

## RESEARCH HIGHLIGHT

## Z-DNA marks the spot for AIRE

Yael Goldfarb<sup>1</sup>✉ and Jakub Abramson<sup>1</sup>✉

© The Author(s) under exclusive licence to Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences 2024

Cell Research (2024) 0:1–2; https://doi.org/10.1038/s41422-024-00964-x



**Autoimmune regulator (AIRE) is a unique transcriptional regulator that induces ectopic expression of thousands of tissue-specific genes in the thymus, a step critical for the establishment of immunological tolerance to self. In their recent publication, Fang et al. provide novel mechanistic insights into AIRE's *modus operandi*, by highlighting Z-DNA as a key *cis*-regulatory element, critical for guiding AIRE to its target genes.**

Autoimmune regulator (AIRE) is a transcriptional regulator specifically expressed in medullary thymic epithelial cells (mTECs), which play an indispensable role in establishment of a functional, yet self-tolerant, T cell repertoire. The physiological significance of AIRE is best illustrated by its dysfunction in humans, which results in a rare devastating autoimmune syndrome, characterized by lymphocyte-driven destruction of various peripheral tissues. For over two decades, researchers have attempted to shed light on AIRE's *modus operandi* and to explain how a single factor can have such a massive impact on gene expression, which in mTECs entails upregulation of thousands of genes. Interestingly, most of the AIRE-induced targets represent tissue-specific transcripts (e.g., insulin, etc.) and their AIRE-dependent ectopic expression by mTECs ensures that as broad a repertoire of self-antigens as possible is presented to developing T cells. This is key to effective elimination of self-reactive T cell clones in the thymus, either through their clonal deletion or through their conversion into Foxp3<sup>+</sup> regulatory T cells, which can actively suppress self-reactive lymphocyte escapees in the immune periphery (reviewed in ref.<sup>1</sup>).

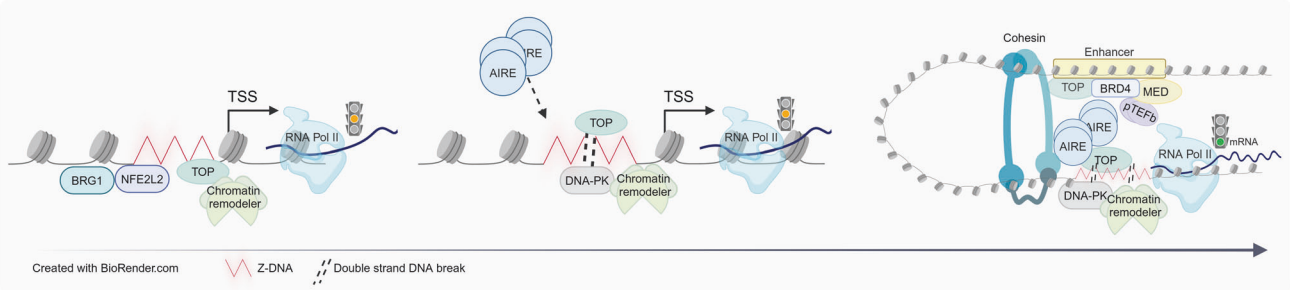
Interestingly, unlike classical transcription factors, AIRE was found to lack a conventional DNA-binding motif, which would allow it to bind to specific genomic sequences. For this reason, most of the previous studies have focused on epigenetic clues that could guide AIRE to its targets. Indeed, AIRE was found to favor silenced chromatin, as it was reported to preferentially target loci that lack active histone marks at their promoters and are instead decorated by repressive chromatin modifications.<sup>2</sup> Moreover, AIRE was found to specifically recognize and bind to H3K4me0,<sup>3,4</sup> a hallmark of silent chromatin, further supporting the notion that AIRE is recruited to transcriptionally silent genes to “wake them up”. Furthermore, while transcriptionally silent, AIRE targets were later found to be demarcated by stalled RNA polymerase II (RNA Pol II) at their transcription start sites (TSSs),<sup>5</sup> which consequently tags them as poised. Remarkably, AIRE was recently shown to unleash the stalled RNA Pol II from pausing<sup>5</sup> through cohesin-dependent looping of its target genes with H3K27ac-rich superenhancer regions<sup>6</sup> and subsequent recruitment of the Mediator complex and other proteins critical for RNA Pol II release to the TSS proximity.<sup>7</sup> Interestingly, AIRE was also

found to interact with several proteins linked to the DNA damage response<sup>8</sup> including topoisomerases,<sup>6</sup> suggesting that AIRE-dependent gene expression may also be associated with DNA breaks. While all these studies resolved a significant portion of the AIRE puzzle, some of its parts have remained incomplete. Recently, a landmark study by Fang et al.<sup>9</sup> published in *Nature*, added several key pieces to this constantly evolving riddle and provided an important breakthrough in our understanding of how AIRE finds its target genes and how breaks in inherently fragile Z-DNA structures (that were found to be enriched in AIRE promoters) may play a pivotal role in this process.

In this study, the authors sought to revisit a long-standing question wondering which *cis*-regulatory elements discriminate between AIRE-responsive vs AIRE-neutral genes. To this end, they utilized most recent advances in deep learning algorithms based on convolutional neural networks, which were trained to distinguish between extended promoter regions of AIRE-induced genes and those of AIRE-neutral genes in mouse mTECs. In parallel, the authors also used *in vivo* analyses of bulk RNA-seq and ATAC-seq of mTECs from F1-hybrid mice from B6 and NOD strains, to pinpoint possible association between strain-dependent genetic variations for various transcription factor-binding motifs and allelic imbalances in the chromatin accessibility and expression of AIRE-induced genes. Strikingly, both unrelated and original approaches highlighted Z-DNA and NFE2-MAF motifs as novel *cis*-regulatory elements that may determine AIRE's target choice.

Z-DNA is a dynamic left-handed non-conventional form of the double helix, which was previously found to be enriched in gene promoter regions and thus implicated in gene regulation.<sup>10</sup> In general, Z-DNA structures are formed from sequences that are rich in alternating purine-pyrimidine tracts, most notably consisting of (CA)<sub>n</sub> repeats.<sup>10</sup> Indeed, comparison of promoter sequences from AIRE-induced vs AIRE-neutral genes revealed that AIRE-induced genes are enriched for longer (CA)<sub>n</sub> repeats (i.e., putative Z-DNA motifs) compared to AIRE-neutral genes and that the Z-DNA-forming motif distance from the TSS was inversely correlated with AIRE targeting. Furthermore, *in silico* mutation of AIRE-neutral gene promoter sequences by replacing part of the sequence with Z-DNA-forming (CA)<sub>n</sub> repeats or NFE2-MAF-binding motifs, increased their likelihood to be detected as AIRE-induced genes, further supporting the hypothesis that they serve as key *cis*-regulatory elements for AIRE targeting. In addition, the authors also confirmed the presence of the actual Z-DNA structures in mTEC promoters *in vivo* by ChIP-seq and found a positive correlation between AIRE-induced gene expression and Z-DNA ChIP-seq signal strength at promoters. Specifically, Z-DNA motifs were enriched in AIRE peaks and vice versa with AIRE binding

<sup>1</sup>Department of Immunology and Regenerative Biology, Weizmann Institute of Science, Rehovot, Israel. ✉email: yael.goldfarb@weizmann.ac.il; jakub.abramson@weizmann.ac.il



**Fig. 1 Z-DNA directs AIRE to its target genes.** The promoters of AIRE target genes are enriched with Z-DNA and NFE2-MAF motifs and with poised RNA Pol II at TSS. In a small fraction of cells, the promoter regions are already accessible and harbor DSBs in the Z-DNA regions due to the inherent fragility of the Z-DNA. This in turn facilitates recruitment of various members of the DNA damage machinery (e.g., DNA-PK), cohesin complex and AIRE homo-oligomers, which then facilitate chromatin looping with corresponding enhancer regions (bound by BRD4, pTEFb and the Mediator complex). This eventually results in release of the paused RNA Pol II and expression of the given gene.

levels also correlated with Z-DNA signals. By utilizing spermidine to stabilize Z-DNA *in vivo*, the authors further showed enhanced AIRE-induced gene expression and promoter chromatin accessibility in wild-type (WT) mTECs and a partial rescue of the mTEC transcriptome in AIRE knockout cells. These results suggested that Z-DNA enables the transcription machinery to function to a certain extent even in the absence of AIRE. Interestingly, NFE2L2, a member of the bZIP-family of transcription factors, which was identified in both screens as having *cis*-regulatory potential on AIRE action, has previously been shown to recruit the chromatin remodeler BRG1 to stabilize Z-DNA.<sup>11</sup> Notably, BRG1 was previously shown to organize the chromosomal architecture in mTECs upstream of AIRE.<sup>12</sup> Indeed, in the current study open chromatin regions upregulated by BRG1 were shown to be enriched with longer Z-DNA-forming motifs, as well as NFE2L2-binding motifs.

Remarkably, Z-DNA structures are also energy intensive, rigid and inherently fragile, making them prone to double stranded breaks (DSBs). Indeed, the authors demonstrated that DSB hotspots were enriched in Z-DNA ChIP-seq peaks and vice versa in both WT and AIRE-deficient mTECs, suggesting that Z-DNA promotes the formation of DSBs independently of AIRE. While AIRE was previously suggested to induce expression of its target genes via stabilizing topoisomerase-dependent generation of DSBs at TSS,<sup>6</sup> subsequent studies suggested that AIRE is rather recruited to superenhancers, where it facilitates their activation via stabilization of TOP1-dependent DSBs.<sup>6</sup> In both cases, the DSBs were suggested to provoke recruitment of DNA damage machinery, that would facilitate local chromatin remodeling, eventually leading to RNA Pol II pause release and subsequent elongation. Most recent findings by Fang et al. and others, however, suggest that DSBs and chromatin remodeling at AIRE targets occur upstream and independently of AIRE and are a prerequisite for the recruitment of AIRE to these targets (Fig. 1). The formation of the DSB at the Z-DNA motifs may be to a high extent probabilistic and may thus provide a better explanation to

the stochastic and noisy nature of AIRE-dependent gene regulation. While it remains unclear whether AIRE can directly recognize Z-DNA structures or is rather recruited to DSBs indirectly via the DNA damage response machinery, there is a clear overlap between the Z-DNA structures, DSBs and AIRE's occupancy on the promoter regions. Increased recruitment of AIRE homo-oligomers to these predetermined regulatory regions, seems to enhance the probability of cohesin-based looping of these regions with their corresponding distal regulatory elements, thus boosting the expression of AIRE's targets (Fig. 1).

## REFERENCES

- Klein, L., Robey, E. A. & Hsieh, C. S. *Nat. Rev. Immunol.* **19**, 7–18 (2019).
- Sansom, S. N. et al. *Genome Res.* **24**, 1918–1931 (2014).
- Koh, A. S. et al. *Proc. Natl. Acad. Sci. USA* **105**, 15878–15883 (2008).
- Org, T. et al. *EMBO Rep.* **9**, 370–376 (2008).
- Giraud, M. et al. *Proc. Natl. Acad. Sci. USA* **109**, 535–540 (2012).
- Bansal, K., Yoshida, H., Benoist, C. & Mathis, D. *Nat. Immunol.* **18**, 263–273 (2017).
- Bansal, K. et al. *Proc. Natl. Acad. Sci. USA* **118**, e2110991118 (2021).
- Abramson, J., Giraud, M., Benoist, C. & Mathis, D. *Cell* **140**, 123–135 (2010).
- Fang, Y., Bansal, K., Mostafavi, S., Benoist, C. & Mathis, D. *Nature* **628**, 400–407 (2024).
- Wang, G. & Vasquez, K. M. *Nat. Rev. Genet.* **24**, 211–234 (2022).
- Zhang, J. et al. *Mol. Cell Biol.* **26**, 7942–7952 (2006).
- Koh, A. S. et al. *Nat. Immunol.* **19**, 162–172 (2018).

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to Yael Goldfarb or Jakub Abramson.

**Reprints and permission information** is available at <http://www.nature.com/reprints>