expression restricted to less than five tissues were designated as tissue restricted. Among these genes, expression in a single tissue was rare (e.g., *Csnk*), whereas expression in two to four tissues represented the most cases (e.g., *Mep1a*, *Tff2*, and *Calb1*).

Gene mapping and bioinformatic analysis. The overexpressed genes were mapped to the genome using the MapIt program (unpublished data). MapIt is an automated database driven tool designed to identify and localize the absolute position of a given list of genes to the organism's specific genome. We performed queries across databases such as Locuslink, UniGene, and Ensembl, implemented under the Sequence Retrieval System using the gene symbol as input. The quality of the data was ascertained by applying checks to minimize the annotation inconsistencies across different databases and by reducing the redundancy from the given list of entries. We queried Locuslink and Unigene databases to obtain the gene-related information. Taking into account the gene information and the input identifier, we retrieved the relative start and end location of each gene from the Ensembl database. The absolute start and end locations were calculated and the genes (geneID) were then mapped to the genome.

To determine the number of clusters for a given set of differentially expressed genes, all genes represented on the array were aligned according to their physical position on the chromosomes. We then counted the number of genes from the set of differentially expressed genes in a moving window of 10 genes and recorded the largest clusters. As it turned out that in some experiments clusters of differentially expressed genes were immediately neighbored, we appended an assembly step, where clusters were joined, when they were <10 genes apart. The significance of the clustering was determined by repeating the same procedure 1,000 times in each case with a list of random genes of the same length as the experimental dataset, and we compared the results with the number of clusters found. This simulation yielded the empirical null distribution which allows p-values to be derived. If, for example, the number of clusters of size 3 was five for a particular gene set, and this value was reached or exceeded only twice in 1,000 simulations, then the empirical p-value would be $2/1,000 = 0.002$.

In this analysis we did not exclude homologous genes or gene families, as expression of individual members of such gene families may also reflect promiscuous gene expression (7). We independently assessed gene clustering by calculating the number of pairs of genes 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 a given cell type were compared with those obtained from 1,000 random lists. p-values were calculated from the empirical distribution according to Roy et al. (42).

To check for diversity of gene expression in different tissues, we investigated a panel of gene expression measurements from mouse liver, DCs, and macrophages, as well as cortical and medullary thymic cells, by ANOVA. A one-way design with tissue type as the only factor was applied. p-values from the F test were corrected for multiple testing by applying the procedure of Benjamini and Hochberg (43). Genes with adjusted p-values

<0.01 were considered to be significantly differentially expressed. These genes were ordered by the tissue type in which they displayed characteristically high expression. Gene expression of these genes was visualized by heat maps of the z-transformed values. The z-transform brings values for each gene to zero mean and unit variance. Rows and columns of the gene expression matrix visualized by the heat map have been reordered by hierarchical clustering of euclidean distances using the complete linkage algorithm (44). All calculations were performed in R version 2.0.1 (http://www.R-project.org) with the extension package multitest, version 1.5.2 (45).

Online supplemental material. Fig. S1 shows validation of the criteria for the identification of genes identified as overexpressed in mTECs as compared with cTECs by real-time PCR (normalized to β-actin) of RNA isolated from purified thymic APCs of young adult male and female C57BL/6 mice. Fig. S2 shows the tissue representation in mutual comparisons of global gene expression between mTECs, DCs, and macrophages. Fig. S3 shows the frequencies of neighbored genes in the gene pools defined by the comparison of various thymic stromal cell subsets.

Table S1 lists all genes overexpressed in mTECs versus cTECs according to chosen criteria described in Material and methods. Table S2 lists all genes overexpressed in the CD80^{hi} versus CD80^{lo} mTEC subset. Table S3 lists all genes overexpressed in the $Aire^{-/-}$ CD80^{hi} versus $Aire^{-/-}$ CD80^{lo} mTEC subset. Table S4 lists cathepsins overexpressed in mTECs versus cTECs, CD80^{hi} versus CD80^{lo} mTECs, or $Aire^{-/-}$ CD80^{hi} versus $Aire^{-/-}$ CD80^{lo} mTEC subsets. Table S5 lists chemokines overexpressed in mTECs versus cTECs, CD80^{hi} versus CD80^{lo} mTECs, or $Aire^{-/-}$ CD80^{hi} versus $Aire^{-/-}$ CD80^{lo} mTEC subsets. Online supplemental material is available at http://www.jem.org/cgi/content/full/jem.20050471/DC1.

Note added in proof. A recent study reports clustering of the subset of promiscuously expressed genes controlled by Aire (Johnnidis, J.B., E.S. Venanzi, D.J. Taxman, J.P. Ting, C.O. Benoist, and D.J. Mathis. 2005. *Proc. Natl. Sci. USA.* 102:7233–7238).

REFERENCES

1. Kyewski, B., and J. Derbinski. 2004. Self-representation in the thymus: an extended view. *Nat. Rev. Immunol.* 4:688–698.

2. Sakaguchi, S. 2004. Naturally arising CD4+ regulatory t cells for immunologic self-tolerance and negative control of immune responses. *Annu. Rev. Immunol.* 22:531–562.

3. 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.

4. Avichezer, D., R.S. Grajewski, C.C. Chan, M.J. Mattapallil, P.B. Silver, J.A. Raber, G.I. Liou, B. Wiggert, G.M. Lewis, L.A. Donoso, and R.R. Caspi. 2003. An immunologically privileged retinal antigen elicits tolerance: major role for central selection mechanisms. *J. Exp. Med.* 198:1665–1676.

5. 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.

6. 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.

7. Gotter, J., B. Brors, M. Hergenhahn, and B. Kyewski. 2004. Medullary epithelial cells of the human thymus express a highly diverse selection of tissue-specific genes colocalized in chromosomal clusters. *J. Exp. Med*. 199:155–166.

8. Pitkanen, J., and P. Peterson. 2003. Autoimmune regulator: from loss of function to autoimmunity. *Genes Immun*. 4:12–21.

9. 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.

10. Soderbergh, A., A.G. Myhre, O. Ekwall, G. Gebre-Medhin, H. Hedstrand, E. Landgren, A. Miettinen, P. Eskelin, M. Halonen, T. Tuomi, et al. 2004. Prevalence and clinical associations of 10 defined autoantibodies in autoimmune polyendocrine syndrome type I. *J. Clin. Endocrinol. Metab*. 89:557–562.

11. Klein, L., and B. Kyewski. 2000. "Promiscuous" expression of tissue antigens in the thymus: a key to T-cell tolerance and autoimmunity? *J. Mol. Med*. 78:483–494.

12. Liston, A., D.H. Gray, S. Lesage, A.L. Fletcher, J. Wilson, K.E. Webster, H.S. Scott, R.L. Boyd, L. Peltonen, and C.C. Goodnow. 2004. Gene dosage–limiting role of Aire in thymic expression, clonal deletion, and organ-specific autoimmunity. *J. Exp. Med*. 200:1015–1026.

13. Chin, R.K., J.C. Lo, O. Kim, S.E. Blink, P.A. Christiansen, P. Peterson, Y. Wang, C. Ware, and Y.X. Fu. 2003. Lymphotoxin pathway directs thymic Aire expression. *Nat. Immunol*. 4:1121–1127.

14. 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.

15. Gordon, J., V.A. Wilson, N.F. Blair, J. Sheridan, A. Farley, L. Wilson, N.R. Manley, and C.C. Blackburn. 2004. Functional evidence for a single endodermal origin for the thymic epithelium. *Nat. Immunol*. 5:546–553.

16. Rijnkels, M., D.A. Wheeler, H.A. de Boer, and F.R. Pieper. 1997. Structure and expression of the mouse casein gene locus. *Mamm. Genome*. 8:9–15.

17. Korley, R., F. Pouresmaeili, and R. Oko. 1997. Analysis of the protein composition of the mouse sperm perinuclear theca and characterization of its major protein constituent. *Biol. Reprod*. 57:1426–1432.

18. Nelson, A.J., S. Hosier, W. Brady, P.S. Linsley, and A.G. Farr. 1993. Medullary thymic epithelium expresses a ligand for CTLA4 in situ and in vitro. *J. Immunol*. 151:2453–2461.

19. Heino, M., P. Peterson, N. Sillanpaa, S. Guerin, L. Wu, G. Anderson, H.S. Scott, S.E. Antonarakis, J. Kudoh, N. Shimizu, et al. 2000. RNA and protein expression of the murine autoimmune regulator gene (Aire) in normal, RelB-deficient and in NOD mouse. *Eur. J. Immunol*. 30:1884–1893.

20. 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.

21. Ramalho-Santos, M., S. Yoon, Y. Matsuzaki, R.C. Mulligan, and D.A. Melton. 2002. "Stemness": transcriptional profiling of embryonic and adult stem cells. *Science*. 298:597–600.

22. Hurst, L.D., C. Pal, and M.J. Lercher. 2004. The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet*. 5:299–310.

23. Reik, W., and J. Walter. 2001. Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet*. 2:21–32.

24. DeChiara, T.M., E.J. Robertson, and A. Efstratiadis. 1991. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell*. 64:849–859.

25. Hemberger, M., C. Redies, R. Krause, J. Oswald, J. Walter, and R.H. Fundele. 1998. H19 and Igf2 are expressed and differentially imprinted in neuroectoderm-derived cells in the mouse brain. *Dev. Genes Evol*. 208:393–402.

26. Charalambous, M., T.R. Menheniott, W.R. Bennett, S.M. Kelly, G. Dell, L. Dandolo, and A. Ward. 2004. An enhancer element at the Igf2/H19 locus drives gene expression in both imprinted and non-imprinted tissues. *Dev. Biol*. 271:488–497.

27. Farr, A.G., J.L. Dooley, and M. Erickson. 2002. Organization of thymic medullary epithelial heterogeneity: implications for mechanisms of epithelial differentiation. *Immunol. Rev*. 189:20–27.

28. 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.

29. 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.

30. Sillanpaa, N., C.G. Magureanu, A. Murumagi, A. Reinikainen, A. West, A. Manninen, M. Lahti, A. Ranki, K. Saksela, K. Krohn, et al. 2004. Autoimmune regulator induced changes in the gene expression profile of human monocyte-dendritic cell-lineage. *Mol. Immunol*. 41:1185–1198.

31. Miyamoto, T., H. Iwasaki, B. Reizis, M. Ye, T. Graf, I.L. Weissman, and K. Akashi. 2002. Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. *Dev. Cell*. 3:137–147.

32. Zipori, D. 2004. The nature of stem cells: state rather than entity. *Nat. Rev. Genet*. 5:873–878.

33. Jackson-Grusby, L., C. Beard, R. Possemato, M. Tudor, D. Fambrough, G. Csankovszki, J. Dausman, P. Lee, C. Wilson, E. Lander, and R. Jaenisch. 2001. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat. Genet*. 27:31–39.

34. Nil, A., E. Firat, V. Sobek, K. Eichmann, and G. Niedermann. 2004. Expression of housekeeping and immunoproteasome subunit genes is differentially regulated in positively and negatively selecting thymic stroma subsets. *Eur. J. Immunol*. 34:2681–2689.

35. Lorenz, R.G., and P.M. Allen. 1989. Thymic cortical epithelial cells lack full capacity for antigen presentation. *Nature*. 340:557–559.

36. Yoneyama, H., S. Narumi, Y. Zhang, M. Murai, M. Baggiolini, A. Lanzavecchia, T. Ichida, H. Asakura, and K. Matsushima. 2002. Pivotal role of dendritic cell–derived CXCL10 in the retention of T helper cell 1 lymphocytes in secondary lymph nodes. *J. Exp. Med*. 195:1257–1266.

37. Langbein, L., U.F. Pape, C. Grund, C. Kuhn, S. Praetzel, I. Moll, R. Moll, and W.W. Franke. 2003. Tight junction-related structures in the absence of a lumen: occludin, claudins and tight junction plaque proteins in densely packed cell formations of stratified epithelia and squamous cell carcinomas. *Eur. J. Cell Biol*. 82:385–400.

38. Cabral, A., P. Voskamp, A.M. Cleton-Jansen, A. South, D. Nizetic, and C. Backendorf. 2001. Structural organization and regulation of the small proline-rich family of cornified envelope precursors suggest a role in adaptive barrier function. *J. Biol. Chem*. 276:19231–19237.

39. Ramsey, C., O. Winqvist, L. Puhakka, M. Halonen, A. Moro, O. Kampe, P. Eskelin, M. Pelto-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.

40. Kenzelmann, M., R. Klaren, M. Hergenhahn, M. Bonrouhi, H.J. Grone, W. Schmid, and G. Schutz. 2004. High-accuracy amplification of nanogram total RNA amounts for gene profiling. *Genomics*. 83:550–558.

41. Su, A.I., T. Wiltshire, S. Batalov, H. Lapp, K.A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, et al. 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA*. 101:6062–6067.

42. 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.

43. Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser. B*. 57:289–300.

44. Eisen, M.B., P.T. Spellman, P.O. Brown, and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*. 95:14863–14868.

45. Pollard, K.S., S. Dudoit, and M.J. van der Laan. 2004. Multiple testing procedures: R multtest package and applications to genomics. *U.C. Berkeley Division of Biostatistics Working Paper Series*. Working Paper 164. http://www.bepress.com/ucbiostat/paper/164.