The variable immunological self: Genetic variation and nongenetic noise in Aire-regulated transcription

Emily S. Venanzi, Rachel Melamed, Diane Mathis*, and Christophe Benoist*

Section on Immunology and Immunogenetics, Joslin Diabetes Center; Department of Medicine, Brigham and Women’s Hospital; and Harvard Medical School, Boston, MA 02215

Contributed by Christophe Benoist, August 18, 2008 (sent for review July 18, 2008)

The Aire transcription factor plays an important role in immunological self-tolerance by mediating the ectopic expression of peripheral self-antigens by thymic medullary epithelial cells (MECs), and the deletion of thymocytes that recognize them. In Aire-deficient humans or mice, central tolerance is incomplete and multiorgan autoimmune disease results. We examined the variability of Aire’s effects on ectopic transcription among individual mice of three different inbred strains. Aire’s function was, overall, quite similar in the three backgrounds, although generally stronger in C57BL/6 than in BALB/c or NOD mice, and a minority of Aire-regulated genes did show clear differences. Gene expression profiling of wild-type MECs from single mice, or from the two thymic lobes of the same mouse, revealed significantly greater variability in Aire-controlled ectopic gene expression than in Aire-independent transcripts. This “noisy” ectopic expression did not result from parental or early developmental imprinting, but from programming occurring after the formation of the thymic anlage, resulting from epigenetic effects or from the stochastic nature of Aire activity. Together, genetic and nongenetic variability in ectopic expression of peripheral antigens in the thymus make for differences in the portion of self determinants presented for tolerance induction. This variable self may be beneficial in preventing uniform holes in the T-cell repertoire in individuals of a species, but at the cost of variable susceptibility to autoimmunity.

Imposition of self-tolerance on the T-cell repertoire entails the exposure of differentiating thymocytes to a large variety of self-antigens. For self-antigens characterized by expression that is normally restricted to one or a few organs (peripheral tissue antigens [PTAs]), exposure takes place through their expression by epithelial cells of the thymus. A segment of this ectopic gene expression is under the control of the Aire transcriptional regulator, a segment of great import because the absence of Aire leads to multiorgan autoimmunity in both knockout (KO) mice and human patients. Aire is expressed primarily by thymic medullary epithelial cells (MECs) (1–4). The connection between Aire and PTA expression was formally demonstrated by analysis of the global gene-expression profiles of MECs isolated from Aire-KO mice. Relative to those of wild-type (WT) controls, Aire-deficient MECs displayed absent or reduced expression of mRNAs encoding such PTAs as insulin, salivary gland proteins, and caseins (4, 5).

Aire-KO mice exhibit an autoimmune disease similar to human autoimmune-polyendocrinopathy-candidiasis-ectodermal-dystrophy (APECED), characterized by lymphocytic infiltrates of, and autoantibodies against, numerous peripheral tissues (4, 6). APECED disease manifestations vary widely from patient to patient, with even siblings exhibiting disparate disease profiles (7). In mice, the situation seems similar, in that the original KO strain also displayed heterogeneity of disease severity and target tissues (4, 6). Part of this variability may be due to environmental influences, although a recent study uncovered no evidence of a significant role for innate immune system stimulation (8). Another part of the variability is undoubtedly of genetic origin. For instance, Aire-deficient mice on the NOD genetic background developed an extreme exocrine pancreatitis that was not found on the C57BL/6 (B6) or BALB/c backgrounds. NOD and BALB/c Aire-KOs also developed severe autoimmune gastritis, which was observed only rarely on the B6 background. The modifier genes controlling these disease phenotypes were traced to several genomic regions, different for each target organ, including both MHC and non-MHC elements (9). In human patients, there is also evidence that APECED disease manifestations are linked to HLA genotypes (10). Some of the disease heterogeneity is also likely to be of epigenetic or stochastic origin. For example, in mice, there is substantial individual variation even among animals of the same inbred strain housed together (9).

There are many possible ways in which genetics may influence the disease caused by Aire deficiency. One interesting possibility is that PTA expression by MECs is somehow influenced by the genetic background. Experiments in several murine models have demonstrated that genetically regulated thymic expression levels of antigens such as insulin, myelin P0 and interphotoreceptor retinoid-binding protein correlate with tolerance to each particular antigen (11–15). Similarly, in humans, polymorphisms in the promoter region of the insulin gene that correlate with susceptibility to type 1 diabetes also correlate with levels of insulin gene transcripts in the thymus (16, 17).

Taubert et al. recently reported that expression levels of Aire and PTAs in MECs varied widely from person to person. For insulin and several other PTAs, mRNA levels correlated closely with those of Aire itself (18). Given the outbred nature of the human population, it is impossible to know whether these variations reflect genetic, epigenetic, or environmental differences between individuals.

Here, we tackled this question in mice, by examining Aire’s impact on ectopic gene expression in MECs of individual mice from several inbred strains, allowing us to parse out the fraction of variation due to genetic effects from that due to interindividual variability.

Results

Gene-Expression Profiling. We chose to profile MEC gene expression in three strains wherein the phenotypic effects of the Aire-KO were quite different. B6 can be considered a reference strain, in which the autoimmune manifestations provoked by the Aire KO mutation were quite mild, dominated by retinal degeneration (9, 15). In contrast, the disease induced by Aire deficiency on the NOD/Lt background was severe, even lethal, by 10–12 weeks of age, and resulted in much higher titers of
autoantibodies as well as destruction of the exocrine pancreas and lung (9, 19). Aire KO-BALB/c mice were again different, exhibiting a florid stomach disease (9, 19). All mice examined were males of 3–4 weeks of age, before the onset of any extensive pathology. MECs were sorted from thymi of individual mice (~20,000 MECs per mouse; 3 to 4 independent KO/wild-type littermate pairs for each strain); and RNA was prepared, amplified, labeled and hybridized to Affymetrix Mouse Genome 430 2.0 microarrays, representing ~30,000 genes. Data (NCBI GEO accession no. GSE8564.) were processed and normalized in the GenePattern suite using the RMA algorithm (20, 21). For interstrain comparisons, mean expression values were calculated by averaging the replicates within each strain.

**Similar Overall Patterns of Aire-Controlled Gene Expression in MECs of Different Mouse Strains.** Overall, the effects of Aire on gene expression in MECs of the various strains were quite similar (Fig. 1). Comparing WT and KO expression values in MECs of each strain showed many genes to exhibit higher expression in WT MECs, i.e., were Aire activated), whereas fewer had elevated levels of expression in Aire-deficient MECs (Aire repressed). The probes highlighted in Fig. 1 are those commonly regulated by Aire, with a WT/KO fold-change (FC) ratio greater than 2 (red) or less than 0.5 (green) in all three strains, demonstrating substantial overlap among them. A “consensus signature” of changes shared between all three strains accounted for 96.8%, 96.7%, and 99.4% of alterations on the B6, NOD and BALB/c backgrounds, respectively (for an arbitrary FC of >3 in the primary strain, and of >1.5 in the other two strains). Direct comparison of Aire’s effects in the three strains by plotting the WT/KO FCs against each other supported this notion (Fig. 2A). The overall diagonal pattern of the probes in each strain-to-strain comparison confirmed that most of the same genes were regulated by Aire in all three strains.

**Subtle Differences in Aire’s Regulation of MEC Gene Expression in the Three Strains.** Closer examination of Figs. 1 and 2 revealed that even though the bulk of gene regulation by Aire was the same on the B6, BALB/c, and NOD backgrounds, there were also rare but clear differences. First, there was a minority of genes whose regulation by Aire was specific to, or specifically missing in, one strain. These genes fall along the FC = 1 horizontal or vertical dotted lines of Fig. 2A, and the most specific examples are listed in Table 1; a more comprehensive list of strain-specific probes can be found in supporting information (SI) Table S1. Genes with this characteristic were quite rare overall: for instance, only 53 probes were uniquely activated by Aire in B6 but not BALB/c or NOD mice (using FC thresholds of 2 and 1.25). We queried genomic and gene ontology databases to see whether these strain-specific genes might be connected to the strain-specific autoimmune manifestations of Aire deficiency (9). However, we failed to uncover any clear patterns: they were distributed over all of the chromosomes, were expressed peripherally in a broad spectrum of tissues, and were a mix of intracellular, secreted and membrane proteins. In particular, no gene specific to the exocrine pancreas was found in the NOD-only list; nor was any stomach-specific gene found in the BALB/c-only list.

Second, for the majority of genes affected in all strains, Aire’s impact was not equivalent in the different strains. This feature was evidenced by diversion from the x = y diagonal of the cloud of genes in the comparisons in Fig. 2A, with a greater degree of Aire-related changes in B6 than in the two other strains. Fitting by local regression indicated an average slope of 0.54 and 0.56 in the NOD/B6 and BALB/c/B6 plots, respectively. This trend was confirmed by calculating, for all genes affected by Aire, the ratio of FCs (calculated for simplicity by subtracting the log2-transform of the FCs). The histograms of Fig. 2B indicate that this feature was indeed a general trend, globally affecting Aire-responsive genes: the bulk of Aire-activated genes were more strongly up-regulated in B6 than in NOD or BALB/c mice,
and repressed genes were also more strongly down-regulated in B6 than in the other strains (mean negative value for the difference in FCs in the latter case). In general, and as might be expected, the stronger effect of Aire in B6 MECs was reflected in higher expression values for Aire-activated genes in B6 than in NOD or BALB/c MECs. Of the 1586 genes induced more than twofold by Aire in at least one strain, 1206 were expressed at a higher level in B6 than in NOD MECs, and 1055 were expressed at a higher level in B6 than in BALB/c MECs. This difference was no longer apparent when the KOs were compared (data not shown), substantiating the conclusion that Aire had strong activating influences on the B6 background.

The overall differences in the strength of Aire’s effect in the various strains could be explained quite simply by different levels of Aire expression, or by different proportions of Aire-expressing cells among the sorted MHC class II	extsuperscript{hi} MECs. To determine whether the strength of Aire regulation correlated with Aire protein levels, we performed flow cytometric analysis of gated CD45	extsuperscript{Ly51intMHC-II	extsuperscript{hi}} cells; representative of three independent mice for each inbred strain. Numbers are the mean percentage intensity, standard deviation, of Aire staining.

### Table 1. Genes regulated by Aire uniquely in one strain

<table>
<thead>
<tr>
<th>Strain Pref.</th>
<th>WT</th>
<th>KO</th>
<th>WT</th>
<th>KO</th>
<th>NOD</th>
<th>KO</th>
<th>WT/KO Ratio</th>
<th>Expression Value</th>
<th>Symbol</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>B6</td>
<td>624</td>
<td>27</td>
<td>62</td>
<td>26</td>
<td>67</td>
<td>24</td>
<td>23.1</td>
<td>5115 ± 55</td>
<td>Sbp</td>
<td>1418307_s_at</td>
</tr>
<tr>
<td>B6</td>
<td>863</td>
<td>58</td>
<td>201</td>
<td>92</td>
<td>219</td>
<td>109</td>
<td>14.93</td>
<td>2.19 ± 2.02</td>
<td>Gstm</td>
<td>1423569_at</td>
</tr>
<tr>
<td>B6</td>
<td>1035</td>
<td>124</td>
<td>72</td>
<td>42</td>
<td>55</td>
<td>38</td>
<td>8.33</td>
<td>1.72 ± 1.43</td>
<td>Ndg1</td>
<td>1455423_at</td>
</tr>
<tr>
<td>B6</td>
<td>732</td>
<td>54</td>
<td>595</td>
<td>238</td>
<td>507</td>
<td>908</td>
<td>13.7</td>
<td>2.5 ± 0.56</td>
<td>Ear11</td>
<td>1425295_at</td>
</tr>
<tr>
<td>B6</td>
<td>645</td>
<td>116</td>
<td>152</td>
<td>123</td>
<td>145</td>
<td>123</td>
<td>5.96</td>
<td>1.24 ± 1.18</td>
<td>Ocpp3</td>
<td>1447456_s_at</td>
</tr>
<tr>
<td>B6</td>
<td>486</td>
<td>98</td>
<td>1294</td>
<td>1077</td>
<td>373</td>
<td>348</td>
<td>4.74</td>
<td>1.07 ± 0.97</td>
<td>Gabrp</td>
<td>1451424_at</td>
</tr>
<tr>
<td>B6</td>
<td>405</td>
<td>118</td>
<td>49</td>
<td>56</td>
<td>56</td>
<td>56</td>
<td>3.44</td>
<td>0.87 ± 0.86</td>
<td>Mtnr7</td>
<td>1447831_s_at</td>
</tr>
<tr>
<td>B6</td>
<td>587</td>
<td>64</td>
<td>580</td>
<td>295</td>
<td>1565</td>
<td>648</td>
<td>9.09</td>
<td>1.96 ± 2.42</td>
<td>Ms46d</td>
<td>1419599_s_at</td>
</tr>
<tr>
<td>B6</td>
<td>1438</td>
<td>669</td>
<td>665</td>
<td>85</td>
<td>671</td>
<td>308</td>
<td>2.15</td>
<td>7.87 ± 2.18</td>
<td>Chi33</td>
<td>1419764_at</td>
</tr>
<tr>
<td>B6</td>
<td>967</td>
<td>469</td>
<td>581</td>
<td>383</td>
<td>315</td>
<td>111</td>
<td>0.95</td>
<td>1.16 ± 4.85</td>
<td>Aplh</td>
<td>1446777_at</td>
</tr>
<tr>
<td>B6</td>
<td>228</td>
<td>234</td>
<td>291</td>
<td>250</td>
<td>541</td>
<td>201</td>
<td>0.95</td>
<td>0.35 ± 3.37</td>
<td>Stclo4c1</td>
<td>1460016_at</td>
</tr>
<tr>
<td>NOD</td>
<td>123</td>
<td>129</td>
<td>201</td>
<td>572</td>
<td>507</td>
<td>908</td>
<td>13.7</td>
<td>2.5 ± 0.56</td>
<td>Ear11</td>
<td>1425295_at</td>
</tr>
<tr>
<td>BALB/c</td>
<td>44</td>
<td>20</td>
<td>115</td>
<td>14</td>
<td>139</td>
<td>14</td>
<td>2.2</td>
<td>8.4 ± 9.71</td>
<td>Defb5</td>
<td>1421808_at</td>
</tr>
<tr>
<td>not B6</td>
<td>4239</td>
<td>712</td>
<td>899</td>
<td>688</td>
<td>3017</td>
<td>43</td>
<td>24.3</td>
<td>1.29 ± 1.29</td>
<td>Mup1/2</td>
<td>1204655_s_at</td>
</tr>
<tr>
<td>not BALB/c</td>
<td>102</td>
<td>11</td>
<td>17</td>
<td>11</td>
<td>151</td>
<td>11</td>
<td>9.4</td>
<td>1.51 ± 13.7</td>
<td>Sper1-ps1</td>
<td>1452794_x_at</td>
</tr>
<tr>
<td>not BALB/c</td>
<td>877</td>
<td>209</td>
<td>358</td>
<td>427</td>
<td>242</td>
<td>43</td>
<td>4.2</td>
<td>0.84 ± 5.68</td>
<td>Fabp1</td>
<td>1446764_at</td>
</tr>
<tr>
<td>not BALB/c</td>
<td>771</td>
<td>225</td>
<td>429</td>
<td>493</td>
<td>383</td>
<td>71</td>
<td>3.42</td>
<td>0.87 ± 5.41</td>
<td>Fabp1</td>
<td>1417556_at</td>
</tr>
<tr>
<td>not BALB/c</td>
<td>1632</td>
<td>252</td>
<td>541</td>
<td>323</td>
<td>2090</td>
<td>324</td>
<td>6.49</td>
<td>1.66 ± 6.45</td>
<td>Gstm3</td>
<td>1427474_s_at</td>
</tr>
<tr>
<td>not BALB/c</td>
<td>759</td>
<td>123</td>
<td>419</td>
<td>254</td>
<td>415</td>
<td>67</td>
<td>6.21</td>
<td>1.85 ± 6.25</td>
<td>Apoa1</td>
<td>1438840_x_at</td>
</tr>
<tr>
<td>not BALB/c</td>
<td>1830</td>
<td>417</td>
<td>3737</td>
<td>3536</td>
<td>1437</td>
<td>363</td>
<td>4.39</td>
<td>1.08 ± 3.82</td>
<td>Lor</td>
<td>1448745_s_at</td>
</tr>
<tr>
<td>not NOD</td>
<td>481</td>
<td>36</td>
<td>235</td>
<td>26</td>
<td>11</td>
<td>13</td>
<td>13.51</td>
<td>8.93 ± 0.83</td>
<td>Ear11</td>
<td>1421802_at</td>
</tr>
<tr>
<td>not NOD</td>
<td>974</td>
<td>165</td>
<td>1009</td>
<td>278</td>
<td>1599</td>
<td>1910</td>
<td>5.92</td>
<td>3.62 ± 0.84</td>
<td>Fabp6</td>
<td>1450682_at</td>
</tr>
<tr>
<td>not NOD</td>
<td>157</td>
<td>23</td>
<td>139</td>
<td>19</td>
<td>34</td>
<td>22</td>
<td>6.71</td>
<td>7.3 ± 1.56</td>
<td>Defb19</td>
<td>1456076_at</td>
</tr>
<tr>
<td>not NOD</td>
<td>756</td>
<td>292</td>
<td>564</td>
<td>233</td>
<td>1927</td>
<td>2862</td>
<td>2.59</td>
<td>2.42 ± 0.67</td>
<td>Lect1</td>
<td>1460258_at</td>
</tr>
<tr>
<td>not NOD</td>
<td>734</td>
<td>194</td>
<td>542</td>
<td>172</td>
<td>504</td>
<td>1026</td>
<td>4.78</td>
<td>3.14 ± 0.88</td>
<td>Npl</td>
<td>1424265_at</td>
</tr>
</tbody>
</table>

**Fig. 3.** Aire protein levels do not vary among strains. Flow cytometric analysis of Aire expression in MECs of different genetic backgrounds (Aire vs MHC-II	extsuperscript{hi} profiles of gated CD45	extsuperscript{Ly51intMHC-II	extsuperscript{hi}} MECs; representative of three independent mice for each inbred strain). Numbers are the mean percentage of Aire	extsuperscript{+} cells among class II	extsuperscript{hi} MECs, and the mean anti-Aire fluorescence intensity, ± SD, of Aire staining.
distribution of CVs for Aire-affected genes to those of a randomly-selected set of Aire-neutral genes with a matched distribution of expression values.

We then asked whether the increased variability in Aire-induced genes was an inherent property of those particular genes or, rather, was caused by Aire itself. Comparing the CVs in MECs from Aire-WT and -KO thymi (Fig. 4A, compiled in Table 2) showed a clear drop in the distribution of CVs for Aire-activated genes in KO thymi, although not quite to the same level as the Aire-neutral gene set. Thus, the residual expression of these genes was less variable than when Aire was present, i.e., Aire was directly contributing to the variability.

If so, one might predict that the variability in Aire-activated genes would be a specific feature of MECs, and would not be found in peripheral tissues where these genes are normally expressed. Thus we analyzed the variability of the same set of genes in another tissue, the liver, by downloading from NCBI GEO repository the microarray datasets reported by Ackert-Bicknell et al. (23), generated from liver samples of three individual B6 mice (accession #GSE5959). In the liver, there was no difference in mean CV between Aire-activated genes (0.076) and Aire-neutral genes (0.069; Fig. 4C and Table 1).

Do the higher mean CV observed for Aire-activated genes translate into meaningful interindividual differences in expression values of individual genes? Indeed, many Aire-activated genes with high CVs exhibited broad distributions of expression value, with ranges as large as 10- to 20-fold, as illustrated for a few such genes in Fig. 4D. This wide range of expression was found in mice of all three strains. Thus, the particular variability in Aire-regulated transcripts in MECs leads to quantitative variations that are likely to have a direct impact on tolerance induction.

### Origin of the Interindividual Variability in Aire-Regulated Genes

Several hypotheses could account for the variability of ectopic expression for Aire-regulated genes and for the stochastic pat-
terms observed in individual mice: 1) transmissible epigenetic programming, imprinted at the parental level; 2) imprinting at an early stage of embryonic development (e.g., variable patterns of DNA methylation in the first few days of development); 3) epigenetic patterns set during the differentiation of the thymic epithelial cells, around embryonic day 11, after lateralization of the body plan; 4) a continuous reshuffling of ectopic expression patterns, each epithelial precursor or clone thereof adopting a defined pattern. The thymus of an adult mouse is composed of two lobes, whose stroma develops independently from endodermal tissue of pharyngeal pouches, around embryonic day 11 (24).

The hypotheses outlined above lead to testable predictions concerning ectopic expression in the two thymic lobes of a mouse: according to #1 and #2, both lobes should have related profiles, but for #3 and #4 both lobes should behave independently, and be no more similar that lobes from different mice. We thus generated a new set of expression profiles, where MECs were purified independently from the right and left lobes of four wild-type B6 mice, and each (A) Histogram of genewise CVs across all datasets, for Aire-induced genes (FC > 2.5) or all other genes on the array. (B) Relative Euclidian distances were calculated between each individual lobe, integrating expression values for all Aire-induced genes.

Discussion

These data indicate that the transcriptional programs regulated by Aire in MECs of a given individual are conditioned by genetic factors and by non-genetic “noise”: genetic polymorphism affects the levels and fine specificity of ectopic transcripts, and interindividual variability in these ectopic transcripts further shuffles the outcome.

These results are of interest in light of the work of Taubert et al., who described extensive variability in Aire and ectopic transcription in MECs from different humans (18). Our data would suggest that both genetic and nongenetic elements are involved. We did not reproduce, in the mouse, the very wide range of transcript abundance for Aire (up to 50-fold), the variations we observed being much more modest. The range of variability may be different between species. Also, the interpretation of the human data may be complicated by variation in the proportion of Aire+ cells in the samples of human thymus.

In this setting, it was of interest to estimate the relative influence of these factors, using the CV as a metric for variation. Aire-regulated genes had a mean CV of 0.40 when all interindividual and genetic variation was encompassed; interindividual variability alone gave a CV of 0.25 (average of the interindividual CVs within each strain); the genetic component, estimated by calculating a CV between the expression means for the three strains, scored 0.33. For a random sample of Aire-neutral genes with a comparable spread of expression values, the three CVs were 0.18, 0.14 and 0.12, respectively. Taking into account an estimated experimental CV of 0.13, we estimated that the variance in Aire-regulated transcription between any two Mus musculus individuals was approximately two-thirds genetic and one-third nongenetic in origin.

Although the genes regulated by Aire were mostly similar in different strains, Aire-activated genes were expressed at a higher level in B6 MECs than in MECs of BALB/c and NOD animals, which pointed to an impact of the genetic background in modulating Aire’s function as a transcriptional inducer. There are several ways through which this control may occur. We have already ruled out one possibility, that Aire protein levels are intrinsically higher in B6 MECs than in the other strains. Secondly, examination of the Perlegen sequence database turned up no differences in the Aire coding regions between the three strains. It is possible that the levels or functional activities of one or more of Aire’s partner factors are higher or stronger in B6 MECs, or that proteins that inhibit Aire’s function are less effective on the B6 background. Transcription may also differ at a more basic level, with widespread alterations in the accessibility to chromatin remodeling in different inbred strains.

We did not find, within the rare instances of strain-specific (or strain-preferential) Aire control, explanations for the particular tropisms of autoimmunity in the different KOs (9, 25). For example, there was no evidence for effects of the Aire KO particularly on genes active in the exocrine pancreas for NOD mice, nor on genes active in the stomach for BALB/c animals. It seems likely, then, that tissue tropism stems from the combination of MHC restriction and tissue-specific factors. One might speculate, however, that the more robust effect of Aire on transcription profiles observed in B6 thymi might contribute to the general autoimmune-resistant nature of this strain.

In addition to the strain-to-strain differences in Aire’s function, we found marked variability in the expression of Aire-activated genes from one individual to another within the same strain, and between the two thymic lobes of the same mouse. This interindividual variability significantly distinguished Aire-activated genes from other Aire-neutral genes in the genome. This augmentation was specific to MECs, as it was not observed in the liver for the same set of genes. Neither did we observe it with a different transcription factor also involved in controlling autoimmunity, Foxp3. The results of the dual lobe profiling show that the variability is not set by parental or early embryonic imprinting, but is set independently in each thymic lobe. This variability may reflect stochastic mechanisms involved in ectopic transcription in MECs. We and the Kyewski group have recently observed that Aire-controlled expression of ectopically expressed genes in a single MEC is inherently probabilistic, any given cell transcribing a subset of PTA genes, following genespecific probabilities (26 and accompanying paper). One would have assumed that the integration of these probabilities over the 15,000 Aire+ MECs of a thymic lobe would smooth out stochastic effects, resulting in uniform patterns of ectopic expression. This proved not be the case, and the stochastic aspect of ectopic
expression at the single cell level somehow extends to the single organ level. This lack of smoothing and persistent stochasticism in a large cell pool suggests that the MECs present in a thymic lobe in a given time must descend from a small number of precursors, in which the accessibility to Aire of a given locus is determined in a stable and inheritable manner, before MEC clonal expansion and Aire expression. Such a view is consistent with current notions of population dynamics of thymic epithelial cells (8, 27, 28) where Aire+ MECs appear derived from a limited number of precursors, after a phase of cell division and expansion.

Noise in gene expression is an unavoidable consequence of the size and complexity of regulatory pathways and of the low number of molecules involved (29, 30). The immune system may be exploiting this noise to increase individuality in the tolerance to self within a species. Since tolerance to self-antigens can also lead to “holes” in the repertoire of responses to nonself, there would be a distinct advantage, at the population level, in some diversity of the molecular drive to self-tolerance. Such variability would ensure a greater diversity of potential responses within the species. In this respect, interindividual variability in ectopic PTA expression may function as does the wide diversity of MHC alleles, helping to ensure that at least some of the individuals will be apt to survive novel pathogenic challenges. Of course, varying susceptibility to autoimmune deviation is the price to pay for this diversity in the education to self, and the phenomenon may contribute to the proportion of nongenetic determinism of diversity in the education to self, and the phenomenon may contribute to the proportion of nongenetic determinism of autoimmune diseases. Susceptibility to autoimmune diseases has a strong genetic component, but one whose influence is modulated by nongenetic factors (illustrated by the frequent discordance of diabetes or multiple sclerosis in monzygotic twins). The present results suggest that epigenetic modulation of the degree and specificity of self-tolerance may account, at least in part, for this individual re-pegging of disease risk.

Materials and Methods

Mice. Aire-deficient mice (4) used in this study were backcrossed to the B6, NOD and BALB/c strains for at least nine generations before intercrossing. For each knockout line, heterozygotes were bred to each other, and resulting KO and WT littermate control animals were used, unless otherwise noted. All experiments were approved by the HMS IACUC (protocol 2954).

Thymus Digestion and MEC Sorting. Thymic epithelial cells for microarray analysis were prepared according to (4). Individual thymi from young adult (3–4.5- week-old) animals were digested, separated by density-gradient cen-

trifugation, and the thymic epithelial cell-enriched fraction was harvested, stained and sorted as CD45- G8.8. Ly51+Thy1.1+Ab3 (21), RNA was prepared from sorted cells using TRIzol.

RNA Amplification and Microarray Hybridization. Total MEC RNA was amplified using a T7 polymerase-based method. First- and second-strand cDNA synthesis, in vitro transcription of complementary RNA, and a second round of first- and second-strand cDNA synthesis were performed using the MessageAmp aRNA Kit (Ambion, Austin, TX). The resulting cDNA was used in a second round of in vitro transcription, performed using the BioArray High Yield RNA Transcription Kit (Enzo, New York), and labeled using biotinylated ribonuclease fragments of the cRNA. The labeled cRNA was purified, fragmented (94° for 35 min), and hybridized to Affymetrix M430v2 chips.

Microarray Data Analysis. The microarray data are available at GEO under accession no. GSE5856. The raw probe-level chip data (CEL files) were normalized by the robust multiarray average (RMA) algorithm (21) using the ExpressionFileCreator module of the GenePattern 2.0 software package (20), and analyzed using the Multipoint module in GenePattern. Random datasets were generated from a Gaussian distribution extrapolated from log-transformed values for each gene. This distribution was centered on the genewise mean across all samples, with a standard deviation that was the mean of the per-class standard deviations for that gene. The false discovery rate was extrapolated from the frequency at which a given FC value is equal or higher in the random data, for FC increments of 0.1. Genewise Pvalues were calculated using Welch’s modified t test. Calculations of genewise CVs on filtered gene sets were performed in Insightful.

ACKNOWLEDGMENTS. We thank Vincent Auyeung for the splenic Treg microarray data and Dr. Hamish Scott for the anti-Aire antibody. This work was supported by a National Institutes of Health Grant RO1 DK60027, Young Chair funds (to D.M. and C.B.), Joslin Diabetes Center’s NIDDK-funded DERC core facilities, and National Institutes of Health Training Grant T32 DK07260 (to E.V.).