A new mode of transcriptional hub assembly by autoimmune regulator Aire

Yu-San Huoh[1,2,*], Qianxia Zhang[1,2,*], Ricarda Torner[2,3], Sylvan C. Baca[4], Haribabu Arthanari[2,3] & Sun Hur[1,2,^]

[1]Howard Hughes Medical Institute and Program in Cellular and Molecular Medicine, Boston Children's Hospital, MA 02115, USA
[2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA
[3]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA
[4]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA
*These authors contributed equally.
^Correspondence: Sun.Hur@crystal.harvard.edu

Abstract

Transcriptional hubs have emerged as critical structural components in gene expression and regulation[1-4]. Studies on hub assembly have primarily focused on the concept of phase separation of disordered regions within transcriptional regulators (TRs)[2,5-7]. We here introduce an unexplored mode of hub assembly by investigating the Autoimmune regulator (Aire), an indispensable TR in immune homeostasis[8,9] distinguished by its unique capacity to form homopolymers. We demonstrate that Aire forms active transcriptional hubs connecting disparate inter-chromosomal genomic loci, and that this hub assembly relies on both Aire's polymerizing domain CARD and the transcriptional activation domain (TAD). In a departure from the conventional TAD functions, Aire's TAD orchestrates site-specific hub assembly by coupling CARD polymerization with genomic targeting. Through direct interactions with the co-activators CBP/p300, Aire TAD guides Aire to genomic loci enriched with CBP/p300, where CARD polymerization is initiated. To avert unwanted CARD polymerization at inappropriate sites, Aire employs its chromatin reader domain PHD1 to disperse Aire outside of its target sites and confine Aire polymerization to CBP/p300-rich loci. Our work thus reveals previously unrecognized mechanism by which a TAD and a chromatin reader domain collaborate to discern genomic loci and facilitate TR polymerization, unveiling a multi-layered assembly process for transcriptional hubs.

Introduction

Controlled protein polymerization underlies diverse biological processes, including cell death and inflammation where polymeric architectures serve as "hubs" that amplify upstream signals[10-12]. One family of protein domains that mediate polymerization is Caspase Activation Recruitment Domain (CARD), which belongs to the death domain superfamily and plays important roles in cell death and inflammatory signaling pathways[13]. CARDs often self-polymerize into filamentous assemblies and have intrinsic abilities to spontaneously form filaments in isolation[13]. Accordingly, proper function of CARD-containing proteins requires tight regulatory mechanisms to ensure that CARD polymerization occurs in the right place at the right time[13]. While CARD domains are well-studied in cytosolic signaling pathways, they are also found in transcriptional regulators like Aire and Speckle Proteins (SPs)[14,15]. However, the specific functional implications and mechanisms underlying CARD polymerization in these transcriptional regulators remain poorly understood.

Aire plays a critical role in central T cell tolerance[8,9]. Aire transcriptionally upregulates the expression of thousands of peripheral tissue antigens (PTAs) in medullary thymic epithelial cells (mTECs)[16,17]. These PTAs are displayed on the cell surface to recognize auto-reactive T cells for their negative selection or diversion into regulatory T cells[8,16,18,19]. Consequently, mutations in human *AIRE* or knock-out of mouse *Aire* result in multi-organ autoimmunity, including autoimmune polyendocrinopathy syndrome type 1 (APS-1)[16,20]. Initially, Aire was regarded as a conventional transcription factor (TF) directly binding PTA gene promoters to induce expression[21,22]. Recent studies, however, propose that Aire largely upregulates PTA expression indirectly by amplifying the actions of various lineage-defining TFs ectopically expressed in mTECs[23-27]. Precise molecular mechanisms underlying Aire's cooperation with these TFs remain unclear.

Aire is a chromatin-binding TR, rather than a sequence-dependent, direct DNA binder[28,29]. Although it possesses a putative DNA binding domain, SAND (Figure 1A), this domain lacks essential DNA binding residues[30]. Instead, Aire features two potential chromatin reader domains—PHD1 and PHD2 (Figure 1A). PHD1 binds histone H3 (H3) with non-methylated lysine 4 (H3K4me0), a histone mark typically associated with repressive chromatin states at inactive loci[31]. Conversely, PHD2 exhibits little histone binding activity, and its functions remain unknown[28,29,32]. At first, PHD1's specificity for H3K4me0 was thought to guide Aire to inactive PTA gene loci[33,34]. However, recent research indicates that Aire primarily binds genomic sites that are already enriched with a permissive histone mark, H3K27ac, such as super-enhancers (SEs)[35]. Notably, regions rich in active H3K27ac marks tend to lack H3K4me0[31,36], raising questions about how Aire specifically targets H3K27ac-rich sites and the precise role of PHD1 in this context.

One of the most intriguing properties of Aire as a TR is its ability to form homopolymers using its N-terminal CARD[37], which in cells—both human and murine mTECs along with model cell lines—manifests as nuclear foci[14,37-40]. These foci are distinct from other known nuclear bodies like PML and Cajal bodies[41,42]. While the importance of CARD has been well-established[14,37], the precise functions of Aire foci—whether they serve as active transcription sites, inactive storage depots, or suppressive compartments—is a matter of controversy[35,43,44]. Furthermore,

studies have demonstrated that Aire mutants with altered foci localization exhibit impaired transcriptional functions and can exert a dominant negative effect on the wild-type allele by affecting its localization[37,45,46]. These findings collectively suggest the existence of yet-to-be-discovered mechanisms ensuring that Aire polymerization occurs in the right place for transcriptional regulation.

In this study, we demonstrate that Aire foci represent active transcriptional hubs and reveal an intricate hub assembly mechanism that coordinates the self-polymerization of CARD with genomic target recognition.

Results

Aire foci are sites of transcriptional activation.

Aire is known to be expressed in a miniscule subset of mTECs with temporal dynamics[24-26,47,48], which has made mechanistic studies using mTECs challenging. Therefore, we generated a model system where Aire is ectopically expressed in human thymic epithelial cell line (4D6) through a doxycycline (Dox)-inducible promoter. This 4D6 model system recapitulated *AIRE* expression level in human mTECs (Figure S1A), Aire localization to H3K27ac-enriched sites including SEs (Figures S1B-S1C, S1E), CARD-dependent nuclear foci formation (Figure 1B), Aire-induced broad transcriptomic changes (Figure S1D) and the impact of loss-of-function APS-1 mutations (Figure S1F)—characteristics that were all previously observed in mTECs[21,35,40,48-51]. We thus utilized 4D6 cells to investigate the mechanism of Aire polymerization in the nucleus.

Given the controversy of whether Aire has activating or suppressing functions at bound genomic loci[35,43,44], we first asked whether Aire foci are the functional sites of active transcription using nascent RNA fluorescence in situ hybridization (FISH) combined with immunofluorescence (IF) (Figures 1B and S2A). We investigated three genes, *RIC8A, SETD1B*, and *UBTF*, which were all up-regulated and bound by Aire (Figures S1D-S1E). In ~80% of the cells examined, Aire expression resulted in the appearance of 1 or 2 nascent RNA-FISH spots for *RIC8A* (Figures 1C and S2B). Aire-dependent increase in the frequency of RNA-FISH spots were also observed with *SETD1B* and *UBTF* (Figures 1C and S2B). RNA-FISH spots and Aire foci had significant overlap as demonstrated by the averaged Aire fluorescence intensities centered on RNA-FISH spots (Figures 1D and S2C) and individual distance between the centers of the closest RNA-FISH and Aire foci (Figure S2D). Consistent with the notion that Aire foci represent active transcriptional hubs, Aire foci also colocalized with coactivators MED1, CBP and p300 (Figures 1E and S3A).

Among all Aire-positive nuclei, 49% showed both *RIC8A* foci and *SETD1B* foci within the same nucleus; within these nuclei containing both *RIC8A* and *SETD1B* FISH spots, ~20% of the *RIC8A* and *SETD1B* foci shared an Aire focus with each other (Figure S2E). We observed a similar frequency of sharing the same Aire focus between *RIC8A* and *UBTF* foci (Figure S2F). While these observations do not support obligatory pairings among these genes, given that *RIC8A, SET1B* and *UBTF* are located on different chromosomes, their frequent contacts through the same Aire foci suggests that each cluster of Aire is likely connecting distinct inter-chromosomal genomic loci to form a transcriptional hub.

Aire hub formation requires the TAD-like domain C-terminal tail (CTT).

We next examined the mechanism by which Aire forms transcriptional hubs. CARD was indispensable for Aire foci formation (consistent with previous reports[14,37]), whereas SAND, PHD1 and PHD2 were dispensable (Figures 1F and S3B). Intriguingly, deletion of C-terminal tail (CTT, residues 482-545 in human and 480-552 in mouse) impaired foci formation of both human and mouse Aire (Figures 1F and S3B) despite having intact CARD. Unlike the CARD domain that spontaneously multimerizes *in vitro* and forms nuclear foci[37], CTT on its own did not show a similar multimerization property (see Figure S5) or foci formation (Figure S3C), suggesting that it may have a role in modulating CARD polymerization. Further truncation analysis suggests that residues 510-521 (R2, human residue numbering) and 522-535 (R3) within CTT were required for Aire foci formation, but residues 499-509 (R1) of CTT were not (Figure 1G-1H). Consistent with this data, R2 and R3 are more conserved than R1 and the rest of the CTT (Figure 1G).

Aire CTT is known to have transcription activation domain (TAD)-like activity[52,53]. We thus asked whether CTT's ability to modulate CARD polymerization and thereby support foci formation stems from its TAD-like activity. Deletion of R2 or R3 abrogated Aire's transcriptional activity regardless of the target genes (Figure 1I). CTTΔR1, on the other hand, showed gene-specific behaviors, suggesting more complex functions for R1 (Figure 1I). A TAD-activity reporter assay using Gal4 DNA-binding domain ($Gal4^{DBD}$)[52] also showed the importance of R2 and R3 in the TAD-like activity of CTT (Figure S3D). Collectively, these results suggest that CTT R2 and R3 are critical not only for Aire foci formation, but also for CTT's TAD-like activity.

CBP/p300 directly bind Aire CTT and mediate its TAD-like activity.

To understand how CTT regulates Aire hub formation, we investigated genetic and physical interaction partners of CTT. The requirement of CTT for Aire hub formation mirrored CTT's TAD-like activity (Figures 1H and S3D), so we took advantage of this observation and designed a genome-wide CRISPR screen based on the TAD reporter assay. The screening reporter mKate2, under the control of upstream activation sequences (UAS), was stably incorporated into the 4D6 genome and induced upon expression of $Gal4^{DBD}$-CTT fused with the expression reporter GFP via a self-cleavable peptide 2A (Figures 2A-2B). In parallel, we performed GST pull-downs of 293T lysate using purified GST-tagged CTT (GST-CTT) and analyzed the co-purified proteins by mass-spectrometry (Figure 2C). From these two independent analyses, we identified the histone acetyltransferases (HATs) CBP and p300 as common hits (Figures 2B-2C and S4A). CBP and p300 are highly homologous HATs that are responsible for a large fraction of H3K27ac[54], and colocalize with Aire foci in both mTECs[14,39] and 4D6 cell line (Figures 1E and S3A). No other HATs were found in either screen.

Using recombinant CTT and p300, we found that the interaction was direct (Figure S4B). In comparison to the intact CTT, CTTΔR2 and ΔR3 showed reduced CTT–CBP/p300 interaction, while CTTΔR1 displayed similar binding (Figure 2D). This dependence on R2 and R3, but not R1, mirrored the requirement for Aire foci formation (Figure 1H). Domain truncation analysis of

p300 revealed that CTT interacted with the TAZ2 and IBiD domains of p300 (Figure S4C-S4D). Isolated CBP TAZ2 could be recombinantly purified and showed direct binding to purified CTT (Figure S4E), while isolated IBiD was insoluble, precluding more detailed binding analysis.

To further characterize the TAZ2–CTT interaction, we performed ^{1}H-^{15}N HSQC NMR spectroscopy on CTT with and without TAZ2. In the absence of TAZ2, isolated CTT displayed largely disordered characteristics (Figure S5A), although R2 and R3 displayed a moderate tendency to form alpha helices (Figures 2E and S5B). Upon incubation with CBP TAZ2, R2 and R3 underwent significant chemical shifts and peak broadening (Figures S5C and S5D), indicative of TAZ2 binding. Utilization of R2 and R3 for TAZ2 binding was further supported by AlphaFold prediction (Figure 2E) and sequence conservation (Figure 1G). Mutations in the putative CTT interface (A536D and D530K) impaired the CTT–CBP interaction (Figure 2F) and the TAD-like activity as measured by the Gal4^{DBD}-CTT luciferase reporter assay (Figure 2G). These results show that CBP/p300 directly interact with Aire CTT and mediate CTT's TAD-like activity.

CBP/p300 are crucial for Aire hub formation and functions.

To examine whether CBP/p300 play an important role in Aire hub formation and functions, we assessed how CTT interface mutations and pharmacological inhibition of CBP/p300 would affect Aire foci formation and transcriptional activity. Mutations in the putative CTT interface (A536D and D530K) compromised Aire foci formation (Figures 2H and S6A) and transcriptional activity (Figure S6B). Although all nuclei examined showed foci for D530K, they were significantly smaller than the WT Aire foci (Figure S6A). Treatment with dCBP-1[55] for 4 hrs degraded most of CBP/p300 and significantly reduced the levels of H3K27ac (Figure 2I). The acetyltransferase catalytic inhibitor A-485[56] also reduced the levels of H3K27ac, but without diminishing the CBP/p300 protein levels (Figure 2I). Both dCBP-1 and A-485 significantly impaired Aire foci formation (Figure 2I). Again, smaller foci were still visible with dCBP-1 and A-485 treatments, but they were transcriptionally inactive, as evidenced by reduced frequency of nascent RNA-FISH foci (Figure S6C) and nascent RNA qPCR of Aire-dependent genes (Figure 2J). Similarly, dCBP-1 and A-485 decreased the Gal4^{DBD}-CTT reporter activity, despite slightly increasing the level of Gal4^{DBD}-CTT (Figure S6D-S6E). Overall, these results suggest that CTT–CBP/p300 interaction and the catalytic activity of CBP/p300 are required for Aire to form transcriptionally active hubs.

CTT guides Aire to CBP/p300-enriched loci.

We next asked how CTT–CBP/p300 interaction facilitates Aire foci formation and transcriptional functions. Notably, Aire foci are associated with genomic regions that are already enriched with H3K27ac marks (Figures S1B, S1C, S1E and 3A). Considering that CBP/p300 is often found at H3K27ac-rich sites[57], we hypothesized that CBP/p300 could play a pivotal role in recruiting Aire to its target sites, and that this recruitment might be crucial for Aire's polymerization process.

ChIP-seq analysis of CBP/p300 revealed that Aire target sites exhibit a high density of CBP/p300 occupancy even before Aire expression (Figure 3A). When we compared Aire ChIP-

seq signals with CBP/p300 or H3K27ac ChIP signals, Aire binding correlated more closely with the pre-existing CBP/p300 occupancy than with H3K27ac levels (Figure 3B, left panels). This observation suggests that CBP/p300 likely play a more direct role in Aire targeting compared to H3K27ac, as the histone modification can be generated by other HATs in addition to CBP/p300[54,58].

To investigate the role of CTT–CBP/p300 interaction on Aire's target site selection, we compared ChIP-seq of WT Aire to ΔCTT.R3, a variant deficient in CBP/p300 binding. Strikingly, ΔCTT.R3 exhibited a more dispersed genomic occupancy pattern, characterized by approximately tenfold more Aire-bound peaks and relatively uniform ChIP-seq signal intensities compared to WT Aire (Figure 3A). Moreover, the ChIP-seq signals of ΔCTT.R3 displayed weaker correlation to CBP/p300 ChIP-seq signals when compared to WT Aire (Figure 3B). These distinctions became even more apparent when examining the top 500 WT-preferred versus ΔCTT.R3-preferred sites (as defined by the WT-to-ΔCTT.R3 ChIP-seq signal ratio in Figure 3A); WT-preferred sites exhibited a pre-existing high density of CBP/p300 occupancy, whereas ΔCTT.R3-preferred sites were largely devoid of CBP/p300 (Figures 3C-3E). Examination of SE binding also showed more efficient recruitment of WT Aire than ΔCTT.R3 (Figure S7A). Consistent with the notion that CBP/p300 play important roles in Aire targeting to active loci, both degrader dCBP-1 and catalytic inhibitor A-485 impaired Aire binding to SEs (Figures 3F and 3G). This suggests that Aire may preferentially home in on the catalytically active form of CBP/p300 that often form clusters and accumulate at active genetic loci[2,59,60].

Altogether, these results demonstrate that Aire's localization to H3K27ac-rich regions is facilitated by the CTT–CBP/p300 interaction. Importantly, any condition that disrupted Aire foci formation, including mutations in CTT and the degradation or catalytic inhibition of CBP/p300 (Figures 1 & 2), also disrupted Aire's genomic targeting. These results indicate that Aire targeting and polymerization are tightly coupled.

At CBP/p300-rich loci, CARD polymerization amplifies Aire recruitment and stabilizes hubs.

To further elucidate the interplay between Aire foci formation and its genomic targeting, we turned our attention to ΔCARD, a variant with an intact CTT but impaired foci formation and transcriptional functions (Figures 1F and S7B). Compared to WT and ΔCTT.R3, ΔCARD exhibited an intermediate genomic binding behavior. ΔCARD's ChIP-seq pattern showed a partial correlation with p300 ChIP signals but did not reach the levels observed with WT Aire (Figures 3A-3B). ΔCARD also favored WT-preferred sites, albeit this preference was less pronounced compared to WT Aire (Figures 3C-3D). Moreover, ΔCARD's SE localization was in-between that of WT and ΔCTT.R3 as well (Figure S7A). These results suggest that Aire's bias towards CBP/p300-rich sites is largely driven by CTT–CBP/p300 interaction and that this bias is independent of polymerization.

ΔCARD's less efficient recruitment to these sites and its more scattered genomic distribution compared to WT Aire highlight the crucial role of CARD polymerization in enhancing Aire's selectivity. CARD polymerization likely recruits additional Aire molecules to CBP/p300-rich loci, leading to feedback amplification of Aire's target selectivity. This feedback mechanism

extends to CBP/p300, as evidenced by the high density of CBP/p300 at Aire foci and lack thereof in the absence of Aire (Figures 1E and S3A). This amplification seems essential for establishing stable Aire–chromatin interactions, as shown by nuclear fractionation analysis. Both WT Aire and ΔCTT exhibited MNase-dependent solubilization, indicating chromatin binding akin to histones (Figure S7C). By contrast, ΔCARD remained predominantly soluble regardless of MNase treatment, suggesting weaker chromatin binding. Collectively, these results support a model wherein Aire targeting via CTT and polymerization via CARD are intricately linked. This synergy enables the stable formation of Aire hubs specifically at CBP/p300-rich loci.

Requirement for CTT in Aire foci formation stems from PHD1-mediated suppression.

Our observation that ΔCTT is unable to form foci despite having intact CARD (Figures 1F and S3B) appears to contradict the previous observation that isolated CARD can form foci in cells and spontaneously polymerize *in vitro*[37]. This discrepancy led us to hypothesize that Aire CARD may be subjected to certain regulatory mechanisms in the context of full-length Aire, which necessitates CTT for Aire polymerization. To identify the domain responsible for CARD regulation, we truncated individual domains in the AireΔCTT background with the expectation that deleting the CARD-suppressive domain in AireΔCTT would restore foci formation. Indeed, deleting PHD1, but no other domains, restored foci formation of AireΔCTT (Figure 4A). Additionally, direct fusion of PHD1 with CARD was sufficient to suppress CARD polymerization, while no other domains had similar impact (Figure 4B). Furthermore, this PHD1-mediated suppression could be reversed by attaching CTT (Figure 4B), recapitulating the stimulatory effect of CTT in full-length Aire foci formation. Collectively, these findings establish PHD1 as the domain responsible for suppressing CARD polymerization and explains the requirement for CTT in Aire foci formation.

Next, we asked how PHD1 suppresses Aire polymerization. Since PHD1 is a histone binding domain with specific preference for H3K4me0[28,29] (K_D = 4.4μM, Figure S8A), we first examined whether the ability of PHD1 to bind H3K4me0 is important. The D299A mutation in Aire PHD1, which diminishes the affinity for H3K4me0 and transcriptional activity of Aire[28,29,33], compromised its ability to suppress Aire CARD polymerization (Figure 4C). Additionally, fusing Aire CARD with another H3K4me0-specific PHD from an unrelated protein, CHD4[61] (Figure S8B), also suppressed Aire CARD polymerization, while CTT restored it (Figure 4D). This suggests that Aire PHD1 suppresses CARD polymerization indirectly through H3K4me0 binding, rather than directly through PHD1–CARD interaction. In additional support of this, nuclear foci formation of homologous CARD from Sp110 was suppressed upon fusion with Aire PHD1, but not with Aire PHD2 (Figure S8C). This implies that PHD1's suppression occurs by distributing Aire across numerous genome sites with H3K4me0, which would dilute Aire and prevent nucleation, whereas CTT counters this effect by concentrating Aire at CBP/p300-rich loci.

PHD1 has a dual function of suppressing spurious CARD polymerization while anchoring Aire to chromatin for productive hub formation.

Our finding that PHD1 suppresses CARD polymerization and, by extension, transcriptional hub formation, raised the question of whether PHD1 also suppresses Aire's transcriptional activity.

While the PHD1-deletion mutant (ΔPHD1) robustly formed nuclear foci (Figure 1F), it was transcriptionally inactive, as measured by RT-qPCR (Figure 5A) or visualized by nascent RNA-FISH of Aire target genes (Figure 5B). In addition, the levels of co-activators MED1 and CBP at Aire foci were significantly decreased in ΔPHD1 compared to WT (Figure 5C). Thus, seemingly contrary to its ability to suppress CARD polymerization, PHD1 is in fact critical for Aire transcriptional activity.

To understand why ΔPHD1 is transcriptionally inactive despite forming nuclear foci, we analyzed ΔPHD1's interaction with chromatin. ΔPHD1 ChIP-seq revealed that PHD1 deletion nearly completely abolished Aire's chromatin binding regardless of genomic locations (Figure 5D). Nuclear fractionation study also corroborated this notion, revealing that ΔPHD1 existed solely in insoluble fractions independent of MNase (Figure S7C), indicating the formation of insoluble aggregates that are untethered to chromatin. These results suggest that PHD1 is the main driver bringing Aire to chromatin, which differs from the role of CTT and CARD in modulating Aire's target specificity. Thus, functional defect in ΔPHD1 foci is due to the lack of chromatin tethering.

Given PHD1's critical role in Aire targeting, we asked how PHD1, with its preference for H3K4me0, can interact with active CBP/p300-rich sites that are generally presumed to lack H3K4me0[31,36]. Indeed, ChIP-qPCR experiments showed an inverse correlation between H3K4me0 and H3K27ac levels; Aire's preferred target sites (WT-preferred) had lower H3K4me0 levels but higher H3K27ac levels compared to non-preferred sites (ΔCTT.R3-preferred) (Figure 5E). Interestingly, both WT Aire and ΔCTT.R3 depended on PHD1 for chromatin interaction but favored different genomic sites with distinct histone modifications. This suggests that CTT may influence PHD1's histone preference. In the case of ΔCTT.R3, PHD1's inherent preference for H3K4me0 likely dominates Aire's target selection. However, for WT Aire, CTT might enhance PHD1's ability to bind substrates with lower affinities, like H3K4me1 (Figure S8A), which is more abundant in WT-preferred sites compared to ΔCTT.R3 sites (Figure S8D). This suboptimal histone tail binding may be compensated with the binding energies gained from CTT–CBP/p300 interaction and subsequent CARD polymerization. In summary, PHD1 serves a dual role: PHD1 anchors Aire to chromatin in conjunction with CTT to ensure precise targeting while also suppressing spurious polymerization outside target regions via its H3K4me0 specificity (Figure 5F).

Discussion

Recent studies revealed that transcriptional activation often occurs in bursts through dynamic formation of "hubs" containing high concentrations of transcriptional co-activators, such as Mediator and CBP/p300[1-4]. Molecular mechanisms underlying hub assembly, however, have predominantly centered around the idea that it is driven by phase separation of disordered regions within co-activators or TRs[2,5-7]. Here, we unveil a new mechanism of transcriptional hub assembly employed by Aire. While Aire utilizes its seemingly disordered CTT to facilitate hub assembly, CTT functions as a targeting domain that guides nucleation of the self-polymerizing domain CARD to ensure the formation of functional hubs at the correct genomic sites. Aire further utilizes PHD1 to limit unwanted polymerization outside its target loci, allowing tight coupling of CARD polymerization to genomic targeting. In the absence of CTT or at genomic

sites with low levels of CBP/p300, PHD1's preference for H3K4me0 dominates; Aire interacts with chromatin in a dispersed manner and is unable to polymerize or stably accumulate. Conversely, at appropriate target sites where CTT can engage with active CBP/p300, CTT modifies PHD1's specificity to recognize other histone marks and nucleate CARD polymerization, possibly by locally concentrating Aire (Figure 5G). Importantly, CTT's role in stimulating Aire polymerization is only relevant when PHD1-mediated suppression is present, suggesting that CTT does not directly participate in polymerization or phase separation, as seen with other TADs[2,5,6]. Once CARD polymerizes, an Aire nucleation site then acts as a sink to recruit more Aire and CBP/p300 molecules, setting up a positive feedback loop for transcriptional hub assembly (Figure 5G). In essence, PHD1-mediated chromatin anchoring makes Aire polymerization strictly dependent on CTT and restricted to activated CBP/p300-rich environments, such as SEs.

Aire's use of pre-established CBP/p300 clusters as a nucleation platform raises questions as to how this process ties into Aire's ultimate function in regulating PTA expression. While it may initially appear redundant to form hubs at already active or weakly-active genomic sites, this aligns with recent studies showing that Aire primarily assists and amplifies the actions of various lineage-defining TFs that in turn directly mediate PTA expression[24-26,47]. Aire's role may be to cement the changes in the chromatin landscape initiated by these TFs, which are potentially acting sub-optimally in the non-native cellular environment of mTECs. These Aire functions could make Aire dispensable once the chromatin landscape is more established, accounting for the loss of Aire expression in later mTEC developmental stages[24-26,47]. Alternatively, Aire's ability to form hubs may coincide with the mobilization of cohesins to facilitate chromatin loop formation[62]. Aire-mediated chromatin looping could then lead to the recruitment of inactive PTA loci at Aire hubs for subsequent activation. These possibilities are not mutually exclusive. In either case, the use of CBP/p300-rich environment as a scaffold for Aire polymerization and hub assembly offers a previously unrecognized mechanism by which TRs modulate global transcriptional output.

Our discovery that Aire employs the histone-binding domain PHD1 for controlling CARD polymerization also has broad implications beyond transcription. While the importance of CARD and the need for its precise regulation are widely acknowledged in various cell death and inflammatory signaling pathways[10-12], the mechanisms governing self-polymerization of CARD to prevent undesirable outcomes remain largely unresolved. A rare example where CARD-suppression mechanism is understood is the case of cytoplasmic viral RNA sensor RIG-I. Here, RIG-I CARD remains masked by an intramolecular interaction until it is released upon viral RNA binding[63]. By contrast, we demonstrate that Aire's PHD1 regulates CARD not through direct contact but by associating with a ubiquitous histone mark, which prevents Aire nucleation by dispersing and diluting Aire. Such mechanism of "regulation-by-dispersion" presents a new mode of regulating protein polymerization.

In summary, our findings provide a new framework of understanding for transcriptional hub assembly and regulation of protein polymerization.

References

1 Cho, W. K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412-415 (2018). https://doi.org:10.1126/science.aar4199

2 Ma, L. *et al.* Co-condensation between transcription factor and coactivator p300 modulates transcriptional bursting kinetics. *Mol Cell* **81**, 1682-1697 e1687 (2021). https://doi.org:10.1016/j.molcel.2021.01.031

3 Rodriguez, J. & Larson, D. R. Transcription in Living Cells: Molecular Mechanisms of Bursting. *Annu Rev Biochem* **89**, 189-212 (2020). https://doi.org:10.1146/annurev-biochem-011520-105250

4 Sabari, B. R., Dall'Agnese, A. & Young, R. A. Biomolecular Condensates in the Nucleus. *Trends Biochem Sci* **45**, 961-977 (2020). https://doi.org:10.1016/j.tibs.2020.06.007

5 Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855 e1816 (2018). https://doi.org:10.1016/j.cell.2018.10.042

6 Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361** (2018). https://doi.org:10.1126/science.aar2555

7 Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361** (2018). https://doi.org:10.1126/science.aar3958

8 Mathis, D. & Benoist, C. Aire. *Annu Rev Immunol* **27**, 287-312 (2009). https://doi.org:10.1146/annurev.immunol.25.022106.141532

9 Proekt, I., Miller, C. N., Lionakis, M. S. & Anderson, M. S. Insights into immune tolerance from AIRE deficiency. *Curr Opin Immunol* **49**, 71-78 (2017). https://doi.org:10.1016/j.coi.2017.10.003

10 Nozaki, K., Li, L. & Miao, E. A. Innate Sensors Trigger Regulated Cell Death to Combat Intracellular Infection. *Annu Rev Immunol* **40**, 469-498 (2022). https://doi.org:10.1146/annurev-immunol-101320-011235

11 Fu, J. & Wu, H. Structural Mechanisms of NLRP3 Inflammasome Assembly and Activation. *Annu Rev Immunol* **41**, 301-316 (2023). https://doi.org:10.1146/annurev-immunol-081022-021207

12 Cadena, C. & Hur, S. Filament-like Assemblies of Intracellular Nucleic Acid Sensors: Commonalities and Differences. *Mol Cell* **76**, 243-254 (2019). https://doi.org:10.1016/j.molcel.2019.09.023

13 Huoh, Y. S. & Hur, S. Death domain fold proteins in immune signaling and transcriptional regulation. *FEBS J* **289**, 4082-4097 (2022). https://doi.org:10.1111/febs.15901

14 Ferguson, B. J. *et al.* AIRE's CARD revealed, a new structure for central tolerance provokes transcriptional plasticity. *J Biol Chem* **283**, 1723-1731 (2008). https://doi.org:10.1074/jbc.M707211200

15 Fraschilla, I. & Jeffrey, K. L. The Speckled Protein (SP) Family: Immunity's Chromatin Readers. *Trends Immunol* **41**, 572-585 (2020). https://doi.org:10.1016/j.it.2020.04.007

16 Anderson, M. S. *et al.* Projection of an immunological self shadow within the thymus by the aire protein. *Science* **298**, 1395-1401 (2002). https://doi.org:10.1126/science.1075958

17 Sansom, S. N. *et al.* Population and single-cell genomics reveal the Aire dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. *Genome Res* **24**, 1918-1931 (2014). https://doi.org:10.1101/gr.171645.113
18 Malchow, S. *et al.* Aire Enforces Immune Tolerance by Directing Autoreactive T Cells into the Regulatory T Cell Lineage. *Immunity* **44**, 1102-1113 (2016). https://doi.org:10.1016/j.immuni.2016.02.009
19 Yang, S., Fujikado, N., Kolodin, D., Benoist, C. & Mathis, D. Immune tolerance. Regulatory T cells generated early in life play a distinct role in maintaining self-tolerance. *Science* **348**, 589-594 (2015). https://doi.org:10.1126/science.aaa7017
20 Husebye, E. S., Anderson, M. S. & Kampe, O. Autoimmune Polyendocrine Syndromes. *N Engl J Med* **378**, 1132-1141 (2018). https://doi.org:10.1056/NEJMra1713301
21 Giraud, M. *et al.* Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. *Proc Natl Acad Sci U S A* **109**, 535-540 (2012). https://doi.org:10.1073/pnas.1119351109
22 Meredith, M., Zemmour, D., Mathis, D. & Benoist, C. Aire controls gene expression in the thymic epithelium with ordered stochasticity. *Nat Immunol* **16**, 942-949 (2015). https://doi.org:10.1038/ni.3247
23 Bautista, J. L. *et al.* Single-cell transcriptional profiling of human thymic stroma uncovers novel cellular heterogeneity in the thymic medulla. *Nat Commun* **12**, 1096 (2021). https://doi.org:10.1038/s41467-021-21346-6
24 Bornstein, C. *et al.* Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells. *Nature* **559**, 622-626 (2018). https://doi.org:10.1038/s41586-018-0346-1
25 Michelson, D. A., Hase, K., Kaisho, T., Benoist, C. & Mathis, D. Thymic epithelial cells co-opt lineage-defining transcription factors to eliminate autoreactive T cells. *Cell* **185**, 2542-2558 e2518 (2022). https://doi.org:10.1016/j.cell.2022.05.018
26 Miller, C. N. *et al.* Thymic tuft cells promote an IL-4-enriched medulla and shape thymocyte development. *Nature* **559**, 627-631 (2018). https://doi.org:10.1038/s41586-018-0345-2
27 Wells, K. L. *et al.* Combined transient ablation and single-cell RNA-sequencing reveals the development of medullary thymic epithelial cells. *Elife* **9** (2020). https://doi.org:10.7554/eLife.60188
28 Koh, A. S. *et al.* Aire employs a histone-binding module to mediate immunological tolerance, linking chromatin regulation with organ-specific autoimmunity. *Proc Natl Acad Sci U S A* **105**, 15878-15883 (2008). https://doi.org:10.1073/pnas.0808470105
29 Org, T. *et al.* The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. *EMBO Rep* **9**, 370-376 (2008). https://doi.org:10.1038/sj.embor.2008.11
30 Bottomley, M. J. *et al.* The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. *Nat Struct Biol* **8**, 626-633 (2001). https://doi.org:10.1038/89675
31 Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-318 (2007). https://doi.org:10.1038/ng1966

32 Yang, S., Bansal, K., Lopes, J., Benoist, C. & Mathis, D. Aire's plant homeodomain(PHD)-2 is critical for induction of immunological tolerance. *Proc Natl Acad Sci U S A* **110**, 1833-1838 (2013). https://doi.org:10.1073/pnas.1222023110
33 Koh, A. S., Kingston, R. E., Benoist, C. & Mathis, D. Global relevance of Aire binding to hypomethylated lysine-4 of histone-3. *Proc Natl Acad Sci U S A* **107**, 13016-13021 (2010). https://doi.org:10.1073/pnas.1004436107
34 Org, T. *et al.* AIRE activated tissue specific genes have histone modifications associated with inactive chromatin. *Hum Mol Genet* **18**, 4699-4710 (2009). https://doi.org:10.1093/hmg/ddp433
35 Bansal, K., Yoshida, H., Benoist, C. & Mathis, D. The transcriptional regulator Aire binds to and activates super-enhancers. *Nat Immunol* **18**, 263-273 (2017). https://doi.org:10.1038/ni.3675
36 Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**, 897-903 (2008). https://doi.org:10.1038/ng.154
37 Huoh, Y. S. *et al.* Dual functions of Aire CARD multimerization in the transcriptional regulation of T cell tolerance. *Nat Commun* **11**, 1625 (2020). https://doi.org:10.1038/s41467-020-15448-w
38 Matsumoto, M. *et al.* AIRE illuminates the feature of medullary thymic epithelial cells in thymic carcinoma. *Cancer Med* **12**, 9843-9848 (2023). https://doi.org:10.1002/cam4.5777
39 Pitkanen, J. *et al.* Cooperative activation of transcription by autoimmune regulator AIRE and CBP. *Biochem Biophys Res Commun* **333**, 944-953 (2005). https://doi.org:10.1016/j.bbrc.2005.05.187
40 Su, M. A. *et al.* Mechanisms of an autoimmunity syndrome in mice caused by a dominant mutation in Aire. *J Clin Invest* **118**, 1712-1726 (2008). https://doi.org:10.1172/JCI34523
41 Bjorses, P. *et al.* Localization of the APECED protein in distinct nuclear structures. *Hum Mol Genet* **8**, 259-266 (1999).
42 Rinderle, C., Christensen, H. M., Schweiger, S., Lehrach, H. & Yaspo, M. L. AIRE encodes a nuclear protein co-localizing with cytoskeletal filaments: altered sub-cellular distribution of mutants lacking the PHD zinc fingers. *Hum Mol Genet* **8**, 277-290 (1999). https://doi.org:10.1093/hmg/8.2.277
43 Goldfarb, Y. *et al.* Mechanistic dissection of dominant AIRE mutations in mouse models reveals AIRE autoregulation. *J Exp Med* **218** (2021). https://doi.org:10.1084/jem.20201076
44 Koh, A. S. *et al.* Rapid chromatin repression by Aire provides precise control of immune tolerance. *Nat Immunol* **19**, 162-172 (2018). https://doi.org:10.1038/s41590-017-0032-8
45 Abbott, J. K. *et al.* Dominant-negative loss of function arises from a second, more frequent variant within the SAND domain of autoimmune regulator (AIRE). *J Autoimmun* **88**, 114-120 (2018). https://doi.org:10.1016/j.jaut.2017.10.010
46 Oftedal, B. E. *et al.* Dominant Mutations in the Autoimmune Regulator AIRE Are Associated with Common Organ-Specific Autoimmune Diseases. *Immunity* **42**, 1185-1196 (2015). https://doi.org:10.1016/j.immuni.2015.04.021

47 Givony, T. *et al.* Thymic mimetic cells function beyond self-tolerance. *Nature* (2023). https://doi.org:10.1038/s41586-023-06512-8
48 Park, J. E. *et al.* A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367** (2020). https://doi.org:10.1126/science.aay3224
49 Abramson, J., Giraud, M., Benoist, C. & Mathis, D. Aire's partners in the molecular control of immunological tolerance. *Cell* **140**, 123-135 (2010). https://doi.org:10.1016/j.cell.2009.12.030
50 Bjorses, P. *et al.* Mutations in the AIRE gene: effects on subcellular location and transactivation function of the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy protein. *Am J Hum Genet* **66**, 378-392 (2000). https://doi.org:10.1086/302765
51 Nagamine, K. *et al.* Positional cloning of the APECED gene. *Nat Genet* **17**, 393-398 (1997). https://doi.org:10.1038/ng1297-393
52 Zumer, K., Plemenitas, A., Saksela, K. & Peterlin, B. M. Patient mutation in AIRE disrupts P-TEFb binding and target gene transcription. *Nucleic Acids Res* **39**, 7908-7919 (2011). https://doi.org:10.1093/nar/gkr527
53 Meloni, A., Incani, F., Corda, D., Cao, A. & Rosatelli, M. C. Role of PHD fingers and COOH-terminal 30 amino acids in AIRE transactivation activity. *Mol Immunol* **45**, 805-809 (2008). https://doi.org:10.1016/j.molimm.2007.06.156
54 Weinert, B. T. *et al.* Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/p300 Acetylome. *Cell* **174**, 231-244 e212 (2018). https://doi.org:10.1016/j.cell.2018.04.033
55 Vannam, R. *et al.* Targeted degradation of the enhancer lysine acetyltransferases CBP and p300. *Cell Chem Biol* **28**, 503-514 e512 (2021). https://doi.org:10.1016/j.chembiol.2020.12.004
56 Lasko, L. M. *et al.* Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours. *Nature* **550**, 128-132 (2017). https://doi.org:10.1038/nature24028
57 Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112 (2009). https://doi.org:10.1038/nature07829
58 Kuo, Y. M. & Andrews, A. J. Quantitating the specificity and selectivity of Gcn5-mediated acetylation of histone H3. *PLoS One* **8**, e54896 (2013). https://doi.org:10.1371/journal.pone.0054896
59 Ibrahim, Z. *et al.* Structural insights into p300 regulation and acetylation-dependent genome organisation. *Nat Commun* **13**, 7759 (2022). https://doi.org:10.1038/s41467-022-35375-2
60 Ortega, E. *et al.* Transcription factor dimerization activates the p300 acetyltransferase. *Nature* **562**, 538-544 (2018). https://doi.org:10.1038/s41586-018-0621-1
61 Jain, K. *et al.* Characterization of the plant homeodomain (PHD) reader family for their histone tail interactions. *Epigenetics Chromatin* **13**, 3 (2020). https://doi.org:10.1186/s13072-020-0328-z
62 Bansal, K. *et al.* Aire regulates chromatin looping by evicting CTCF from domain boundaries and favoring accumulation of cohesin on superenhancers. *Proc Natl Acad Sci U S A* **118** (2021). https://doi.org:10.1073/pnas.2110991118

63 Kowalinski, E. *et al.* Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* **147**, 423-435 (2011). https://doi.org:10.1016/j.cell.2011.09.039

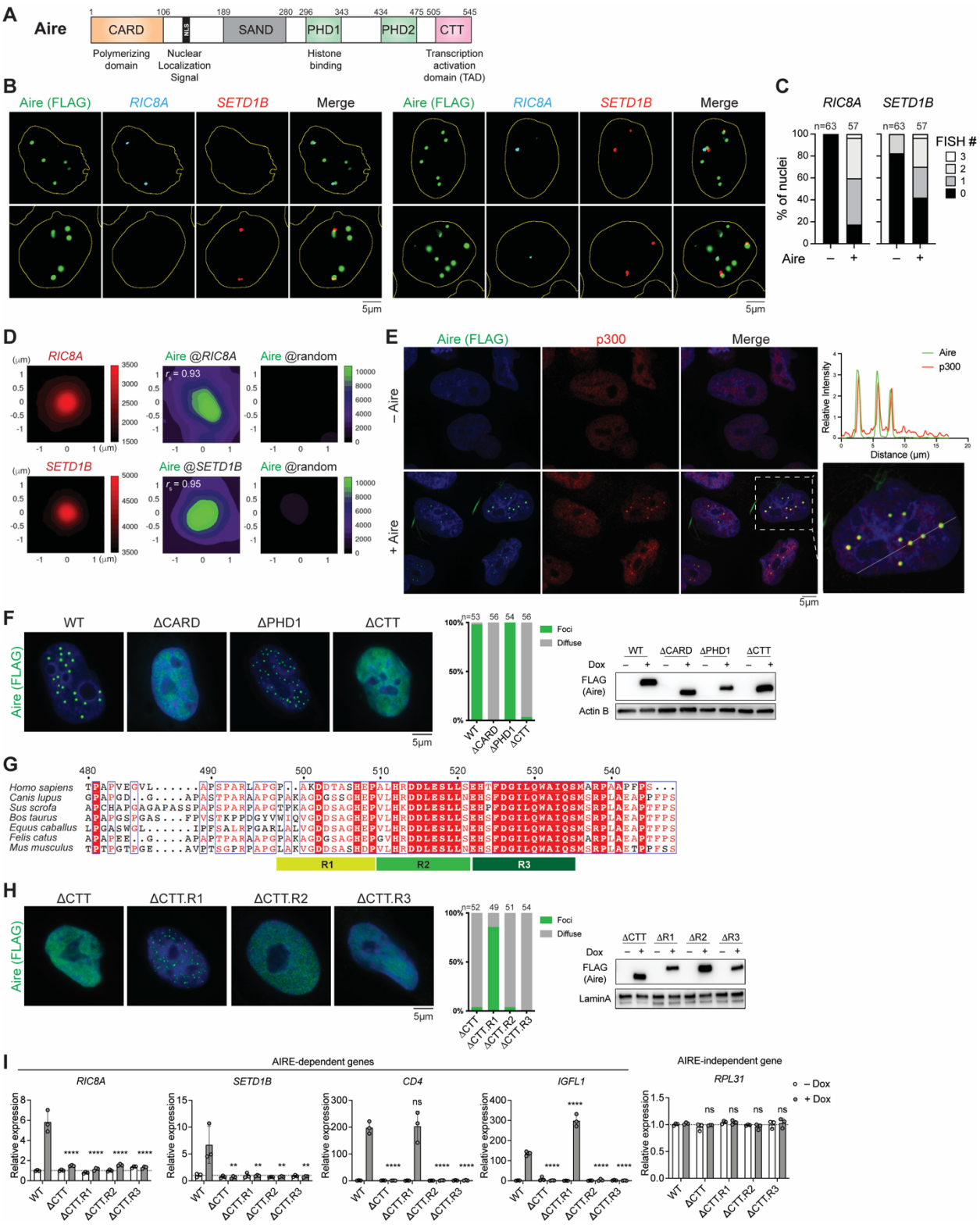

Figure 1. Aire foci are transcriptional hubs that require the TAD-like domain CTT for their assembly.

(A) Schematic of Aire domain architecture with previously characterized functions of individual domains. The numbers above denote amino acid residuals of human Aire (hAire).

(B) Nascent RNA-fluorescence in situ hybridization (FISH) coupled with immunofluorescence (IF) images of stable 4D6 cells expressing hAire-FLAG under a Dox-inducible promoter. Cells were induced with 1 μg/ml Dox (+Aire) for 24 hrs before staining with FISH probes and anti-FLAG. RNA-FISH probes were designed to hybridize with the intronic regions of Aire-dependent targets. Yellow outline represents the boundaries of nuclei from DAPI staining. Representative images highlight nuclei containing Aire foci with *RIC8A, SETD1B* or both RNA-FISH spots simultaneously. See Figure S2A for staining of cells that were not induced with Dox (–Aire) in addition to a zoomed-out image of +Aire cells.

(C) Quantitation of percent of nuclei that have 0, 1, 2 or 3 RNA-FISH spots in 4D6 cells before and after hAire expression as shown in (B) and Figure S2A. n indicates the number of nuclei examined.

(D) Spatial relationship between RNA-FISH spots and hAire foci. Shown are average signals of RNA-FISH (left), Aire IF centered on the indicated FISH spots (center) and Aire IF centered on randomly selected nuclear positions (right). r_s denotes the Spearman's correlation coefficient between RNA-FISH and Aire IF signals.

(E) Representative immunofluorescence images of endogenous p300 in stable 4D6 cells expressing Aire. Cells were not (– Aire) or were induced with 1 μg/ml Dox (+ Aire) for 24 hrs before immunostaining with anti-FLAG and anti-p300. Right: zoomed-in view of a nucleus enclosed by a white-dashed box along with measured fluorescence intensities across a drawn solid white line. See Figure S3A for IF images of endogenous CBP and MED1 in stable 4D6 cells expressing Aire.

(F) Representative immunofluorescence images of hAire ΔCARD, ΔPHD1 and ΔCTT variants in Dox-inducible stable 4D6 cells. Middle: percentage of nuclei with Aire foci vs. diffuse Aire staining. n represents the number of nuclei examined for each sample. Right: Western blot (WB) showing the levels of FLAG-tagged hAire variants.

(G) Sequence alignment of Aire CTT domains from various species. CTT.R1, R2 and R3 indicate aa 499-509, aa 510-521, aa 522-535, respectively, in hAire numbering.

(H) Representative immunofluorescence images of hAire ΔCTT, ΔCTT.R1, ΔCTT.R2, and ΔCTT.R3 in 4D6 cells. Experiments were done as in (F).

(I) Transcriptional activity of WT hAire or various CTT deletion mutants, as measured by the relative mRNA levels of Aire-dependent genes, *RIC8A, SETD1B*, *CD4,* and *IGFL1* in Dox-inducible stable 4D6 cells. An Aire-independent gene, *RPL31*, was examined as a negative control. All genes were normalized against the internal control *RPL18*. Data are presented as mean ± SD, n = 3. *p*-values (one-way ANOVA with Dunnett's multiple comparisons test) were calculated in comparison to WT Aire $^{**}p < 0.01$; $^{****}p < 0.0001$; $p > 0.05$ is not significant (ns).

All data are representative of at least three independent experiments.

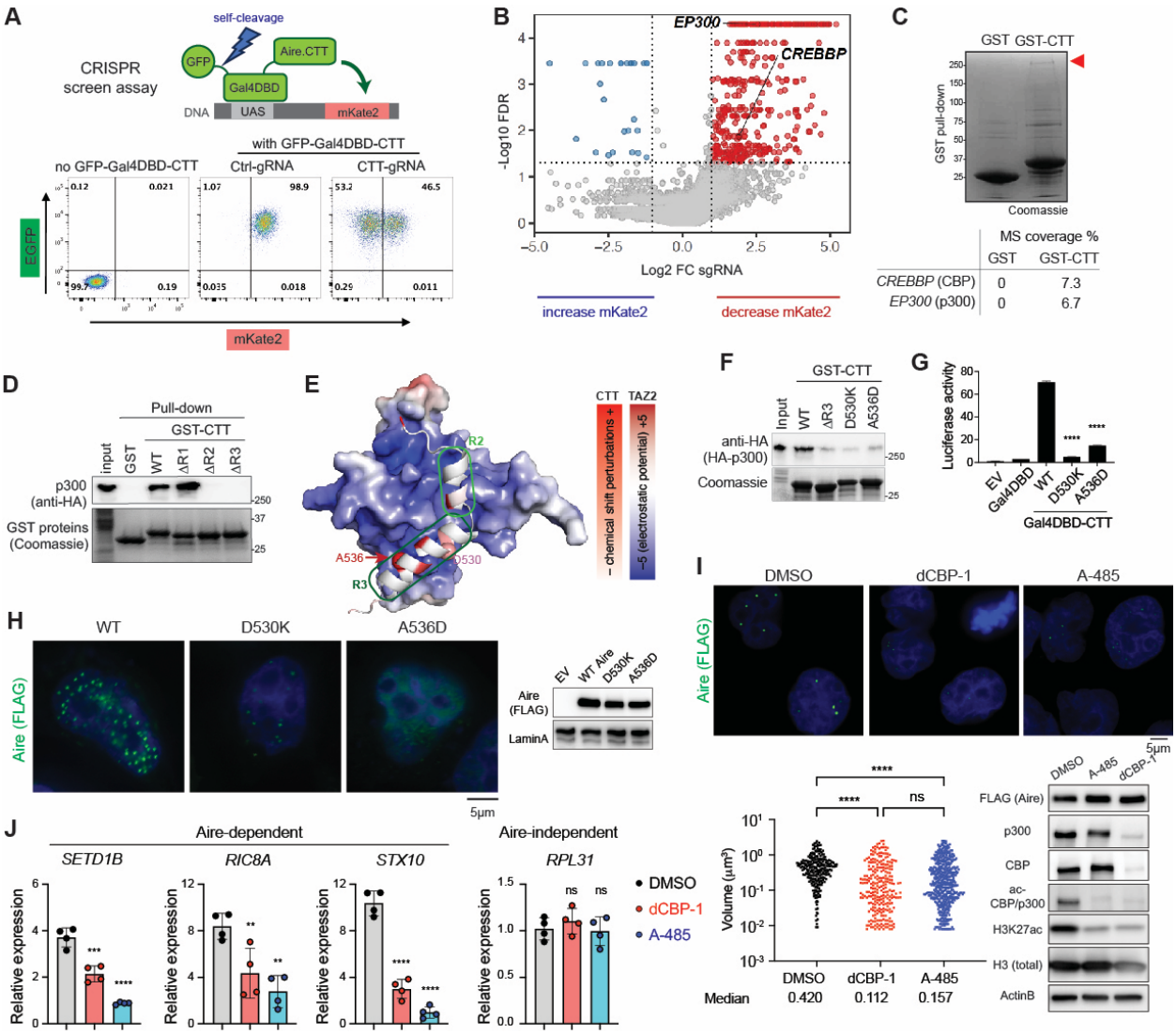

Figure 2. CBP/p300 directly interact with Aire CTT and are required for Aire hub formation.

(A) Gal4-reporter based CRISPR screening assay in stable 4D6 cells. Gal4-DNA binding domain (Gal4DBD)-CTT from Figure S3D was fused with GFP via a self-cleavable peptide (P2A). Dox-inducible expression of GFP-Gal4DBD-CTT leads to induction of mKate2 under the control of UAS. Bottom: validation of the system by flow cytometry. Ctrl, control; gRNA, guide RNA.

(B) Single gRNA (sgRNA) enrichment in 4D6 cells with decreased (red) versus increased (blue) expression of mKate2. FDR, false discovery rate. FC, fold-change.

(C) Coomassie Blue-stained SDS-PAGE gel of GST pull-down eluates. Purified recombinant His_6-GST-CTT (or His_6-GST) was incubated with 293T nuclear extracts prior to pull-down. Red arrow denotes the location of the ~250 kDa gel band that was cut and analyzed by mass spectrometry analysis. Bottom: sequence coverages determined by mass spectrometry (MS) of p300 and CBP pulled-down with GST or GST-CTT. See Figure S4A for the comprehensive list of top hits.

(D) His_6-GST and His_6-GST-CTT variant pull-downs of HA-tagged p300 transiently expressed in 293T nuclear extracts.

(E) NMR results incorporated into the top AlphaFold model[1] of the CBP–CTT complex using CBP TAZ2 (aa 1764-1855) and mAire CTT (aa 480-552). CTT is colored based on the degree of NMR chemical shift perturbation upon TAZ2 interaction. CBP TAZ2 is colored by electrostatic potential (PyMol, APBS module). mAire CTT residues D530 and A536 within R3 are highlighted with spheres.

(F) His_6-GST-CTT variant pull-downs of HA-tagged p300 transiently expressed in 293T nuclear extracts.

(G) TAD-like activity of WT mAire.CTT and CTT point mutants as measured by the Gal4-luciferase assay in Figure S3D. Data are presented as mean ± SD, n = 3. *p*-values (one-way ANOVA with Dunnett's multiple comparisons test) were calculated in comparison to Gal4DBD-CTT WT.

(H) Representative immunofluorescence images of FLAG-tagged mAire variants in 4D6 cells. Cells were transfected with mAire-FLAG expression plasmids 24 hrs prior to fixation. See Figure S6A for quantitative analysis.

(I) Representative immunofluorescence images of FLAG-tagged WT hAire in stable 4D6 cells (top), and quantification of Aire foci volumes (bottom left) in the presence of the p300/CBP degrader dCBP-1 (0.25 µM) or the catalytic inhibitor A-485 (3 µM) in comparison to vehicle control DMSO. Dox was added to induce expression of hAire for 8 hrs prior to fixation, and DMSO, A-485 or dCBP-1 were added 4 hrs prior to fixation. Aire foci n = 249, 259, 502 in DMSO-, dCBP-1 and A-485-treated cells, respectively. *p*-values (two-tailed Mann-Whitney test) were calculated in comparison to DMSO-treated cells.

(J) Transcriptional activity of hAire in the presence of dCBP-1 or A-485, as measured by nascent RNA RT-qPCR of *SETD1B*, *RIC8A* and *STX10*. Cells were treated as in (I), except for the metabolic labeling of nascent transcripts with 5'-ethylene uridine (see Methods). Data are presented as mean ± SD, n = 4. *p*-values (one-way ANOVA with Dunnett's multiple comparisons test) were calculated in comparison to DMSO-treated cells (+ Dox for 24hr).

All data are representative of at least three independent experiments. $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$; $p > 0.05$ is not significant (ns).

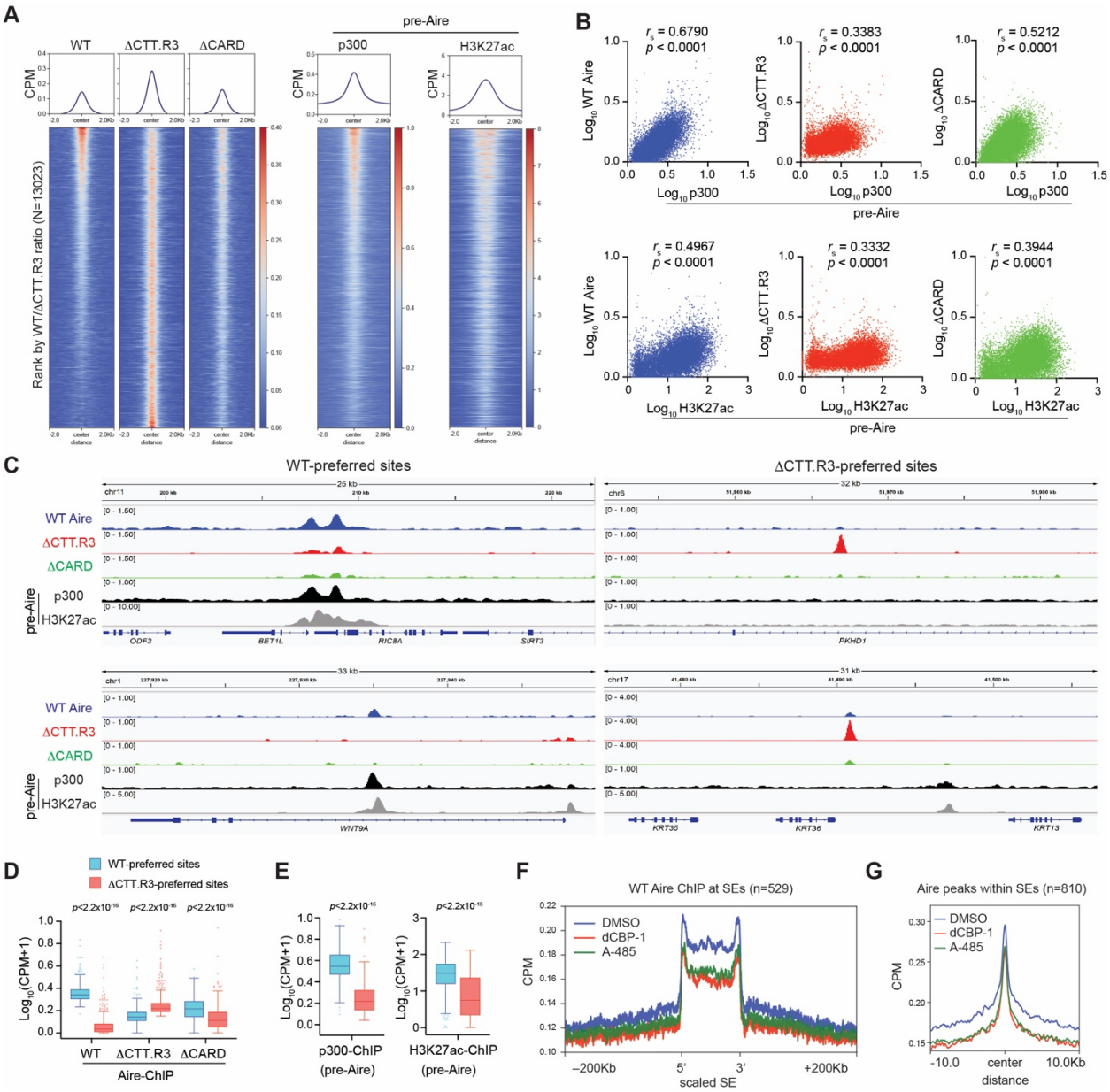

Figure 3. CTT and CARD domains bias Aire towards CBP/p300-enriched genomic loci

(A) Heatmaps of normalized ChIP-seq signals (Counts Per Million, CPM) for indicated proteins in stable 4D6 cells. Heatmaps are centered on all Aire peaks (n = 13023) and ranked by the ratio of WT hAire over ΔCTT.R3 ChIP-seq signals. hAire ChIP-seq signals were subtracted of background noise in corresponding input controls. p300 and H3K27ac ChIP-seq signals are from stable 4D6 cells prior to hAire expression (pre-Aire).

(B) Correlation between WT hAire, ΔCTT.R3, or ΔCARD with p300 (top panels) or H3K27ac (bottom panels) ChIP-seq signals at all Aire peaks (n=13,023). r_s, Spearman's correlation coefficient.

(C) Genome browser views of normalized ChIP-seq profiles for indicated proteins at exemplar WT-preferred sites versus ΔCTT.R3-preferred sites in stable 4D6 cells. hAire ChIP-seq signals shown were subtracted of background noise from corresponding input controls. Numbers to the left of each panel indicate the ranges of normalized CPM for ChIP-seq.

(D-E) Quantification of normalized ChIP-seq signals (CPM) for indicated proteins at WT-preferred sites versus ΔCTT.R3-preferred sites (top and bottom 500 peaks from A, respectively). ChIP-seq signals were normalized using Trimmed Mean of M-values (TMM, see Methods). *p*-values were calculated using Wilcoxon rank sum test.

(F-G) Contribution of CBP/p300 on Aire localization at super-enhancers (SEs). Stable 4D6 cells were treated with Dox for 24 hrs, and DMSO, 0.25 µM dCBP-1 or 3 µM A-485 were added 4 hrs prior to harvest. Average hAire ChIP-seq profiles (normalized CPM) from cells treated with p300 inhibitors compared to DMSO spanning 200Kb up- or downstream of H3K27ac-delimited SEs (n = 529, defined in Figure S1C) (F) or centered at Aire peaks located within SEs (n=810) (G). *p*-values were calculated using Wilcoxon rank sum test where $p < 2.2 \times 10^{-16}$ for DMSO vs. dCBP-1 and DMSO vs. A-485 in (F-G).

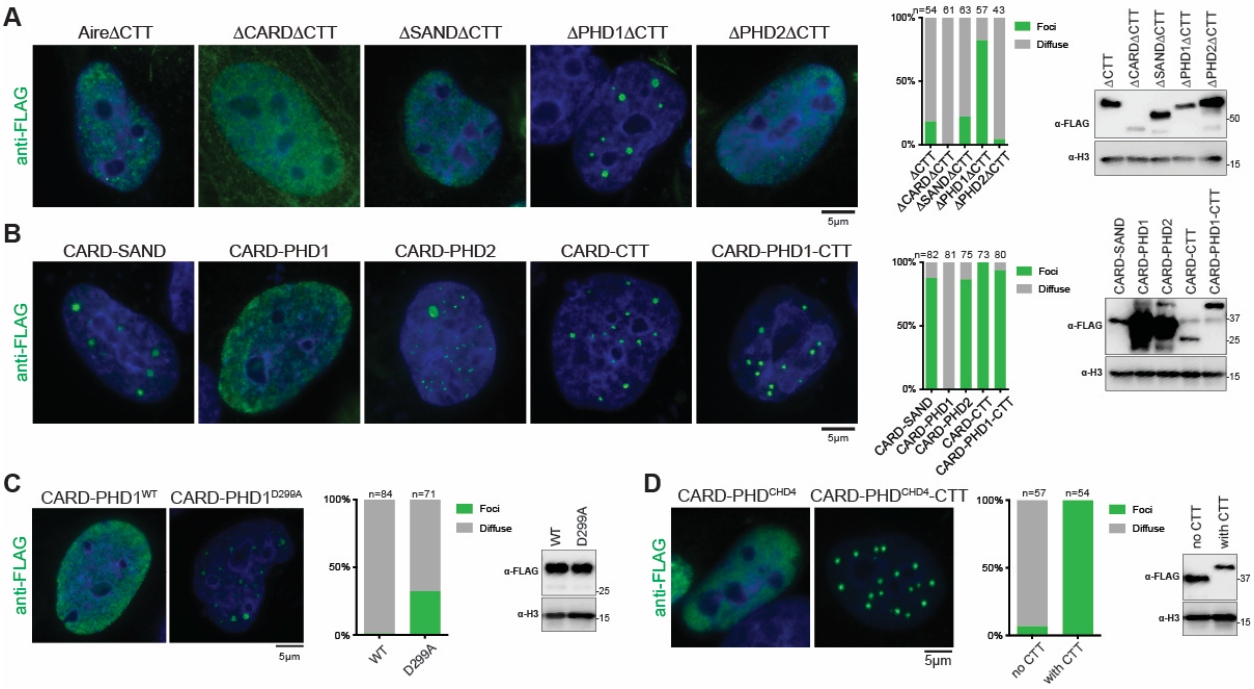

Figure 4. PHD1-mediated chromatin interaction suppresses CARD polymerization.

(A) Representative immunofluorescence images of mAire domain deletion mutants in 4D6 cells. Cells were transfected with mAire-FLAG expression plasmids 24 hrs prior to fixation. Center: percentage of nuclei with Aire foci vs. diffuse Aire staining. n represents the number of nuclei examined.

(B-C) Representative immunofluorescence images of mAire CARD-fusion variants. Experiments and analyses were done as in (A).

(D) Representative immunofluorescence images of FLAG-tagged mAire CARD fused with CHD4 PHD2 with and without CTT. Experiments and analyses were done as in (A).

All data are representative of at least three independent experiments.

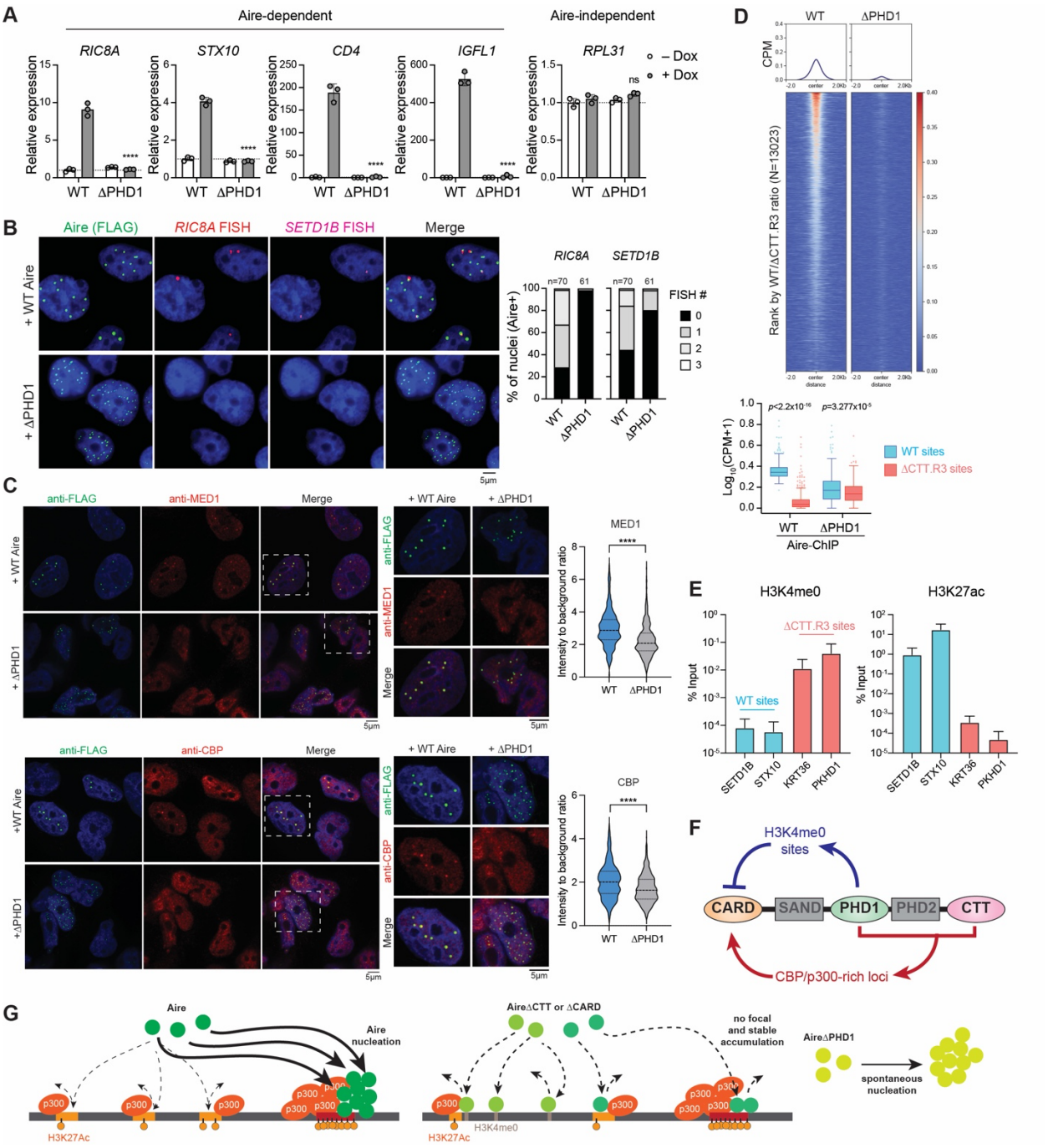

Figure 5. PHD1 cooperates with CTT to enable correct targeting and transcriptional hub formation of Aire.

(A) Transcriptional activity of WT hAire or ΔPHD1 mutant, as measured by the relative mRNA levels of Aire-dependent genes, *RIC8A*, *STX10*, CD4, and *IGFL1*, expressed in Dox-inducible stable 4D6 cells. An Aire-independent gene, *RPL31*, was also examined as a negative control. All genes were normalized against the internal control *RPL18*. Data are representative of at least three independent experiments and presented as mean ± SD,

n = 3. *p*-values (two-tailed unpaired t-test) were calculated in comparison to WT hAire (+Dox). ****$p < 0.0001$; $p > 0.05$ is not significant (ns).

(B) Representative images showing IF of hAire-FLAG and nascent RNA-FISH of Aire-induced genes, *RIC8A* and *SETD1B*, after 24 hrs Dox-induction of WT hAire or ΔPHD1. Right: percent of nuclei that have 0, 1, 2 or 3 RNA-FISH spots (left: *RIC8A*; right: *SETD1B*).

(C) Representative IF images of MED1 and CBP in stable 4D6 cells expressing hAire-FLAG WT or ΔPHD1. Middle: zoomed-in views of nuclei enclosed within white-dashed boxes. Right: quantitation of the average intensities of MED1/CBP staining at Aire foci. In each nucleus, MED1/CBP intensities at Aire foci were normalized to the average MED1/CBP intensity within the entire nucleus. ****$p<0.0001$ using a Mann-Whitney test.

(D) Heatmaps of normalized ChIP-seq signal (CPM) for indicated proteins in stable 4D6 cells. Top: heatmaps are centered on all Aire peaks (n = 13,023) and ranked by the ratio of WT hAire over ΔCTT.R3 ChIP-seq signals as in Figure 3A. Bottom: quantification of TMM-normalized ChIP-seq signal (CPM) at WT-preferred vs. ΔCTT.R3-preferred sites. Aire ChIP-seq signals shown were subtracted of background noise from corresponding input controls. *p*-value was calculated using Wilcoxon rank sum test. For clarity, WT Aire ChIP-seq heatmap was reproduced from Figure 3A.

(E) Anti-H3K4me0 and anti-H3K27ac ChIP-qPCR of hAire WT- or ΔCTT.R3-preferred target sites in stable 4D6 in the absence of hAire. Each pull-down sample amount was normalized to its respective input and reported as a % of input. Data are presented as the mean ± SD, n=4.

(F) Dual functions of PHD1. As a negative regulator, PHD1 disperses Aire to numerous H3K4me0 sites across the entire genome, thereby "diluting" Aire and preventing CARD polymerization at inappropriate locations. At Aire-target sites, however, PHD1 cooperates with CTT to recognize CBP/p300-rich loci, leading to the local concentration of Aire and promotion of Aire CARD polymerization.

(G) Model for Aire-mediated transcriptional hub assembly. Aire CTT directly binds the histone acetyltransferases CBP/p300 and utilizes CBP/p300-rich genomic loci as a nucleation platform for Aire's polymerizing domain CARD. CARD polymerization then enables focal recruitment of more Aire to its target sites, thereby further enhancing CBP/p300-driven transcriptional activity. In the absence of CTT, PHD1's inhibitory effect prevails, preventing CARD multimerization [see (F)]. In the absence of CARD, Aire recruitment is still biased towards CBP/p300-rich loci, but fails to amplify this bias and establish stable transcriptional hubs. In the absence of PHD1, Aire is not chromatin-bound, and CARD regulation is lost, resulting in unanchored and transcriptionally inactive Aire aggregates.

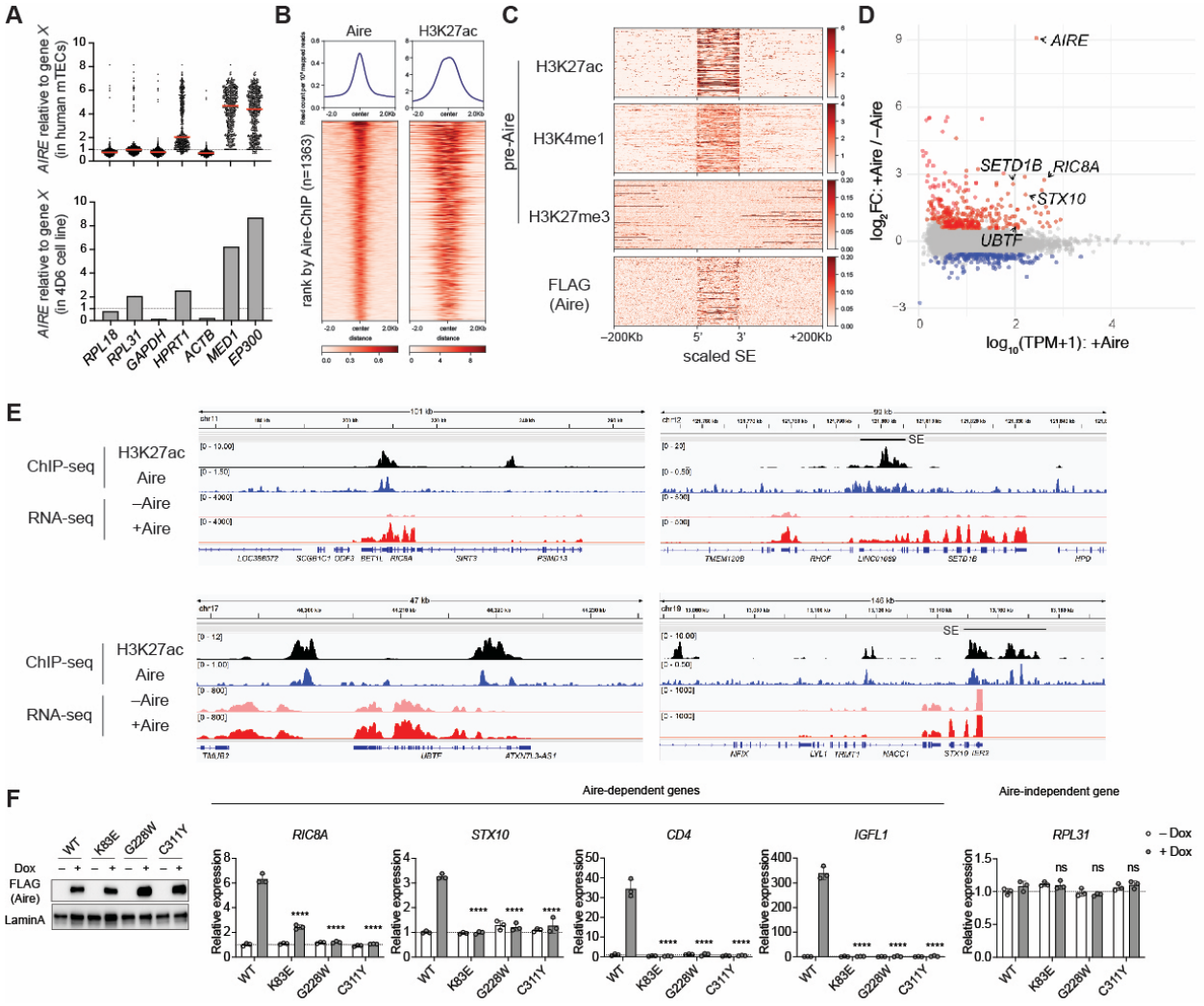

Figure S1. Thymic epithelial cell line (4D6) with Dox-inducible Aire recapitulates Aire transcriptional activity and function.

(A) *AIRE* transcripts relative to other gene transcripts in human *AIRE*$^+$ mTECs (top panel) and Dox-inducible *AIRE*-expressing 4D6 cells (bottom panel). Bulk RNA-seq was performed on 4D6 cells 24 hrs post-induction of *AIRE* expression. Data for single-cell RNA-seq performed on human thymic epithelial cells (n = 477 *AIRE*$^+$ mTECs) were from ref[2].

(B) Heatmaps of normalized Aire and H3K27ac ChIP-seq signals (Counts Per Million, CPM) in stable 4D6 cells. Heatmaps are centered on all Aire peaks (n = 1363) and ranked by Aire ChIP-seq intensity. Aire ChIP-seq was performed on 4D6 cells stably expressing WT human Aire-FLAG (hAire-FLAG) under a doxycycline (Dox)-inducible promoter. H3K27ac ChIP-seq signal was from stable 4D6 cells prior to hAire expression (pre-Aire).

(C) Aire ChIP heatmaps of normalized ChIP-seq signal (CPM) for indicated proteins 200Kb up- or downstream of H3K27ac-delimited super-enhancers (n = 529). Histone marks are shown in stable 4D6 cells without hAire expression (pre-Aire).

(D) Gene expression (transcripts per million, TPM) in Aire-expressing 4D6 cells (+Aire) versus expression changes between cells with and without Aire (+Aire/–Aire). Bulk RNA-seq was performed on Dox-inducible hAire-expressing stable 4D6 cells without or with 24 hr Dox treatment. Red, genes upregulated by Aire (>1.5-fold, FDR<0.05); blue, genes downregulated by Aire (<1.5-fold, FDR<0.05). FC, fold-change. FDR, false discovery rate.

(E) Genome browser views of normalized ChIP-seq and RNA-seq profiles at exemplar Aire-induced genes in doxycycline (Dox)-inducible hAire-expressing stable 4D6 cells. H3K27ac ChIP-seq and RNA-seq were shown in cells prior to hAire expression. Numbers to the left indicate the ranges of normalized reads for RNA-seq or counts per million (CPM) for ChIP-seq. SE, super-enhancer.

(F) Transcriptional activity of WT hAire or APS-1 mutants, as measured by the relative mRNA levels of Aire-dependent genes, *RIC8A*, *STX10*, *CD4* and *IGFL1*, in Dox-inducible 4D6 cells. An Aire-independent gene, *RPL31*, was also examined as a negative control. All genes were normalized against the internal control *RPL18*. Data are presented as mean ± SD, n = 3. *p*-values (one-way ANOVA with Dunnett's multiple comparisons test) were calculated in comparison to WT hAire (+ Dox). ****$p < 0.0001$; $p > 0.05$ is not significant (ns). Left: a western blot (WB) showing the nuclear levels of FLAG-tagged hAire proteins compared to endogenous levels of Lamin A.

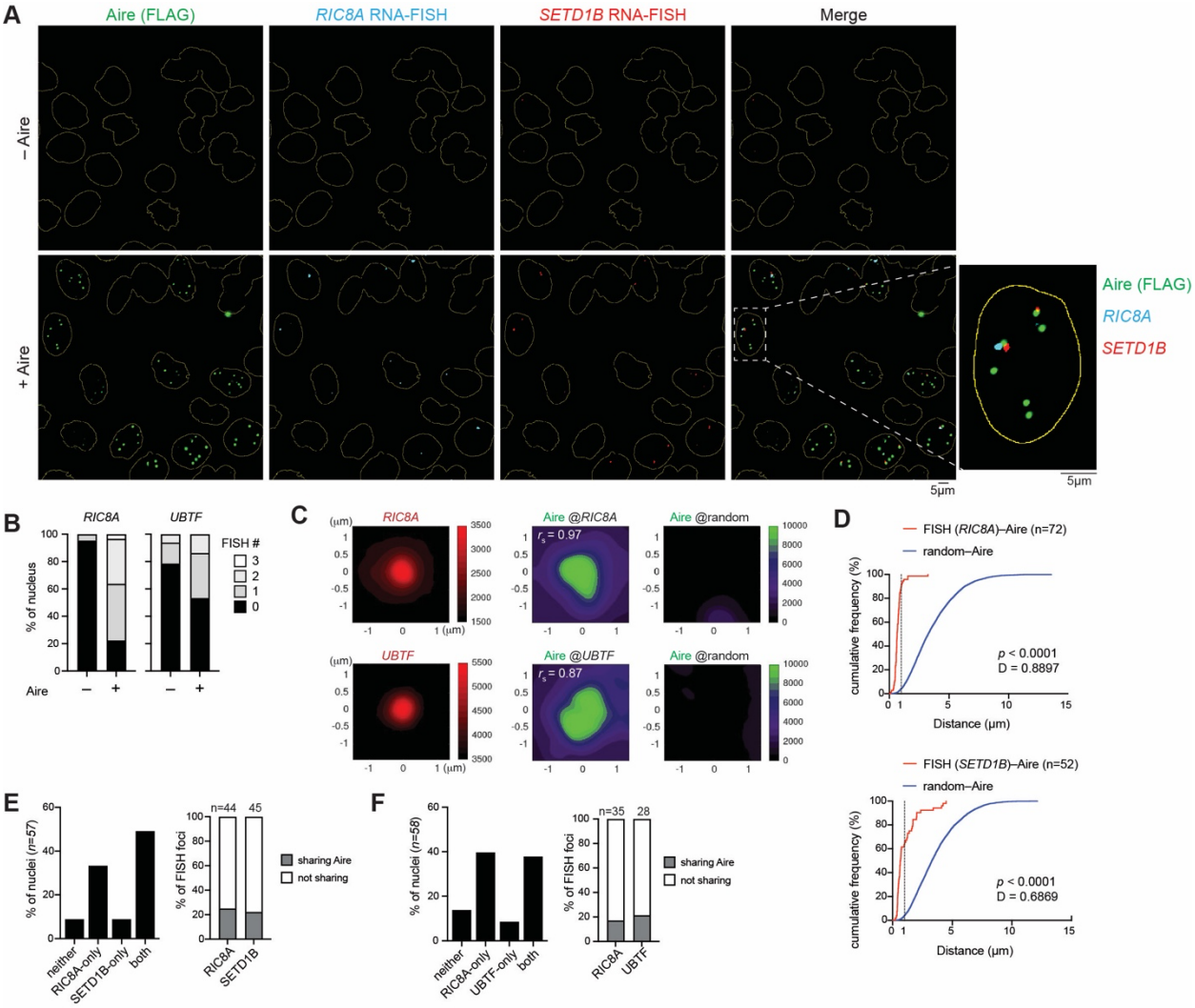

Figure S2. Aire links distinct inter-chromosomal genomic loci into active transcription hubs

(A) Nascent RNA-fluorescence in situ hybridization (FISH) coupled with immunofluorescence (IF) images of stable 4D6 cells expressing hAire-FLAG under a Dox-inducible promoter. Cells were not (–Aire) or were induced with 1 μg/ml Dox (+Aire) for 24 hrs before staining with FISH probes and anti-FLAG. RNA-FISH probes were designed to hybridize the intronic regions of Aire-dependent targets. Yellow outline represents the boundaries of nuclei from DAPI staining. Bottom right: zoomed-in view of the white-dashed box. For clarity, the zoomed-in image of a nuclei with *RIC8A* and *SETD1B* RNA-FISH spots sharing the same Aire focus was reproduced from Figure 1B.

(B) Quantitation of percent of nuclei that have 0, 1, 2 or 3 RNA-FISH spots (left: *RIC8A*; right: *UBTF*). A total of 65 and 58 nuclei were examined in cells treated without (–Aire) and with Dox (+Aire), respectively.

(C) Quantitation of average intensity signals from RNA-FISH combined with IF of hAire-FLAG in 4D6 cells. Shown are average signals of RNA-FISH (left), Aire IF centered on FISH spots (center) and Aire IF centered on randomly selected nuclear positions (right). r_s denotes the Spearman's correlation coefficient between RNA-FISH or IF signals.

(D) Cumulative distribution of the minimum center-to-center distances between Aire foci and observed RNA-FISH spots (FISH–Aire) or between Aire foci and simulated random positions (random–Aire). D indicates Kolmogorov-Smirnov test D statistic. *p*, *p*-value. While RNA-FISH and Aire foci often did not show perfect concentric colocalization, this is in line with recent studies showing that active genetic elements can be partially (within ~1 µm), rather than completely, overlapping with the target genes by imaging[3-5]. The close association between Aire foci and the target gene RNA-FISH spots therefore is consistent with the notion that Aire foci directly activate transcription.

(E) Quantitation of percent of nuclei harboring RNA-FISH spots. Left: statistics of *RIC8A* and *SETD1B* RNA-FISH spots from (A) and Figure 1B. Right: frequency of RNA-FISH spots sharing Aire foci, when measuring within nuclei showing RNA-FISH spots of both genes.

(F) Quantitation of percent of nuclei harboring RNA-FISH spots. Left: statistics of *RIC8A* and *UBTF* RNA-FISH spots from (B). Right: frequency of RNA-FISH spots sharing Aire foci, when measuring within nuclei showing RNA-FISH spots of both genes.

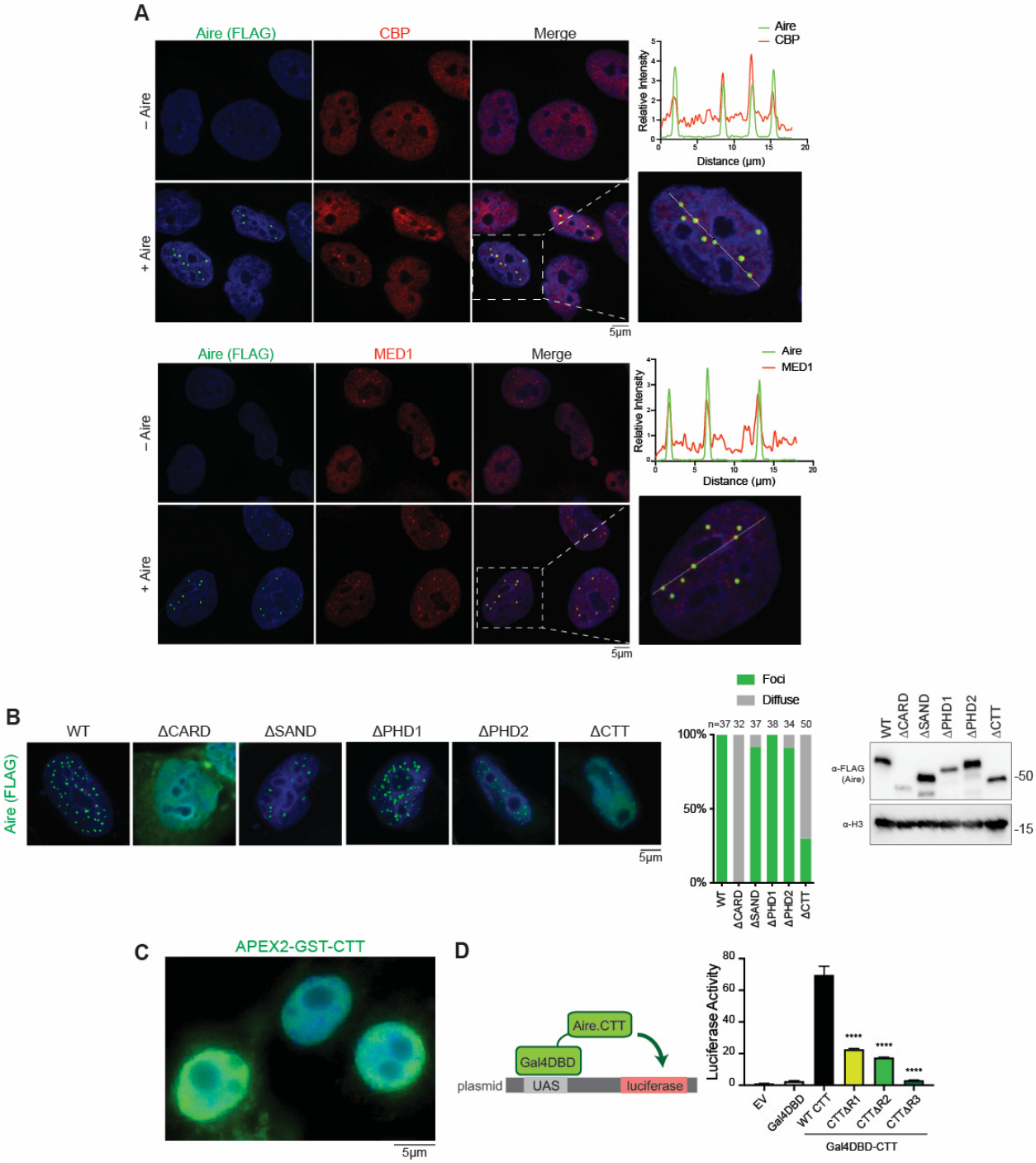

Figure S3. Aire foci represent sites of active transcription and require Aire CTT.

(A) Representative immunofluorescence images of endogenous p300, CBP and MED1 in 4D6 cells before and after hAire-FLAG expression. Cells were not (– Aire) or were induced with 1 μg/ml Dox (+ Aire) for 24 hrs before immunostaining with anti-FLAG, anti-CBP and anti-MED1. Right: zoomed-in views of nuclei enclosed within white-dashed boxes along with measured fluorescence intensities across drawn solid white lines.

(B) Representative fluorescence images of mouse Aire (mAire) WT and domain deletion variants in 4D6 cells. Cells were transfected with mAire-FLAG expression plasmids 24 hrs prior to fixation. Center: percentage of nuclei with Aire foci versus diffuse Aire staining. n represents the number of nuclei examined. Right: WB showing the levels of the indicated proteins.

(C) Representative fluorescence image of APEX2-GST-CTT in 293T. Cells were transfected with APEX2-GST-NLS-mAire.CTT-FLAG expression plasmid for 24 hrs prior to fixation and staining with anti-FLAG.

(D) TAD-like activity of mAire.CTT and various CTT deletion mutants. Top: schematic of the assay. CTT was fused with Gal4 DNA-binding domain (Gal4DBD), which binds upstream activation sequences (UAS) and controls the expression level of the reporter luciferase. The effect of Gal4DBD-CTT on the reporter activity was measured in 4D6 cells. Right: luciferase activities were shown relative to that of empty vector (EV)-transfected cells. Data are representative of at least three independent experiments and presented as mean ± SD, n = 3. *p*-values (one-way ANOVA with Dunnett's multiple comparisons test) were calculated in comparison to WT CTT, ****$p < 0.0001$.

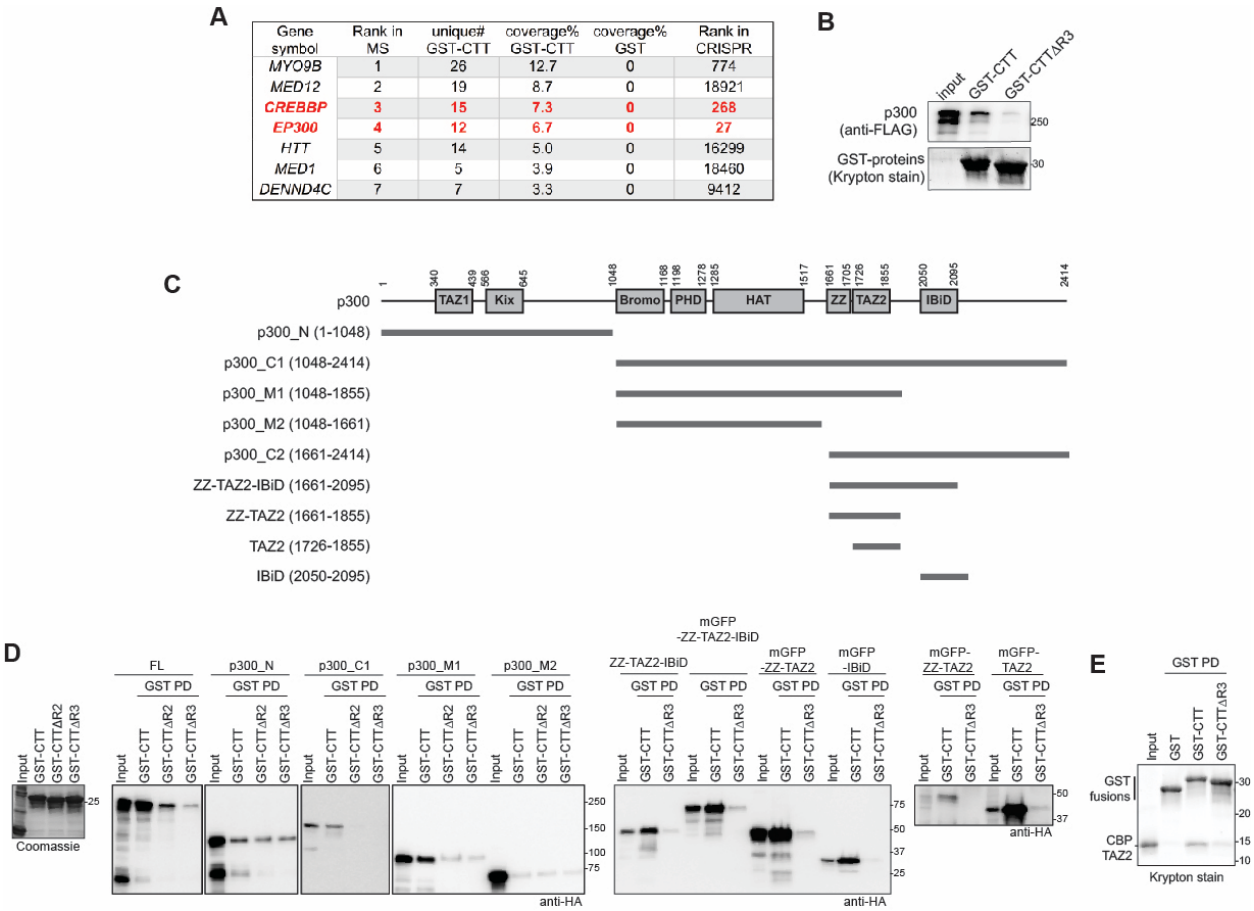

Figure S4. CBP/p300 directly interact with Aire CTT.

(A) LC-MS/MS analysis results of the GST-CTT pull-down in Figure 2C. ~250 kDa band in the GST-CTT-bound fraction and the equivalent region in the GST control were analyzed.

(B) His_6-GST-CTT pull-downs of full-length FLAG-tagged p300. Both baits and prey were purified recombinant proteins.

(C) Domain architecture of p300 and the truncation variants used in (D). Note that the homolog CBP has the same domain architecture.

(D) His_6-GST-CTT variant pull-downs of HA-tagged p300 transiently expressed in 293T nuclear extracts. Far left: SDS-PAGE gel of captured GST-CTT variant proteins captured onto glutathione beads used for the pull-down assay. Right: WBs of HA-p300 variants pulled-down with GST-CTT WT vs. GST-CTTΔR2 or GST-CTTΔR3.

(E) His_6-GST-CTT variant pull-downs of recombinantly expressed and purified CBP TAZ2.

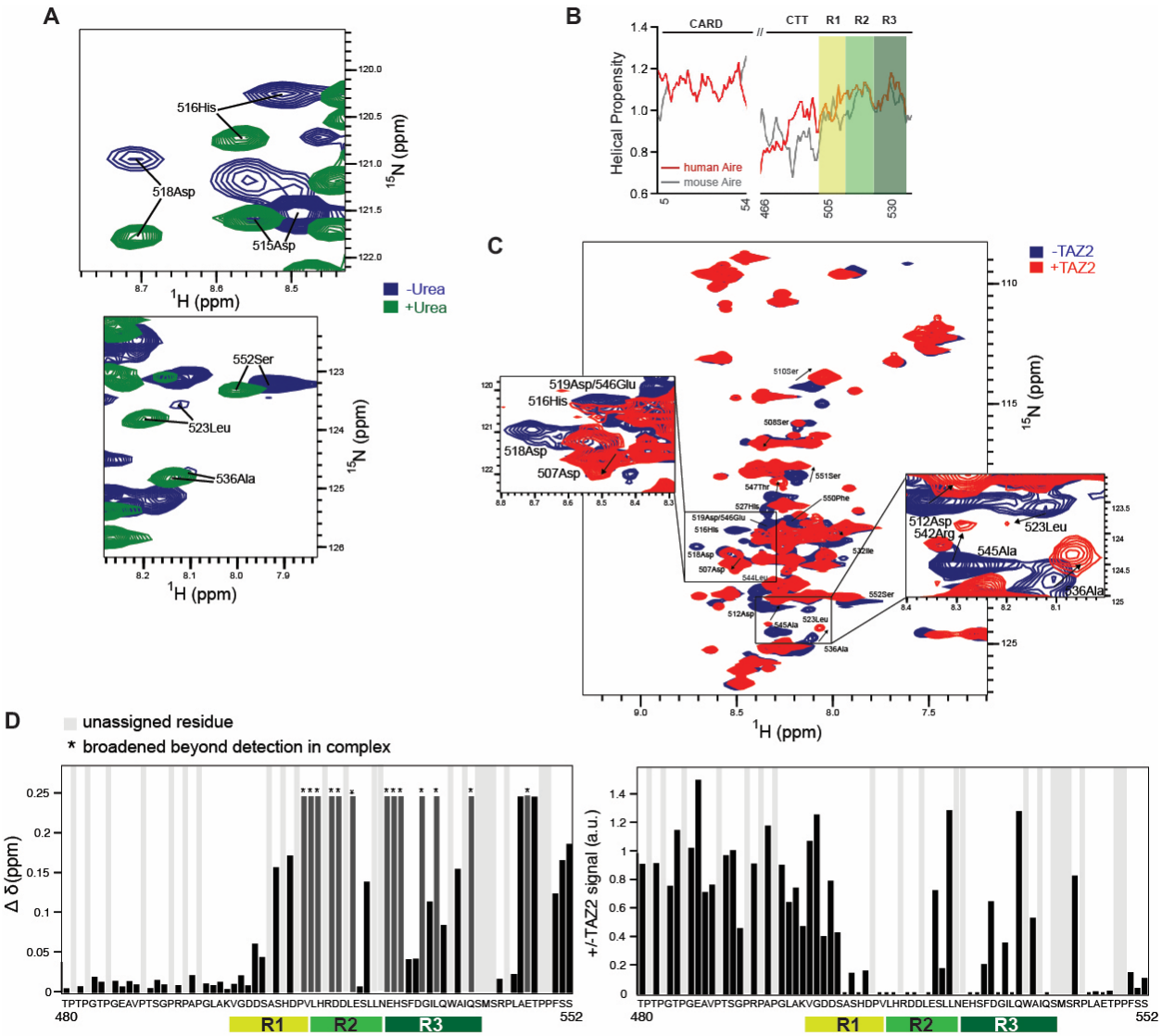

Figure S5. Aire CTT R2 and R3 bind CBP TAZ2.

(A) NMR spectroscopy of mAire CTT (aa 480-552) in the absence or presence of the chaotrope urea to identify regions with structural propensity, as measured by exchange broadening. 2D-^{15}N-HSQC of ^{15}N^{13}C-labeled mAire CTT was recorded at 15°C (288K) on an 800 MHz spectrometer equipped with a cryoprobe. To assign the CTT regions with exchange broadening, urea was added to ^{15}N^{13}C-labeled mAire CTT. Aire residues in both panels are severely exchange broadened in the absence of urea (blue), suggesting secondary structure propensity. Upon addition of urea (green), the structural propensity is lost, and the highlighted peaks gain in intensity, allowing for their assignment. The results show that, in the absence of TAZ2, isolated CTT is largely disordered, as evident from low dispersion in the proton chemical shift and secondary chemical shifts. R2 and R3, on the other hand, alternate between disordered and alpha helical characteristics, as evidenced by strong peak broadening. The helical propensities of R2 and R3 are

further supported by secondary structure predictions (B) and AlphaFold modeling (Figure 2E).

(B) Helical propensity of CTT from hAire and mAire (ProtScale[6], using Chou & Fasman parameters). Aire CARD, which has high α-helical propensity, was used in comparison.

(C) NMR spectroscopy of $^{15}N^{13}C$-labeled mAire CTT (aa 480-552) in the absence or presence of CBP TAZ2. 2D-^{15}N-HSQC of $^{15}N^{13}C$-labeled mAire CTT alone (blue) and with 1:1.75 molar ratio of $^{15}N^{13}C$-labeled mAire CTT:CBP TAZ2 (red), recorded at 15°C (288K) on an 800 MHz spectrometer equipped with a cryoprobe. Close-ups at a lower contour level show the extent of peak broadening and chemical shift perturbation upon complex formation in select regions.

(D) Plots of detected chemical shift perturbations (CSPs) $\Delta\delta$ [calculated as $\Delta\delta$=sqrt(0.14$*\Delta N^2$+ΔH^2)][7] and the signal loss upon addition of CBP TAZ2 (1:1.75 molar ratio of mAire CTT:CBP TAZ2). mAire CTT residues marked with an asterisk were broadened beyond detection in the mAire–CBP complex.

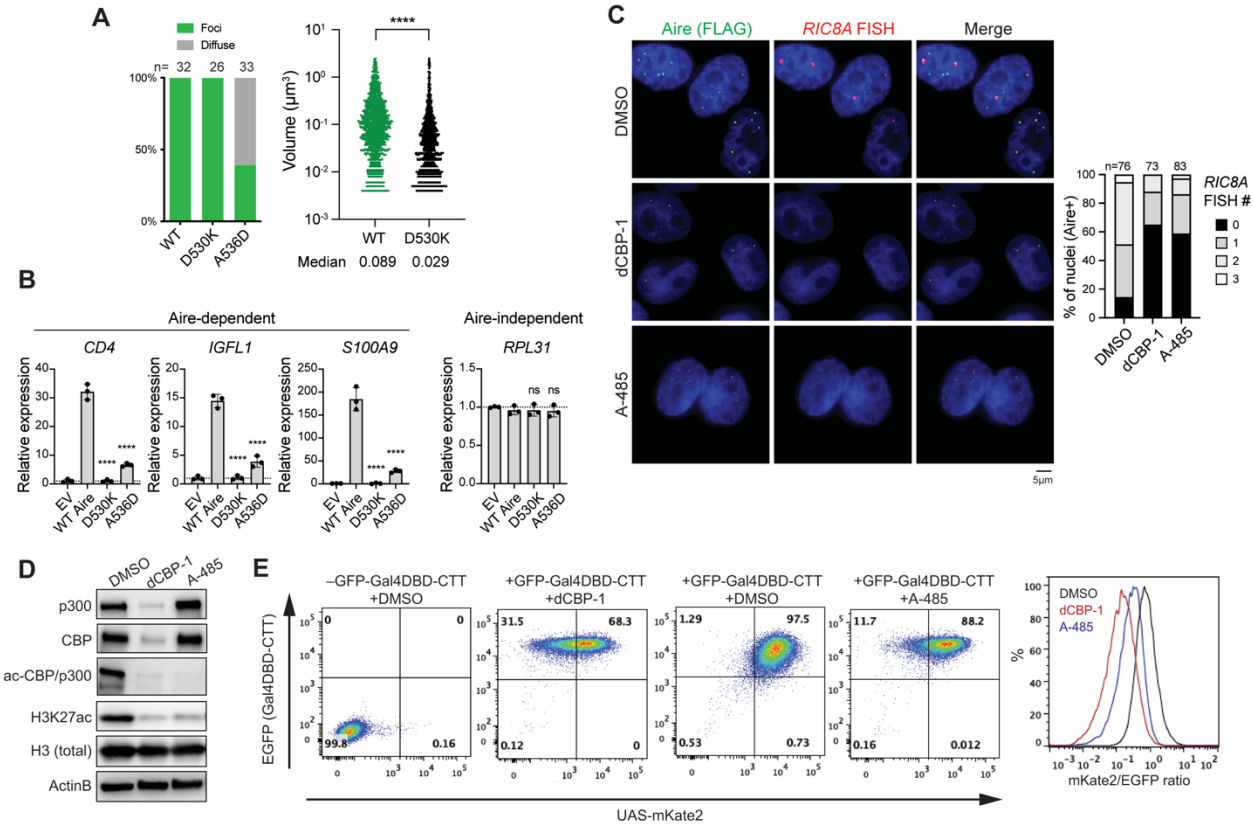

Figure S6. CBP/p300 are essential for Aire hub formation.

(A) Quantitative analysis of the immunofluorescence images of mAire variants in Figure 2H. Left: percentage of nuclei with Aire foci vs. diffuse Aire staining. n represents the number of nuclei examined for each sample. Right: quantification of mAire WT and D530K foci volumes. Each data point represents individual Aire foci, n = 1597 and 1369, for cells expressing WT and D530K, respectively. ****$p < 0.0001$ (two-tailed Mann-Whitney test).

(B) Transcriptional activity of WT mAire or CTT point mutants, as measured by the relative mRNA levels of Aire-dependent genes, *CD4*, *IGFL1* and *S100A9*, in 4D6 cells transiently transfected with plasmids expressing mAire variants. An Aire-independent gene, *RPL31*, was also examined as a negative control. All genes were normalized against the internal control *RPL18*. Data are presented as mean ± SD, n = 3. *p*-values (one-way ANOVA with Dunnett's multiple comparisons test) were calculated in comparison to WT Aire. ****$p < 0.0001$; $p > 0.05$ is not significant (ns).

(C) Representative images showing IF of hAire-FLAG and nascent RNA-FISH of Aire-induced gene *RIC8A* in the presence of dCBP-1 and A-485 in stable 4D6 cells. Right: percent of nuclei that have 0, 1, 2 or 3 RNA-FISH spots; n represents the number of nuclei examined for each sample. Drug treatments of cells were done as in Figure 2I.

(D) WB showing the levels of indicated proteins post-24hr treatment of the p300/CBP degrader dCBP-1 (0.25 μM) or the catalytic inhibitor A-485 (3 μM) in comparison to vehicle control DMSO in stable 4D6 cells.

(E) Flow cytometry plots showing expression levels of the mKate2 reporter versus GFP-Gal4DBD-CTT (left) or the distribution of their ratios (right) in stable 4D6 cells. Cells were treated with Dox to induce GFP-Gal4DBD-CTT expression in the presence of DMSO, 0.25 µM dCBP-1 or 3 µM A-485 for 24 hrs prior to flow cytometry examination.

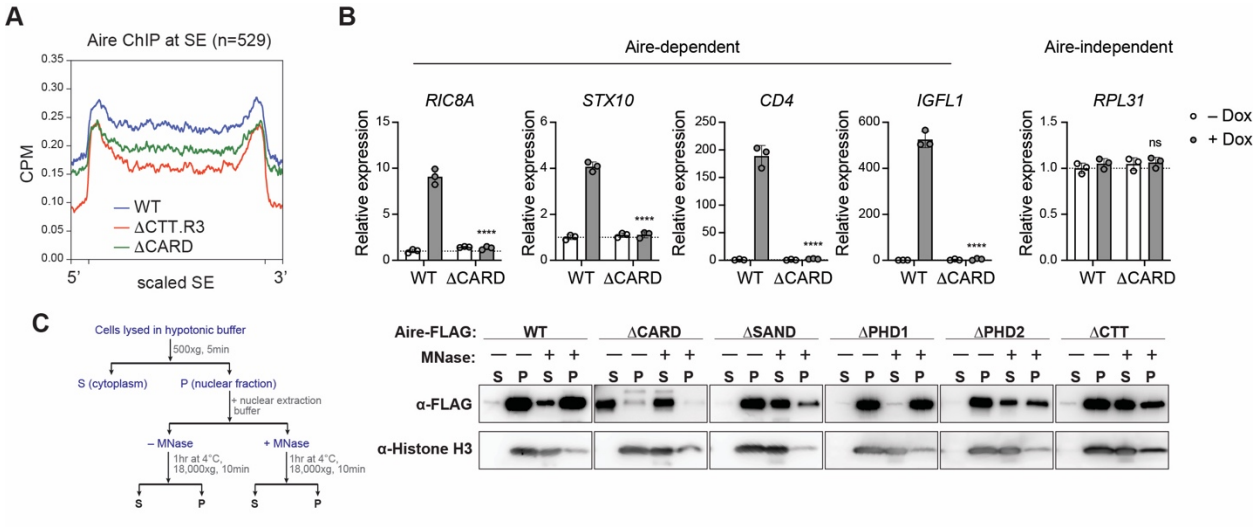

Figure S7. CARD is essential for robust Aire localization to super enhancers and mediating transcriptional activity.

(A) Average WT hAire, ΔCTT.R3, and ΔCARD ChIP-seq profiles (CPM) at H3K27ac-delimited super-enhancers (SEs, n = 529, defined in Fig S1C). *p*-values were calculated using Wilcoxon rank sum test ($p < 2.2\text{x}10^{-16}$ for both WT vs. ΔCTT.R3 and WT vs. ΔCARD).

(B) Transcriptional activity of WT hAire or ΔCARD mutant, as measured by the relative mRNA levels of Aire-dependent genes, *RIC8A*, *STX10*, CD4, and *IGFL1*, in Dox-inducible 4D6 cells. An Aire-independent gene, *RPL31*, was also examined as a negative control. Data are representative of at least three independent experiments and presented as mean ± SD, n = 3. *p*-values (two-tailed unpaired t-test) were calculated in comparison to WT hAire (+Dox). ****$p < 0.0001$; $p > 0.05$ is not significant (ns).

(C) Chromatin fractionation analysis of WT and domain deletion variants of mAire. Left: schematic of chromatin fractionation analysis of Aire. 293T cells were transfected with mAire expressing plasmids for 48 hrs before harvesting. Solubility of Aire and chromatin (as measured by Histone H3) before and after MNase treatment was analyzed by WB. Independently of MNase, ΔCARD was soluble, while ΔPHD1 was insoluble. Solubilities of all other variants were dependent on MNase, consistent with their chromatin interaction.

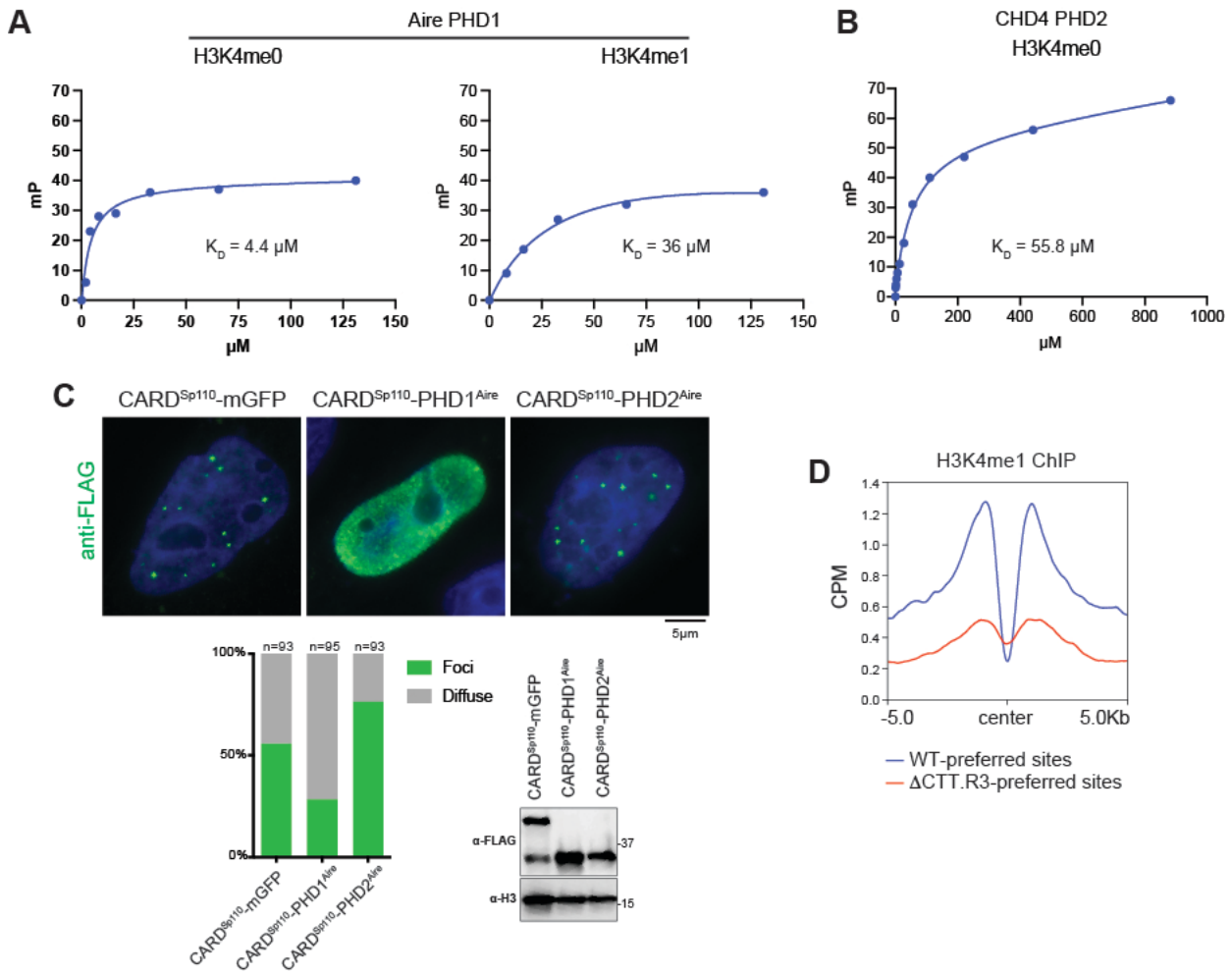

Figure S8. PHD1's ability to bind H3K4me0 is required for anchoring Aire to chromatin.

(A) Binding of the indicated fluorescein-labeled histone H3 tail-derived peptides to increasing concentrations of mAire PHD1. Binding is detected by a change in the fluorescence polarization (mP) of fluorescein. A representative experiment is shown for each peptide. The curves indicate the fit to a simple binding isotherm for the data sets shown. Data are representative of three independent experiments and presented as mean ± SD. H3K4me0 K_D = 5.25 ± 2.59 μM; H3K4me1 K_D = 23.18 ± 11.8 μM.

(B) Binding of the fluorescein-labeled histone H3 tail-derived peptide (H3K4me0) to increasing concentrations of the CHD4 PHD2. Data are representative of three independent experiments and presented as mean ± SD. K_D = 45.3 ± 10.2 μM.

(C) Representative immunofluorescence images of FLAG-tagged Sp110 CARD fused with mGFP (monomeric GFP), mAire PHD1 or mAire PHD2. Experiments and analyses were done as in Figure 4A.

(D) Average H3K4me1 ChIP profiles (normalized CPM) centered at WT-preferred sites versus ΔCTT.R3-preferred sites (top and bottom 500 peaks from Figure 3A, respectively). H3K4me1 ChIP was performed in 4D6 cells prior to hAire expression. *p*-value was calculated using Wilcoxon rank sum test ($p < 2.2\text{x}10^{-16}$).

Figure Legend References

1 Richard Evans, M. O. N., Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, Demis Hassabis. Protein complex prediction with AlphaFold-Multimer. *bioRxiv* (2021).

2 Park, J. E. *et al.* A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367** (2020). https://doi.org:10.1126/science.aay3224

3 Benabdallah, N. S. *et al.* Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol Cell* **76**, 473-484 e477 (2019). https://doi.org:10.1016/j.molcel.2019.07.038

4 Cho, W. K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412-415 (2018). https://doi.org:10.1126/science.aar4199

5 Gu, B. *et al.* Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* **359**, 1050-1055 (2018). https://doi.org:10.1126/science.aao3136

6 Elisabeth Gasteiger, C. H., Alexandre Gattiker, S'everine Duvaud, Marc R. Wilkins, Ron D. Appel, Amos Bairoch Protein Identification and Analysis Tools on the ExPASy Server. *John M. Walker (ed): The Proteomics Protocols Handbook*, 571-607 (2005).

7 Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* **73**, 1-16 (2013). https://doi.org:10.1016/j.pnmrs.2013.02.001