Cell

Aire's Partners in the Molecular Control of Immunological Tolerance

Jakub Abramson,[1] Matthieu Giraud,[1] Christophe Benoist,[1,2,*] and Diane Mathis[1,2,3,*]
[1]Department of Pathology, Harvard Medical School
[2]The Broad Institute
[3]The Harvard Stem Cell Institute
77 Avenue Louis Pasteur, Boston, MA 02115, USA
*Correspondence: cbdm@hms.harvard.edu (C.B., D.M.)
DOI 10.1016/j.cell.2009.12.030

SUMMARY

Aire induces the expression of a battery of peripheral-tissue self-antigens (PTAs) in thymic stromal cells, promoting the clonal deletion of differentiating T cells that recognize them. Just how Aire targets and induces PTA transcripts remains largely undefined. Screening via Aire-targeted coimmunoprecipitation followed by mass spectrometry, and validating by multiple RNAi-mediated knockdown approaches, we identified a large set of proteins that associate with Aire. They fall into four major functional classes: nuclear transport, chromatin binding/structure, transcription and pre-mRNA processing. One set of Aire interactions centered on DNA protein kinase and a group of proteins it partners with to resolve DNA double-stranded breaks or promote transcriptional elongation. Another set of interactions was focused on the pre-mRNA splicing and maturation machinery, potentially explaining the markedly more effective processing of PTA transcripts in the presence of Aire. These findings suggest a model to explain Aire's widespread targeting and induction of weakly transcribed chromatin regions.

INTRODUCTION

Negative selection of self-reactive T cells is essential for the establishment of an immune system that is capable of defending against foreign pathogens but tolerates "self" and thereby avoids autoimmunity (Starr et al., 2003). A long-standing tenet was that T cells reactive to ubiquitous or blood-circulating self-antigens are deleted "centrally" in the thymus, while tolerance to antigens expressed exclusively in parenchymal organs (liver, brain, pancreas, etc.) is imposed by "peripheral" mechanisms. However, it is now known that thymic cells, specifically stromal medullary epithelial cells (MECs), transcribe thousands of genes whose expression was thought to be restricted to parenchymal organs (Kyewski and Klein, 2006). Hence, transcription of peripheral-tissue-antigen (PTA) genes in the thymus "foreshadows" the self-antigens that T cells will encounter once they reach maturity and are released into the body. Many of these ectopic transcripts are regulated by the product of a single gene, *Aire*, as mice with a mutation at this locus express only a fraction of the PTA repertoire (Anderson et al., 2002). As a result, these animals develop antibodies and immune infiltrates directed at multiple peripheral tissues, resembling the multiorgan autoimmune disorder characteristic of humans with a mutated *AIRE* gene, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) (Mathis and Benoist, 2009).

A number of observations argue that Aire functions as a transcriptional regulator (Mathis and Benoist, 2009). First, it is localized in the nucleus, typically in punctate structures. Second, Aire can modulate gene expression in a variety of transfection assays. Accordingly, it has been reported to bind to a transcription factor, CREB-binding protein (CBP), in vitro and to associate with a regulator of transcriptional elongation, p-TEFb. Lastly, Aire contains several functional domains characteristic of transcriptional regulators. For example, there is a SAND domain, thought to function as a DNA-binding module in other proteins, but lacking the critical DNA-binding residues (the KDWK motif) in Aire (Bottomley et al., 2001). There are also two plant homeodomain (PHD) zinc fingers analogous to those found in several transcriptional coactivators and chromatin-modulating factors. Aire's PHD1 binds to histone-3 molecules with a lysine at position 4 (H3K4) that is unmethylated, generally a mark for transcriptionally repressed loci, suggesting that this interaction might be important for recruiting Aire to regions of relatively inactive chromatin, where it might have an activating influence (Org et al., 2008; Koh et al., 2008). PHD2 is actually more reminiscent of a RING finger, and its functional significance remains under debate.

The precise molecular mechanisms Aire employs to regulate transcription are still poorly understood, but several unusual features suggest that it does not operate as a classical transcription factor, binding to promoters and driving mRNA initiation. For example, Aire impacts on thousands of genes, including loci with vastly different geographical and temporal modes of regulation in the periphery (Johnnidis et al., 2005). Relatedly, Aire influences a similarly large number of genes when introduced into diverse cell types, but the repertoire of loci is different in each case (Guerau-de-Arellano et al., 2008). Another unusual property is that each MEC expresses only a subset of

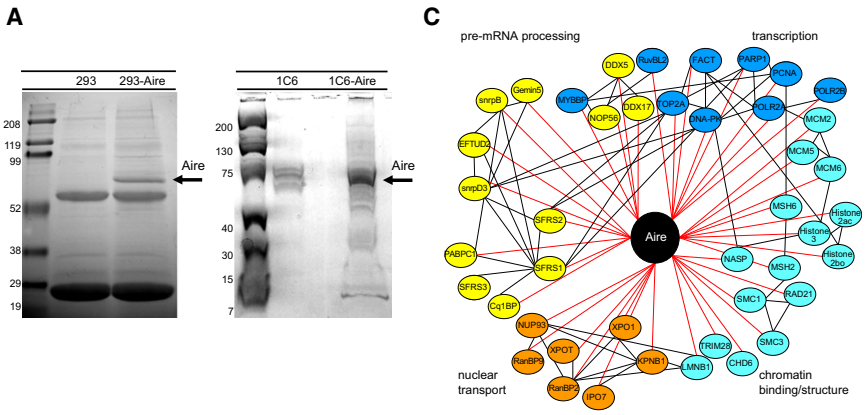

Figure 1. Identification of Putative Aire-Associated Proteins

(A) Aire appears to be associated with a multitude of proteins. Nuclear lysates from cells expressing Aire-Flag (293Aire and 1C6Aire) or from control cells (293 and 1C6) were IPed with anti-Flag mAb. Protein complexes were eluted, separated by SDS-PAGE and blue-stained. Excised bands were analyzed by MS.

(B) List of Aire-associated proteins identified by MS. The number of unique peptides identified for each protein in each pull-down is shown, as well as the approximate MW of the band in which the given protein was identified.

(C) Interaction network of Aire-associated proteins. The network was constructed from protein-protein interaction databases. Red lines: the protein was detected in at least 2 (out of 4) independent Aire IPs; black lines: a previously reported interaction between proteins. See also Figure S1.

B

gene symbol	other name	293T			1C6	avg #hits	band's mw (kD)	activity	process
		1	2	3	4				
PRKDC	DNA-PK	340	3	25	46	104	300-500	S/T kinase	Gene exp., DDR
PARP1	PARP-1	8	2	10	1	5	90-130	ribosyl transferase	Gene exp., DDR
TOP2A	TOP2a	1	0	2	13	4	150-200	DNA topoisomerase	Gene exp., DDR
SUPT16H	FACT	1	0	12	0	3	90-130	elongation factor	Gene exp., DDR
POLR2A	RPB1	1	0	0	2	1	200-300	RNA polymerase	Gene exp.
POLR2B	RPB2	0	4	0	3	2	120-150	RNA polymerase	Gene exp.
RUVBL2	TIP49B	13	0	0	5	5	40-55	DNA-helicase (ATP)	Gene exp., DDR, Chromat. struct.
MSH2	MSH2	2	0	0	2	1	70-100	ATPase	Gene exp., DDR, CC
MSH6	MSH6	0	14	0	2	4	120-150	ATPase	Gene exp., DDR, CC
PCNA	PCNA	3	0	3	2	2	120-150	cyclin	DDR, CC
SMC1A	SMC1	0	1	3	2	2	150-200	ATPase	Chromat. struct., CC, DDR, Gene exp.
SMC3	SMC3	0	14	3	3	6	150-200	ATPase	Chromat. struct., CC, DDR, Gene exp.
Rad21	SCC1	1	0	5	0	2	55-70	ATPase	Chromat. struct., CC, DDR, Gene exp.
MCM2	MCM2	0	52	0	1	13	90-130	DNA-helicase (ATP)	Chromat. struct., CC, Gene exp.
MCM5	MCM5	0	11	5	0	4	90-130	DNA-helicase (ATP)	Chromat. struct., CC, Gene exp.
MCM6	MCM6	4	14	0	0	5	70-100	DNA-helicase (ATP)	Chromat. struct., CC, Gene exp.
TRIM28	KAP-1	6	14	7	16	11	120-150	E3 sumo ligase	Chromat. struct., Gene exp.
CHD6	CHD6	0	0	10	2	3	300-500	DNA-helicase (ATP)	Chromat. struct., Gene exp.
LMNB1	LMNB1	0	1	8	0	2	55-70	structural protein	Chromat. struct., Gene exp.
NASP	NASP	19	12	0	0	8	90-130	linker histone binding	Chromat. struct., Gene exp.
HIST3H3	Histone H3	0	2	0	13	4	10–15	DNA-binding	Chromat. struct., Gene exp.
HIST1H2BO	Histone H2bo	0	3	46	32	20	10–15	DNA-binding	Chromat. struct., Gene exp.
HIST1H2AC	Histone H2ac	0	0	28	18	12	10–15	DNA-binding	Chromat. struct., Gene exp.
CAND1	TIP120	5	30	0	0	9	90-130	protein binding	Unknown
KPNB1	IPO1	4	29	2	8	11	90-130	transport	Nuclear transport
IPO7	RanBP7	4	5	0	1	3	90-130	transport	Nuclear transport
RANBP2	NUP358	16	2	7	2	7	300-500	E3 Sumo; transport	Nuclear transport
RANBP9	RanBP9	0	7	0	1	2	90-130	transport	Nuclear transport
NUP93	NUP93	2	6	1	9	5	90-130	transport	Nuclear transport
XPO1	XPO1	10	15	1	4	8	70-100	transport	Nuclear transport
XPOT	XPOT	12	7	2	22	11	70-100	transport	Nuclear transport
MYBBP1A	P160	7	9	4	39	15	150-200	protein binding	mRNA processing, Gene exp.
DDX5	P68	4	2	1	8	4	55-70	RNA-helicase	mRNA processing
DDX17	P72	10	0	0	7	4	55-70	RNA-helicase	mRNA processing
NOP56	NOL5A	0	3	0	7	3	55-70	Spliceosome binding	mRNA processing
GEMIN5	Gemin5	0	6	0	1	2	200-300	Spliceosome binding	mRNA processing
SNRPB	snrpB	0	2	1	2	1	20-35	Spliceosome binding	mRNA processing
SNRPD3	snrpD3	3	0	1	2	2	20-35	Spliceosome binding	mRNA processing
EFTUD2	snrp116	6	0	0	6	3	90-130	GTPase	mRNA processing
SFRS1	ASF/SF2	2	0	0	12	4	20-35	Spliceosome binding	mRNA processing
SFRS2	SC35	1	3	0	1	1	20-35	Spliceosome binding	mRNA processing
SFRS3	SFRS3	2	3	0	2	2	20-35	Spliceosome binding	mRNA processing
C1QBP	SF2P32	0	18	0	10	7	20-35	Spliceosome binding	mRNA processing
PABPC1	PABPC1	40	17	1	1	15	55-70	polyA binding	mRNA processing
GCN1L	GCN1	1	4	39	8	13	300-500	protein binding	Activation of translation, Unknown

Aire-induced genes, in a deterministic fashion (Villaseñor et al., 2008; Derbinski et al., 2008). Lastly, Aire-induced transcription depends on cofactors and start-sites different from those characteristic of the relevant peripheral tissue(s) (Villaseñor et al., 2008).

As one approach to addressing these issues, we sought to identify molecules that associate with Aire, either directly or indirectly, and impact on its activities. We performed a broad screen via Aire-targeted coimmunoprecipitations (co-IPs) followed by mass spectrometry (MS), and evaluated the functional relevance of candidate Aire-associated proteins by a number of RNAi-based and other assays. The data obtained suggest a model that accounts for several of Aire's properties in controlling MEC gene expression.

RESULTS

Identification of Aire-Associated Proteins by Coimmunoprecipitation and MS Analyses

Our strategy for identifying Aire's partners entailed MS screening of proteins that co-IP with it. The number of cells needed for this approach (~5 × 10^8/IP) precluded the use of isolated MECs because of their rarity (~2 × 10^4/mouse thymus). Instead, we chose to work with the human embryonic kidney (HEK) epithelial cell line, 293T, as an easily transfectable cell type and a well-established system for studying protein-protein interactions; as well as with two cell lines derived from the thymic epithelium, 1C6 (murine) and 4D6 (human), as more physiological, but experimentally more challenging, systems. To validate these lines as suitable for assaying Aire's function, in particular its interactions with other proteins, we first established that Aire is correctly expressed in nuclear speckles in these cells (Figure S1A available online), can be isolated from the nucleus by a relatively gentle technique capable of preserving protein-protein interactions (Figure S1B), and is functional, i.e., induces widespread changes in the gene-expression profile (Figure S1C).

In four independent experiments, 293T or 1C6 cells were either mock-transfected or transfected with the pCMV-Aire-Flag expression vector; 48 hr later, the cells were lysed, and their nuclei were isolated and digested with micrococcal nuclease (MNase). Nuclear extracts were incubated with Sepharose beads to which Flag-Tag-specific monoclonal antibody (mAb) had been conjugated; bound proteins were then eluted, separated by SDS-PAGE and stained with colloidal blue. For both lines, Aire-expressing cells gave rise to more bands than did the corresponding control cells (Figure 1A). The strongest band present in the former and absent from the latter migrated at 60–65 kDa (arrow), the expected size of Aire-Flag, suggesting that Aire was IPed together with a number of allied proteins. Individual bands showing a clear difference in the presence or absence of Aire were excised from both the control and experimental lanes, and were analyzed by MS. Each band harbored a number of proteins of discernable identity. Importantly, Aire

was the most abundant protein from the 60–65 kDa band, serving as a positive control for the validity of the approach.

The combined raw data sets of the four independent experiments yielded a large number of putative Aire-associated proteins. To allow us to focus on the most robust candidates and to eliminate false positives, we removed from further consideration all proteins that (1) were also detected in the relevant control lane, (2) were identified in only one IP, and (3) corresponded to known contaminants/false positives (e.g., keratins, albumins, etc.). This filtering approach narrowed the list of putative Aire-interacting proteins to approximately 45 best candidates (Figure 1B).

An immediate question was whether this set of proteins included elements of one or more multimolecular complexes containing Aire, especially since it has been described to partake in high-molecular-weight (>670 kDa) conglomerates (Halonen et al., 2004). Several protein-protein interaction databases were queried in order to construct a potential interaction network (Figure 1C). Many of the Aire-associated candidates were indeed previously reported to interact with each other, enabling us to cluster them into several putative functional groups (with multiple overlaps). Four groups stood out, representing proteins involved in nuclear transport [e.g., exportin (XPO)1, nucleoporin (NUP)93, and karyopherin (KPN)B1], which are likely to mediate the shuttling of Aire into or out of the nucleus; chromatin binding/structure (e.g., histones, the cohesin complex, etc.), which could affect the structure or spatial organization of chromatin in response to Aire or serve as an access point for Aire binding; postinitiation RNA polymerase (RNAP)II-mediated transcription events [e.g., a putative complex of DNA-dependent protein kinase (DNA-PK), poly(ADP-ribose) polymerase (PARP)-1, Topoisomerase (TOP)2a, Facilitates Chromatin Transcription (FACT) and RNAPII], which may modify the nucleosome environment in response to progressing RNAPII; and pre-mRNA processing (including several splicing factors and RNA helicases), which likely affect the processing of pre-mRNAs into mature transcripts, and thereby can impact their expression levels.

Validation of Candidate Aire-Interacting Proteins by Reciprocal Coimmunoprecipitation

To confirm that the interactions between Aire and its putative partners were specific ones, we attempted to verify the associations by reciprocal co-IP, i.e., antibody targeting of a particular candidate followed by western blotting (WB) for Aire. Since it was cost-prohibitive to verify the interactions for all of the ~45 Aire-associated candidates, we chose at least two representatives from each functional group (or putative multiprotein complex). In addition, to serve as negative controls, a number of nuclear proteins not picked up in the screen [e.g., Chromodomein helicase (CHD)3, CHD4, CCTC-binding factor (CTCF)] were tested. Lastly, we analyzed several factors that were recently reported to bind to Aire [Cyclin-dependent kinase (CDK)9, Cyclin (Cyc)T1, CBP] (Pitkänen et al., 2000; Oven et al., 2007) but were not detected in our MS analyses. Candidate proteins either endogenously expressed (if a specific antibody was available) or expressed via transfection (tagged with a peptide sequence) were IPed from nuclear lysates of Aire-expressing 293T cells, followed by WB with an Aire-tag-specific mAb.

Most (~2/3) of the candidates evaluated by this procedure, did indeed associate with Aire (Figure 2A), suggesting that these proteins are true Aire-interactors (whether direct or indirect). Aire was not immunoprecipitated with a smaller fraction of candidates (Figure 2B), implying that their interaction is either extremely weak or nonexistent. None of the negative controls showed detectable binding to Aire (Figure 2C). Consistent with the MS results, the reciprocal co-IPs did not reveal any association of CycT1, CDK9, or CBP with Aire (Figure 2C), all previously reported to interact with it. These proteins may have only a very weak liaison with Aire or they may interact with it under conditions different than those used in our studies.

The strong convergence of the MS and reciprocal co-IP data argue for the validity of the candidate-filtering process. It is likely that most of the highlighted candidates interact with Aire in some way, either directly or as common participants in a multi-protein complex.

Functional Relevance of Aire-Interacting Proteins on Expression of Aire-Dependent Transcripts

An interaction between Aire and a given candidate protein does not necessarily imply that the protein is also involved in Aire's biological activities. Functional associations were tested in a series of RNAi-mediated knockdown experiments in Aire-transfected cells. We utilized a pLKO.1-hosted library containing, for each targeted candidate, five short-hairpin RNAs (shRNAs) expressed under the ubiquitously active U6 promoter. Individual RNAi-encoding constructs and Aire-expression plasmids or control plasmids were cotransfected into 293T cells, and the effect of knockdown of each candidate's expression on mRNA levels of *keratin (KRT)14* (an Aire-dependent transcript in both MECs and 293T cells) was quantified by RT-PCR in three independent experiments. Not all of the shRNAs contained in this library are effective in dampening expression of their target proteins; we considered the knockdown to have a significant effect if at least two different hairpins provoked an average change of at least 40% in the level of the Aire-induced transcript(s). Indeed, for those loci that influenced Aire-mediated transcription, we found a good correlation between the shRNA knock-down efficiency and the impact on Aire-induced transcription (Figure S2).

In about half of the 42 cases tested, lowering the expression of an Aire-associated candidate had a significant and consistent effect on the level of Aire-induced transcripts (Figure 3A and Table S1). The candidates with the greatest impact included representatives of all four functional groups (Figure 3B). Strikingly, dampening of the expression of most of the members of the putative DNA-PK/PARP-1/TOP2a/FACT/RNAPII complex led to a drop in Aire-induced transcripts. Similarly, knockdown of several proteins involved in pre-mRNA processing such as Splicing factor arginine/serine rich (SFRS)1, SFRS3, DEAD box (DDX)5, DDX17, Myb-binding protein (MYBBP)1a and PolyA binding protein (PABP)C1 had a strong negative effect. Reduced expression of certain nuclear-pore-complex (NPC)-associated proteins such as XPO1, NUP93, KPNB1, and Ran-binding protein (RANBP)2 also suppressed the levels of Aire-dependent transcripts, while knockdown of others [e.g., importin (IPO)7, exportin tRNA (XPOT)] had no effect.

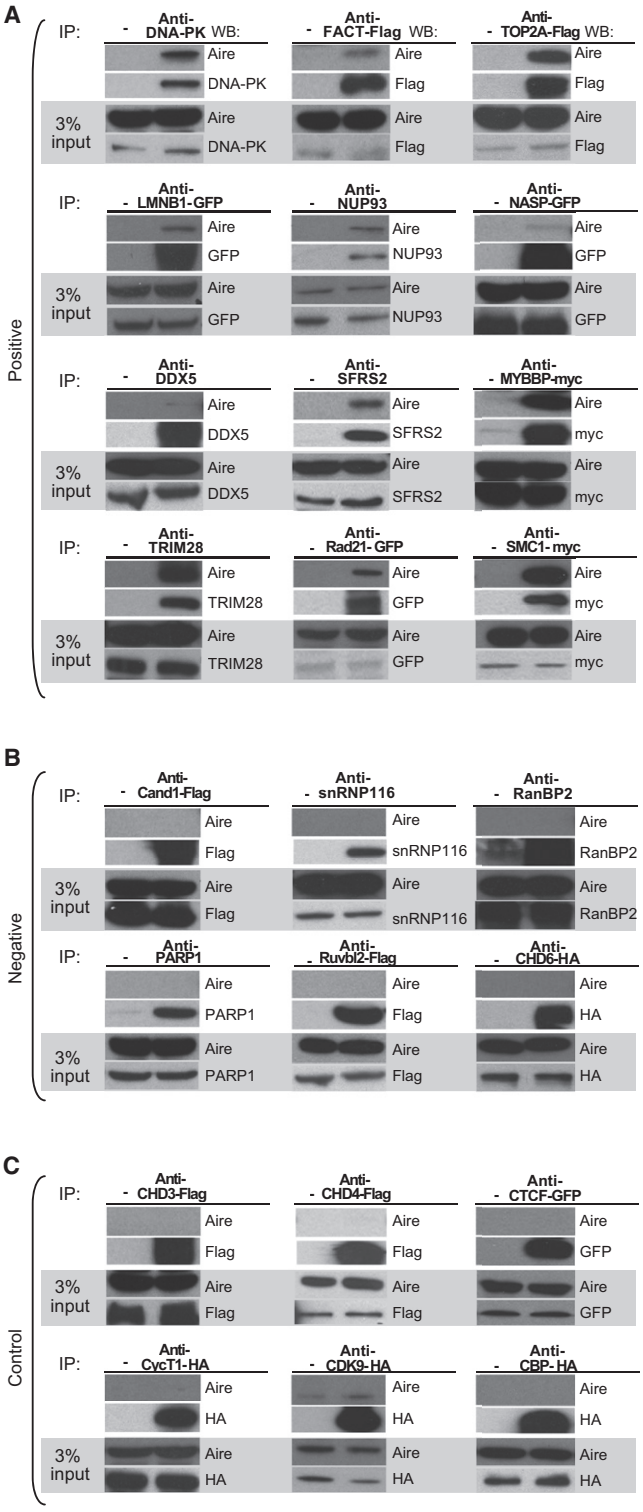

Figure 2. Confirmation of Candidates by Reciprocal Co-IP

293T cells expressing Aire and a given candidate interactor, expressed either endogenously or exogenously as an HA, Flag, c-*myc*, or GFP-tagged protein, were lysed and their nuclear extracts were IPed with an antibody directed at the candidate or with a control non-specific antibody. Association between Aire and the candidate was analyzed by sequential Western blotting using indicated antibodies. (A) Candidates confirmed to associate with Aire; (B) Candidates not confirmed to associate with Aire; (C) Analysis of interactions between Aire and other proteins not detected by MS.

Interestingly, underexpression of Lamin (LMN)B1 increased, while its overexpression reduced, the levels of Aire-dependent transcripts (Figure 3C), suggesting that this protein, a component of the nuclear lamina/matrix, may somehow inhibit Aire's activities. These results may not be so surprising given several recent studies indicating that nuclear lamina/matrix provides a platform for sequestering transcription factors away from chromatin (Heessen and Fornerod, 2007).

As expected, shRNAs corresponding to the negative-control proteins (NUP155, CTCF, and CHD3) did not influence levels of Aire-induced transcripts (Figure 3B and data not shown), nor did those targeting CycT1, previously reported to be important in Aire's effect on gene expression (Oven et al., 2007). However, this prior study used trichostatin-A (TSA)-treated transfectants, which is likely to change the chromatin structure and dynamics.

To determine whether the impact of inhibiting the different putative partners was indeed Aire-specific and did not merely reflect generic influences, we analyzed the effects of the panel of shRNAs on three additional genes that are not regulated by Aire in 293T cells (*SERPD*, *SERPB2*, and *SPIC*) but have similar expression values (C_T ~25–30). With the exception of those targeting *POLR2A*, *POLR2B*, and *SNRNPB*, the shRNAs did not consistently induce downregulation of these Aire-independent genes (Table S1), indicating that most of the putative partners are specifically involved in the regulation of Aire's activities.

To strengthen the validity of the above results, we also examined the capacity of the candidate Aire-associated proteins to influence Aire-induced transcription in an independent system based on an episomal reporter (M.G., D.M., and C.B., unpublished data). 4D6 cells were transduced by shRNA-bearing lentiviruses targeting a given candidate, subjected to puromycin selection, and cotransfected with an Aire-expression plasmid and the pGL4.23 reporter vector containing a minimal promoter driving luciferase expression. The influence of each shRNA on Aire-induced luciferase expression was measured in three independent experiments and was considered to be meaningful if at least two different hairpins provoked significant changes (quantified as detailed in the Experimental Procedures). Knockdown of the expression of 12 candidate Aire-associated proteins had an influence on Aire-induced transcription according to this assay (Table S1 and Figure 3A); 11 of these also showed an effect in the RT-PCR-based test described above, indicating substantial overlap. Other candidates that influenced Aire-induced *KRT14* transcript levels did not show an effect on Aire activity in the episomal-reporter-based assay. Such discrepancies were not surprising, given the many differences between these two test systems, e.g., transcription in the context of spatially organized chromatin with specific epigenetic DNA and protein marks and a given level of compaction versus transcription of a reporter driven by a minimal promoter on a transfected episomal plasmid. Such differences are likely to explain the discordant results for most of the candidates in the "transcription" class, i.e., RNAPII, TOP2a, etc. (Mondal and Parvin, 2001). However, the critical point is that results from this assay validated at least one

A

gene symbol	gene	Aire-mediated transcription: endogenous target	Aire-mediated transcription: transfected target	Aire's nuclear localization
PRKDC	DNA-PK	+	+	-
PARP1	PARP-1	+	-	-
TOP2A	TOP2a	+	-	-
SUPT16H	FACT	-	-	-
POLR2A	POLR2B	+	-	-
POLR2B	POLR2A	+	-	-
RUVBL2	RUVBL2	-	-	-
MSH2	MSH2	NA	-	-
MSH6	MSH6	+	-	-
PCNA	PCNA	-	-	-
SMC1A	SMC1	+	-	-
SMC3	SMC3	-	-	-
Rad21	Rad21	-	-	-
MCM2	MCM2	-	-	-
MCM5	MCM5	-	-	-
MCM6	MCM6	-	+	-
TRIM28	TRIM28	-	-	-
CHD6	CHD6	-	-	-
LMNB1	LMNB1	+	-	-
NASP	NASP	-	-	-
CAND1	CAND	-	-	-
KPNB1	KPNB1	+	+	+
IPO7	IPO7	-	-	-
RANBP2	RanBP2	+	+	+
RANBP9	RanBP9	-	-	-
NUP93	NUP93	+	+	+
XPO1	XPO1	+	+	+
XPOT	XPOT	-	-	-
MYBBP1A	MYBBP	+	+	-
DDX5	DDX5	+	+	-
DDX17	DDX17	+	+	-
NOP56	NOP56	-	-	-
GEMIN5	Gemin5	+	-	-
SNRPB	snrpB	+	-	-
SNRPD3	snrpD3	-	-	-
EFTUD2	EFTUD2	-	-	-
SFRS1	SFRS1	+	+	-
SFRS2	SFRS2	-	-	-
SFRS3	SFRS3	+	+	-
C1QBP	C1QBP	-	-	-
PABPC1	PABPC1	+	+	-
GCN1L	GCN1L	-	-	-

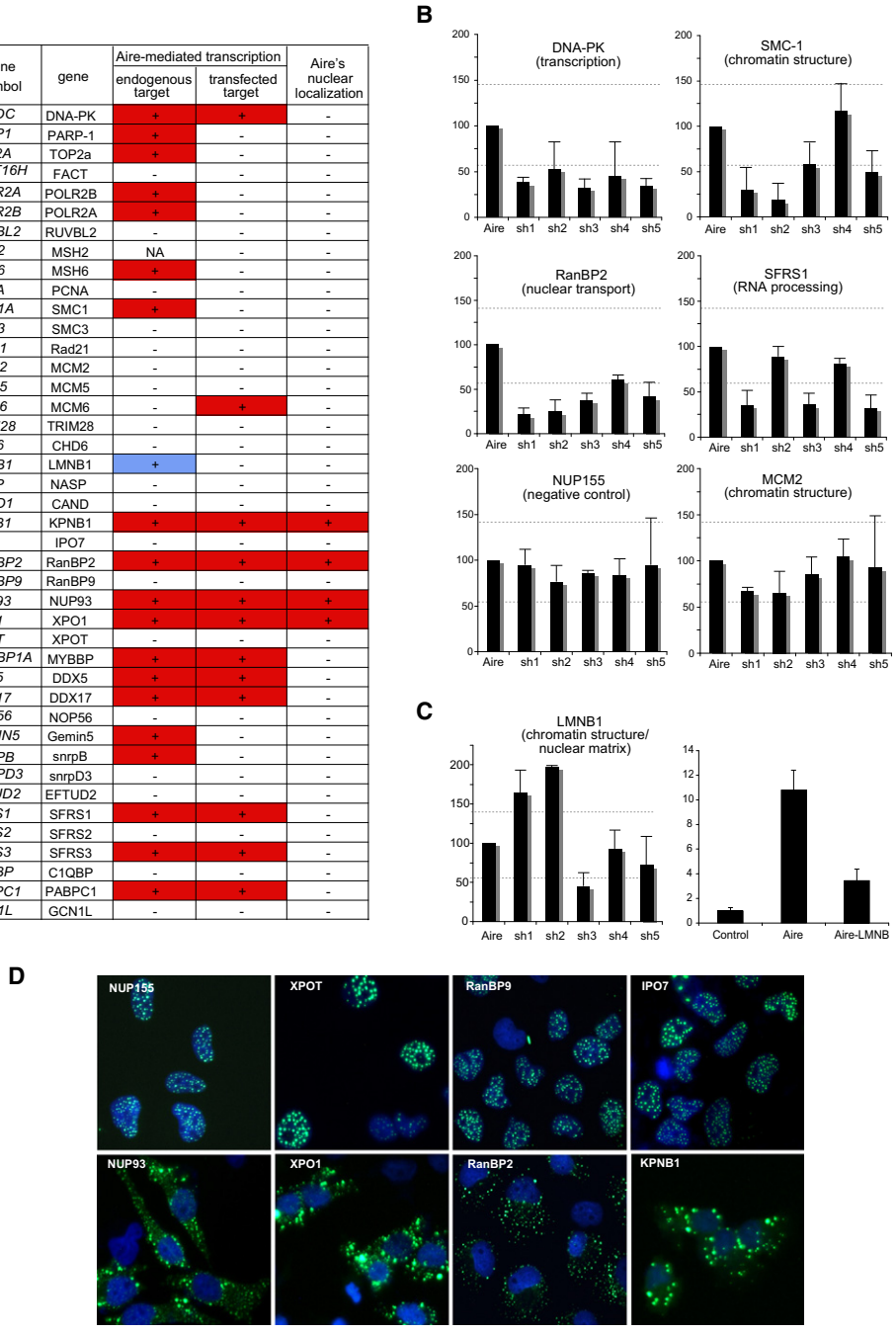

Figure 3. Functional Analysis of Individual Aire-Interacting Candidates by shRNA-Mediated Knockdown

(A) Summary of the knockdown impact of Aire-associated molecules on Aire localization and function (as indicated). White: no effect; red: inhibitory effect; blue: stimulatory effect. (See also Table S1.)

(B) RT-PCR analysis of the effect of candidate interactor knockdowns on Aire-induced expression in 293T cells. Relative expression of *KRT14* (as an average of three independent experiments ± SD) in cells cotransfected either with a pCMV-Aire and an empty pLKO plasmid (Aire) or a pCMV-Aire and a pLKO vector containing specified shRNA hairpins (sh1 - sh5) for each gene. Representatives of the four functional groups are shown; data for the other candidates appear in Table S1.

(C) LMNB1 negatively regulates Aire-induced transcription. RT-PCR analysis of the effect of either LMNB1-knockdown (left panel) or overexpression (right panel) on Aire-induced expression of the *KRT14* gene in 293T cells.

(D) The influence of Aire-interacting candidates on its nuclear localization. A typical fluorescent microscopic image of Aire-GFP-expressing 4D6 cells infected with a virus, carrying an shRNA capable of knocking down expression of the designated gene. The examples in the top row had no effect on nuclear localization of Aire, while those below suppressed it. Nuclei stained with DAPI (blue).

See also Figure S2.

member of each of the four functional groups highlighted by the co-IP/MS approach.

Some NPC-Associated Proteins Affect the Nuclear Localization of Aire

The influence of NPC-associated proteins (such as XPO1, NUP93, etc.) on the level of Aire-induced transcripts is most likely an indirect one, reflecting their ability to control the transport of nuclear factors into or out of the nucleus. To test this notion, we performed RNAi-mediated knockdowns in a 4D6 transfectant stably expressing an Aire-GFP fusion protein. Each candidate was targeted by five different shRNA hairpins, each having a different knockdown capacity. While dampening expression of most of the candidates did not have an effect on the nuclear localization of Aire, nor on its speckled disposition therein, down-modulation of NUP93, XPO1, KPNB1, and RanBP2 resulted in Aire retention in the cytosol rather than the nucleus (Figures 3A and 3D). Therefore, these four NPC-associated proteins are important for nuclear shuttling of Aire, and their effect on Aire-dependent gene expression is likely an indirect one.

DNA-PK Controls Expression of PTA Genes In Vivo

Results from the interaction and functional analyses converged to suggest that three functional groups might participate in Aire's regulation of gene expression: chromatin-associated proteins, transcriptional control elements, and factors involved in pre-mRNA processing. A functional link between Aire and chromatin has already been provided by recent studies demonstrating that its PHD1 binds specifically to H3 tails, especially those devoid of methylation on K4 residues (Org et al., 2008; Koh et al., 2008). Therefore, we chose to further explore the latter two functional groups. In particular, we were struck by the fact that several members of a complex known to regulate transcriptional elongation (DNA-PK, PARP-1, TOP2a, and FACT) were identified in the MS screening, most of which survived the functional filters. Indeed, DNA-PK, the core of the putative complex, received the highest score in the MS analysis, influenced Aire activity in both of the RNAi-based functional assays, and has been previously reported to phosphorylate Aire in vitro (Liiv et al., 2008).

To evaluate the relevance of DNA-PK in vivo, we turned to the NOD.CB17-Prkdc^{scid} severe combined immunodeficiency (SCID) mouse mutant, which has an inactive truncated form of DNA-PK. As this kinase is essential for T cell receptor and Ig gene rearrangements, T and B cell differentiation are abrogated in these mice. Consequently SCID thymi are usually small and are largely devoid of MECs, as mature thymocytes are essential for differentiation of this stromal cell type (Surh et al., 1992). To evaluate the role of DNA-PK in the regulation of PTA expression, we had to circumvent this problem by reconstituting mature thymocytes and thereby MECs. Bone-marrow cells from wild-type (WT) donors were transferred into either SCID or recombination activating gene (RAG)-KO recipients (both lacking mature thymocytes and MECs), generating chimeric animals with a WT lymphocyte compartment and stromal cells bearing either the DNA-PK mutation (SCID) or its WT counterpart (RAG-KO). Mutant thymi were fully reconstituted by 6–8 weeks after transfer, displaying normal thymocyte (Figure S3A) and stromal cell (Figure S3B) populations, in particular MECs.

PCR titration of mRNA transcripts revealed that purified mature (MHCII^{hi}) MECs from the three types of Aire-positive animals transcribed the *Aire* gene at similar levels, but that there was a striking decrease in Aire-dependent PTA transcripts in the SCID, but not RAG-KO or WT MECs, though the reduction was less than with Aire-KO MECs (Figure 4A). The generality of this phenomenon was established by global gene-expression profiling, focusing on the most relevant comparison: reconstituted SCID versus reconstituted RAG-KO individuals, i.e., mice without versus with MECs expressing functional DNA-PK (Figure 4B, left). Superposition of the repertoire of Aire-induced genes on the RAG-KO versus SCID differential revealed a highly significant overlap ($p = 1.5 \times 10^{-11}$), signifying poor induction of Aire-regulated genes in the reconstituted SCID mice. Interestingly, the overlap was not significant when the Aire-repressed loci were similarly analyzed (Figure 4B, right). Flow cytometric assessments confirmed that the frequencies of Aire-expressing MECs and the levels of Aire they expressed were not significantly different in the two types of reconstituted mice, i.e., the most relevant comparison (Figure S3C). Though Aire expression in WT cells was slightly higher and the frequency of Aire-positive cells slightly lower compared with those of the reconstituted SCIDs and RAG-KOs, this 20%–30% difference was not able to explain the often 5- to 10-fold differential in Aire-regulated gene expression found for reconstituted SCIDs. The nuclear localization and punctate disposition of Aire were also normal in both types of reconstituted animals (Figure S3D).

To test whether the deficient PTA expression by SCID MECs provoked a failure in central tolerance, we assayed serum autoAbs. Although we did not find as great a repertoire of anti-self specificities as usually found in Aire-KO mice, prolonged exposure times revealed a few rather weak bands (eye, stomach) in the reconstituted SCID mice, but not in the WT or reconstituted RAG-KO animals (Figure S3E). This reduced response is not entirely surprising given that the drops in PTA levels were weaker in the reconstituted SCIDs than in the Aire-KOs.

In addition, we further characterized the putative complex hosting Aire and DNA-PK. First, co-IPs on 293T cells identified some new participants: H2AX, Ku80, and (weakly) Ku70 (Figure 4C and data not shown). Indeed, DNA-PK, TOP2a, PARP1, H2AX, Ku80, and Aire all associated with each other (Figure 4C). Interestingly, all of these proteins are early-acting elements of the nonhomologous end-joining DNA repair machinery. In contrast, a late-acting element, XRCC4, did not associate with Aire (Figure 4D). Also absent from this complex were the Aire-associated pre-mRNA processing factors (SFRS2, DDX5) (Figure S3F). Supporting the notion of independent complexes, shRNA-driven knock-down of DNA-PK compromised the binding of Aire to PARP-1 (Figure 4E), while knockdown of TOP2a did not destroy the interaction between Aire and SFRS2 (Figure S3G).

Aire Promotes TOP2a-Initiated DNA Double-Stranded Breaks

As an alternative to shRNA knockdowns, we looked into available small-molecule inhibitors of putative Aire partners. An interesting option was etoposide, an anticancer drug known to stabilize TOP2-introduced single-stranded breaks in DNA by "freezing"

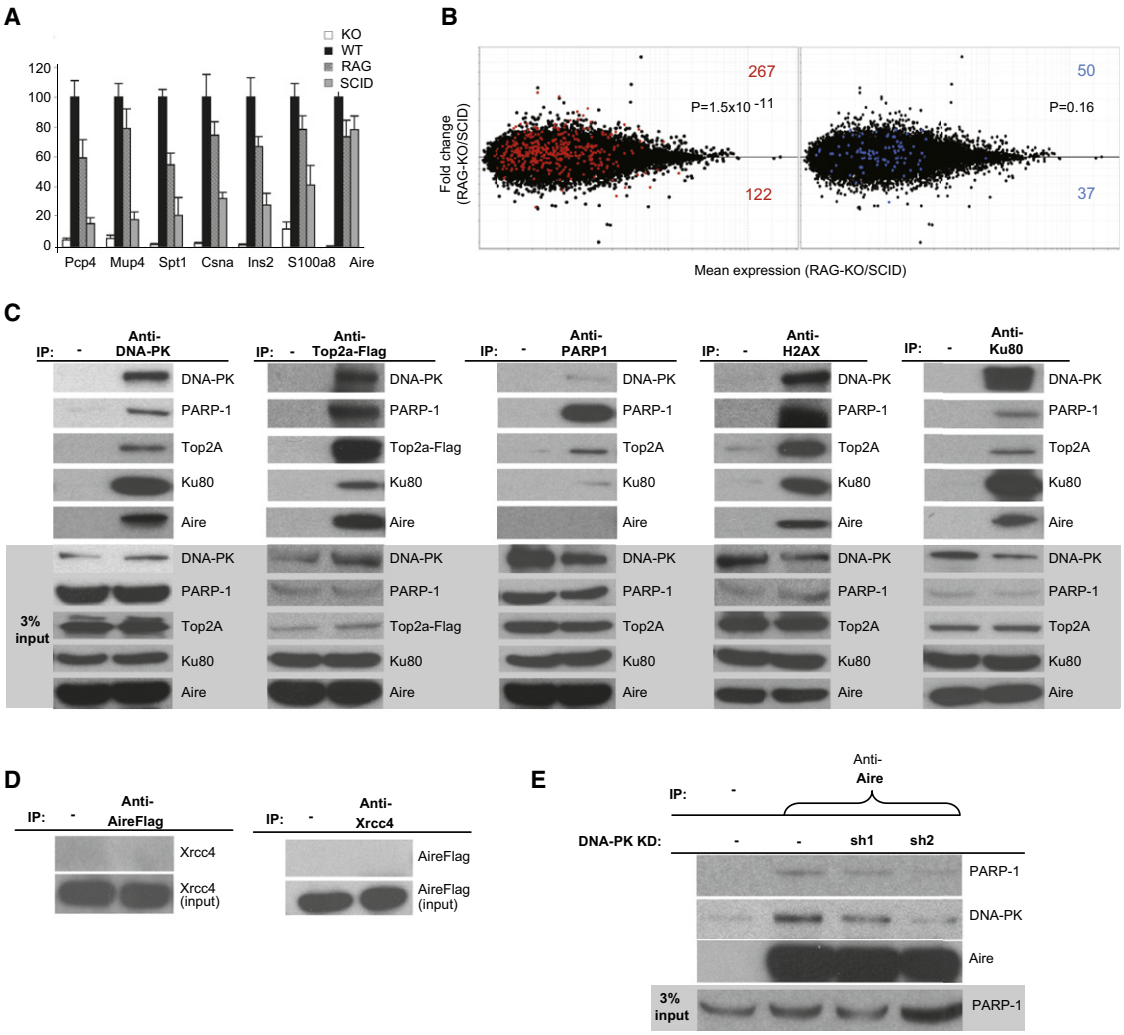

Figure 4. DNA-PK Controls PTA Gene Expression In Vivo

(A) DNA-PK dependence of Aire-induced PTAs. Relative PTA expression levels in MHCII^{hi} MEC populations isolated from Aire-KO, WT, reconstituted RAG-KO or reconstituted SCID mice, measured by RT-PCR analysis.

(B) The global impact of DNA-PK on expression of Aire-dependent genes in MHCII^{hi} MECs. A scatterplot of expression values averaged for two BM-reconstituted RAG-KO or BM-reconstituted SCID mice. Left: highlighted in red are genes whose expression is increased by > 2-fold in WT versus Aire-KO MECs. Right: the same parameters are displayed for genes downregulated > 2-fold in WT versus Aire-KO MECs (in blue). P values from a χ^2 test are indicated.

(C) Characterization of the DNA-PK-based complex. 293T cells expressing Aire and a given putative partner (endogenous or Flag-tagged) were lysed and co-IPs and WBs were performed using the specified Abs (as per Figure 2).

(D) Lack of Aire association with XRCC4. Aire/XRCC4 were IPed from 293T cells and their association was analyzed by WB using the specified Abs.

(E) A requirement for DNA-PK expression to permit Aire to bind to PARP-1. Aire was IPed from 293T cells expressing either normal or shRNA-down-modulated levels of DNA-PK. The PARP-1/Aire association was then analyzed by WB using an anti-PARP-1 Ab. (See also Figure S3.)

the enzyme covalently bound at the site it just clipped, thereby inhibiting religation; these breaks are converted to double-stranded breaks (DSBs) when the replication or transcriptional machineries pass through. The overall effect, then, is to promote DSBs at loci undergoing some degree of replication or transcription (Muslimović et al., 2009). We treated 293T cells with etoposide (or H_2O_2, which induces DSBs by a different mechanism) and assessed the effect on gene expression 24 hr later. To our surprise, there appeared to be a close correspondence between the ability of etoposide (but not of H_2O_2) and of Aire to induce expression of our standard panel of Aire-dependent genes, including *KRT14*, *S100A8*, *ALOX12*, and *IGFL1* (Figure 5A and data not shown). Interestingly, the effects of the two agents were not additive or synergistic (Figure 5A), suggesting that they might impinge on the same pathway. Microarray analysis highlighted the striking correspondence between Aire- and etoposide-induced ($p = 1 \times 10^{-128}$) and repressed ($p = 4 \times 10^{-36}$) loci (Figure 5B, left), confirmed by the largely diagonal disposition of the major cloud in the Aire-versus-control/etoposide-versus-control fold-change/fold-change plot (Figure 5B, right).

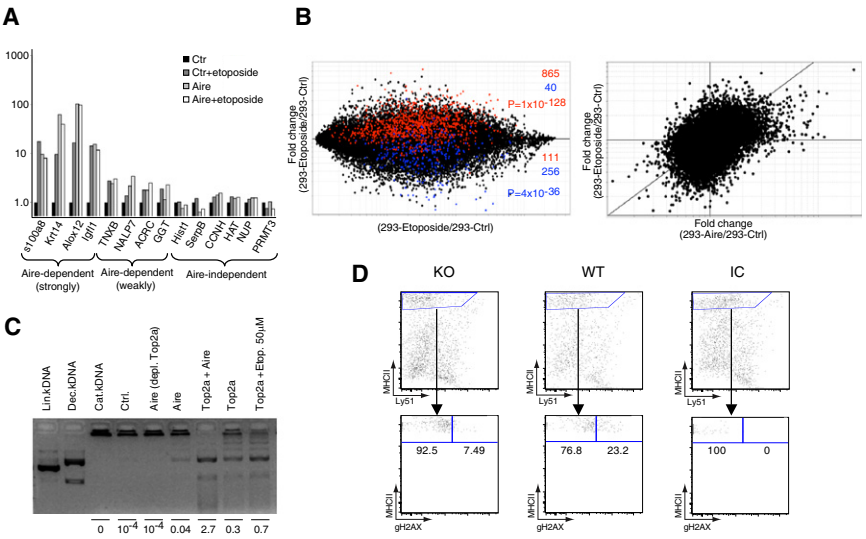

Figure 5. Aire Promotes TOP2a-Introduced DSBs

(A) RT-PCR analysis of the effect of etoposide and Aire on relative levels of mRNA expression for a panel of Aire dependent or -independent genes in 293T cells. Genes are grouped according to the degree of Aire induction.

(B) Global comparison of Aire- and etoposide-induced gene expression in 293T cells. Left: A scatterplot comparing averaged expression values (n = 3) in 293T cells treated either with etoposide (50 μM) or a control medium. Highlighted are genes whose expression was increased (red) or suppressed (blue) by > 2-fold by Aire in 293T cells. P values from a χ^2 test are indicated. Right: A fold-change/fold-change plot comparing the effects of Aire (*x* axis) and etoposide (*y* axis) in 293T cells.

(C) Nuclear lysates prepared either from 293T cells expressing Aire-Flag, TOP2a-Flag, TOP2a-Flag/Aire-myc or from mock-transfected cells (Ctrl) were first pre-cleared or TOP2a-depleted (depl. TOP2a) and then IPed with anti-Flag mAb. The beads containing the respective IPed complexes were then incubated with catenated kDNA (and with either 0 or 50 μM etoposide) for 10 min at 37°C. Samples were resolved on a 1% agarose gel. Numbers below indicate the calculated ratio of decatenated/catenated kDNA for each sample.

(D) WT, but not Aire-KO, MHCII^{hi} MECs contain a subpopulation that has higher γH2AX levels. MHCII^{hi} MECs from either WT or Aire-KO mice were stained either with a γH2AX-specific Ab or with an isotype control (IC) and analyzed by flow cytometry. Displayed dot plots are representative of three experiments.

These findings prompted us to hypothesize that at least part of Aire's mechanism of action might be analogous to that of etoposide: interacting with TOP2a and somehow inhibiting its religation function, thereby promoting the stability of single-stranded clips in the DNA, which degrade to DSBs on passage of the replication or transcriptional machineries. Given that MECs are postmitotic (Gray et al., 2007), the DSBs in this context would be located primarily in regions of transcribing chromatin. To test this hypothesis, we employed a well-established in vitro assay for TOP2-mediated changes in DNA structure based on decatenation of kinetoplast DNA (kDNA). Given the difficulty of producing full-length recombinant Aire, we IPed Aire, TOP2a, or both from transfected 293T cells. As illustrated in Figure 5C, Aire by itself induced slight decatenation of kDNA, but this activity was abolished by preclearing nuclear lysates with anti-TOP2a Abs, indicating that it was not Aire-autonomous and reflected the presence of some TOP2a in the IPed Aire complex. TOP2a by itself induced more extensive decatenation, and this was greatly augmented when Aire was coexpressed with it, even more so than 50 μM etoposide. Thus, Aire does seem to exert an influence analogous to that of etoposide.

Lastly, we used a flow cytometric assay to explore Aire's influence on the quantity of DSBs in isolated MECs ex vivo and monitored levels of γH2AX - an accepted marker of DSBs. Mature (MHCII^{hi}) MECs from Aire-positive mice had a 3- to 4-fold larger population of cells highly expressing γH2AX then did their Aire-negative littermates (Figure 5D).

Aire Regulates Gene Expression at the Level of Pre-mRNA Processing

Our co-IP/MS screen indicated that Aire associates with a number of proteins involved in the regulation of pre-mRNA processing and, more importantly, that specific knockdown of a number of these factors influenced Aire-induced gene expression. This was an unexpected observation that also merited more profound pursuit.

As several mRNA-processing molecules reside in nuclear speckles (Lamond and Spector, 2003) that resemble those containing Aire, we wondered whether the two sets of nuclear bodies overlap. Thymus sections from normal mice were stained with an anti-Aire mAb together with an antibody specific for a marker of nuclear speckles, snRNP116/EFTUD2, a protein also detected in our initial screen (although reciprocal co-IP analysis could not confirm this association). Confocal microscopy revealed that many of the Aire-containing dots did indeed overlap with snRNP-containing nuclear speckles (Figure 6A), consistent with the hypothesis that Aire might be functionally linked to pre-mRNA processing.

That Aire regulates the expression of a large number of genes in MECs and other cell types has by now been shown by several groups. However, all of these reports quantified gene transcripts using probes that detect fully mature mRNAs. To compare Aire's effect on spliced (mature mRNA) versus unspliced (pre-mRNA) transcripts, we transfected 293T cells with either an Aire-expression plasmid or a control plasmid, and 48 hr later isolated total RNA and treated it extensively with DNase I. Spliced and unspliced transcripts of several Aire-dependent genes were quantified by RT-PCR using primers within particular exons or spanning splice junctions, respectively (Figure 6B). Strikingly, while Aire substantially increased the levels of all of the mature transcripts examined (e.g., *S100A8, KRT14*), it had no or only minimal effect on the corresponding immature forms (Figure 6C). To determine whether this observation was also true in vivo, we followed a similar protocol using MECs isolated from thymi of Aire-KO mice versus their WT littermates. The results were very similar: for all four PTA genes examined, Aire strongly

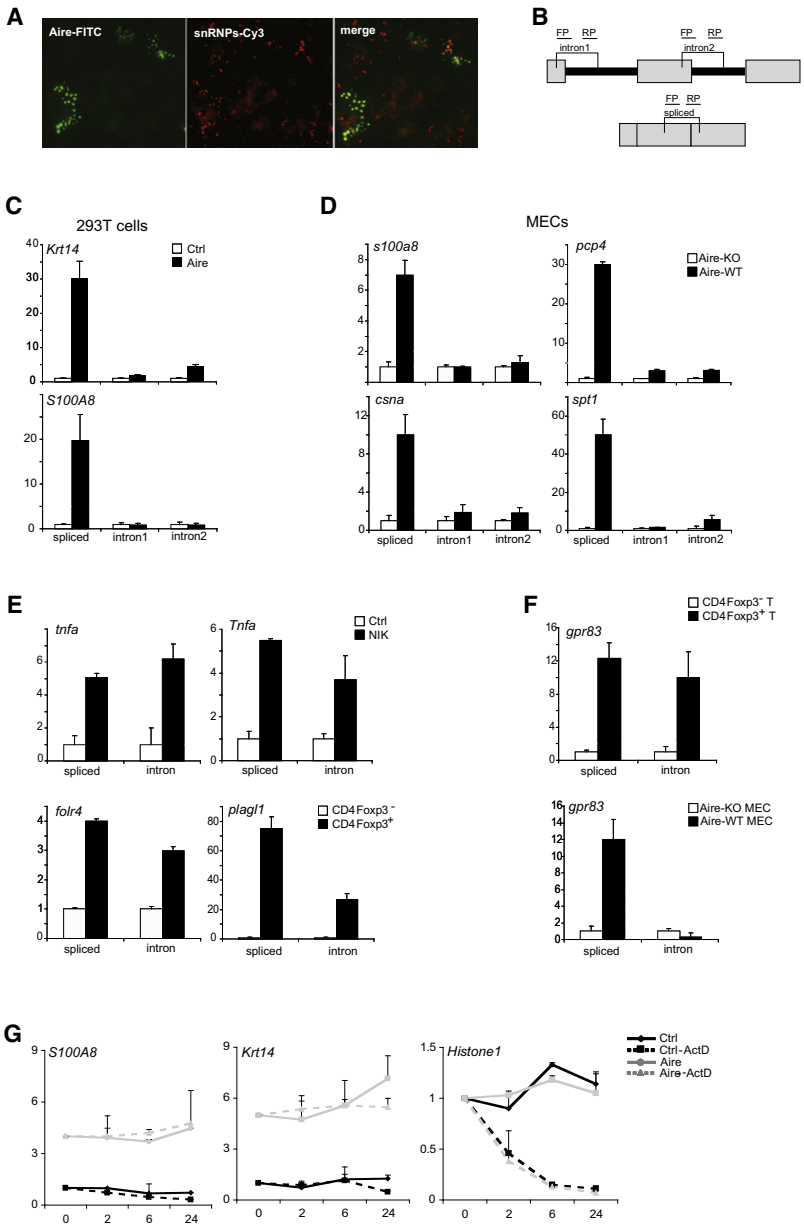

Figure 6. Aire regulates Gene Expression at the Level of Pre-mRNA Processing

(A) Colocalization of Aire with nuclear speckles containing pre-mRNA processing proteins. Imunohistochemical staining of thymus sections for Aire (green) and snRNP116/EFTUD2 (red).

(B) Schematic of probe localizations.

(C) RT-PCR analysis of the relative expression of spliced (mature mRNA) and unspliced (pre-mRNA) forms of the Aire-dependent genes *KRT14* and *S100A8* in 293T cells either expressing or not expressing Aire.

(D) RT-PCR analysis of the relative expression of spliced and unspliced forms of the Aire-dependent genes *S100A8, Pcp4, Csna*, and *Spt1* in MECs isolated from WT or Aire-KO mice.

(E) Spliced versus unspliced transcripts for NFkB- and FoxP3-dependent genes. Upper panel: Expression levels of transcripts for the *TNF-a* gene in 293T (right) and 1C6 (left) cell lines either expressing or not expressing NIK. Lower panel: Expression levels of Foxp3-dependent transcripts for *Folr4* (right) and *Plagl1* (left) genes in sorted $CD4^+$ cells either expressing or not expressing Foxp3GFP.

(F) Divergent mechanisms of Aire and Foxp3 control of expression of the same gene, *Gpr83*. Expression levels of *Gpr83* in sorted $CD4^+$ cells (left) and MECs (right) either expressing or not expressing Foxp3 or Aire, respectively.

(G) RT-PCR analysis of the decay of Aire-dependent (*KRT14* and *S100A8*) and Aire-independent (*Histone 1*) mRNA transcripts in 293T cells either expressing or not expressing Aire subsequent to treatment with either ActD or with control medium for indicated time periods (in hrs; *x* axis). See also Figure S4.

increased levels of the spliced form of the mRNA, but minimally affected the unspliced versions (Figure 6D). We subsequently performed careful quantitation of pre-mRNA and mRNA levels for a few of the genes (Figure S4), which revealed that the levels of mature mRNA were dramatically higher (30- to 400-fold) than those of pre-mRNA. There was a slight bias for transcripts picked up by probes at the 5′, rather than 3′, end, but we do not know whether or not this will prove to be a general phenomenon.

To confirm that this effect was specific to Aire and was not an experimental artifact or a general feature of all transcriptional regulators, we also measured mature mRNA and pre-mRNA levels for genes induced by two proteins, NFkB and Foxp3, considered to function as conventional transcription factors. In both the 293T and 1C6 cell lines, both the spliced and unspliced forms of *Tnfa* and *Ccl2* transcripts were augmented by transfection of a construct encoding the NFkB-specific activator, NIK (Figure 6E and data not shown). Similarly, both mature and immature mRNA transcripts of the Foxp3-dependent genes, *Folr4* and *Plagl1*, were increased in sorted Foxp3-positive $CD4^+$ regulatory T cells (Figure 6E). Lastly, we analyzed the spliced versus unspliced transcript levels for a gene, *Gpr83*, whose expression is controlled by Aire in MECs and by Foxp3 in regulatory T cells. Foxp3 and Aire induced *Gpr83* mature mRNA levels to a similar extent, i.e., ~12-fold; however, only Foxp3 significantly induced *Gpr83* pre-mRNA levels (Figure 6F).

One possible interpretation of these data is that Aire controls the half-lives of the fully mature mRNA transcripts of the genes it regulates. We monitored the kinetics of decay of several Aire-induced transcripts by treating 293T cells, either expressing or not expressing Aire, with Actinomycin D (ActD) for the indicated time periods, and quantifying mRNA levels by RT-PCR (Figure 6G). Interestingly, transcripts of all of the Aire-dependent genes examined (*KRT14*, *S100A8*, and *IGFL1*) showed relatively long mRNA half-lives of >12 hr, in contrast to those of Aire-independent genes such as *histone 1*, which (as expected) had a half-life of ~2 hr. At the 24 hr time point, Aire did somewhat enhance the stability of the transcripts from Aire-dependent genes but,

the relatively long half-lives of these mRNAs argue that this mechanism cannot account for their several-fold (>7) upregulation in the <48 hr time period encompassed in the experiments described above.

DISCUSSION

Although there have been many indications that Aire is some kind of transcriptional regulator, it does not appear to be a conventional transcription factor that binds to promoters and induces or represses initiation of pre-mRNA transcripts. To further decipher Aire's mechanism of action, we set out to identify its partners, directly or indirectly associated, and to evaluate their functional significance. We were confronted with a wealth of putative Aire-associated proteins, consistent with a previous report that it partakes in large multiprotein complexes >670 kDa in size (Halonen et al., 2004). Validation and functional studies highlighted two clusters of protein partners, yielding important insights into how Aire operates.

An Aire Complex Centered on DNA-PK

A putative Aire-containing complex focused on DNA-PK attracts attention for a number of reasons. First, DNA-PK was scored by far the highest in our composite MS analysis (Figure 1B). Second, a group of proteins known to interact with each other (DNA-PK, PARP-1, TOP2a, FACT, Ku80, Ku70 and H2AX) (Ju et al., 2006; Heo et al., 2008) was found to form a multimolecular complex with Aire (Figure 4C and data not shown). Third, functional tests revealed most of these proteins to be involved in Aire-mediated regulation of endogenous gene expression (Figures 3A and 3B and Table S1). Importantly, analysis of MECs isolated from reconstituted thymi of SCID mice (i.e., MECs expressing an inactive form of the enzyme) demonstrated that DNA-PK is indeed required for efficient expression of Aire-induced PTA transcripts (Figures 4A and 4B). And finally, analogous to the anticancer drug, etoposide, Aire was found to promote TOP2a-initiated DSBs, which are able in turn to recruit and activate multiple members of this complex, such as DNA-PK, Ku80, PARP-1, and H2AX, as well as several other Aire-associated molecules, including SMC1, TRIM28, MSH2/6, and RUVBL2. All of the proteins in this putative Aire-containing complex have been reported to function as potent transcriptional regulators, often acting in concert within a multiprotein complex. However, the exact molecular mechanisms underlying their influence(s) are still poorly understood. Our lack of understanding may arise from the fact that their functions have been studied primarily in processes other than transcription, such as DNA repair or DNA replication. Given the limited number of studies investigating a possible mode of action of this putative multi-molecular complex (or of etoposide) in transcriptional regulation, one can only speculate about how it drives Aire-dependent gene expression.

TOP2a-introduced DNA breaks ease superhelical tensions generated by advancing RNAPII during transcriptional elongation, which helps it to proceed more smoothly through the unwinding helix, resulting in more efficient mRNA synthesis (Mondal and Parvin, 2001). Such a scenario could explain why most of this group of Aire associates were important for regulation of endogenous genes, while only DNA-PK had a detectable effect on expression of the episomal reporter (Figure 3A). Moreover, TOP2a-introduced DNA DSBs are known to activate DNA-PK, PARP-1 and many other Aire-associated proteins, which may then induce posttranslational modifications of histones and other chromatin-associated proteins in their vicinity (e.g., by phosphorylation or poly-ADP-ribosylation), leading to local chromatin relaxation, which would result in increased efficiency of transcription. An analogous scenario has been proposed for DNA-PK-induced phosphorylation, and the subsequent replacement of histone H2AX, during DNA-damage repair. In this context, DNA-PK acts as part of a complex with PARP1 and FACT (Heo et al., 2008), which have been shown in a number of studies to act as an "eviction complex," promoting gene transcription through histone displacement in front of elongating RNAPII (Belotserkovskaya et al., 2003; Krishnakumar et al., 2008). Alternatively or additionally, the putative Aire/DNA-PK/PARP-1/TOP2a conglomerate may operate in a manner analogous to that recently reported for a DNA-PK/PARP-1/TOP2b complex in the context of hormone stimulation (Ju et al., 2006). TOP2b generates DSBs in the DNA, which activates PARP-1 and DNA-PK, resulting in displacement of nucleosomal linker H1 and local relaxation of the chromatin structure. Therefore, given the reported activities of DNA-PK/TOP2a and their associated proteins in other contexts related to transcriptional regulation, it seems logical to propose that Aire partners with them to relax surrounding chromatin as the transcriptional machinery proceeds along genes.

An Aire Complex Regulating Pre-mRNA Processing Events

Another interesting set of proteins highlighted in our analyses are factors involved in pre-mRNA processing—13 of the top 45 candidates, including RNA helicases, spliceosome constituents, etc. (Figures 2B and 2C). This observation is not entirely surprising given that Aire often resides in nuclear speckles (Figure 6A) (Su et al., 2008), subnuclear structures thought to be sites of assembly, modification and/or storage of pre-mRNA processing and several transcription factors (Lamond and Spector, 2003). That pre-mRNA processing may indeed be an important point of impact in Aire's regulation of gene expression was further supported by our data showing that while Aire controls the level of fully spliced mRNAs from a number of genes, it has strikingly little effect on the corresponding unspliced pre-mRNAs, very different from the performance of Foxp3 and NFkB in parallel assays (Figures 6B–6E). The differential influence of Aire could not be readily explained as an effect on mature mRNA stability, and was therefore likely to reflect an impact on splicing and/or related pre-mRNA processing events. This could signify reactivation of RNAPII stalled at the 5′ end, as previously suggested (Oven et al., 2007), suppression of improperly initiated transcripts, and/or stabilization of short-lived pre-mRNAs through proper splicing, as recently observed for lipopolysaccharide induction of macrophage transcripts (Hargreaves et al., 2009).

In weighing the significance of Aire's association with this set of proteins, it is important to keep in mind that transcript elongation and pre-mRNA processing occur simultaneously in

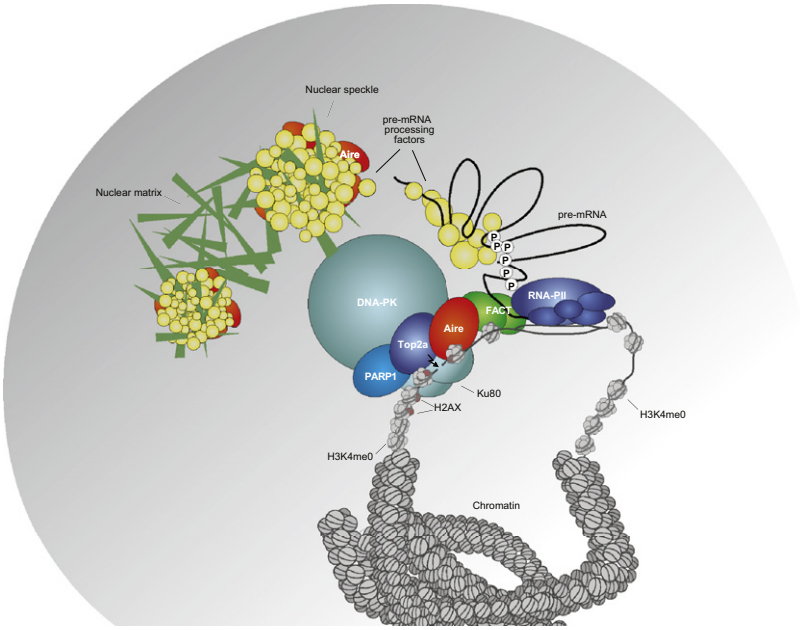

Figure 7. A Model for Aire-Mediated Regulation of Gene Expression

After arriving in the nucleus through the nucleopore complex, Aire preferentially localizes to relatively transcriptionally inert chromatin regions, binding to hypomethylated H3 tails. Within these regions, it interacts with TOP2a to promote DNA double-stranded breaks, activating DNA-PK and other partners, in turn attracting chromatin remodeling complexes. Several of the same Aire-associates (including DNA-PK and TOP2a) might also participate in the so-called "eviction complex" that removes an H2A-H2B dimer in front of RNAPII as it proceeds along nucleosome-packaged DNA, and reassembles the octamer behind, thereby enhancing elongation efficiency.

Another Aire-containing complex would promote the accumulation of fully mature mRNA by re-activating RNAPII stalled at the 5′ end, by suppressing improperly initiated transcripts, and/or by stabilizing short-lived pre-mRNAs through proper splicing.

eukaryotes, and that they are tightly coupled spatially (Moore and Proudfoot, 2009). As a consequence, these two processes influence each others' rate and/or efficiency, clearly indicating that pre-mRNA processing is not simply downstream of transcriptional elongation. Physical intercoupling of the two processes is well illustrated by a number of studies documenting interactions between pre-mRNA processing factors (including SFRs and snRNPs) and the large subunit of RNAPII (CTD) as well as other complexes regulating elongation (Das et al., 2007). Functional intercoupling operates in both directions. On the one hand, truncation of CTD prevented targeting of the splicing machinery to a transcription site and severely reduced pre-mRNA processing (Misteli and Spector, 1999). On the other hand, depletion of the splicing factor SFRS2 (SC35) provoked a dramatic decrease in transcripts from several genes (Lin et al., 2008).

A Model for Aire Regulation of Gene Expression

Aire has been shown to bind specifically to H3 tails unmethylated at the K4 residue, a mark of relatively inactive chromatin. We propose that Aire induces these chromatin regions to "wake-up" transcriptionally and to bridge it to factors that promote transcript elongation and pre-mRNA processing. Such a mechanism seems very plausible given recent findings on CHD1, which activates gene expression by specifically binding to H3K4me3 via its tandem chromodomains and bridging it to several factors (e.g., FACT, snRNPs, DDX9, etc.) involved in transcript elongation and pre-mRNA processing (Sims et al., 2007).

A speculative model illustrating such a bridging function for Aire is depicted in Figure 7. According to this scenario, Aire would preferentially localize to chromatin regions that are relatively transcriptionally inert, binding to their unmethylated H3K4 tails and being excluded from active chromatin decorated with multi-methylated H3K4 tails (Org et al., 2008; Koh et al., 2008). Within these regions, Aire would interact with TOP2a and, analogous to the mechanism of action of etoposide, would "freeze" the enzyme covalently bound to the single-stranded clips it just introduced; these breaks would be converted to DSBs upon passage of the transcriptional machinery (Muslimović et al., 2009). These steps readily explain how Aire can target a battery of weakly expressed genes, differing in identity in different cell-types, as recently highlighted. The DSBs would mobilize γH2AX, and activate DNA-PK/Ku80, PARP-1 and other proteins found to associate with Aire (e.g., TRIM28, the cohesin complex, RUVBL2, etc.), processes known to attract chromatin remodeling complexes and thereby alter chromatin architecture (van Attikum and Gasser, 2009). Several of the same Aire-associates (DNA-PK, TOP2, PARP-1, FACT, and Ku) might also participate in the so-called "eviction complex" that removes an H2AX-H2B dimer in front of RNAPII as it proceeds along nucleosome-packaged DNA, and re-assemble the octamer behind, thereby enhancing elongation efficiency (Mahaney et al., 2009). Increasing accumulation of DSBs would also incite the DNA-damage response, putatively providing an explanation for Aire's induction of MEC death (Gray et al., 2007). Another Aire-containing complex would promote the accumulation of fully mature mRNA by re-activating RNAPII stalled at the 5′ end (Oven et al., 2007), by suppressing improperly initiated transcripts, and/or by stabilizing short-lived pre-mRNAs through proper splicing (Hargreaves et al., 2009).

Such a model could rationalize many of the odd features previously noted about Aire control of gene expression. Specifically, it would readily explain how so many loci can be induced or repressed, why the particular genes affected by Aire differ from cell type to cell type, and why different factors and start-sites can be used in Aire-dependent and independent expression of the same gene. Nonetheless, many aspects of the model remain speculative at this point, and beg for experimental evaluation.

EXPERIMENTAL PROCEDURES

Antibodies and Plasmids

Antibodies and expression plasmids were bought commercially or were gifts from investigators, as specified in the Extended Experimental Procedures.

Gene-Expression Profiling

Total RNA, isolated from either transfected cells or sorted MECs was processed into biotinylated cRNA, which was then hybridized to Affymetrix GeneChips. Raw data were processed with the RMA algorithm for probe-level normalization and were analyzed using GenePattern software.

IP and MS of Aire-Containing Molecular Complexes

Postnuclear supernatants obtained from MNase-treated nuclei of Aire-Flag-expressing or control cells were incubated with anti-Flag-coupled Sepharose beads. Following extensive washes, bound proteins were eluted, separated on a 10% SDS-PAGE and stained with Colloidal Blue. Several bands (typically 12–16) were excised from the control and experimental lanes, and were analyzed by LC-MS/MS using an LTQ mass spectrometer (as described in the Extended Experimental Procedures).

Co-IP and Western Blotting

Nuclear extracts from MNase-treated nuclei of transfected cells were incubated with Protein-G Sepharose beads coupled to specified antibodies. Bound proteins were eluted by boiling, separated by SDS-PAGE, and electro-transferred to PVDF membranes, which were blotted with the specified antibodies.

shRNA-Mediated Knockdown of Aire-Associated Protein Candidates

293T cells were cotransfected with pCMV-Aire (or the empty pCMV-2B) vector and a lentivirus vector pLKO.1 containing shRNAs for the specified Aire-associated candidates. cDNA was prepared from total RNA and then used for qPCR analysis of several "diagnostic" Aire-dependent genes (e.g., *Krt14, S100A8,* etc.), which were then normalized against *Hprt* levels in the same sample. Data were quantified using the comparative C_t method.

Mice

Aire-deficient and –sufficient mice (Anderson et al., 2002) were on the B6 and NOD genetic backgrounds. Foxp3-GFP mice were a gift from Dr. A. Rudensky. NOD.CB17-Prkdc^{scid} (SCID) and RAG1 knockout (RAG-KO) mice were purchased from the Jackson Laboratory. Mice were housed at the ARCM center at Harvard Medical School under IACUC approved procedures.

Bone-Marrow Cell Transfers

Bone marrow cells were harvested from femurs and tibiae of 6-week-old standard NOD mice, and then injected intraperitoneally into 2- to 4-day-old SCID or RAG-KO recipients, 6 hr postirradiation at 200 rad. Reconstituted thymi were analyzed 6–8 weeks after transfer.

Analysis of Pre-mRNA versus mRNA

Primers for unspliced forms of a given mRNA were designed as illustrated in Figure 6B. Total RNA, isolated either from transfected cell lines or from sorted cells, was extensively treated with DNaseI, to avoid contamination by genomic DNA. RT (using random primers) and subsequent qPCR of spliced and unspliced transcripts was done as described above.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and two tables and can be found with this article online at doi:10.1016/j.cell.2009.12.030.

ACKNOWLEDGMENTS

We thank Drs. N. Hacohen, D. Gray, D. Mayerova, S. Haxhinasto, and E. Husebye for materials, techniques, and/or discussions; and N. Asinovski, K. Leatherbee, Dr. A. Ortiz-Lopez, C. Campbell, G. Buruzula, J. LaVecchio, and J. Chen for technical assistance. This work was supported by the Joslin Diabetes Center's NIDDK-funded DERC core facilities. J.A. was supported by fellowships from the Juvenile Diabetes Research Foundation (JDRF#: 10-2007-601 and 03-2005-1138) and M.G. by fellowships from the Campbell and Hall Charity Fund, The Harold Whitworth Pierce Charitable Trust, and the Philippe Foundation.

Received: April 2, 2009
Revised: November 13, 2009
Accepted: December 7, 2009
Published: January 7, 2010

REFERENCES

Anderson, M.S., Venanzi, E.S., Klein, L., Chen, Z., Berzins, S.P., Turley, S.J., von Boehmer, H., Bronson, R., Dierich, A., Benoist, C., and Mathis, D. (2002). Projection of an immunological self shadow within the thymus by the aire protein. Science *298*, 1395–1401.

Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., and Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. Science *301*, 1090–1093.

Bottomley, M.J., Collard, M.W., Huggenvik, J.I., Liu, Z., Gibson, T.J., and Sattler, M. (2001). The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. Nat. Struct. Biol. *8*, 626–633.

Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P., and Reed, R. (2007). SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. Mol. Cell *26*, 867–881.

Derbinski, J., Pinto, S., Rösch, S., Hexel, K., and Kyewski, B. (2008). Promiscuous gene expression patterns in single medullary thymic epithelial cells argue for a stochastic mechanism. Proc. Natl. Acad. Sci. USA *105*, 657–662.

Gray, D., Abramson, J., Benoist, C., and Mathis, D. (2007). Proliferative arrest and rapid turnover of thymic epithelial cells expressing Aire. J. Exp. Med. *204*, 2521–2528.

Guerau-de-Arellano, M., Mathis, D., and Benoist, C. (2008). Transcriptional impact of Aire varies with cell type. Proc. Natl. Acad. Sci. USA *105*, 14011–14016.

Halonen, M., Kangas, H., Rüppell, T., Ilmarinen, T., Ollila, J., Kolmer, M., Vihinen, M., Palvimo, J., Saarela, J., Ulmanen, I., and Eskelin, P. (2004). APECED-causing mutations in AIRE reveal the functional domains of the protein. Hum. Mutat. *23*, 245–257.

Hargreaves, D.C., Horng, T., and Medzhitov, R. (2009). Control of inducible gene expression by signal-dependent transcriptional elongation. Cell *138*, 129–145.

Heessen, S., and Fornerod, M. (2007). The inner nuclear envelope as a transcription factor resting place. EMBO Rep. *8*, 914–919.

Heo, K., Kim, H., Choi, S.H., Choi, J., Kim, K., Gu, J., Lieber, M.R., Yang, A.S., and An, W. (2008). FACT-mediated exchange of histone variant H2AX regulated by phosphorylation of H2AX and ADP-ribosylation of Spt16. Mol. Cell *30*, 86–97.

Johnnidis, J.B., Venanzi, E.S., Taxman, D.J., Ting, J.P., Benoist, C.O., and Mathis, D.J. (2005). Chromosomal clustering of genes controlled by the aire transcription factor. Proc. Natl. Acad. Sci. USA *102*, 7233–7238.

Ju, B.G., Lunyak, V.V., Perissi, V., Garcia-Bassets, I., Rose, D.W., Glass, C.K., and Rosenfeld, M.G. (2006). A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. Science *312*, 1798–1802.

Koh, A.S., Kuo, A.J., Park, S.Y., Cheung, P., Abramson, J., Bua, D., Carney, D., Shoelson, S.E., Gozani, O., Kingston, R.E., et al. (2008). Aire employs a histone-binding module to mediate immunological tolerance, linking

chromatin regulation with organ-specific autoimmunity. Proc. Natl. Acad. Sci. USA *105*, 15878–15883.

Krishnakumar, R., Gamble, M.J., Frizzell, K.M., Berrocal, J.G., Kininis, M., and Kraus, W.L. (2008). Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. Science *319*, 819–821.

Kyewski, B., and Klein, L. (2006). A central role for central tolerance. Annu. Rev. Immunol. *24*, 571–606.

Lamond, A.I., and Spector, D.L. (2003). Nuclear speckles: a model for nuclear organelles. Nat. Rev. Mol. Cell Biol. *4*, 605–612.

Liiv, I., Rebane, A., Org, T., Saare, M., Maslovskaja, J., Kisand, K., Juronen, E., Valmu, L., Bottomley, M.J., Kalkkinen, N., and Peterson, P. (2008). DNA-PK contributes to the phosphorylation of AIRE: importance in transcriptional activity. Biochim. Biophys. Acta *1783*, 74–83.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. Nat. Struct. Mol. Biol. *15*, 819–826.

Mahaney, B.L., Meek, K., and Lees-Miller, S.P. (2009). Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. Biochem. J. *417*, 639–650.

Mathis, D., and Benoist, C. (2009). Aire. Annu. Rev. Immunol. *27*, 287–312.

Misteli, T., and Spector, D.L. (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. Mol. Cell *3*, 697–705.

Mondal, N., and Parvin, J.D. (2001). DNA topoisomerase IIalpha is required for RNA polymerase II transcription on chromatin templates. Nature *413*, 435–438.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688–700.

Muslimović, A., Nyström, S., Gao, Y., and Hammarsten, O. (2009). Numerical analysis of etoposide induced DNA breaks. PLoS ONE *4*, e5859.

Org, T., Chignola, F., Hetényi, C., Gaetani, M., Rebane, A., Liiv, I., Maran, U., Mollica, L., Bottomley, M.J., Musco, G., and Peterson, P. (2008). The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. EMBO Rep. *9*, 370–376.

Oven, I., Brdicková, N., Kohoutek, J., Vaupotic, T., Narat, M., and Peterlin, B.M. (2007). AIRE recruits P-TEFb for transcriptional elongation of target genes in medullary thymic epithelial cells. Mol. Cell. Biol. *27*, 8815–8823.

Pitkänen, J., Doucas, V., Sternsdorf, T., Nakajima, T., Aratani, S., Jensen, K., Will, H., Vähämurto, P., Ollila, J., Vihinen, M., et al. (2000). The autoimmune regulator protein has transcriptional transactivating properties and interacts with the common coactivator CREB-binding protein. J. Biol. Chem. *275*, 16802–16809.

Sims, R.J., 3rd, Millhouse, S., Chen, C.F., Lewis, B.A., Erdjument-Bromage, H., Tempst, P., Manley, J.L., and Reinberg, D. (2007). Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. Mol. Cell *28*, 665–676.

Starr, T.K., Jameson, S.C., and Hogquist, K.A. (2003). Positive and negative selection of T cells. Annu. Rev. Immunol. *21*, 139–176.

Su, M.A., Giang, K., Zumer, K., Jiang, H., Oven, I., Rinn, J.L., Devoss, J.J., Johannes, K.P., Lu, W., Gardner, J., et al. (2008). Mechanisms of an autoimmunity syndrome in mice caused by a dominant mutation in Aire. J. Clin. Invest. *118*, 1712–1726.

Surh, C.D., Ernst, B., and Sprent, J. (1992). Growth of epithelial cells in the thymic medulla is under the control of mature T cells. J. Exp. Med. *176*, 611–616.

van Attikum, A.H., and Gasser, S.M. (2009). Crosstalk between histone modifications during the DNA damage response. Trends Cell Biol. *19*, 207–217.

Villaseñor, J., Besse, W., Benoist, C., and Mathis, D. (2008). Ectopic expression of peripheral-tissue antigens in the thymic epithelium: probabilistic, monoallelic, misinitiated. Proc. Natl. Acad. Sci. USA *105*, 15854–15859.