An RNAi screen for Aire cofactors reveals a role for Hnrnpl in polymerase release and Aire-activated ectopic transcription

Matthieu Giraud[a,b], Nada Jmari[b], Lina Du[a], Floriane Carallis[b], Thomas J. F. Nieland[c], Flor M. Perez-Campo[d], Olivier Bensaude[e], David E. Root[c], Nir Hacohen[c], Diane Mathis[a,1], and Christophe Benoist[a,1]

[a]Division of Immunology, Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115; [b]Department of Immunology, Institut Cochin, Institut National de la Santé et de la Recherche Médicale (INSERM) U1016, Université Paris Descartes, 75014 Paris, France; [c]The Broad Institute of MIT and Harvard, Cambridge, MA 02142; [d]Department of Internal Medicine, Hospital U.M. Valdecilla-Instituto de Formación e Investigación Marqués de Valdecilla, University of Cantabria, 39008 Santander, Spain; and [e]Ecole Normale Supérieure, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8197, INSERM U1024, 75005 Paris, France

Contributed by Christophe Benoist, December 19, 2013 (sent for review December 3, 2013)

Aire induces the expression of a large set of autoantigen genes in the thymus, driving immunological tolerance in maturing T cells. To determine the full spectrum of molecular mechanisms underlying the Aire transactivation function, we screened an AIRE-dependent gene-expression system with a genome-scale lentiviral shRNA library, targeting factors associated with chromatin architecture/function, transcription, and mRNA processing. Fifty-one functional allies were identified, with a preponderance of factors that impact transcriptional elongation compared with initiation, in particular members of the positive transcription elongation factor b (P-TEFb) involved in the release of "paused" RNA polymerases (CCNT2 and HEXIM1); mRNA processing and polyadenylation factors were also highlighted (HNRNPL/F, SFRS1, SFRS3, and CLP1). Aire's functional allies were validated on transfected and endogenous target genes, including the generation of lentigenic knockdown (KD) mice. We uncovered the effect of the splicing factor Hnrnpl on Aire-induced transcription. Transcripts sensitive to the P-TEFb inhibitor flavopiridol were reduced by *Hnrnpl* knockdown in thymic epithelial cells, independently of their dependence on Aire, therefore indicating a general effect of Hnrnpl on RNA elongation. This conclusion was substantiated by demonstration of HNRNPL interactions with P-TEFb components (CDK9, CCNT2, HEXIM1, and the small 7SK RNA). Aire-containing complexes include 7SK RNA, the latter interaction disrupted by *HNRNPL* knockdown, suggesting that HNRNPL may partake in delivering inactive P-TEFb to Aire. Thus, these results indicate that mRNA processing factors cooperate with Aire to release stalled polymerases and to activate ectopic expression of autoantigen genes in the thymus.

autoimmunity | negative selection | hnRNP

The transcription factor Aire plays a very specific role in the management of immunological tolerance, promoting the ectopic expression, in medullary epithelial cells (MECs) of the thymus, of a wide array of peripheral-tissue antigens (PTA) (1). Differentiating T cells are thus exposed to the self-antigens that they will encounter later on, and potentially self-reactive cells are deleted or deviated to regulatory phenotypes as a result. *Aire*-deficient mice and human patients develop autoantibodies and multiorgan autoimmune infiltration, similarly to human patients with a mutation at the *AIRE* gene locus (reviewed in ref. 2).

Aire is an unusual transcriptional regulator. It impacts thousands of PTA-encoding genes whose expression is very differently controlled in parenchymal tissues (3), does not have a clear DNA binding motif, adjusts its targets according to the cell in which it is expressed (4), and takes as cues nonspecific marks of inactive chromatin such as unmethylated H3K4 (5, 6). Indeed, Aire-interacting proteins include factors involved in chromatin structure/modification, transcriptional elongation, and pre-mRNA processing (7). Several lines of investigation have shown that Aire primarily impacts the elongation steps of transcription, in particular by releasing promoter-bound polymerases that remain paused after abortive initiation (8, 9). Correspondingly, Aire has been shown to interact with subunits of the key controller of polymerase release, the positive transcription elongation factor b (P-TEFb) (8, 10, 11). P-TEFb is recruited to stalled initiation complexes in many systems of regulated transcription. It phosphorylates Ser2 on the C-terminal domain (CTD) of Pol-II and the elongation inhibitors DRB sensitivity-inducing factor (DSIF) and negative elongation factor (NELF), dislodging NELF and converting DSIF into an activator (12–14). P-TEFb itself exists in an inactive form when bound to the 7SK small nuclear RNA or to HEXIM1 (15, 16).

To better understand the molecular mechanisms that Aire exploits to promote ectopic gene expression, and to complement the biochemical interaction analyses (7), we undertook a systematic large-scale RNAi screen to identify Aire's transcriptional allies. The results confirmed the importance of factors involved in transcriptional elongation and in RNA processing and identified several previously unrecognized elements of the Aire pathway, which we validated in RNAi knockdown (KD) lentigenic mice. We also identified physical interaction and functional involvement of RNA splicing regulators, in particular Hnrnpl, in elongation control, helping to integrate the diverse classes of Aire-associated proteins previously uncovered.

Results

Identification of AIRE's Functional Allies by a High-Throughput RNAi Lentivirus Screen. To perform the intended RNAi screen for Aire allies, we needed to develop a robust assay for Aire-induced gene expression, with a strong signal of Aire activity, minimal noise to

Significance

The transcription factor Aire controls an unusual mode of transcriptional regulation, important to establish immune tolerance to self, which allows the ectopic expression in the thymic epithelium of RNA transcripts normally restricted to defined tissues. Through a genome-scale RNAi, we identify 51 functional partners of Aire, reaffirming a role for Aire in unleashing stalled transcription, interestingly through involvement of RNA processing factors.

Author contributions: M.G., N.J., T.J.F.N., F.M.P.-C., O.B., D.E.R., N.H., D.M., and C.B. designed research; M.G., N.J., L.D., F.C., T.J.F.N., and F.M.P.-C. performed research; M.G., D.M., and C.B. analyzed data; and M.G., D.M., and C.B. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. E-mail: cb@hms.harvard.edu or dm@hms.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1323535111/-/DCSupplemental.

avoid masking true signals, and amenable to high throughput. It was impossible to use primary MECs, given their very low yield from a mouse thymus, so we opted for transient transfection of an AIRE-expression construct into the AIRE-negative human MEC line 4D6 (17). An assay of AIRE-induced transcription based on a paired chemiluminescence readout (firefly/*Renilla* luciferase) was implemented. Expanding from our earlier experience (18), we tested a panel of luciferase reporter plasmids cotransfected with *AIRE* into 4D6 cells and found that AIRE's impact was highest for plasmids with largely inactive minimal promoters (human CHRNA1, or an isolated TATA box; Fig. 1*A*). In contrast, AIRE did not increase luciferase expression driven by strong eukaryotic promoters (SV40 or CMV).

To validate the specificity of the assay, we verified that transactivation of minP-luciferase luminescence by AIRE did correspond to increased reporter mRNA (Fig. S1*A*), and that Aire expression constructs bearing deletions of certain of the Aire functional domains were indeed impaired in luciferase induction (Fig. S1*B*). The screen strategy is outlined in Fig. 1*B*: infection of 4D6 cells by a lentivirus library in the LKO.1 vector encoding short hairpin RNAs (shRNAs) under the U6 promoter (one virus per well in 384-well plates); selection of transduced cells by puromycin treatment; cotransfection with the AIRE-expressing construct, an AIRE-dependent (minimal promoter driven) firefly luciferase reporter and an AIRE-insensitive (CMV promoter driven) *Renilla* luciferase reporter; and quantitation of luminescence from the two reporters. *Renilla* luciferase provided an internal control for shRNAs that would generically affect transcription or cell viability and served as a normalization for transfection efficiency, which varied with the density of puromycin-surviving cells and thus with the titer of each lentivirus (Fig. S1*C*). The primary screen targeted 2,994 human genes (15,065 nontoxic shRNAs) including basal transcription factors, transcription activators, chromatin modifiers, and RNA-binding proteins. The libraries included 5 different shRNAs for each gene, 55% of which provoke a >70% reduction of their target mRNA, and each of the 384-well screening plates included 40 wells with irrelevant controls or empty shRNAs for a robust estimate of assay noise and plate- or batch variance. *Renilla* and firefly luminescence levels for each plate were fit with a locally smoothed regression, the residuals from the fit for each data point reflecting the effect of the corresponding shRNA on AIRE activity. To estimate the significance of these effects over the whole screen, a null distribution of the firefly/*Renilla* residuals was determined for a set of 340 control shRNAs, and a Z score was computed for each test shRNA. As illustrated in Fig. 1*C*, 1,315/15,065 hairpins met a Z-score criterion of $P < 0.05$, indicating that the screen was indeed identifying shRNAs with a negative impact on AIRE-induced gene expression ($P < 0.007$, Fisher's exact), many inhibitory targets being identified by 2 or more shRNAs (Fig. S1*D*). The screen also identified shRNAs with a positive impact on AIRE activity, but with a distribution similar to that of control shRNAs. To conservatively flag robust candidates in this primary screen and to avoid false positives due to assay noise and off-target effects, we selected those genes identified by 2 shRNAs passing the Z-score criterion, or one passing if the aggregate of the 5 shRNAs for that gene showed a significantly skewed distribution in a Student *t* test. Applying this rule, we identified 227 candidate genes (~9.2% hit rate) whose knockdown diminished AIRE-activated transcription (Table S1). We then performed a secondary replication screen for these 227 genes (using only those shRNAs that scored in the primary screen). The Z scores showed an excellent replication of the primary screen, confirming the reliability of the high-throughput system (Fig. 1*D*).

It was theoretically possible that these cofactors influenced AIRE-activated transcription only when AIRE was present, or that they modified the transcriptional structure on which AIRE

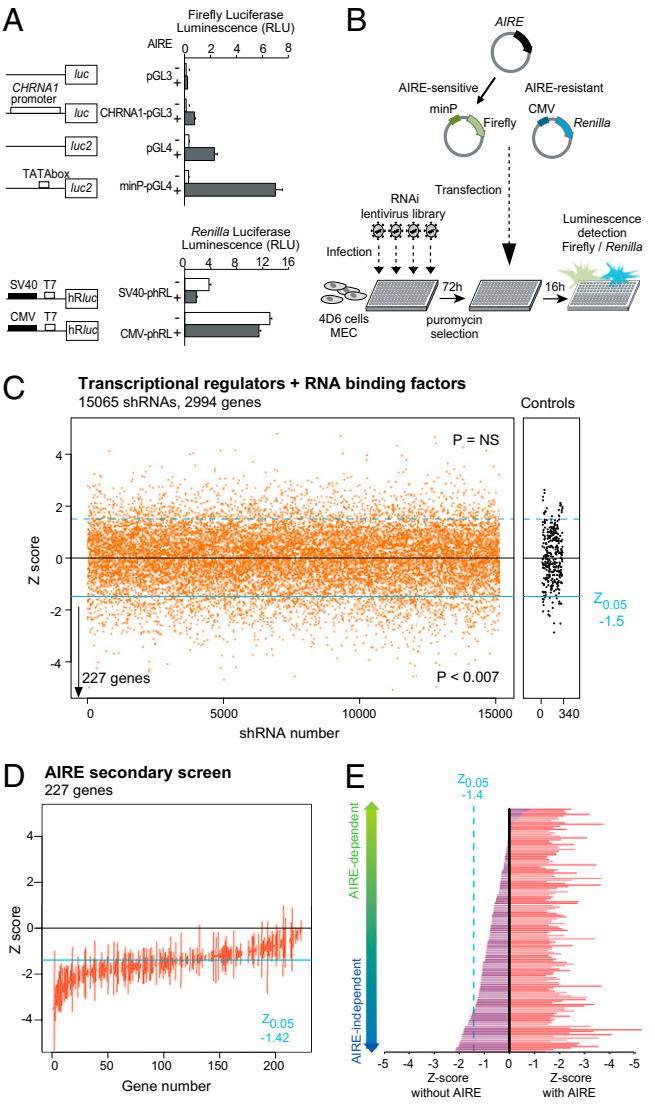

Fig. 1. Luciferase-based RNAi screening for AIRE transactivation partners. (*A*) Effect of AIRE expression in human MEC 4D6 cells transfected with firefly or *Renilla* luciferase reporter vectors, driven by no, minimal, or full promoters. Data are average ± SEM of four transfections. (*B*) Schematic of the screening strategy. The 4D6 cells were plated in 384-well plates, infected with lentivirus-containing shRNAs, selected for puromycin resistance, and transfected with *AIRE*, minP-pGL4, and CMV-phRL. (*C*) Primary screen summary. Z scores for >15,065 shRNAs targeting 2,994 genes or irrelevant controls (gray); negative scores indicate a decrease of AIRE transactivation for the given shRNA. (*D*) Replication screen on the two or more "hit shRNAs" targeting the 227 genes identified in the primary screen. The vertical line indicates the range of Z scores observed. (*E*) Screen for AIRE dependence performed on hit shRNAs targeting the 77 genes most robustly confirmed in the replication screen; values on the *Left* represent the effect of each shRNA when the AIRE-expression vector is replaced by an empty plasmid. There is a spectrum of "AIRE dependence," but genes listed below as AIRE independent are those for which the effect in the absence of AIRE meets the $Z_{0.05}$ criterion.

docks before its recruitment, in which case one might also expect an impact of the knockdown in the absence of AIRE. Thus, we repeated the analysis of shRNAs for the 77 AIRE "allies" identified, but without AIRE. Although the effects of the shRNAs were never as strong as in the presence of AIRE, two types of results were observed (Fig. 1*E*): some shRNAs lost all effect in the absence of AIRE, whereas others did have marked activity even

without AIRE, presumably acting on a baseline state, which AIRE amplified. To further increase the likelihood that the candidates were not false positives, we selected genes with significant expression in 4D6 cells, based on microarray analysis (Table S1), and also dropped from further consideration genes whose knockdown resulted in sequestration of Aire in the cytoplasm (Fig. S1*E* and Table S1). This RNAi screen ultimately led to the identification of 51 genes most likely involved in the Aire-transactivation function; comfortingly, these molecules included several Aire-interaction factors highlighted in our previous screen (7), e.g., DDX5, SFRS1, SFRS3, and MCM6 (Fig. 2*A*).

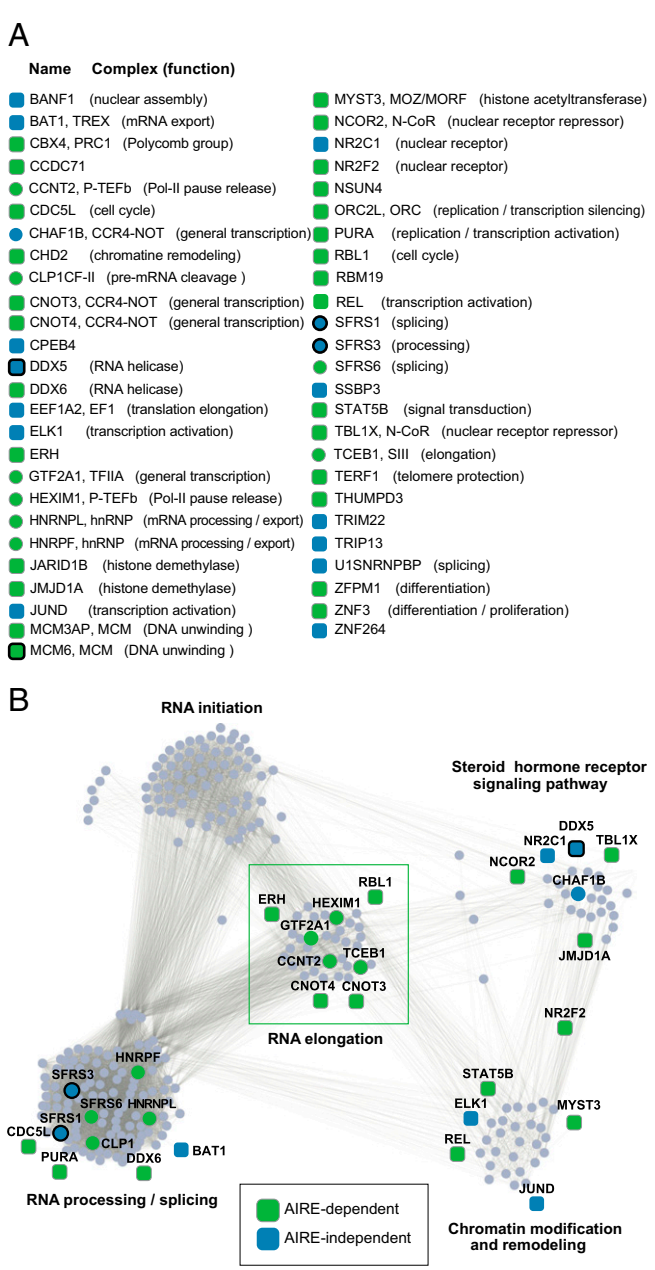

Fig. 2. AIRE's functional allies. (*A*) List of the 51 AIRE allies most robustly identified, the molecular complex to which they belong, and their currently ascribed molecular function. (*B*) AIRE functional allies were positioned on a transcription-related network where the function of each cluster is represented by Gene Ontology terms (RNA elongation, Pcorr = 3.10^{-33}; RNA initiation, Pcorr = 1.10^{-31}; RNA processing/splicing, Pcorr = 1.10^{-200}; steroid signaling pathway, Pcorr = 4.10^{-10}; and chromatin modification and remodeling, Pcorr = 4.10^{-11}). AIRE allies that belong to a cluster are shown as circles, those with predicted interactions with members of each cluster as squares. A thick line denotes factors found to physically interact with Aire in our earlier Mass Spec screen.

AIRE's Functional Allies Are Primarily Involved in RNA Elongation. We then investigated the nature of the 51 confirmed candidates, categorizing the molecular steps of transcription in which they are involved. The STRING database (19), which groups and scores all known molecular and genetic interactions between gene products, was used to generate a network of transcription-related genes. The network was clustered based on the frequency and validity of the interactions by the MCODE algorithm, and the inferred biological function of each cluster was defined by a frequency of Gene Ontology identifiers. Although there were distinct cross-relationships between clusters, as expected, this process grouped distinct clusters predominantly involved in specific functions such as transcriptional initiation, elongation, or pre-mRNA processing (Fig. S2). We mapped onto this framework 38 AIRE allies (Fig. 2*B*; 13 of the 51 were unlisted in the STRING database). Remarkably, 8 of them, all of which belonged to the AIRE-dependent group, mapped to the "elongation" cluster, whereas none mapped to the "initiation" cluster (Fisher's exact $P = 0.0001$). A high proportion of AIRE allies also mapped to the "processing/splicing," "chromatin modification" and "nuclear receptor signaling pathway" clusters. Thus, the functional screen dovetailed with the physical association screen in showing that AIRE impacts several steps in the formation of mature transcripts, and some of the factors were hits in both screens (symbols outlined by thick lines in Fig. 2 *A* and *B*).

Validation of AIRE's Functional Allies. The reporter system used in our screen, although well-suited for a high-throughput process, was rather artificial, so we felt it important to confirm the validity of the candidates by evaluating their influence on AIRE's induction of its normal endogenous targets. After performing the same lentiviral shRNA infection and subsequent *AIRE* transfection as above, we measured the expression of two of AIRE's endogenous targets in 4D6 cells, *KRT14* and *S100A9* (6). Knockdown of the candidates had a clear dampening effect on their induction by AIRE, with a good correlation between the reduction in *KRT14* or *S100A9* mRNA and the magnitude of the shRNA's effect in the luciferase assay (Fig. 3*A*).

To address the functional relevance of these candidates in the natural mouse thymus, we first selected a set of 10 genes for which we could identify shRNAs that effectively reduced AIRE-activated transcription in the 1C6 mouse thymic epithelial cell line (Fig. S3*A* and Table S1). High-titer lentiviruses containing U6-driven shRNAs, and the phosphoglycerate kinase-GFP as a marker of activity, were used to microinfect fertilized oocytes, which were reimplanted into pseudopregnant females (Fig. 3*B*). Of the 15 hairpins tested, 2 pups with lentigene expression and target knockdown >50% were obtained for only *Myst3* and *Hnrnpl* (Table S2). No pups, or only pups with complete extinction of the lentigene, or with little to no effect on the target transcript, were obtained for shRNAs targeting most of our candidates, likely reflecting a lethal effect or competitive disadvantage of the knockdown (success rates for control lentigenics were far greater). GFP$^+$ MECs were sorted from these mice, gene-expression profiles were generated, and the changes induced by each knockdown (relative to averages from three control lentigenics) were plotted in relation to Aire's transcriptional impact (Fig. 3*C*). As indicated by the vertical shift of transcripts with an *Aire*-WT/KO ratio >2 (Fig. 3*C*, red dots), a clear relationship was observed with knockdowns for *Myst3* and *Hnrnpl*, as the vast majority of genes activated by Aire were underexpressed in the corresponding knockdowns (although not all transcripts sensitive to Myst3 and Hnrnpl were Aire responsive).

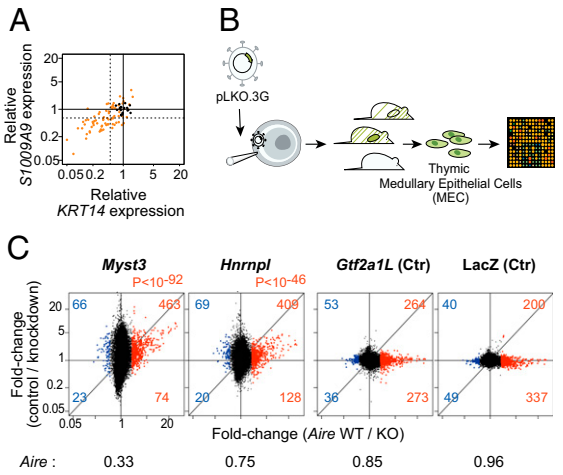

Fig. 3. Validation of AIRE's functional allies. (*A*) Effect of hit shRNAs on AIRE transactivation of two AIRE targets in 4D6 cells (*KRT14* and *S100A9*); orange dots are for test shRNAs, black dots for control shRNAs. (*B*) Schematic of the lentigenic knockdown strategy. Each selected shRNA was transferred to PLKO.3G, packaged to $>10^9$ viral titer for microinfection of fertilized oocytes under the zona pellucid. Resulting offspring were selected for activity of the lentigene in >10% of somatic cells, and GFP$^+$ MECs were sorted and profiled. (*C*) MEC gene expression profiles for mice with knockdown expression of *Myst3* and *Hnrnpl*, or Aire-irrelevant controls *Gtf2a1L* and LacZ. Control shRNA lentigenes targeting RFP were used as reference, and the results are shown as a FoldChange relative to RFP controls (*y* axis) vs. Aire's transcriptional footprint (*x* axis; from a comparison of *Aire*-KO and WT MECs). *Myst3* and *Hnrnpl* KD effects on Aire-induced genes are shown in red. *Aire* expression values from microarray profiling of the knockdown lentigenics are shown normalized to those of the RFP control.

This deviation was highly significant (χ^2 $P < 10^{-46}$ overall); the significance of the effects on individual genes was also estimated by computing their probabilities based on the variance observed among control lentigenics (Fig. S3*B*). In contrast, no deviation in the expression of Aire-responsive genes was observed in MECs expressing shRNAs targeting the *Gt2a1l* or LacZ control genes. In addition the expression of *Aire* itself proved to be impaired in the *Myst3* but not in the *Hnrnpl* knockdown (Fig. 3*C*). To confirm the specificity of the effects observed in the *Hnrnpl* knockdown, we compared the changes to those reported by Huang et al. (20) after knockdown of *HNRNPL* in HeLa cells; the results showed a clear reciprocal enrichment of the genes impacted by Hnrnpl (Fig. S3*C*).

In addition, we were able to test one conventional knockout for a factor identified in the screen, Stat5 (21). Gene-expression profiling of MECs from *Stat5a/b* double-deficient mice and matched controls revealed a moderate but significant impact of Stat5 on Aire-dependent transcripts (Fig. S3*D*).

An Hnrnpl Effect on Transcriptional Elongation. HNRNPL is an RNA processing factor from the heterogeneous nuclear ribonucleoprotein (hnRNP) family, with a well-defined role in alternative splicing and in recognition of 3′ UTR motifs (22–25), and it was intriguing that it had such a broad impact on Aire's in vivo footprint. Other members of the hnRNP family, like HNRNPK, impact on transcriptional elongation through interactions with P-TEFb, via binding to the small 7SK RNA (26). We hypothesized that HNRNPL has similar properties and interacts with P-TEFb and its subunits CCNT2 and CDK9 (CCNT2 being a validated hit in our screen; Fig. S3*A*). We observed that there was, among the transcripts affected by *Hnrnpl* knockdown in MECs, a strong overlap with the genes controlled by CCNT2 (as determined by profiling of *CCNT2* knockdown 4D6 cells) and by CDK9 (determined by inhibition of CDK9 by Flavopiridol) (Fig. 4*A*). In addition, we could show an interaction between endogenous HNRNPL and tagged versions of CDK9 and CCNT2 in transduced HEK293 cells (7) (Fig. 4*B*). Binding of HNRNPL to the endogeneous P-TEFb complex was also found by reciprocal coimmunopreciptations between HNRNPL and CDK9 (Fig. 4*C*).

P-TEFb is inhibited by binding to the 7SK small nuclear RNA (15, 16) and/or to HEXIM1, which also interacts with the small 7SK RNA on the inactive P-TEFb. The deinhibition of P-TEFb at the transcriptional start sites (TSS) can be regulated by various factors, notably c-Myc, BRD4, or SFRS2, the latter a serine/arginine-rich (SR) splicing factor, reflecting the functional link between the splicing machinery and transcriptional elongation (27–29). We asked whether HNRNPL also interacted with P-TEFb inhibitory elements. Reciprocal coimmunoprecipitation was found between HNRNPL and HEXIM1 (Fig. 4*C*). In

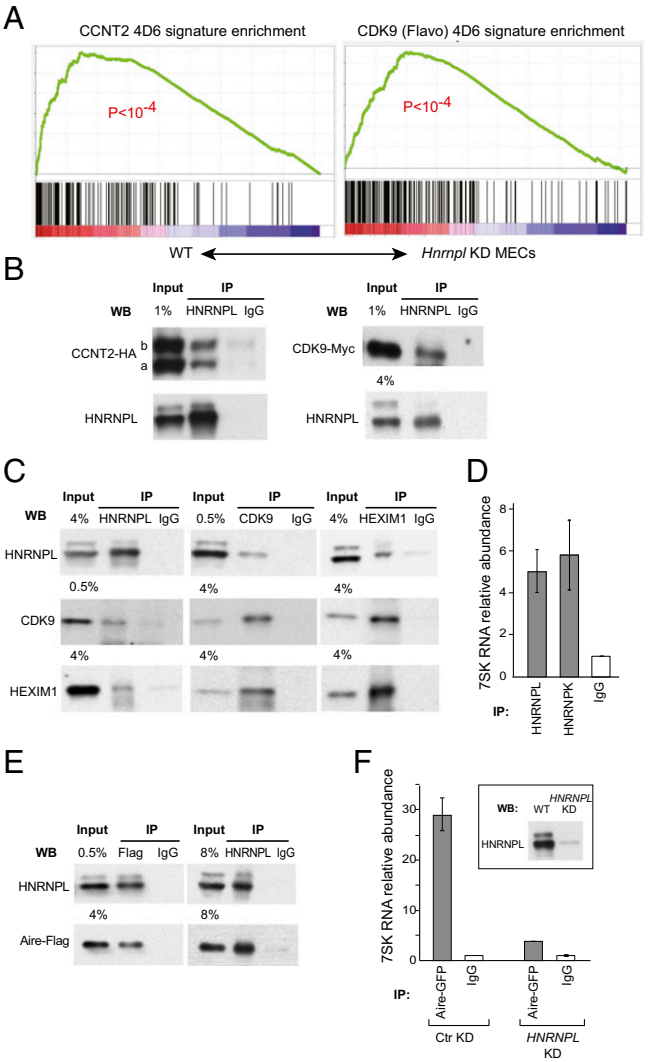

Fig. 4. Cooperation of Hnrnpl with the inactive P-TEFb and Aire. (*A*) Gene Set Enrichment Analysis in the WT vs. *Hnrnpl* KD ranked expression dataset from primary MECs, of the genes impacted by Flavopiridol or *CCNT2* knockdown in 4D6 cells. (*B*) Coimmunoprecipitation of HNRNPL and Myc-tagged CDK9 or HA-tagged CCNT2 in HEK293 cells. CCNT2 is expressed as two isoforms, CCNT2a and CCNT2b. (*C*) Reciprocal coimmunoprecipitations of endogenous HNRNPL, CDK9, and HEXIM1 in HEK293s. (*D*) RT-PCR quantitation of 7SK RNA in HNRNPL and HNRNPK immunoprecipitates, relative to control IgG. Average ± SEM of three experiments. (*E*) Reciprocal coimmunoprecipitation of HNRNPL with the Flag-tagged Aire. (*F*) Quantitation of 7SK RNA bound to Aire complex, as affected by *HNRNPL* knockdown, in stably transduced Aire-GFP 4D6 cells.

addition, 7SK RNA was enriched in HNRNPL immunoprecipitates (Fig. 4*D*) and in HNRNPK precipitates as expected (30). The amount of 7SK associated with HNRNPL increased upon transcription blockade, as it does for HNRNPK (30).

If Hnrnpl is a necessary Aire ally, and if it interacts with transcriptional control components that are also important for Aire's activity, how do Aire and Hnrnpl interact? Reciprocal coimmunoprecipitations showed a physical interaction between HNRNPL and Aire in HEK293 cells (Fig. 4*E*). Importantly, we found that anti-Aire coprecipitated a readily detectable amount of 7SK RNA, but that this was decreased when *HNRNPL* was knocked down (Fig. 4*F*). Thus, HNRNPL plays an important role in the interaction of Aire with 7SK-containing complexes. That Aire is functionally involved with inactive P-TEFb is consistent with the deleterious effect of *HEXIM1* knockdown on AIRE-induced expression in the screen.

Discussion

Aire-dependent ectopic gene expression in MECs operates on an array of genes that are normally controlled by a wide variety of transcription factors in diverse tissues, thus posing an unusual challenge to gene regulation. To identify the cast of molecules that cooperate with Aire to perform this function, we conducted an RNAi screen that combined a high-throughput in vitro component with validation in lentigenic mice. The results further connect the activity of Aire with elongation control, and to lifting of the paused polymerases, and help bridge the relation between splicing- and elongation-control factors.

The disparity in the number of initiation and elongation factors identified in our shRNA screen was quite striking. In hindsight, this is perhaps not surprising, because Aire is specifically effective on reporters with enhancer-less minimal promoters, or on promoters for which key initiation factors are absent. Our earlier work had proposed a potential role for Aire in boosting elongation, through the recruitment of the DNA-PK/TOP2a/FACT "eviction" complex, thought to facilitate transcription by relieving torsional constraints during Pol-II progression (7). Oven et al. (8) had also shown a preferential effect of Aire on elongation, through interaction with P-TEFb, as had our detailed analysis of Aire-regulated transcripts in MECs (9). To ensure the specificity of the factors identified in the RNAi screen, we cotransfected an AIRE-insensitive reporter, which ruled out factors with generic effects on all transcription. Thus, some ubiquitously active factors that are also important for AIRE activity may have been missed, and this may explain why some important factors of transcriptional elongation, such as CDK9 or CCNT1, did not score as hits.

If Aire and its allies partake in pause release, how is this accomplished at the biochemical level? Two allies identified in the screen are at the center of Pol-II release. CCNT2, a subunit of P-TEFb, was highlighted in the screen, and was found to interact with Aire. The elongation inhibitor HEXIM1 was also a hit in the screen, two shRNA hairpins strongly decreasing AIRE's activity on both the transfected reporter and endogenous targets. This result may seem paradoxical at first, as HEXIM1 actually inhibits P-TEFb by sequestering it within the nucleoplasm, and Oven et al. (8) reported the opposite effect: activation of Aire activity by transient RNAi repression of Hexim1 in 1C6 cells versus stable RNAi in our screen, a difference which may be related to different limiting factors in the two systems. A requirement for HEXIM1 is compatible with a model in which AIRE, like the TAT protein, targets for activation the inactive P-TEFb, bound to 7SK RNA and HEXIM1 (31). Increasing the pool of free active P-TEFb by knocking down *HEXIM1* would lead to a relative decrease in AIRE's activity. The necessity for AIRE to recruit the inactive 7SK RNA-associated P-TEFb complex to activate transcription is consistent with the fact that the 7SK RNA coprecipitates with Aire, as do CDK9 and HEXIM1.

A number of mRNA processing factors were identified in our screen, therefore confirming the involvement of the splicing machinery in the Aire transactivation function, as first demonstrated in a proteomic interaction screen for Aire's partners (7). It is well established that tight connections between splicing and transcription exist, in particular with the release of Pol-II pausing (27–29). Of the 10 splicing factors that were validated in our screen, two were also identified in the Aire interactome, the two SR-proteins SFRS1 and SFRS3. Relevant to a role for RNA processing factors in transcriptional elongation, SFRS1 and SFRS2 were reported to recruit the inactive 7SK-associated P-TEFb complex to the polymerase and to promote the dissociation of 7SK RNA from P-TEFb (29). The role of SFRS1 in recruiting and activating the inactive P-TEFb complex is consistent with the notion that 7SK RNA-associated P-TEFb is the substrate of the Aire-dependent machinery to activate paused polymerases. A similar explanation may well hold for HNRNPL, whose function was validated in MECs of lentigenic mice. HNRNPL is primarily known for its major known role in alternative splicing and in the recognition of 3′ UTR motifs (22, 24, 25). It has also been shown to interact with the Mediator complex, therefore coupling mRNA processing to transcription (20). Here, analyses of the expression data from *Hnrnpl* knockdown mice revealed that Hnrnpl preferentially stimulates the expression of genes (some of which are Aire insensitive) whose transcription is also affected by inhibition of CDK9 (flavopiridol inhibition), i.e., genes under dominant elongation control. The interaction of HNRNPL and P-TEFb was confirmed biochemically, including to the 7SK inhibitory RNA (in keeping with the demonstration that several hnRNPs can trap 7SK RNA and alter the balance of active to inactive P-TEFb) (30, 32).

A key to conceptually integrating the interplay between Aire and P-TEFb factors and inhibitors may lie in the observation (Fig. 4*F*) that Aire interacts with 7SK-containing complexes, and that this interaction requires HNRNPL, leading to the following speculative scenario: Aire binds to 7SK- and HEXIM1-containing inactive P-TEFb complexes at the position of stalled polymerases, which it helps activate in concert with other factors, and perhaps in combination with short promoter-associated transcripts, by analogy to HIV TAT and to the model of Ji et al. (29). HNRNPL might help recycle P-TEFb from the the site of transcription termination where hnRNP factors were shown to facilitate pre-mRNA 3′ end processing (33) also facilitated by CLP1, a cleavage and polyadenylation factor, which was also a hit in the RNAi screen.

Thus, and although much work remains to elucidate the interplay between Aire's allies in enabling ectopic transcription, the set of characters that has been identified here should be instrumental in piecing the puzzle together.

Experimental Procedures

Full methods are detailed in *SI Experimental Procedures*.

Luciferase-Based RNA-Interference Screening. The screen was performed using the lentiviral library prepared, titered,and arrayed as described (34) as schematized in Fig. 1*B*, in 384-well plates and in duplicate, using automated pipetting robotics.

Lentigenic Mouse Generation. Selected shRNAs from the mouse screen were transferred from pLKO.1 to pLKO.3G with GFP as a reporter gene. Lentivirus stocks were generated in 293-FT cells, concentrated by ultracentrifugation to very high titer ($>10^9$ infectious unit/mL) for microinfection into fertilized oocytes (20–40 per construct) (35). Pools of 30 transduced oocytes were reimplanted into pseudopregnant females. Newborns with GFP expression were selected and the shRNA present in each pup was identified by PCR amplification and sequencing (Table S3).

Network Modeling. A transcription-related network was generated using the STRING database (http://string-db.org/). This search was seeded, blind to AIRE cooperants, by introducing factors known to be important in various transcriptional processes, secondarily incorporating all other factors with which

these seeds interact. Within this network, proteins with a high density of interaction were clustered and the biological function of these clusters identified by a Gene Ontology analysis.

Gene Expression Profiling. Total RNA, isolated from sorted MECs of 4-wk-old individual mice were profiled as described on Affymetrix Mouse Gene ST1.0 arrays.

Coimmunoprecipitations. HEK293 cells were transfected with CCNT2, CDK9, and Aire expression vectors. Nuclear extracts were prepared after 48 h, protein complexes captured with specific antibodies on magnetic beads (ActiveMotif) followed by immunoblotting. For 7SK RNA quantitation, the immunoprecipitated RNA was extracted by adding TRIzol directly on the beads and retrotranscribed with a specific RT primer.

ACKNOWLEDGMENTS. We thank Drs. J. Ihle for mice; N. Fujikado, J. Lopes, and S. Silver for materials or advice; H. Le for help with the screen; and S. Davis and H. Paik for computation. This work was supported by National Institutes of Health DK060027 and AI088204 (to C.B. and D.M.), ANR-2011-CHEX-001-R12004KK and PCIG9-GA-2011-294212 (to M.G.), National Institute of Allergy and Infectious Diseases U01AI074575 (to N.H.). M.G. received fellowships from the Campbell and Hall Charity Fund, the Harold Whitworth Pierce Charitable Trust, the Philippe Foundation, and "la Fondation pour la Recherche Médicale."

1. Kyewski B, Klein L (2006) A central role for central tolerance. *Annu Rev Immunol* 24: 571–606.
2. Mathis D, Benoist C (2009) Aire. *Annu Rev Immunol* 27:287–312.
3. Venanzi ES, Melamed R, Mathis D, Benoist C (2008) The variable immunological self: Genetic variation and nongenetic noise in Aire-regulated transcription. *Proc Natl Acad Sci USA* 105(41):15860–15865.
4. Guerau-de-Arellano M, Mathis D, Benoist C (2008) Transcriptional impact of Aire varies with cell type. *Proc Natl Acad Sci USA* 105(37):14011–14016.
5. Org T, et al. (2008) The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. *EMBO Rep* 9(4):370–376.
6. Koh AS, 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(41):15878–15883.
7. Abramson J, Giraud M, Benoist C, Mathis D (2010) Aire's partners in the molecular control of immunological tolerance. *Cell* 140(1):123–135.
8. Oven I, et al. (2007) AIRE recruits P-TEFb for transcriptional elongation of target genes in medullary thymic epithelial cells. *Mol Cell Biol* 27(24):8815–8823.
9. Giraud M, et al. (2012) Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. *Proc Natl Acad Sci USA* 109(2):535–540.
10. Žumer K, Plemenitaš A, Saksela K, Peterlin BM (2011) Patient mutation in AIRE disrupts P-TEFb binding and target gene transcription. *Nucleic Acids Res* 39(18): 7908–7919.
11. Yang S, Bansal K, Lopes J, Benoist C, Mathis D (2013) Aire's plant homeodomain(PHD)-2 is critical for induction of immunological tolerance. *Proc Natl Acad Sci USA* 110(5): 1833–1838.
12. Ivanov D, Kwak YT, Guo J, Gaynor RB (2000) Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Mol Cell Biol* 20(9):2970–2983.
13. Fujinaga K, et al. (2004) Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol Cell Biol* 24(2):787–795.
14. Yamada T, et al. (2006) P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol Cell* 21(2):227–237.
15. Nguyen VT, Kiss T, Michels AA, Bensaude O (2001) 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* 414(6861):322–325.
16. Michels AA, et al. (2003) MAQ1 and 7SK RNA interact with CDK9/cyclin T complexes in a transcription-dependent manner. *Mol Cell Biol* 23(14):4859–4869.
17. Fernández E, et al. (1994) Establishment and characterization of cloned human thymic epithelial cell lines. Analysis of adhesion molecule expression and cytokine production. *Blood* 83(11):3245–3254.
18. Giraud M, et al. (2007) An IRF8-binding promoter variant and AIRE control CHRNA1 promiscuous expression in thymus. *Nature* 448(7156):934–937.
19. Jensen LJ, et al. (2009) STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 37(Database issue):D412–D416.
20. Huang Y, et al. (2012) Mediator complex regulates alternative mRNA processing via the MED23 subunit. *Mol Cell* 45(4):459–469.
21. Teglund S, et al. (1998) Stat5a and Stat5b proteins have essential and nonessential, or redundant, roles in cytokine responses. *Cell* 93(5):841–850.
22. Oberdoerffer S, et al. (2008) Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPLL. *Science* 321(5889):686–691.
23. Wu Z, et al. (2008) Memory T cell RNA rearrangement programmed by heterogeneous nuclear ribonucleoprotein hnRNPLL. *Immunity* 29(6):863–875.
24. Hui J, Stangl K, Lane WS, Bindereif A (2003) HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat Struct Biol* 10(1):33–37.
25. Ergun A, et al.; ImmGen Consortium (2013) Differential splicing across immune system lineages. *Proc Natl Acad Sci USA* 110(35):14324–14329.
26. Hogg JR, Collins K (2007) RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. *RNA* 13(6):868–880.
27. Rahl PB, et al. (2010) c-Myc regulates transcriptional pause release. *Cell* 141(3): 432–445.
28. Lin S, Coutinho-Mansfield G, Wang D, Pandit S, Fu XD (2008) The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol* 15(8):819–826.
29. Ji X, et al. (2013) SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* 153(4):855–868.
30. Barrandon C, Bonnet F, Nguyen VT, Labas V, Bensaude O (2007) The transcription-dependent dissociation of P-TEFb-HEXIM1-7SK RNA relies upon formation of hnRNP-7SK RNA complexes. *Mol Cell Biol* 27(20):6996–7006.
31. D'Orso I, Frankel AD (2010) RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation. *Nat Struct Mol Biol* 17(7):815–821.
32. Van Herreweghe E, et al. (2007) Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. *EMBO J* 26(15):3570–3580.
33. Millevoi S, et al. (2009) A physical and functional link between splicing factors promotes pre-mRNA 3′ end processing. *Nucleic Acids Res* 37(14):4672–4683.
34. Moffat J, et al. (2006) A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* 124(6):1283–1298.
35. Chen Z, Stockton J, Mathis D, Benoist C (2006) Modeling *CTLA4*-linked autoimmunity with RNA interference in mice. *Proc Natl Acad Sci USA* 103(44):16400–16405.

Supporting Information

Giraud et al. 10.1073/pnas.1323535111

SI Experimental Procedures

AIRE-Transient Transfection. ***Luciferase assay.*** The 4D6 cell line, originally derived from human thymic medullary epithelium from children undergoing cardiac surgery (clone P1.4D6) (1), was maintained in RPMI-1640 medium complemented with 10% (vol/vol) FBS, L-glutamate, and pen/strep antibiotics. Cells were seeded in 96-well plates at the density of 10,000 cells per well or at the specified density. The next day and before transfection, the medium was changed to Opti-MEM I Reduced-Serum Media (Invitrogen). Transient transfection assays were conducted using the Lipofectamine 2000 transfection reagent (Invitrogen) following the manufacturer's instructions, with four replicate wells per condition per experiment in at least three independent experiments. Transfections were carried out in Opti-MEM I Reduced-Serum Media with 10 ng of the pcDNA3.1 vector harboring the cDNA encoding human AIRE (2), murine Aire, or the latter deleted for some functional domains (3). Firefly luciferase vectors: pGL3 (pGL3-Basic), pGL4 (pGL4.10), minP-pGL4 (pGL4.23) (Promega), or the CHRNA1-pGL3 construct (4) were added at 200 ng to the cotransfection mix. The *Renilla* vectors CMV-phRL or SV40-phRL (Promega) were added at 0.4 ng. After 16 h, the transfection medium was removed, the cells were lysed using Glo Lysis Buffer (Promega), and the lysate was transferred to a 96-well white flat-bottom plate (Corning). Luciferase activity was measured with the Dual-Glo Luciferase Reporter Activity system (Promega) using the VICTOR2 Multilabel Plate Reader (PerkinElmer).

Luciferase messenger quantification. The 4D6 cells were seeded in six-well plates at the density of 340,000 cells per well. Transfection was performed as above with the AIRE-expressing vector (200 ng) and the firefly luciferase vector minP-pGL4 (4 μg). Sixteen hours posttransfection, the cells were harvested and total RNA was isolated by TRIzol extraction (Invitrogen). Residual contaminating DNA was removed by DNaseI treatment (TurboDNase; Ambion), followed by phenol–chloroform extraction. First-strand cDNA was synthesized using SuperScript II Reverse Transcriptase (Invitrogen) and oligo-(dT) primers. cDNA was used for subsequent RT-PCR amplification using the Mx3000P QPCR system (Stratagene) and SYBR Green PCR Master Mix kit (Applied Biosystems). Primers for Luciferase and *GAPDH* used for normalization are listed in Table S3.

Luciferase-Based RNA-Interference Screening. ***Primary screen.*** High-throughput screening was performed in duplicate at the RNAi Platform of the Broad Institute of Massachusetts Institute of Technology and Harvard University using two lentiviral shRNA libraries (the RNAi consortium, TRC) directed against 2,129 transcription regulators and 865 RNA-binding proteins (www.broadinstitute.org/rnai/trc). These two libraries consisted of five shRNAs per targeted gene prearrayed in 46 384-well plates. In addition to these plates, we also screened an independent plate of control shRNAs. The 4D6 cells were seeded in 384-well white flat-bottom plates (Corning) at the density of 500 cells per well (40 μL) using a MicroFill Microplate Dispenser (BioTek). The following day, 10 μL of medium containing polybrene was added to reach a final polybrene concentration of 8 μg/mL using a JANUS liquid handling automated workstation (PerkinElmer). Subsequently, 1 μL of lentivirus was added to each well of each screening plate using an Evolution P3 precision dispenser (PerkinElmer). Plates were incubated at 37 °C with 5% CO_2 for 24 h without spinning. The medium was changed with the JANUS to 50 μL of fresh medium containing puromycin (2 μg/mL). We let the puromycin selection of infected cells occur for 72 h and then proceeded to *AIRE* transfection. The medium was changed to a mix (75 μL) containing the Lipofectamine 2000 transfection reagent, 5 ng of the AIRE-expressing vector, 100 ng of the minP-pGL4 vector, and 0.2 ng of the CMV-phRL vector in Opti-MEM I Reduced-Serum Media without Phenol Red (Invitrogen). After 16 h, the transfection mix was removed and 20 μL of the Glo Lysis Buffer was added to each well. After lysis of the cells, 10 μL of the Dual-Glo Luciferase Reagent was added to each well and the Firefly luminescence was quantified using the enhanced luminescence detection mode of the EnVision Multilabel Plate Reader (PerkinElmer). The *Renilla* luminescence was detected similarly after addition of 10 μL of the Dual-Glo Stop and Glo Reagent.

Secondary screens. We performed the replication screen on 227 genes whose knockdown (KD) diminished the AIRE-induced transcription in the primary screen. For these 227 genes, we selected only the shRNAs that scored in the primary screen. Fresh lentiviral stocks were prepared by the TRC of the Broad Institute and were rearrayed in 384-well plate format leaving a free quadrant on each plate to include 96 control shRNAs. An independent control plate was also screened and we followed the same experimental procedure as we did for the primary screen. We also performed an AIRE dependence screen in transfecting the empty pcDNA3.1 vector (Invitrogen) instead of the AIRE-expressing vector.

Screen data analysis. For each 384-well plate screened, a local regression (loess) was performed, in R (www.r-project.org/), on the luminescence data of the shRNAs (primary screen) or of the control shRNAs included in all screening plates of the secondary screen. The Z scores were computed as the loess residuals (the vertical distance from the shRNA data points to the fit curve) over the SD of the residuals in the independent control plate. From the Z-score distribution in the independent control plate, two thresholds ($Z_{0.05}$ and $Z_{0.025}$) were calculated, corresponding to the limit under which 5% and 2.5% of the control shRNA Z scores lie, respectively. For the primary screen, a candidate was selected if (*i*) at least one of its targeting hairpins has a Z score below $Z_{0.025}$ and (*ii*) it has a second hairpin with a Z score below $Z_{0.05}$ or if the distribution of its five hairpins is significantly skewed toward the low Z scores, as tested by a *t* test. For the secondary screens, a candidate is confirmed if at least two of its hairpins have a Z score below $Z_{0.05}$. In the case of just one hairpin being rescreened, the candidate is confirmed if its sole hairpin has a Z score below $Z_{0.05}$.

Presence–Absence Analysis in 4D6 Cells. Gene expression profiling in 4D6 cells had previously been performed on Human Genome U133Plus2.0 microarrays (Affymetrix) (3). We reprocessed the raw probe-level data (.CEL files) by the guanine cytosine robust multiarray algorithm (5) using the ExpressionFileCreator module of the GenePattern 3.0 software package (6). Presence–absence calls for expression of probesets were made using the PANP Bioconductor package (www.bioconductor.org).

Aire Nuclear Localization Assay. The 4D6 cells stably expressing an Aire-GFP fusion protein were seeded in 96-well plates at the density of 2,500 cells per well (7). The next day, polybrene was added to the cell medium at 8 μg/mL final concentration and transduction was performed with 2 μL of lentiviral stocks encoding hit shRNAs corresponding to the 77 AIRE cofactors confirmed in the replication screen. Twenty-four hours post-infection, we changed the medium to a puromycin-containing

medium (2 μg/mL) and let the selection of infected cells occur for 72 h. Cells were then washed in PBS and fixed in BD Cytofix/Cytoperm solution (BD Biosciences) for 1 h at room temperature. Cells were washed in BD Perm-wash buffer (BD Biosciences), and were incubated in PBS-containing DAPI (Invitrogen) for 10 min. After a final wash with BD Perm-wash buffer, images were acquired with the florescence microscope Axiovert 200 (Zeiss) equipped with filters matching the spectral excitation and emission characteristics of DAPI and GFP.

Network Modeling. A nonexhaustive set of factors known to be important in various transcriptional processes were used as "seeds" to search the STRING 8.2 database (http://string-db.org/) in order to capture common interactors and construct a transcription-related network. For the search, we unselected "Textmining" in the Active Prediction Methods, used a high confidence score (0.7) and set −2 for the network depth (only direct neighbors). A detail description of these parameters can be found elsewhere (8). The extracted network was visualized and analyzed further with Cytoscape v2.6.3 (9). To define the molecular subcomplexes of the network, we performed a protein–protein interaction cluster analysis with the Cytoscape plugin MCODE v1.3 (10), using default parameters. To infer the biological functions of the subcomplexes, we did a Gene Ontology (GO) analysis with the Cytoscape plugin ClueGO v1.3 (11) searching the GO_BiologicalProcess_20.10.2008 database and using default parameters. Subsequently we mapped onto this generated framework the AIRE cofactors showing direct and high confidence interactions in the STRING database.

Endogenous RNA-Interference Screening. The 4D6 cells were seeded in 6-well plates at the density of 150,000 cells per well. The following day, polybrene was added at 8 μg/mL final and the transduction was carried out in duplicate with 40 μL of lentiviral stocks encoding hit shRNAs for 43 AIRE cofactors confirmed in the replication screen, which do not impair Aire nuclear speckle localization and which are unambiguously expressed. We also performed control infections with lentiviral stocks of the empty pLKO vector (pLKO_TRC005-nullT). After 72 h of puromycin selection (2 μg/mL), cells were harvested and seeded in 12-well plates at the density of 16×10^4 cells per well (one well of a 6-well plate into one well of a 12-well plate). Transient transfection assays were conducted as described (3) using Lipofectamine 2000 in Opti-MEM I Reduced-Serum Media with 500 ng of the AIRE-expressing construct. Forty hours post-transfection, the cells were harvested, and total RNA was isolated by TRIzol extraction. First-strand cDNA was synthesized from 2 μg of purified RNA using SuperScript II Reverse Transcriptase (Invitrogen) and oligo-(dT) primers. cDNA was used for subsequent SYBR Green RT-PCR amplification using the Mx3000P QPCR System (Stratagene). AIRE-dependent gene expression was assessed by analyzing the mRNA levels of "diagnostic" AIRE-dependent genes (including *KRT14* and *S100A9*) previously identified by microarray analysis in 4D6 cells (3). Transcript levels in each reaction were normalized against the quantity of *HPRT* transcripts in the same sample. The primers used for the reactions were the same as in ref. 3.

Mouse Screen and shRNA Knockdown Validation. The murine 1C6 cell line, originally derived from mouse thymic medullary epithelium (12), was maintained in S-MEM medium complemented with 10% (vol/vol) FBS, 2-mercaptoethanol, Hepes, pen/strep antibiotics. Lentiviral stocks encoding shRNAs targeting mouse orthologs of the AIRE cofactors that have a knockdown effect on both *KRT14* and *S100A9* in human 4D6 cells were prepared by the TRC of the Broad Institute and arrayed in 384-well plate format. The murine screen was performed following the same format and experimental procedure as the secondary screen with the exceptions of (*i*) the seeded cell density: 2,000 cells per well, (*ii*) the volume of lentivirus stocks used for infection: 2 μL, and (*iii*) the volume of Dual-Glo reagents used for luciferase quantification: 20 μL. Data analysis was conducted as for the 4D6 secondary screen. A shRNA was called "hit" when its associated Z score was better than $Z_{0.05} = -1.7$, as defined by the Z-score distribution of the control shRNAs in the independent control plate. We then selected the hit shRNAs to validate their induced-knockdown effect, as well as all of the shRNAs of candidate genes for which no significant reduction of the AIRE-induced luciferase activity was shown. For this knockdown validation assay, 1C6 cells were seeded in 96-well plates at the density of 10,000 cells per well and were transduced in triplicate with 4 μL of the shRNA lentiviral stocks in polybrene-containing medium (8 μg/mL final). Control infections were performed with an shRNA targeting GFP (ClonetechGfp_473), which is used as control for hairpin knockdown validation by the RNAi Consortium at the Broad Institute. After 72 h of puromycin selection (2 μg/mL), mRNA was purified using the TurboCapture 96 mRNA kit (Qiagen). First-strand cDNA was synthesized following the TurboCapture two-step protocol using Sensiscript Reverse Transcriptase (Qiagen) and oligo-(dT) primers. cDNA was used for subsequent SYBR Green RT-PCR amplification with the 7900HT Fast Real-Time PCR system (Applied Biosystems). Transcript levels of *Bat1a*, *Ccdc71*, *Ccnt2*, *Chaf1b*, *Clp1*, *Cnot3*, *Cpeb4*, *Erh*, *Gtf2a1*, *Hexim1*, *Hnrnpl*, *Jarid1b*, *Jmjd1a*, *Jund*, *Mcm3ap*, *Myst3*, *Ncor2*, *Nr2f2*, *Rbl1*, *Rel*, *Sfrs1*, *Stat5b*, and *Trip13* were assayed and normalized against the quantity of *Gapdh* transcripts in the same samples with primers listed in Table S3.

Mice. Aire-deficient mice on the C57BL/6 (B6) and nonobese diabetic (NOD) genetic backgrounds were derived and genotyped as previously described (13). *Stat5a/b* double-deficient mice were a generous gift from J. N. Ihle (St. Jude Children's Research Hospital, Memphis, TN). Mice were housed at the Center for Animal Resources and Comparative Medicine at Harvard Medical School under Institutional Animal Care and Use Committee approved procedures.

Lentigenic Generation. From the mouse screen and the knockdown validation assay, we selected 15 of the most efficient shRNAs (one or two for each of the 10 cofactors in mouse). We also selected three control shRNAs, one targeting the Aire-irrelevant *Gtf2a1l* gene and two targeting RFP (rfp_407) or LacZ (lacZ_1339), which are neutral in the Aire screen. To have GFP as the marker of the shRNA expression, we transferred the selected shRNAs from PLKO.1 (puromycin marker) to PLKO.3G (http://cbdm.hms.harvard.edu/Transgenics/Transgenic.html) by cloning the BamHI-NdeI restriction fragment containing the shRNA. For the production of high-titer viruses, we prepared two 15-cm^2 plates, each seeded with 8×10^6 293-FT cells in DMEM 10% (vol/vol) FCS (20 mL). The next day, we performed the transfection using 100 μL of the TransIT-293 transfection reagent (Mirus) with the shRNA-PLKO.3G plasmid (20 μg) and three packaging plasmids: pMDLg/pRRE (10 μg), RSVRev (5 μg), and CMV-VSVG (5 μg). Forty-eight hours later, the viral supernatant was filtered through a 0.45-μm filter and ultracentrifuged at 25,000 rpm for 90 min at 4 °C using a Beckman SW28 ultracentrifuge rotor (Beckman Coulter). Cold sterile PBS (20 μL) was added to the pellet and left overnight at 4 °C. The virus was then resuspended by pipetting up and down and flash frozen for storage at −80 °C. shRNA lentiviruses were titered by infection of 293-FT cells using standard protocols and those with a titer over 10^9 infectious unit/mL were selected for oocyte microinfection. Fertilized oocytes (B6 × NOD) were microinfected under the zona pellucida with high titer lentiviruses as described (14). Each oocyte was infected with one lentivirus and mixed pools of infected oocytes (up to four different viruses) were

reimplanted in NOD females. Newborns were aged for 3 wk, blood was collected from the tails, lysed by the ACK Lysing Buffer (Lonza), and GFP-reporter expression was analyzed by flow cytometry using an LSRII flow cytometer (BD Bioscience) and FlowJo software (Tree Star). Mice expressing GFP in more than 10% of cells were selected and their integrated shRNA was resequenced with primers matching the PLKO.3G vector (Table S3).

Thymus Digestion and Medullary Epithelial Cell Sorting. Individual thymi from 4-wk-old lentigenics (B6 × NOD), Aire KO mice, and Stat5a/b double-deficient mice were digested. Thymi were dissected and trimmed of fat and connective tissue. Lobes of adult thymi were cut into small pieces with a pair of fine scissors and agitated in RPMI to release thymocytes. The resulting fragments of adult thymi or the pools of neonate thymi were digested 30 min at 37 °C in RPMI medium containing collagenase D (1 mg/mL final) (Roche) and DNase I (1 mg/mL final) (Sigma) and were further agitated with a 1-mL pipette to free more thymocytes. Enzyme mixtures with released thymocytes were removed after fragments had settled. RPMI containing a collagenase/dispase mixture (2 mg/mL final) (Roche) and DNase I (2 mg/mL final) was added, and the mixture was incubated at 37 °C. Every 5 min, cells were agitated using a Pasteur pipette until a single-cell suspension was obtained. Cells were then passed through 70-μm mesh, spun down, and resuspended in staining buffer (PBS containing 1% FBS and 5 mM EDTA). Fluorophore-labeled antibodies CD45-PerCPCy5.5 (1:50) (Biolegend), Ly51-PE (1:800) (Biolegend), and I-A/E-APC (1:1,200) (eBioscience) were added to the samples and incubated with resuspended cells for 20 min at 4 °C. Cells were washed, then resuspended in 400 μL of staining buffer. Sorting medullary epithelial cells (MECs) ($CD45^{-}PE^{-}I\text{-}A/E^{high}$) or GFP + MECs for lentigenics was performed on the Aria cell sorter (BD Bioscience).

Gene Expression Profiling. Total RNA was prepared from sorted MECs (ranging here from 10,000 to 30,000 cells) using TRIzol. Single-stranded DNA in the sense orientation was synthesized from total RNA with random hexamer priming using the GeneChip WT cDNA Synthesis and Amplification kit (Affymetrix) and following the 100-ng total RNA protocol. The DNA was subsequently purified, fragmented, and terminally labeled using the GeneChip WT Terminal Labeling kit (Affymetrix) incorporating biotinylated ribonucleotides into the DNA. The labeled DNA was then hybridized to Mouse Gene ST1.0 microarrays (Affymetrix), washed, stained, and scanned. Raw probe-level data (.CEL files) were normalized by the robust multiarray average (RMA) algorithm (15) and summarized using the R-package aroma.affymetrix (www.aroma-project.org/). Correlation analyzes were done in R.

Coimmunoprecipitation and 7SK RNA Quantitation. The 293-FT cells were seeded in 10-cm^2 plates at the density of 4 million cells per well. The next day, they were either transfected using the TransIT-293 transfection reagent (Mirus) or left untransfected for 48 h. Transfection was performed with the relevant plasmids: pCDNA3-HA-CCNT2b (provided by O.B.), pEF-myc-Gal-CDK9 or pCMV-Aire-Flag. Nuclear extraction and coimmunoprecipitation were performed using the Universal Magnetic Co-IP kit (ActiveMotif). Cells were first lysed with hypotonic lysis buffer and incubated on ice for 15 min. Cell lysates were centrifuged for 1 min at 14,000 × *g*, 4 °C, and the liquid fraction containing cytosolic proteins was discarded. Pelleted nuclei were resuspended in complete digestion buffer, digested with the enzymatic shearing mixture for 10 min at 37 °C, and spun down for 10 min at 14,000 × *g*, 4 °C. Postnuclear supernatants were first incubated with specific antibodies for 4 h, then with Protein-G magnetic beads for 1 h with rotation at 4 °C, and were then washed (four times).

For Western blot analysis, bound proteins were eluted by boiling 5 min in sample buffer, separated by SDS/PAGE, electrotransferred to PVDF membranes, blocked for 30 min with 3% (wt/vol) milk solution, followed by Western blotting with the appropriate antibodies and detection by enhanced chemiluminiscence (ECL). Antibodies used for immunoprecipitation or revelation, are as follows: Flag-tag M2 (Sigma; F1804), myc-tag (Cell Signaling; 9B11), HA-tag (Roche; 3F10), GFP (Santa Cruz; FL), HNRNPL (Abcam; 4D11), HNRNPK (Abcam; 3C2), CDK9 (Santa Cruz; C-20), and HEXIM1 (provided by O.B.).

For 7SK RNA quantitation, beads were resuspended in TRIzol and total RNA was reverse transcribed using SuperScript II Reverse Transcriptase (Invitrogen) with the 7SK-specific RT primer: CACATGCAGCGCCTCATTTG. cDNA was used for subsequent SYBR Green RT-PCR amplification with the 7300HT Real-Time PCR system (Applied Biosystems). The 7SK RNA levels were assayed with primers listed in Table S3.

1. Fernández E, et al. (1994) Establishment and characterization of cloned human thymic epithelial cell lines. Analysis of adhesion molecule expression and cytokine production. *Blood* 83(11):3245–3254.
2. Sillanpää N, et al. (2004) Autoimmune regulator induced changes in the gene expression profile of human monocyte-dendritic cell-lineage. *Mol Immunol* 41(12):1185–1198.
3. Koh AS, 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(41):15878–15883.
4. Giraud M, et al. (2007) An IRF8-binding promoter variant and AIRE control CHRNA1 promiscuous expression in thymus. *Nature* 448(7156):934–937.
5. Wu Z, Irizarry RA (2004) Preprocessing of oligonucleotide array data. *Nat Biotechnol* 22(6):656–658, author reply 658.
6. Reich M, et al. (2006) GenePattern 2.0. *Nat Genet* 38(5):500–501.
7. Abramson J, Giraud M, Benoist C, Mathis D (2010) Aire's partners in the molecular control of immunological tolerance. *Cell* 140(1):123–135.
8. von Mering C, et al. (2005) STRING: Known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res* 33 (Database issue):D433–D437.
9. Shannon P, et al. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504.
10. Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4:2.
11. Bindea G, et al. (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25(8):1091–1093.
12. Mizuochi T, Kasai M, Kokuho T, Kakiuchi T, Hirokawa K (1992) Medullary but not cortical thymic epithelial cells present soluble antigens to helper T cells. *J Exp Med* 175 (6):1601–1605.
13. Anderson MS, et al. (2002) Projection of an immunological self shadow within the thymus by the aire protein. *Science* 298(5597):1395–1401.
14. Chen Z, Stockton J, Mathis D, Benoist C (2006) Modeling *CTLA4*-linked autoimmunity with RNA interference in mice. *Proc Natl Acad Sci USA* 103(44):16400–16405.
15. Irizarry RA, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4(2):249–264.

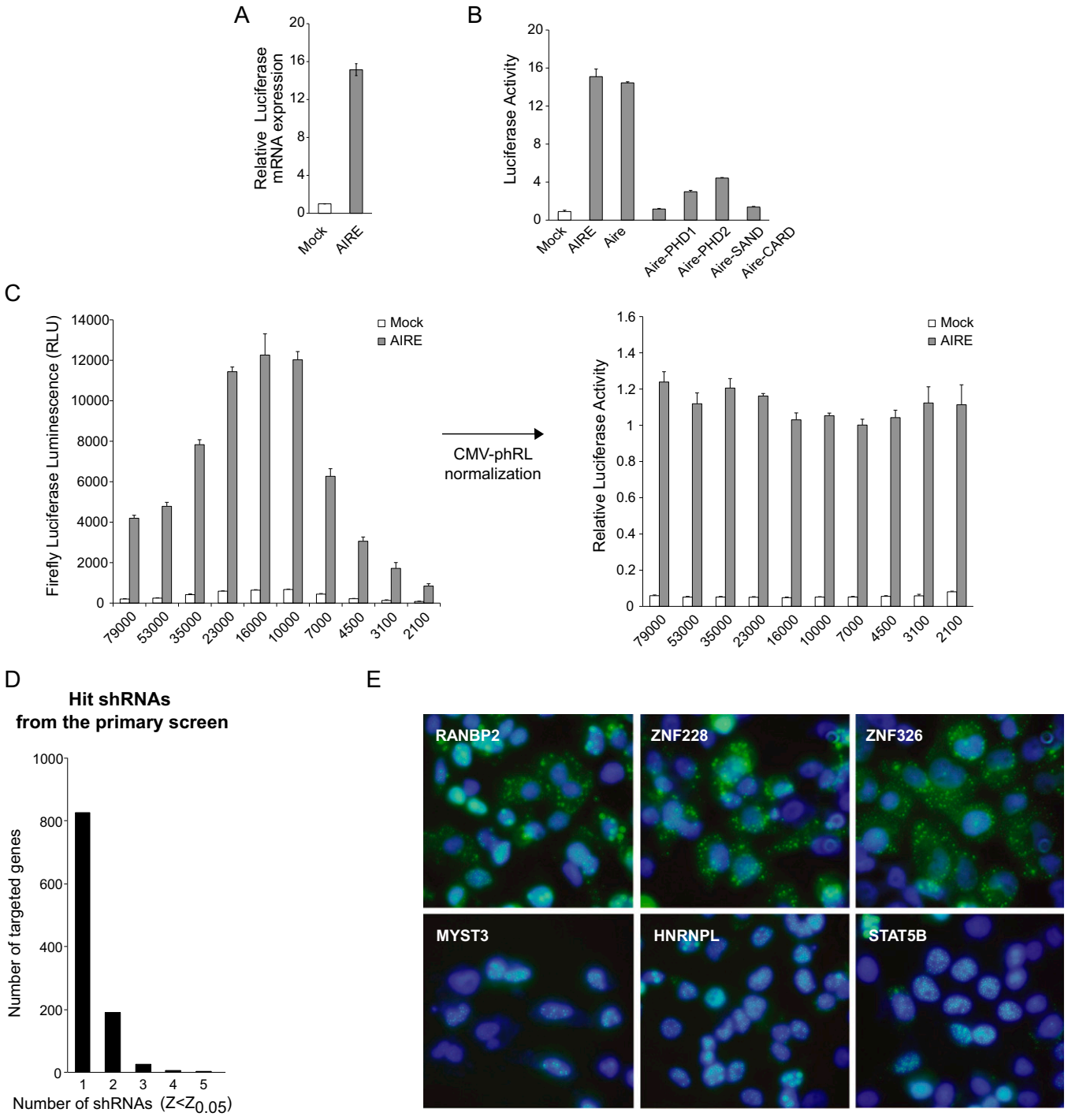

Fig. S1. Validation of the AIRE transactivation assay and of AIRE's functional allies. (*A*) RT-PCR analysis of AIRE's effect on firefly luciferase expression from the minP-pGL4 plasmid transfected in 4D6 cells. (*B*) Luciferase assay of the transactivation function of human AIRE, murine Aire, and some of its functional domain deletion mutants. (*C*) Normalization of the transfection efficiency by the *Renilla* luciferase activity of the CMV-phRL vector cotransfected in 4D6 cells seeded at various densities. (*D*) Distribution of the number of hit shRNAs from the primary screen that target unique genes. (*E*) Influence of AIRE's functional allies 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 three candidates in the upper row suppressed the nuclear localization of Aire, whereas those in the lower row had no effect. Nuclei were stained with DAPI (blue).

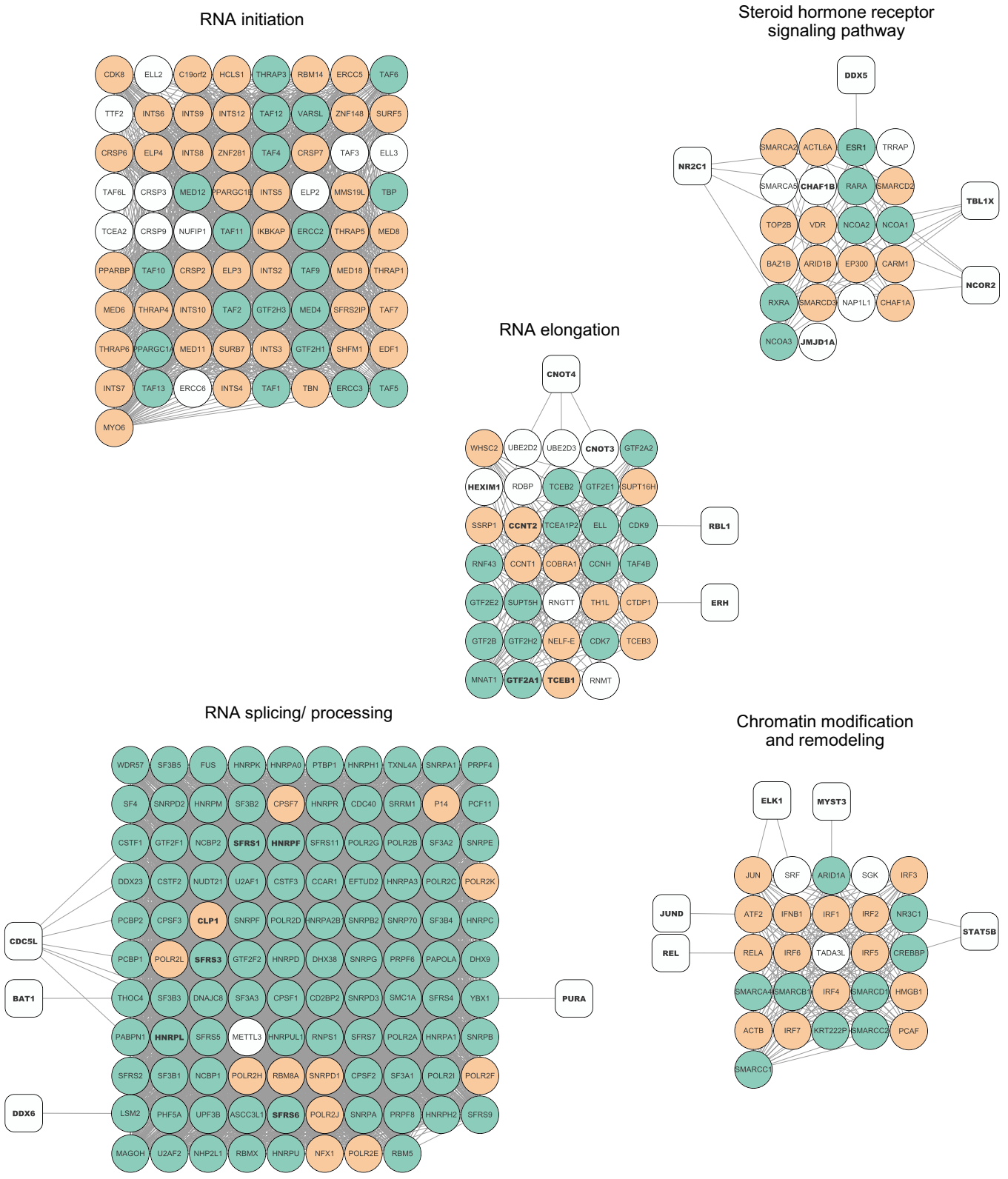

Fig. S2. Identity of the molecular clusters of the transcription-related network. Members of the transcription-related network, which delineate the five protein-interaction clusters, are shown in color. Green is for the members that belong to the group of proteins defining the GO term the most significantly associated with each cluster. Clusters are named with their associated GO terms. AIRE's allies are labeled in bold and are either part of the network (color) or interact with it (white).

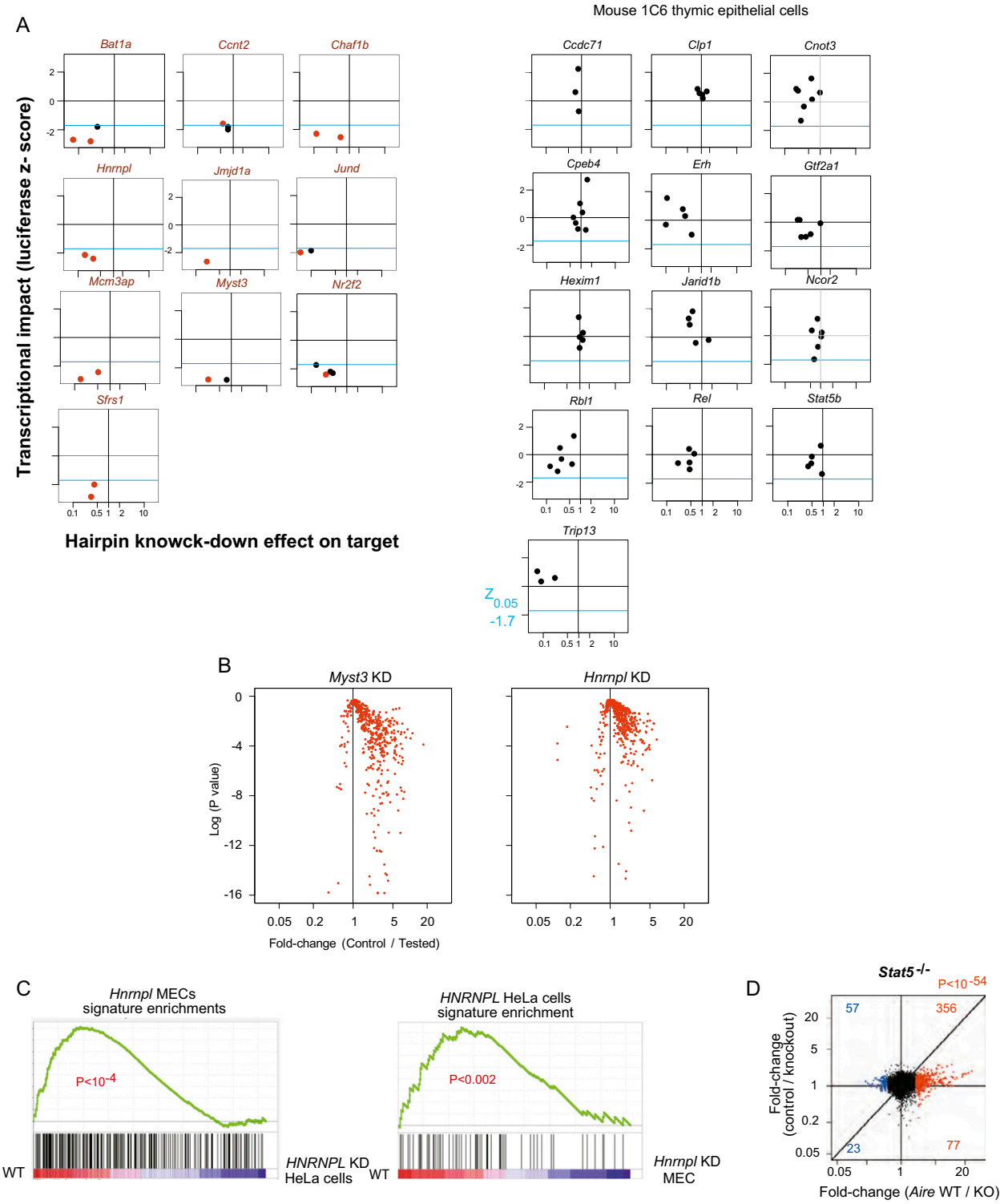

Fig. S3. Significance of the effect of lentigenic knockdowns on Aire-sensitive genes. (*A*) Effect of shRNAs targeting mouse orthologs, tested in murine MEC 1C6 cells using the luciferase-based assay. For each shRNA, the luciferase Z score (*y* axis) is plotted vs. the knockdown level of target expression (*x* axis). The red dots indicate the hairpins selected for in vivo lentigenic validation. (*Left*) Groups together orthologs whose effect is confirmed in 1C6; (*Right*) groups together orthologs showing no effect or low knockdown. (*B*) For each of the two knockdown lentigenics, the probability of each individual Aire-sensitive gene was computed based on the variance observed in control lentigenics (*y* axis). The level of induction is shown (*x* axis). (*C*) Reciprocal enrichment (Gene Set Enrichment Analysis, GSEA) in the WT vs. *HNRNPL* KD-ranked expression dataset from HeLa cells (GEO GSE33771), of the genes impacted by Hnrnpl in primary MECs. (*D*) MEC gene expression profiles for mice with knockout expression of *Stat5a/b*. *Stat5a/b* KO mice are double deficient for *Stat5a* and *Stat5b*. The results are shown as a FoldChange (*y* axis) vs. Aire's transcriptional footprint (*x* axis; from a comparison of *Aire*-KO and WT MECs). *Stat5* KD effect on Aire-induced genes (red): $P < 2.10^{-54}$.

Table S1. List of AIRE's functional cooperants identified by RNAi lentivirus screening

Table S1

The 227 candidates from the primary screen are listed along with quantitative and qualitative observations gathered throughout subsequent steps. Vertical red lines delineate groups of candidates meeting selection criteria. Within a group, genes are sorted in alphabetical order. The first red line highlights the 77 AIRE allies confirmed by the replication screen, the second is for the 54 genes expressed in 4D6 cells, and the third stresses the 51 allies that do not perturb the proper Aire nuclear localization.

Table S2. Summary of the lentigenic production strategy

Table S2

The 10 Aire functional allies selected after a number of validation steps are listed with the name of their specific hairpins that were used for oocyte microinfection. Three genes whose specific hairpins were used as control are also listed. Lentigenics with knockdown expression of the hairpin target gene are shown in red.

Table S3. List of primers

Table S3