Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

| As Published | http://dx.doi.org/10.1073/pnas.1119351109 |
| Publisher | National Academy of Sciences |
| Version | Final published version |
| Accessed | Mon Apr 25 01:35:17 EDT 2016 |
| Citable Link | http://hdl.handle.net/1721.1/72512 |
| Terms of Use | Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use. |
| Detailed Terms | |

DSpace@MIT

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells

Matthieu Giraud1,2, Hideyuki Yoshida3, Jakub Abramson1,2, Peter B. Rahl4, Richard A. Young4,5, Diane Mathis6,7, and Christophe Benoist6,8

1Division of Immunology, Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115; and 2Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Christophe Benoist, November 28, 2011 (sent for review October 31, 2011)

Aire is a transcriptional regulator that induces expression of peripheral tissue antigens (PTA) in thymic medullary epithelial cells (MECs), driving immunological self-tolerance in differentiating T cells. To elucidate its mechanistic pathways, we examined its transcriptional impact in MECs in vivo by microarray analysis with mRNA-spanning probes. This analysis revealed initiation of Aire-activated genes to be comparable in Aire-deficient and wild-type MECs, but with a block to elongation after 50–100 bp in the absence of Aire, suggesting activation by release of stalled polymerases by Aire. In contrast, patterns of activation by transcription factors such as Klf4 were consistent with regulation of initiation. Mapping of Aire and RNA polymerase-II (Pol-II) by ChIP and high-throughput sequencing (ChIP-seq) revealed that Aire bound all Pol-II–rich transcriptional start sites (TSS), irrespective of its eventual effect. However, the genes it preferentially activated were characterized by a relative surfeit of stalled polymerases at the TSS, which resolved once Aire was introduced into cells. Thus, transcript mapping and ChIP-seq data indicate that Aire activates ectopic transcription not through specific recognition of PTA gene promoters but by releasing stalled polymerases.

The capacity of T lymphocytes to react to a virtually infinite spectrum of foreign antigens carries a corollary potential reactivity to self-constituents and, consequently, autoimmunity. Self-reactive T cells are eliminated during maturation in the thymus, requiring that gene products normally expressed only by parenchymal cells of specialized organs be made visible to differentiating thymocytes. The solution to this conundrum seems to be, at least in part, the specialized ability of thymic stromal cells, in particular medullary epithelial cells (MECs), to transcribe a large repertoire of genes encoding these peripheral tissue antigens (PTAs) (1). This ectopic gene expression is controlled in a large part by the Aire transcription factor, which is expressed almost exclusively in highly differentiated MECs. Mice with an Aire gene defect express only a fraction of the PTA repertoire (3). Aire affects the expression of thousands of genes in MECs, which are regulated by very different pathways in peripheral parenchymal tissues (4), with a strong stochastic element as to which individual cell and which of the two chromosomes actually expresses a given PTA (5, 6). This regulator does not have a clear DNA binding motif, and its transcriptional footprint is highly dependent on the cell type in which it is expressed (7, 8). Aire binds chromatin through an interaction between its PHD1 domain and the amino-terminal tail of histone H3, but only in its hypomethylated state, a mark associated with poorly transcribed genes (9–12).

The precise molecular mechanisms that Aire uses to regulate transcription remain elusive. A large-scale screen based on mass spectrometry of communoprecipitated proteins revealed Aire’s connection to factors involved in a number of nuclear processes: chromatin structure/modification, transcriptional elongation, premRNA processing, and nuclear transport (8). A systematic and large-scale RNAi screen of Aire’s transcriptional allies identified a number of previously unrecognized elements of the Aire pathway, with a preponderance of factors involved in transcriptional elongation rather than initiation, consistent with work showing an effect of Aire on elongation in transfection experiments, and an interaction of Aire with P-TEFb (positive transcription elongation factor b) (13).

Elongation factors might be needed for processive elongation of Pol-II through Aire target genes, or because of an effect of Aire on promoter-proximal Pol-II stalling. Also referred to as poised-polymerase, paused-polymerase, abortive elongation premature termination, this mode of transcriptional regulation was initially thought to apply to special loci with very fast responses to inducers (14, 15) but was more recently recognized to be widespread (16–18): initiation occurs constitutively, Pol-II proceeds for 40–50 bp, but is then blocked by the action of dominant pause factors such as DSIF (DRB sensitivity-inducing factor) and NELF (negative elongation factor) (19, 20). Release from this block, which allows Pol-II to proceed along the gene, confers tissue specificity or inducer specificity (21–23). Gene-specific transcription factors can relieve the block by seeding the recruitment or activation of the pause-release factor P-TEFb (24), whose kinase activity phosphorylates DSIF and NELF, dislodging NELF and converting DSIF into an activator (25–27), and phosphorylates Ser2 on the C-terminal domain of Pol-II. The overall result is the unleashing of Pol-II and a number of the other preexpression activities, including H3K4 methylation and assembly of splicing factors (28, 29).

Here, we analyzed the genome-wide impact of Aire on elongation, mapping its chromosomal interaction sites and testing, in primary MECs ex vivo, the hypothesis that Aire affects Pol-II complexes nonproductively stalled at the promoter of its target genes.

Results
Pol-II stalled at the transcriptional start sites (TSS) results in the accumulation of short RNAs (30–33). Because MECs are far too rare for direct analysis of transcriptional elongation, we sought to identify the hallmarks of elongation control through microarray
analysis, with expression profiles generated from sorted MECs of Aire-knockout (KO) and WT mice on the C57BL/6 (B6) × NOD (GEO GSE33878). These experiments used Affymetrix MoGene ST1.0 microarrays, which include many short probes (“features”) spread across the whole length of the mRNA, with several probes for each exon. We reasoned that analysis of the array data

<table>
<thead>
<tr>
<th>Interval start (bp from TSS)</th>
<th>WT/KO FoldChange within intervals, Aire-induced genes</th>
<th>WT/KO FoldChange within intervals, Aire-neutral genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10.0</td>
<td>5.0</td>
</tr>
<tr>
<td>500</td>
<td>5.0</td>
<td>2.5</td>
</tr>
<tr>
<td>1000</td>
<td>2.5</td>
<td>1.0</td>
</tr>
<tr>
<td>1500</td>
<td>1.0</td>
<td>0.5</td>
</tr>
</tbody>
</table>

**Fig. 1.** “Feature-level” analysis of microarray data identifies a role for Aire in releasing paused Pol-II. (A) Expression values for individual microarray features were analyzed relative to whole-transcript averages, and “imbalanced” exons in datasets from MECs from Aire-KO and Aire-WT mice were flagged (PLATA t test P < 0.05). The proportion of imbalanced exons is shown for 420 transcripts most strongly induced by Aire, or an equal number of expression-matched Aire-neutral genes, or 84 transcripts repressed by Aire (Table S1). (B) Exon-level expression was normalized to the whole-transcript average in either WT or Aire KO datasets. Transcripts with imbalanced exon1 representation are shown in red (Left) or imbalanced representation of internal exons in purple (Right). (C) The expression level of each feature for transcripts showing an exon1 imbalance was plotted relative to its distance from the TSS for Aire-induced genes, in Aire KO or WT MECs. (D) Expression level of each feature for all Aire-induced (Left; overall WT/KO ratio >3) or Aire-neutral genes (Right) plotted as a function of distance from the TSS, either individually (Upper) or grouped in 25-bp bins (Lower). (E) The expression level of each feature for transcripts showing an exon1 imbalance was plotted relative to its distance from the TSS for Klf4-induced genes, in Klf4 KO or WT cells (data from ref. 35). A “TSS proximal-distal” index was calculated in the KO samples as the ratio of the medians of expression values in proximal (<200 bp from TSS) vs. distal features.
at the feature level might reveal a footprint of Aire’s effects on elongation, either at the very beginning of the gene (if Aire were to release stalled polymerases) or spread more uniformly throughout the target genes (if it were to generally facilitate elongation). We developed an R-implementation of the PLATA algorithm, a software tool that detects differential splicing through an imbalance of signals from individual exons relative to the entire gene (34). In effect, the algorithm detects in the KO data the over- or underrepresentation of intensity for single features relative to the whole gene, the relative efficiency of the feature being normalized by the feature-gene ratio observed in the control WT dataset. Analysis of ST1.0 profiles of Aire-induced genes in MECs from Aire-KO and WT mice showed a much greater proportion of imbalanced expression for the first than for internal exons (at greater than twofold relative variation from gene-level and f test P < 0.05). Such a skew was observed for Aire-neutral or Aire-repressed genes (Fig. 1A). We then analyzed the directionality of these imbalances by plotting the expression values of each imbalanced exon relative to the whole transcript in WT vs. KO MECs. The directionality of these imbalanced exons was completely skewed, with higher signals in Aire-KO relative to WT MECs (Fig. 1B, Left), indicating that these sequences were always overrepresented in RNA from KO MECs relative to the rest of the transcript. In contrast, there was no such bias with imbalanced internal exons, which likely reflect differential splicing that Aire could potentially influence in either direction (Fig. 1B, Right).

To better understand this imbalanced representation of first exons in Aire-induced genes, we selected transcripts presenting an exon1 imbalance and plotted the expression value at each feature relative to its distance from the TSS. In Aire-KO MECs, a sharp drop was observed between features located within 200 bp of the TSS and those further downstream (Mann-Whitney P = 2.10^{-160}) (Fig. 1C, Left). This drop was absent for the same genes in Aire-WT MECs (Fig. 1C, Right). For a broader perspective on Aire-activated genes, we analyzed all genes induced at least twofold by Aire, calculating the WT/KO ratio for each feature. For many of those features mapping in the first 50–100 bp from the TSS, the ratio remained largely around 1, increasing only further 3′ (Fig. 1D, scatter plot of all features (Upper), or after averaging WT/KO ratios within a sliding window (Lower)). This drop in expression after 50–100 bp in KO MECs, coming after signals comparable to those of WT, suggests that initiation of transcription of Aire target genes by Pol-II occurred normally even in the absence of Aire but that elongation was quickly extinguished, a block that could be lifted by Aire.

As a comparator, we analyzed published ST1.0 datasets from cells for Klf4 (35), a conventional DNA-binding transcription factor whose levels of induction of target genes are comparable to Aire. In Klf4-KO cells, there was no drop in the expression signals distal to the TSS (Fig. 1E), actually just the opposite, which we interpret as reflecting the involvement of additional TSS downstream of the canonical TSS (36).

To validate this means to evaluate the impact of transcriptional regulators on Pol-II elongation, we generated ST1.0 datasets from 4D6 cells treated with flavopiridol, an inhibitor of CDK9 (the enzymatically active half of P-TEFb), which prevents polymerase release and inhibits Aire-driven induction of the four Aire-sensitive genes that we tested (Fig. 2A). Transcripts in flavopiridol-treated cells seemed quite similar to Aire-induced transcripts in MECs, with a drop in average expression after 100 bp (Fig. 2B).

Last, we analyzed published ST1.0 datasets from KOs or knockdowns of sequence-specific transcription initiation factors [Foxl2, LMX1B, and Stat3 (37–39)] or factors involved in the stalling or release of Pol-II (CDK8, which encodes a kinase found in some variants of Mediator, a complex known to regulate the recruitment and activation of P-TEFb (40, 41)). We selected the genes induced by each factor that presented an exon1 imbalance and calculated from the KO datasets a “proximal–distal” index, the difference between the median of the single-probe expression values within 200 bp of the TSSs and those farther away. As Aire, the factors involved in elongation (CDK8, or flavopiridol treatment) showed a positive index, whereas initiation factors had a negative index (Fig. 2C). This difference between the signature of initiation and pause-release factors further substantiates the localization of Aire’s effect to the release of stalled Pol-II.

This inference that Aire effects Pol-II release on its target genes rested primarily on analysis of the resulting transcripts; it seemed important to analyze by independent means the global effect of Aire on the positioning of Pol-II on the genome. We performed ChIP coupled to high-throughput sequencing (ChIP-seq), mapping Aire and Pol-II on the genome. To achieve the high cell numbers needed for this technique, we used 293 cells in which we previously showed Aire to be transcriptionally active
upon transfection (8). Two independent experiments were performed (Table S2), with highly concordant results. Signals for Aire were weaker than for Pol-II, such that only 1,562–11,400 enriched regions were detected at a statistical significance of 10^−5 or better (23) vs. 38,900 for Pol-II. As illustrated for two typical genes, Aire was detected primarily, although not exclusively, at the TSS regions, with a local distribution coinciding with the peak of Pol-II (Fig. 3A). On representations at higher resolution, the peak of Aire binding largely overlapped with that of Pol-II (Fig. 3B and C). Integrated heatmaps of the distributions are shown in Fig. 3D, ranked relative to mRNA expression levels in 293 cells. The strongest Aire binding was seen on genes with the highest expression, and correspondingly the highest Pol-II density at the TSS (Fig. 3D). This general correlation of Aire binding with the density of Pol-II was independent of Aire’s transcriptional effect, because most of the genes displayed in Fig. 3D are not notably influenced by Aire. This distribution of Aire around the TSS was not an immunoprecipitation artifact, as evidenced by the use of an irrelevant control antibody in an independent ChIP-seq experiment, in which Aire binding was tested at two time points after transfection (Fig. 3E). The complex around the TSS actually leads to an underrepresentation of the region in control precipitates.

These results do not agree with an earlier report from a more limited ChIP-on-chip experiment (42), which indicated a random distribution of Aire. This discrepancy likely stems from the lower sensitivity and span of the ChIP-chip, but perhaps also because stable AIRE transfectants were used in the ChIP-chip study. The correlation of Aire and Pol-II density was surprising because it is known that it predominantly influences poorly transcribed genes. To further analyze Aire’s specificity of action, we plotted the expression level for each gene vs. the total Pol-II in a 500-bp window around its TSS, in control-transfected 293 cells (Fig. 4A). As expected, the most active genes had the most Pol-II bound at the TSS. Genes sensitive to Aire induction (determined from microarray analysis of Aire-transfected 293 cells, fold change >2; red dots in Fig. 4A; Table S3) showed a skewed distribution, generally displaced toward higher levels of Pol-II compared with the general distribution of all genes (ANOVA P = 5.2 × 10^-14) or of Aire-repressed genes (fold change <0.5; blue dots in Fig. 4A). This excess of Pol-II relative to the steady-state transcript levels was also seen in the representation of

![Image](https://www.pnas.org/doi/10.1073/pnas.1119351109)
Fig. 4B, where the local density of Pol-II around the TSS was compared for Aire-induced genes and a set of Aire-neutral genes matched for expression levels. The disposition of Pol-II was similar, but the levels were higher for genes susceptible to activation by Aire. Thus, although Aire bound the TSS of all genes, those that responded to its presence had a surfeit of TSS-bound Pol-II relative to their activity, consistent with an impact on stalled Pol-II.

Next, we examined the effect of transfected Aire on the distribution of Pol-II on Aire-induced or Aire-neutral genes, by comparing ChIP-seq datasets from control and Aire-transfected 293 cells. With the same gene expression vs. TSS Pol-II representation as above, there was a shift after Aire transfection, with a normalization of the discrepancy between the density of Pol-II and expression levels in Aire-induced and -neutral genes (Fig. 4C). These patterns were consistent with the conclusion that Aire affects the release of stalled Pol-II on its target genes.

Discussion
The function of Aire-dependent ectopic gene expression in MECs is the display of a wide array of proteins to impose self-tolerance on the differentiating thyrocyte pool, clearly an unusual challenge to gene regulation. The present results provide two important insights into its mechanism. First, that Aire’s activity in MECs is associated with release of stalled polymerases. Second, that there is a disconnect between its physical targeting in the genome: although Aire localizes to all Pol-II–rich TSS, it preferentially affects only a subset of them.

A unique approach to analyze gene expression profiles from MECs of adult mice revealed, in the absence of Aire, a relative overabundance of short transcripts from Aire’s target genes, similar to those typically produced by stalled polymerases (30–32). The validity of this mode of analysis was confirmed by testing in a similar manner the transcriptional footprint of several known transcriptional initiation factors and of flavopiridol, a pharmacological inhibitor of P-TEFb and, thereby, transcriptional elongation. This unique application of feature-level microarray analysis to the detection of products of stalled polymerases may have more general use, because it requires considerably fewer cells than ChIP-seq or RNA-seq and can be applied retrospectively to existing microarray datasets. Interestingly, the profile observed in Aire’s absence (quasi-normal abundance of RNAs immediately downstream of the TSS, followed by a rapid drop) was different from that observed with Klf4, with which TSS-proximal signals in its target genes varied in lockstep with the rest of the gene.

The ChIP-seq analysis also brought a unique perspective: within the limits of sensitivity of our detection, Aire seemed to bind all TSS regions, and the highest amount of Aire was found on the promoters with the most Pol-II. Pol-II–rich promoters are also those with the highest density of NELF (23, 43). Of all of the promoters to which it bound, Aire activated only a subset, and Aire-induced genes were those that had, in the absence of Aire, a surfeit of polymerases relative to their usual challenge to gene regulation. The present results provide evidence particularly open to such effects, and it is important to recall that some degree of PFA expression occurs in MECs, even in Aire’s absence. From a teleological standpoint, generically unleashing sluggish genes with blocked polymerases would seem an effective strategy to achieve widespread induction of silent genes and thus drive the induction of organo-specific tolerance.

Materials and Methods
Gene Expression Profiling. Total RNA, isolated from sorted MECs of 4-wk-old individual mice, was profiled as previously described (44) on Affymetrix Mouse Gene ST1.0 arrays. For feature-level analysis, normalized feature-level expression files, generated by aroma.affymetrix, were processed with an R-implementation of the PLATA algorithm (34), and the expression of a given exon was flagged as imbalanced if the ratio between exon WT/KO fold change to transcript WT/KO fold change was >2 or <0.5, with a test P value <0.05 for a WT vs. KO comparison of expression values of all features mapping to the exon, after normalization to whole-transcript expression. An index was then computed as the ratio of the medians of expression values in promoter-proximal (TSS to 200 bp) vs. distal (>200 bp) features of the microarray.

Chromatin Immunoprecipitation. 293FT cells were transfected for pCMV-Tag2C (control) or pCMV-Tag2C containing FLAG-tagged murine Aire (Aire) with TransIT-293 Transfection Reagent (Minus). The cells were cross-linked by formaldehyde 48 h after the transfaction for ChIP, as previously described (23). Antibodies used are as follows: total Pol-II: Santa Cruz sc-899; FLAG-tagged Aire: Sigma (M2 antibody) F1804; control IgG: Santa Cruz sc-3877. For ChIP-seq analysis, illumina sequencing and analysis were as previously described (45), and sequence counts integrated in 25-bp bins, with plots generated in R. A detailed description of data analysis is provided in SI Materials and Methods and Dataset S1.

Acknowledgments. We thank S. Davis, G. Frampton, and H. Paik for help with computational analysis. This work was supported by National Institutes of Health Grants R01-DK060027 and R01-A088204 (to D.M. and C.B.) and R01-HG002668 and R01CA46455 (to R.A.Y.). M.G. received fellowships from the Campbell and Hall Charity Fund, The Harold Whitworth Pierce Charitable Trust, and the Philippe Foundation. H.Y. was supported by a fellowship from the Human Frontier Science Program, J.A. by Juvenile Diabetes Research Foundation Fellowships 10-2007-601 and 03-2005-1138, and P.B.R. by American Cancer Society Postdoctoral Fellowship 120272-PF-11-042-01-DMC.


