



OPEN ACCESS

EDITED BY

Claudio Pignata,
University of Naples Federico II, Italy

REVIEWED BY

Margaret A. Jordan,
James Cook University, Australia
Emilia Cirillo,
University of Naples Federico II, Italy

*CORRESPONDENCE

Amund Holte Berger
✉ Amund.Berger@uib.no
Eirik Bratland
✉ Eirik.Bratland@uib.no
Stefan Johansson
✉ Stefan.Johansson@uib.no

†These authors have contributed
equally to this work and share
last authorship

RECEIVED 07 February 2025

ACCEPTED 25 March 2025

PUBLISHED 22 April 2025

CITATION

Berger AH, Oftedal BE, Wolff ASB,
Husebye ES, Knappskog PM, Bratland E and
Johansson S (2025) High-resolution
transcriptional impact of AIRE: effects of
pathogenic variants p.Arg257Ter, p.Cys311Tyr,
and polygenic risk variant p.Arg471Cys.
Front. Immunol. 16:1572789.
doi: 10.3389/fimmu.2025.1572789

COPYRIGHT

© 2025 Berger, Oftedal, Wolff, Husebye,
Knappskog, Bratland and Johansson. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

High-resolution transcriptional impact of AIRE: effects of pathogenic variants p.Arg257Ter, p.Cys311Tyr, and polygenic risk variant p.Arg471Cys

Amund Holte Berger^{1,2,3*}, Bergithe Eikeland Oftedal¹,
Anette Susanne Bøe Wolff^{1,4}, Eystein Sverre Husebye^{1,4},
Per Morten Knappskog^{1,3}, Eirik Bratland^{1,3*†}
and Stefan Johansson^{2,5*†}

¹Department of Clinical Science, University of Bergen, Bergen, Norway, ²Mohn Center for Diabetes Precision Medicine, Department of Clinical Science, University of Bergen, Bergen, Norway,

³Department of Medical Genetics, Haukeland University Hospital, Bergen, Norway, ⁴Department of Medicine, Haukeland University Hospital, Bergen, Norway, ⁵Children and Youth Clinic, Haukeland University Hospital, Bergen, Norway

Introduction: The Autoimmune Regulator, AIRE, acts as a transcriptional regulator in the thymus, facilitating ectopic expression of thousands of genes important for the process of negative T-cell selection and immunological tolerance to self. Pathogenic variants in the gene encoding AIRE are causing Autoimmune polyendocrine syndrome type 1 (APS-1), defined by multiorgan autoimmunity and chronic mucocutaneous candidiasis. More recently, Genome Wide Association Studies (GWAS) have also implicated AIRE in several common organ-specific autoimmune diseases including autoimmune primary adrenal insufficiency, type 1 diabetes and pernicious anemia.

Methods: We developed a highly sensitive cell-system approach based on HEK293FT cells transfected with AIRE that allowed us to characterise and functionally evaluate the transcriptional potential of genetic variants in the AIRE gene. By utilizing RNAseq with an average read depth of 100 million reads and 12 replicates per condition we have the statistical power and sensitivity to characterize the AIRE induced transcriptome in depth.

Results: We confirm that our cell system recapitulates the expression of the vast majority of known AIRE induced genes including well-characterised tissue restricted antigens (TRAs). Our approach also increases the total number of identified AIRE induced genes by an order of magnitude compared to previously published strategies, including a comprehensive number of clinically relevant autoantigens.

Discussion: Our cell-system approach differentiates between categories of AIRE variants on the transcriptional level, including the nonsense variant p.R257* (near complete loss of function), the p.C311Y variant associated with dominantly inherited APS-1 (severely impaired function), and the polygenic risk variant p.R471C (slightly increased function) linked to common organ-specific

autoimmunity. The increased activity of p.R471C compared to wildtype indicates different molecular mechanisms for monogenic and polygenic AIRE related autoimmunity. We find that AIRE induced expression is characterised by a small absolute increase in expression levels of genes of both high and low tissue specificity.

KEYWORDS

AIRE, Autoimmune polyendocrine syndrome type, APECED, Autoimmune primary adrenal insufficiency, type 1 diabetes (T1D), pernicious anemia, RNAseq, HEK293FT

1 Introduction

Epithelial cells of the thymus stroma, specifically medullary thymic epithelial cells (mTECs), are instrumental for the generation of a repertoire of self-tolerant T-cells (1). mTECs are able to turn on the expression of thousands of genes normally restricted to cell lineages of specialized tissues, collectively referred to as tissue restricted antigens (TRAs) (2). The precise molecular events that drive this transcriptional plasticity in mTECs are incompletely understood, although studies of the rare monogenic condition Autoimmune polyendocrine syndrome type 1 (APS-1) have identified a crucial role for the AutoImmune Regulator (AIRE) in these processes. APS-1, also known as autoimmune polyendocrinopathy candidiasis-ectodermal dystrophy (APECED, OMIM 240300), is clinically characterized by a wide range of autoimmune manifestations, affecting both endocrine and non-endocrine tissues (3).

APS-1 can be inherited in both autosomal dominant and recessive manners depending on the molecular effect of the pathogenic genetic variants in AIRE, encoding the AutoImmune Regulator (AIRE) protein (4–8). In addition to the monogenic disorder APS-1, genetic variations in AIRE have recently been associated with more common organ-specific autoimmune diseases such as Addison's disease, pernicious anemia and type 1 diabetes (9–11). In particular the p.R471C variant, with a frequency of 1–2% in Northern European populations, have been associated with more than 3-fold increased risk of Addison's disease, almost 2-fold increased risk of pernicious anemia and 1.5-fold increased risk of type 1 diabetes. Thus, the extensive heterogeneity in effects, ranging from autosomal dominant to bi-allelic and polygenic risk-variants within AIRE has created a need for functional characterization and clinical classification of AIRE variants.

AIRE is predominantly expressed in the mTECs where it facilitates the expression of TRAs normally restricted to alternative cell lineages representing most tissues and organ systems (12). It seems that AIRE is indirectly inducing TRA expression by ensuring transcriptional heterogeneity of mTECs (2, 13, 14). This is achieved by the repurposing of general transcriptional mechanisms and by alteration of the chromatin landscape, such as pause-release of stalled RNA polymerase II and enhancer-promoter looping (15–18). At the molecular level, AIRE

is able to carry out these functions through its functional protein domains, including a CARD multimerization domain, a SAND DNA interaction domain, and two PHD domains, where PHD1 is involved in histone binding while PHD2 is involved in interaction with other proteins (19). Recently, it was also discovered that AIRE is preferentially recruited to its target genes by promoters already poised for transcription (independently of AIRE, characterized by Z-DNA enrichment and double-stranded breaks in DNA (20)). Still, in spite of numerous ingenious studies based on mTECs from genetically modified mouse models that has contributed immensely to the knowledge of functional aspects of AIRE, its exact way of action is incompletely understood at several levels. In particular the definition of bona fide AIRE targets is problematic due to the low and promiscuous expression levels of TRAs in the mTECs.

Transcriptomics is particularly well-suited for the illumination of AIRE's molecular properties, given the vast number of induced or perturbed targets in the presence of AIRE in a given cell. However, in order to avoid the sheer complexity in the interpretation of transcriptomics data derived from a highly heterogeneous cellular population such as mouse mTECs, additional complementary cell systems are needed for AIRE-related studies. In fact, cell-line based models may in some ways be better suited for the exploration of the molecular mechanism of AIRE, as they are not limited by the number of AIRE expressing cells in a given tissue (such as mTECs) and can be easily genetically altered. Therefore, we have designed a simple and accessible cell system exploiting the homogeneous nature of the widely used and well characterized cell-line HEK293, and a plasmid conferring high levels of AIRE without leading to overloading of protein synthesis. We have chosen to study three different AIRE variants in addition to the WT, representing different proposed molecular mechanisms and different associated clinical phenotypes: the nonsense variant p.R257* located within the CARD domain (21), the dominant negative variant p.C311Y located within the PHD1 domain (22), and the polygenic risk variant p.R471C located within the PHD2 domain (23).

In the context of the functional characterization of AIRE, this offers a simple solution to the problems represented by the considerable heterogeneity of mTECs and their highly diverse and complex spectrum of TRA expression and transcriptional networks, as well as different maturation stages and subpopulations. We show that our AIRE transfected HEK293FT cell line system can be used to give

increased insight to the molecular function of AIRE and to differentiate between AIRE variants based on their transcriptomic properties.

2 Methods

2.1 Plasmid amplification and mutagenesis

pSF-UB plasmids were purchased from OXGENE with and without *AIRE* cDNA inserts. Empty pSF-UB plasmids were used as negative control, while pSF-UB plasmids containing *AIRE* were used for all *AIRE* WT transfections and mutagenesis to create the *AIRE* mutations p.R257*, p.C311Y and p.R471C. Mutagenesis was performed using QuikChange II Site Directed Mutagenesis kit from Agilent Technologies. Plasmids were amplified using TOP10 competent *E. coli* cells from Thermo Fisher and purified using a QIAprep Spin Miniprep Kit from Qiagen. The sequence of *AIRE* and mutated *AIRE* containing plasmids were confirmed by Sanger sequencing.

2.2 Cell culture and transfection

HEK293FT human embryonic kidney cells (RRID: CVCL_6911) were purchased from ThermoFisher Scientific (cat no R70007) and grown in a medium consisting of Dulbecco's Modified Eagle Medium supplemented with 4.5 g/l D-Glucose, Pyruvate, 10% Fetal Bovine Serum (FBS), and 1% Penicillin- Streptomycin. Mycoplasma testing were performed using the MycoStrip 100 Mycoplasma Detection kit from Invivogen. No mycoplasma was detected. The cells were incubated at 37°C in a 5% CO₂ humidified incubator until reaching 80-100% confluency. The cells were subsequently counted using a Scepter cell counter from Merck and transferred to 6-well plates, where each well was seeded with 6×10^5 cells. After 24h, the cells were transfected using the Lipofectamine 2000 Transfection Reagent from Thermo Fisher with 2.5 μ g DNA and 12 μ l Lipofectamine. All experiments reported in the following were performed at low passage numbers of the HEK293FT cell line.

2.3 Western immunoblotting

Transfected cells grown for 48 hours were removed from the growing surface by flushing, then washed using centrifugation and resuspension in PBS (Cat#: D8537, Merck). Cells were lysed using cComplete lysis buffer (Cat#: 04719956001, Merck) for 30 minutes on ice. LDS-PAGE was performed using NuPAGE MOPS SDS Buffer Kit (Cat#: NP0050) according to its instructions and run on a NuPage 10% Bis-Tris gel (Cat#: NP0301BOX, Thermo Fisher) with SeeBlue Plus2 Pre-Stained Protein Standard (Cat#: LC5925, Thermo Fisher) at 180V for 1 hour and 10 minutes. Samples were transferred to a PVDF membrane using the iBlot dry blotting system (Cat#: IB401002, Thermo Fisher). Washing steps were performed using Pierce TBS Tween 20 (Cat#: 28360, Thermo Fisher) while blocking was performed using SuperBlock Blocking Buffer (Cat#: 37517,

Thermo Fisher) for 1 hour on a shaker. Membranes were incubated with primary antibody in TBS-T with 5% BSA overnight on a shaker in a cold room. The primary antibodies used were mouse α -GAPDH (Cat#: MAB374, RRID: AB_2107445, Merck) in a 1:1000 dilution, goat α -AIRE (Cat#: MAB155, Merck) in a 1:500 dilution. The membranes were rewashed, then incubated with the secondary antibodies in TBS-T with 5% BSA for 1 hour on a plate shaker. The secondary antibody used was goat α -Mouse antibody (Cat#: 626520, RRID: AB_2533947, Thermo Fisher) conjugated with Horse Radish Peroxidase (HRP) in a 1:2000 dilution. After rewashing the membranes were soaked in Pierce ECL Western Blotting Substrate (Cat#: 32106, Thermo Fisher) and imaged using a Bio-Rad Chemidoc MP (RRID: SCR_019037).

2.4 RNA isolation

Transfected cells grown for 48h were harvested by flushing, then transferred to 15ml tubes where they were centrifuged at 300g for 7 minutes before being resuspended in cold PBS. RNA was isolated using Qiagen RNeasy Mini Kit. A Qiagen QIAshredder spin column was used to homogenise the samples, and a Qiagen RNase free DNase solution was used in order to digest any genomic DNA. The NanoDrop microvolume spectrophotometer (Thermo Fisher) was used to check the RNA quantitation. In order to ensure that high-quality RNA was used, the samples were also analyzed using an Agilent Bioanalyzer 2100 (RRID: SCR_018043) with an Agilent RNA 6000 Nano kit. RNA samples used in this study had a RNA Integrity Number (RIN) of between 9.2 and 9.9.

2.5 Real-time quantitative PCR

RNA from one sample of each condition of an independent experiment was analyzed using qPCR consisting of a two-step process using Thermo Fisher Superscript VI VILO cDNA Synthesis Kit (Cat#: 11754050) followed by Thermo Fisher TaqMan Universal PCR Master Mix (Cat#: 4304437) and a variety of TaqMan probes according to product info sheet and user guide. qPCR was performed using Applied Biosystems QuantStudio 5 (RRID: SCR_020240) and the Thermo Fisher Cloud software. Three technical controls for each condition were used, in addition to no template control (NTC) and no reverse transcriptase control (-RT). TaqMan probes consisted of probes for the genes *AIRE* (Cat#: Hs00230829_m1), *CCNH* (Cat#: Hs00236923_m1), *S100A8* (Cat#: Hs00374264_g1), *IGFL1* (Cat#: Hs01651089_g1), *KRT14* (Cat#: Hs00265033_m1), and *GAPDH* (Cat#: Hs99999905_m1). The qPCR results were processed using the $\Delta\Delta C_t$ method in R, and normalized against the *GAPDH* housekeeping gene, while fold change was calculated between *AIRE* transfected and empty plasmid transfected control. Data was prepared in R (24) (v.4.0.2, RRID: SCR_001905) and Rstudio (25) (v.1.4.1103, RRID: SCR_000432) using the tidyverse (26) (v. 1.3.1, RRID: SCR_019186) package, then plotted using ggplot2 (27) (v.3.3.5, RRID: SCR_014601) with the additional packages

RColorBrewer (28) (v. 1.1-2, RRID: SCR_016697) and scales (29) (v. 1.1.1, RRID: SCR_019295).

2.6 RNAseq

Illumina TruSeq Stranded mRNA Library Prep kit together with Illumina TruSeq RNA CD Index Plate index adapters was used for library preparation. Sequencing was performed with 12 biological replicates using an Illumina HiSeq 4000 sequencer (RRID: SCR_016386), with a read depth of approximately 100 million reads per sample.

2.7 RNA sequencing alignment

Quality control of fastq files were performed using FastQC (30) (v.0.11.9, RRID: SCR_014583) and MultiQC (31) (v.1.12, RRID: SCR_014982). Fastq files were subsequently aligned using the Kallisto (32) pseudoaligner (v.0.46.2, RRID: SCR_016582) to the GRCh38.p12 (release 100) reference genome provided by ensembl (33) (RRID: SCR_002344). All subsequent analysis were performed using R (v.4.0.2, RRID: SCR_001905) and Rstudio (25) (v.1.4.1103, RRID: SCR_000432), where tximport (34) (v.1.18, RRID: SCR_016752) was used to import Kallisto output files into R and aggregate transcript data to the gene level using the “hsapiens_gene_ensembl” gene annotation data accessed from biomaRt (35) (v.2.46.3, RRID: SCR_019214).

2.8 Differential expression analysis

Differential expression analysis was performed using DESeq2 (36) (v.1.30.1, RRID: SCR_015687) with the design formula “~mutation” with *AIRE* WT and variant genotypes separately vs empty control and *AIRE* variants separately vs *AIRE* WT samples. Significance testing was performed using a Wald test with an FDR threshold of 5%. Log2foldchange shrinkage was performed using apeglm (37) (v.1.12.0). Result tables were generated using tidy data principles with the tidyverse (26) package (v. 1.3.1, RRID: SCR_019186), and volcano, hexbin, violin and dotplots were generated using ggplot2 (27) (v.3.3.5, RRID: SCR_014601). FDR adjusted p-values of two-sided t-tests comparing *AIRE* WT and variants to control were performed using the packages ggpubr (38) (v. 0.6.0, RRID: SCR_021139) and rstatis (39) (v. 0.7.2, RRID: SCR_021240). Color scales used were generated using the packages RColorBrewer (28) (v. 1.1-2, RRID: SCR_016697) and Viridis (40) (v.0.6.2, RRID: SCR_016696). Plots were also generated with the help of the additional packages ggrepel (41) (v.0.9.1, RRID: SCR_017393), cowplot (42) (v.1.1.1, RRID: SCR_018081), ggtext (43) (v.0.1.1), and scales (29) (v. 1.1.1, RRID: SCR_019295). Venn diagrams were generated using the VennDiagram (44) (v.1.7.0, RRID: SCR_002414) package, but plotted in ggplot using draw_grob from cowplot. Hypergeometric distribution tests of

overlap in the venn diagrams were performed using the phyper function in R with the lower.tail=FALSE option.

2.9 Categorization of genes according to tissue specificity and restriction

As *AIRE* induced gene expression has been considered to predominantly consist of tissue restricted antigens (TRA), i.e. genes only expressed in a small subset of tissues, we classified genes as tissue enriched (45), group enriched (46), or tissue enhanced (47) according to their entries by the Human Protein Atlas (HPA, RRID: SCR_006710) (48, 49). Tissue enriched genes were defined as genes with a fourfold increase in mRNA expression in one tissue compared to any other tissue, group enriched genes as genes with a four-fold increase in mRNA expression in 2–5 tissues compared to all other tissues, and tissue enhanced genes as genes with a four-fold increase in mRNA in one tissue compared to the average level in all other tissues.

3 Results

3.1 Expression of functional *AIRE* in transiently transfected HEK293FT cells

To induce functional *AIRE* expression, HEK293FT cells were transfected with a pSF-UB plasmid containing *AIRE* cDNA. In addition to *AIRE* wildtype (WT), we also included the variants p.R257*, p.C311Y and p.R471C, to investigate whether transient transfection could be used to differentiate between WT and variants. The nonsense variant p.257* was selected as a representative of recessive variants with a large effect, p.C311Y was selected as a dominant variant with a moderate effect and p.R471C was selected as a common variant with small effect. Western blot analyses using a polyclonal anti-*AIRE* antibody targeting the N-terminal of human *AIRE* and protein lysates from cells transfected with pSF-UB plasmids encoding wildtype (WT) *AIRE* and *AIRE* variants revealed bands corresponding to the expected molecular weight for all *AIRE* variants (Figure 1a).

Quantitative PCR on RNA from cells transfected with each of the four variants showed that *AIRE* expression was increased by 10^3 – 10^5 fold for all *AIRE* variants compared to empty vector control (Figure 1b). For the canonical *AIRE* induced genes *S100A8*, *IGFL1* and *KRT14*, expression increased dramatically in WT, p.C311Y and p.R471C transfected cells, while no induction of gene expression could be seen in the cells with the nonsense mutation p.R257*. For *S100A8* and *KRT14*, but not *IGFL1*, induction was greatly reduced in p.C311Y transfected cells compared to WT. Notably, cells transfected with *AIRE* variant p.R471C exhibited consistently higher expression of *AIRE* induced genes than WT. An independent RNAseq experiment replicated the robust expression of the *AIRE* variants in transfected cells (Figure 1c). Furthermore, RNAseq confirmed expression of a range of verified and well-

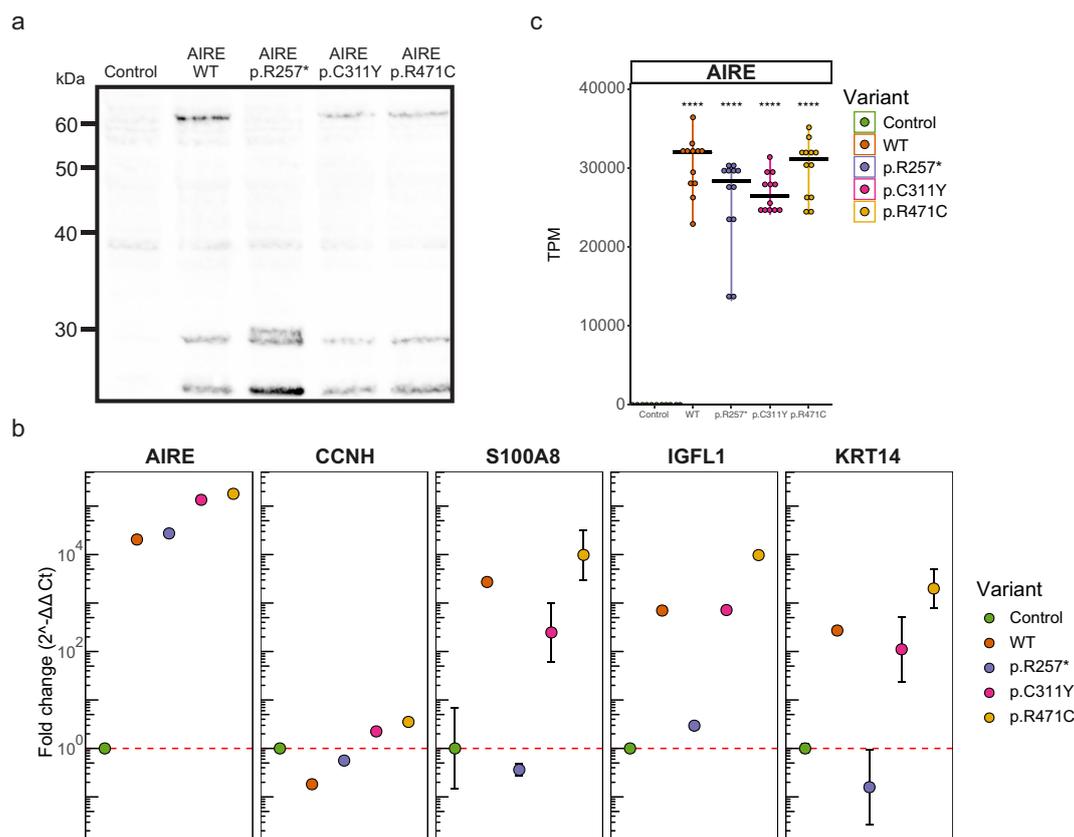


FIGURE 1
 Expression of functional AIRE. **(a)** Western blot of cellular lysates containing AIRE wildtype and variants p.257*, p.C311Y, and p.R471C from transfected HEK293FT cells. AIRE wildtype, p.C311Y and p.R471C exhibit full length bands at around 60 kDa, while the truncated p.R257* exhibits a band at approximately 30kDa. All samples indicate band for the control GAPDH at 36kDa. **(b)** qPCR of RNA from transfected HEK293FT cells. One sample for each condition with three technical replicates. *CCNH* used as negative (i.e. non-AIRE induced) control, *S100A8*, *IGFL1* and *KRT14* used as known (i.e. canonical) AIRE regulated genes. Results normalized against *GAPDH* in the $\Delta\Delta C_t$ method. Error bars indicate 95% confidence intervals based on C_t values of the 3 technical replicates for each condition. Experiment is representative of 3 independent experiments. **(c)** Expression of *AIRE* in transcripts per million (TPM) as determined by an independent RNAseq experiment of RNA from transfected HEK293FT cells with 12 samples for each condition. Significance level of FDR adjusted p-values from two-sided t-test's comparing against control indicated. **** indicates FDR < 0.0001.

characterized molecular partners in AIRE-mediated transcriptional regulation, further supporting the suitability of the cell-system for biochemical and molecular studies of AIRE (Supplementary Figure S1). For most of the molecular partners of AIRE, the overexpression of AIRE did not influence gene expression levels. The exception was tissue enhanced transcription factors such as *HNF1A*, *HNF1B* and *HNF4A*, that were all clearly induced by AIRE in our system.

3.2 AIRE variants induce a spectrum of variant-dependent transcriptional effects, reflecting the associated clinical phenotypes

After observing robust expression of functional AIRE in transfected HEK293FT cells, we wanted to identify the global expression profiles induced by different AIRE variants. RNA sequencing was performed on high quality RNA derived from

HEK293FT cells transfected with empty control pSF-UB plasmid, and pSF-UB plasmids containing cDNA for AIRE WT, p.R257*, p.C311Y, or p.R471C (Figure 1c). To enable the robust discovery of low abundance transcripts, typical for AIRE induced genes, a total of 12 replicates per variant were included at a read depth of approximately 100 million.

Principal component analysis (PCA) revealed a clear separation between the WT and variant transfected cells compared to empty control using PC1 and PC2 (Figure 2a). The overall expression pattern induced by the nonsense mutation p.R257* indicates that it clusters more closely to the empty control, the missense mutation p.C311Y displays intermediate clustering between p.R257* and WT, while WT and p.R471C induced similar expression patterns.

Differential gene expression analysis revealed a massive upregulation of genes induced by WT AIRE compared to empty vector. A total of 6340 genes were significantly higher expressed and 3690 lower expressed at a 5% FDR level (Figure 2b), with up to 15 times log2 fold change (Log2FC) increased expression levels for

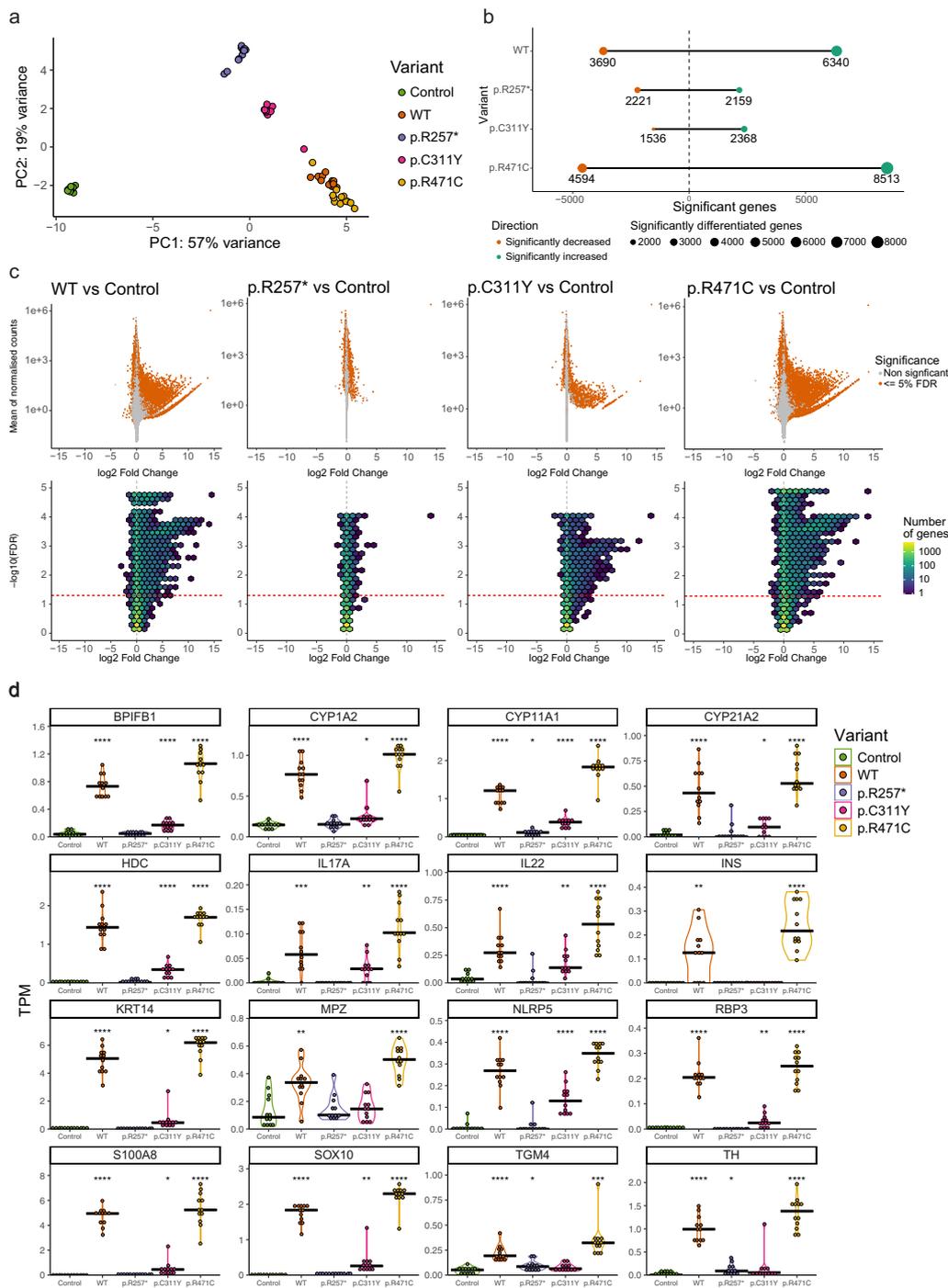


FIGURE 2

AIRE induced expression and transcriptional effects of AIRE variants revealed by RNA sequencing. (a) PCA plot of HEK293FT samples transfected with empty control plasmid, AIRE wildtype or AIRE variants p.R257*, p.C311Y, p.R471C (n = 12 for all conditions, AIRE levels seen in Figure 1c) (b) Number of significant differentially expressed genes of AIRE WT and AIRE variant transfected cells compared to control cells (FDR ≤ 0.05). (c) MA plot (upper row) and Hex bin Volcano plots (lower row) displaying differentially expressed genes induced by AIRE wildtype and variants compared to empty vector control. (d) Differential expression of a selected subset of AIRE-relevant genes (encoding known autoantigens in APS-1 and/or previously reported to be AIRE-regulated) by different AIRE variants, in transcripts per million (TPM). Significance level of FDR adjusted p-values from two-sided t-tests comparing against control indicated. * indicates FDR < 0.05, ** indicates FDR < 0.01, *** indicates FDR < 0.001, **** indicates FDR < 0.0001.

certain genes (Figure 2c). The nonsense mutation p.R257* however, exhibited a different pattern characterized by over-all low log2 fold change and symmetrical effect in directions of expressed genes (2159 vs 2221 respectively) compared to empty vector (Figures 2b,

c). The missense mutation p.C311Y induced significantly increased expression levels of 2368 genes accompanied by significantly reduced expression of 1536 genes, both categories with log2 fold changes similar to those induced by AIRE WT (Figures 2b, c). The

p.R471C variant exhibited a pattern distinct from the two pathogenic variants, with the induction of significantly higher expression of 8513 genes and significantly lower expression of 4594 genes compared to empty vector control, with log₂ fold changes similar to, or higher, than those induced by WT.

Figure 2d shows the absolute expression data in transcript per million (TPM) of a curated list of genes previously shown to be induced by AIRE in both human and murine systems and genes encoding known APS-1 autoantigens representing a wide range of tissues such as adrenal cortex, ovaries, lung, liver, pancreas, and parathyroid. While AIRE WT clearly induced or enhanced the expression of all selected genes, the p.R257* variant induced no or little changes in expression levels. The p.C311Y missense mutation displayed limited ability to induce or enhance gene expression, and was unable to induce any expression of some of the selected AIRE targets above the levels of empty vector control. p.R471C, comparatively, consistently induced or enhanced gene expression greater than AIRE WT.

3.3 The developed cell system recapitulates findings from earlier studies, but expands the range of AIRE induced target genes

To analyze the performance of our AIRE induced expression cell system we compared our data set to data from a published comparable cell system Guha et al. (50). In their study, HEK293 cells were transfected with AIRE WT and compared to untransfected cells. The main focus of the study by Guha and colleagues was the interaction of AIRE with topoisomerase2 (TOP2) in order to induce double-stranded breaks leading to enhanced chromatin access. Treating AIRE transfected cells with the TOP2 inhibitor etoposide was found to increase AIRE induced gene expression. To avoid systematic differences due to different methodology, we reanalysed their raw data using our own differential expression analysis pipeline and only used data from cells not treated with etoposide. This re-analysis of their data identified 507 significantly up-regulated genes and 5 down-regulated genes at an FDR of 5% (Figure 3a). These numbers are slightly lower but comparable to what was described in the original publication (n=691 induced genes with an FDR of 10%). A comparison with our study reveals that the vast majority (88%) of AIRE WT induced genes in Guha et al., 2017 were also identified in our cell system (WT vs control) (Figure 3b). This confirms that the transfection-based approach developed by us recapitulates findings of earlier studies yet offers an additional layer of high-resolution data.

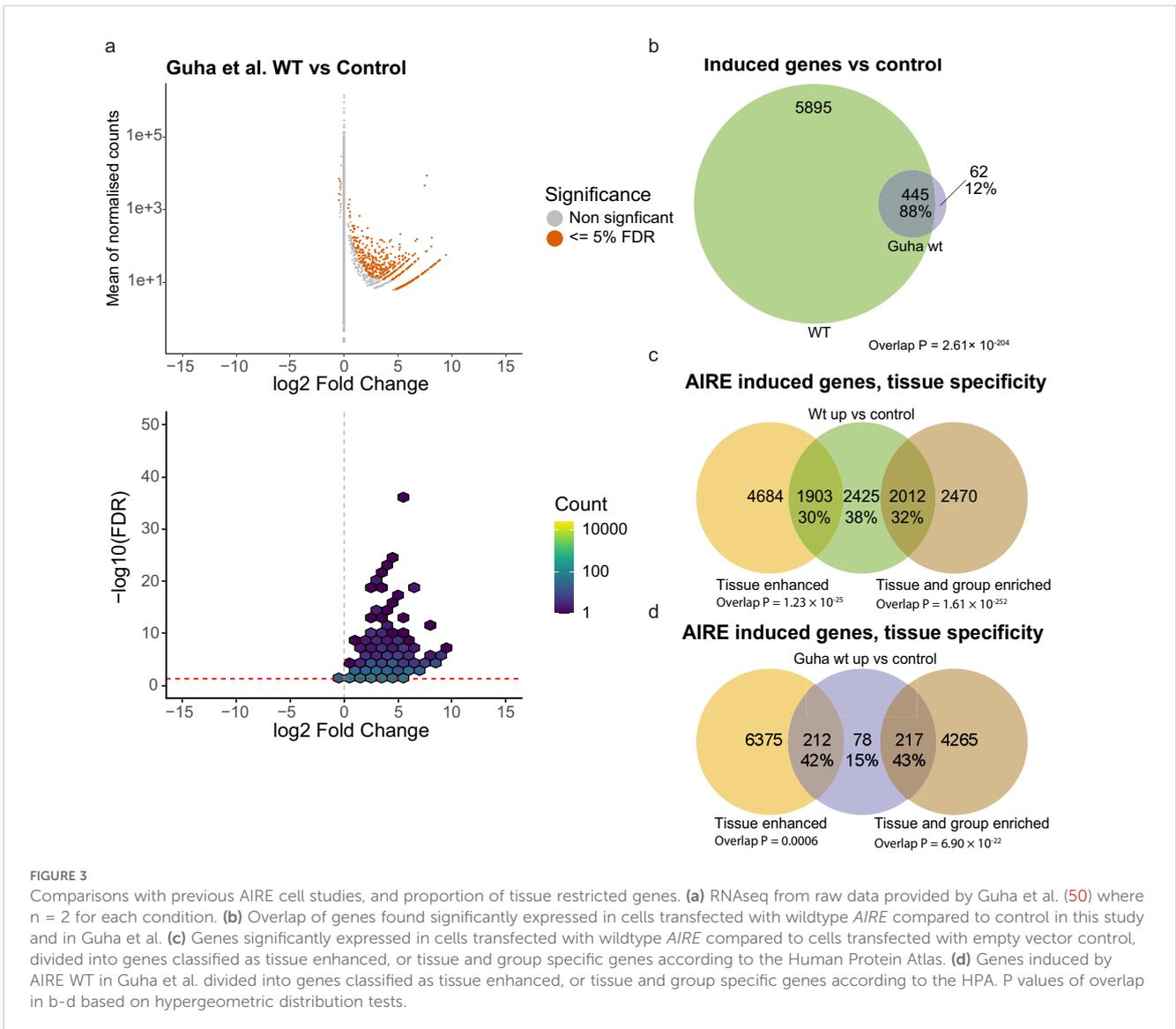
AIRE induced gene expression has been considered to be predominantly confined to Tissue Restricted Antigens (TRA), i.e. genes normally only expressed in a small subset of tissues. To explore this view, we compared the 6340 AIRE induced genes (defined as significantly higher expressed in AIRE WT transfected cells) to proteins classified as tissue enriched, group enriched, or

tissue enhanced by the Human Protein Atlas (HPA). Thirty percent of genes induced by AIRE WT were classified as tissue enhanced by HPA, while 32% were classified as either tissue enriched or group enriched, the remaining 38% were not classified as either (Figure 3c) compared to 27%, 8% and 65% respectively for all genes expressed at levels greater than 1 TPM in empty vector transfected cells (data not shown).

As a comparison, we analyzed the distribution of AIRE WT induced genes in the data set reported by Guha et al. (using our analysis pipeline). Here, a greater proportion of significantly upregulated genes were classified as tissue enhanced (42%) and tissue or group enriched (43%) while only 15% were outside these groups (Figure 3d). The lower proportion of non-tissue restricted genes in Guha et al. compared to the 38% found in our cell-assay might be explained by the higher resolution of our dataset, providing sufficient statistical power to reveal that AIRE not only is able to induce so-called ectopic expression of tissue-specific genes, but also to induce and/or enhance expression of genes not considered tissue-specific.

3.4 AIRE is capable of inducing low level transcription of non-tissue restricted genes

While the classical fold-change measure emphasizes the relative differences between induced genes, the absolute expression in transcripts per million (TPM) allows for a more direct comparison of the natural expression levels of genes induced by AIRE. The absolute expressions are depicted in Figure 4 and shows that AIRE induces low-level increase in gene expression for the majority of targeted genes, with most genes exhibiting an increase below 1 TPM and a median increase of 0.39 TPM (Figures 4a–c) from a median baseline TPM in control of 0.11 to a median TPM of 0.51 in WT. However, there are some exceptions where absolute gene expression is increased by as much as 57 TPM. Comparatively, genes significantly induced by p.C311Y had a median baseline of 0.12 TPM in control and was increased by a median 0.24 TPM to a median level of 0.36 TPM, while significant genes induced by p.R471C started at a median baseline of 0.16 TPM, increased by a median 0.46 TPM to a median level of 0.62 TPM (Figures 4a, c). To further investigate the differences between individual genes classified into the tissue-specific categories, the differential expression data of individual genes in AIRE WT transfected cells and empty vector control HEK293FT cells were further divided into the three groups defined above (Figure 4d). Among these groups, tissue and group enriched genes had the lowest baseline expression with a median TPM of 0.04 in empty vector control cells and a median increase of 0.29 TPM in WT transfected cells to a median level of 0.33 TPM, consistent with their classification as tissue restricted. Tissue enhanced genes had a baseline median expression of 0.14 TPM with a median AIRE induced increase of 0.40 TPM to a median level of 0.53 TPM, while the genes not classified as tissue enriched or enhanced had a baseline median expression level of 0.65 TPM and a median increase of 0.62 TPM to a median level of 1.27



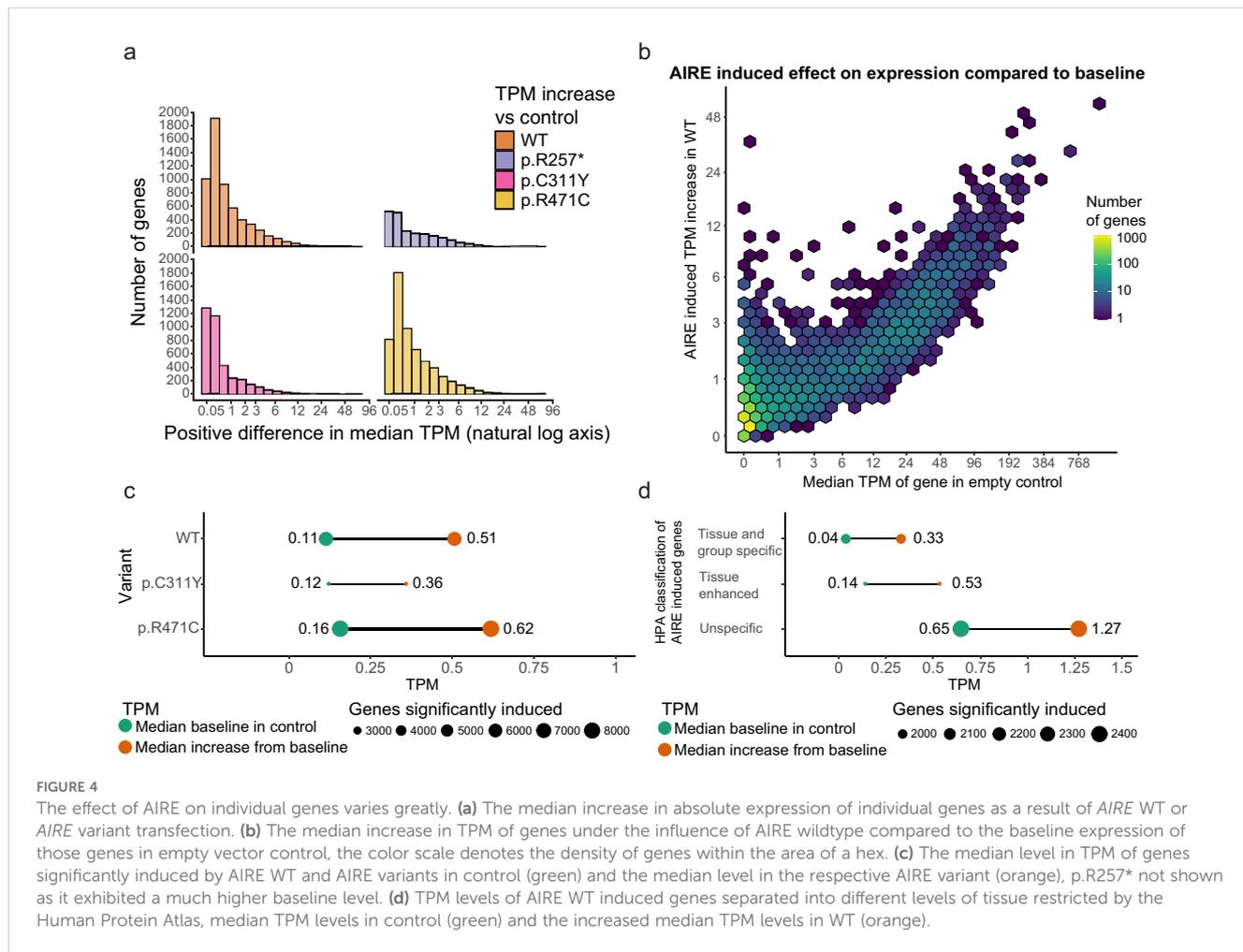
TPM. These results suggest that AIRE may have a magnifying effect not only restricted to the tissue restricted genes, but also on genes already expressed at low to moderate levels in HEK293FT cells. The low absolute induction may explain why earlier analyses are biased toward changes in tissue restricted genes where the fold-change is comparatively higher and thus power is higher.

4 Discussion

The molecular mechanisms behind AIRE-mediated induction of tissue restricted antigens in the thymus are still unclear, despite more than two decades of intense research (51). We have developed a simple, accessible and affordable cell-culture system based on transient transfection of an established human cell-line with plasmids encoding AIRE variants associated with different clinical phenotypes. By combining a large number of replicates and high-depth RNAseq, the developed cell assay allows for in depth analysis

of differential transcriptional effects of AIRE variants, providing detection of both dramatic and rather subtle differences induced by the different variants. Furthermore, we present a comprehensive analysis of the transcriptional effects of AIRE at a higher level of resolution compared to earlier studies. Our results indicate that AIRE is extremely potent as a transcriptional activator across an even broader range of target genes than previously estimated (2, 50). In addition to inducing gene expression of a multitude of tissue-restricted/tissue-enriched antigens, AIRE also seems to induce or enhance the expression of a wide range of genes that are not tissue specific. We believe this observation is in accordance with AIREs ability to act on previously poised non-coding regulatory elements and that it reflects AIREs multifaceted action on both chromatin accessibility and transcriptional initiation and elongation (52).

Numerous ingenious *in vivo*, *in vitro* and *in silico* experimental model systems have been described for molecular and immunological studies of AIRE (2, 12, 17, 20, 53, 54). Together, such models have contributed to better understanding of AIRE's



function during central tolerance and beyond (51). However, advanced model systems including those reliant on animal models might require access to considerable infrastructure not available to most standard-equipped labs. Therefore, we here propose another approach based on transient transfection of a widely used human cell-line with different variants of *AIRE*, in order to establish a manageable and reproducible assay accessible to most experimental laboratories. HEK293FT cells were chosen as cell-line platform because of their high susceptibility to transfection, their rapid growth (55, 56), as well as their ease of use in downstream methods for functional characterization such as flow-cytometry. In addition, HEK293 cells have been shown to express about half the human proteome in relatively high median expression values compared to other cell-lines (57), which may decrease viable *AIRE* targets and strengthen *AIRE* induced RNA expression signals. As we focus on the effect of *AIRE* induced expression as a result of *AIRE* variants in this paper, and we can show that the relevant *AIRE* partners are available in HEK293FT cells (Supplementary Figure S1) leading to *AIRE* induced expression (Figure 2), the biological relevance of the cell type is secondary to our needs. Historically, the human cell-line P1.4D6 (RRID: CVCL_A046) has been utilized as a biologically relevant thymus cell-line, unfortunately this endothelial cell-line has long

been misidentified and as such is not more biologically relevant than HEK293 (58, 59). Also, HEK293 cells have been utilized in a range of critical biochemical experiments fundamental to the characterization of several aspects of *AIRE* function, including *AIRE* dimerization, subnuclear localization, transactivation potential, histone binding and induction of TRAs (15, 60). For transfection of the HEK293FT cells we utilized expression plasmids containing an ubiquitin promoter of moderate strength. Preliminary data from our laboratory (data not shown) indicated superior results from this expression vector compared to more traditional CMV promoter containing vectors, and we were able to achieve robust heterologous expression of *AIRE* in transfected HEK293FT cells, presumably without overloading the protein synthesis machinery of the host cell. *AIRE* expression was evident on both protein and mRNA levels, and *AIRE* activity was confirmed by assessing gene expression of canonical *AIRE* target genes in *AIRE* transfected cells, using both qPCR and RNAseq analyses. In addition to being permissive to *AIRE* expression, we show that HEK293FT cells also express a wide range of genes known to be critical for *AIRE*-mediated transcriptional regulation (16, 51). These include components of gene repression, transcriptional elongation, mRNA splicing, chromatin structure and remodeling, and post-translational modification. A subset

of these genes involved in AIRE-mediated transcriptional regulation were also shown to be induced by AIRE itself in HEK293FT cells, such as members of the hepatocyte nuclear factor (HNF-) family of transcription factors.

Importantly, we verify induction of several clinically relevant autoantigens upon *AIRE* transfection. These included genes encoding highly tissue restricted proteins such as CYP21A2/21OH (adrenal cortex) and NLRP5/NALP5 (parathyroid and ovaries), representing the main autoantigens of Addison's disease and hypoparathyroidism, the predominant endocrinopathies of APS-1 (3, 61, 62). As previously reported (63–65) replicates are more important than read depth for power in RNAseq experiments, so the very high sequencing depth and multiple replicates increased our power to identify much smaller alterations than previous studies and thus reveal more of the AIRE induced global expression changes. This allowed us to demonstrate that many genes encoding clinically relevant autoantigens in the context of AIRE/APS-1 were differentially induced by the different AIRE variants, reflecting their associated clinical phenotypes. To our knowledge, this is the first time AIRE has been shown to induce such a comprehensive set of clinically relevant APS-1 autoantigens. Several other well-established AIRE targeted genes, such as *KRT14* and *S100A8*, previously used as surrogate markers of AIRE activity in both mTECs and transfected cells (16, 66, 67), were also shown to be induced in our system. Still, recognizing the limitations of a highly artificial system as described herein is important, as transfected HEK293FT cells may not be utilized to address several emerging and important roles of AIRE in immune system development and thymus biology, such as extrathymic AIRE expressing cells (e.g. so-called Janus cells) and the differentiation of mimetic cells within the medullary thymus epithelium (51). However, this is also beyond the scope of our current study, as our main goal is to establish a cellular system to investigate the effects of AIRE variants on the AIRE-induced transcriptome.

While p.R257* shows little to no activity in inducing known AIRE target genes, it still presents with a large number of significantly upregulated genes compared to the control when looking at the overall differential expression analysis. These genes have a much higher average expression level in the control compared to the significantly induced genes by the other AIRE variants. As such, it is possible that these changes are the result of an artificial overabundance of a truncated protein variant (Figure 1a) that consists of the CARD and partially of the SAND domain but lacks the PHD1 and PHD2 domains necessary for targeting by AIRE of inactive chromatin regions and partner interaction. It is probable that this variant would undergo nonsense-mediated decay *in vivo*, and as such any transcriptional effects here can be ignored. The p.C311Y variant is able to induce expression of many known AIRE targets, but to a much lower degree than AIRE WT, similar to the nearby variant p.D312A as previously described (68). This is also consistent with its clinical presentation with a later onset and possibly milder phenotype. Unexpectedly, the common variant p.R471C consistently induces more genes and induces those genes to a larger degree than AIRE WT. As this variant is not associated

with APS-1 but rather with the risk of developing common organ-specific autoimmunity (9–11) it is possible that this increased activity disrupts antigen presentation enough to increase the likelihood of escape by autoreactive T-cells in the thymus. We might speculate that AIRE induced antigen expression and antigen presentation in the thymus is a highly tuned system that may be disrupted either by an increased number of different antigens to present, or by increasing the number of the same antigens such that the whole repertoire of antigens is not presented.

Traditionally, the leading view has been that AIRE preferentially induces genes that are tissue specific and that are characterized by low levels of initial expression. ChipSeq studies and other molecular approaches have failed to identify any consensus binding motif in the target DNA. In order to achieve so-called ectopic gene expression, it has therefore been assumed that AIRE interacts with repressive histone marks or repressive chromatin complexes to pinpoint TRAs. Indeed, previous studies have indicated that the majority, but not all, AIRE induced genes have promoters with low levels of H4K3me3 and H3 acetylation (68), but high levels of H3K27me3 (2). More recently, increased understanding of how AIRE induces its target genes have added even more layers of complexity. In mTECs and transfected cell-lines, AIRE may be located at up to 40 000 sites across the genome at transcriptional start sites, superenhancers and introns (8). Here, AIRE appear to act rather late in the transcription cycle by reinforcing the actions of other transcription factors and by triggering transcriptional elongation by RNA polymerase (17). The steps prior to transcriptional elongation, such as chromatin remodulation and accessibility, seem to be AIRE independent (69). Hence, the AIRE target genes appear to be poised for transcription prior to AIRE induction. Recent evidence suggests that this is achieved in an ordered stochastic manner in which AIRE preferentially targets super enhancers enriched with acetylated H3K27 residues and genomic regions enriched with Z-DNA and/or double-stranded DNA breaks (20, 52, 70). We consistently find that AIRE induced expression is affecting a large number of genes with little to no initial expression by a large relative increase, but still with a low absolute increase in transcription. However, AIRE seems to also induce an increase in expression on a number of already expressed genes, and while 2/3's of the genes induced by AIRE WT is categorized as some form of tissue restriction by the Human Protein Atlas, 1/3 is not categorized as such. It is possible the overabundance of AIRE in our transfected system is responsible for these effects, however it may also be secondary expression caused by AIRE induced transcription factors, or that the increase in power of our study allows us to identify effects that may be lost in studies with a lower power focusing on the relative expression of AIRE induced genes.

In summary, we have developed a cell-culture system that allows for in depth analysis of differential transcriptional effects of AIRE variants associated with different clinical phenotypes. By combining a large number of replicates and high-depth RNAseq, we have the statistical power to detect both dramatic and subtle transcriptional differences induced by AIRE WT and how this

induction is disrupted in the activity of the pathogenic variants p.R257* and p.C311Y, as well as the increased activity of the common variant p.R471C linked to common organ specific autoimmunity. Consistent with previous studies, we find that AIRE induces thousands of genes from a level of little or no initial expression by a small increase in expression. However, we also find that AIRE is able to boost expression of some already expressed genes, and while the majority of AIRE expressed genes can be categorized as tissue specific, many genes are not. Finally, as a first, we confirm the AIRE regulated expression of a comprehensive number of clinically relevant autoantigens.

Data availability statement

The RNAseq data underlying this publication (fastq and bam files) has been submitted to the European Nucleotide Archive with accession number PRJEB83008. RNAseq differential expression results in tables are collected as an Excel file in Supplementary file S2 uploaded to Figshare with DOI: 10.6084/m9.figshare.28369352. Additional data is available at request.

Ethics statement

Ethical approval was not required for the studies on humans in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used.

Author contributions

AB: Data curation, Formal Analysis, Funding acquisition, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing, Methodology. BO: Resources, Writing – review & editing. AW: Resources, Writing – review & editing. EH: Conceptualization, Resources, Writing – review & editing. PK: Conceptualization, Methodology, Writing – review & editing. EB: Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. SJ: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported

by grants (to SJ) Helse Vest's Open Research Grant (grants #912250 and F-12144), the Novo Nordisk Foundation (grant NNF19OC0057445), Research Council of Norway (grant #315599), and (to AB) an L. Meltzer Foundation project grant. We would also like to thank Haukeland University Hospital, and the University of Bergen for general funding of this project. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

We would like to thank the Research and Development group at the Medical Genetics department of Haukeland University hospital and the Genomics Core Facility at the University of Bergen for help in library preparation and sequencing for RNAseq in this project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2025.1572789/full#supplementary-material>

References

- Derbinski J, Schulte A, Kyewski B, Klein L. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nat Immunol.* (2001) 2:1032–9. doi: 10.1038/ni723
- Sansom SN, Shikama-Dorn N, Zhanybekova S, Nusspaumer G, Macaulay IC, Deadman ME, 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.* (2014) 24:1918–31. doi: 10.1101/gr.171645.113
- Husebye ES, Perheentupa J, Rautemaa R, Kämpe O. Clinical manifestations and management of patients with autoimmune polyendocrine syndrome type I. *J Intern Med.* (2009) 265:514–29. doi: 10.1111/j.1365-2796.2009.02090.x
- Finnish-German APECED Consortium. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat Genet.* (1997) 17:399–403. doi: 10.1038/ng1297-399
- Nagamine K, Peterson P, Scott HS, Kudoh J, Minoshima S, Heino M, et al. Positional cloning of the APECED gene. *Nat Genet.* (1997) 17:393–8. doi: 10.1038/ng1297-393
- Su MA, Giang K, Zumer K, Jiang H, Oven I, Rinn JL, et al. Mechanisms of an autoimmune syndrome in mice caused by a dominant mutation in aire. *J Clin Invest.* (2008) 118:1712–26. doi: 10.1172/jci34523
- Oftedal BE, Hellesen A, Erichsen MM, Bratland E, Vardi A, Perheentupa J, et al. Dominant mutations in the autoimmune regulator AIRE are associated with common organ-specific autoimmune diseases. *Immunity.* (2015) 42:1185–96. doi: 10.1016/j.immuni.2015.04.021
- Goldfarb Y, Givony T, Kadouri N, Dobeš J, Peligero-Cruz C, Zalayat I, et al. Mechanistic dissection of dominant AIRE mutations in mouse models reveals AIRE autoregulation. *J Exp Med.* (2021) 218:e20201076. doi: 10.1084/jem.20201076
- Eriksson D, Royrvik EC, Aranda-Guillén M, Berger AH, Landegren N, Artaza H, et al. GWAS for autoimmune addison's disease identifies multiple risk loci and highlights AIRE in disease susceptibility. *Nat Commun.* (2021) 12:959. doi: 10.1038/s41467-021-21015-8
- Laisk T, Lepamets M, Koel M, Abner E, Estonian Biobank Research Team, Mägi R. Genome-wide association study identifies five risk loci for pernicious anemia. *Nat Commun.* (2021) 12:3761. doi: 10.1038/s41467-021-24051-6
- Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature.* (2023) 613:508–18. doi: 10.1038/s41586-022-05473-8
- Anderson MS, Venanzi ES, Klein L, Chen Z, Berzins SP, Turley SJ, et al. Projection of an immunological self shadow within the thymus by the aire protein. *Science.* (2002) 298:1395–401. doi: 10.1126/science.1075958
- Dhalla F, Baran-Gale J, Maio S, Chappell L, Holländer GA, Ponting CP. Biologically indeterminate yet ordered promiscuous gene expression in single medullary thymic epithelial cells. *EMBO J.* (2020) 39:e101828. doi: 10.15252/emboj.2019101828
- Nishijima H, Matsumoto M, Morimoto J, Hosomichi K, Akiyama N, Akiyama T, et al. Aire controls heterogeneity of medullary thymic epithelial cells for the expression of self-antigens. *J Immunol.* (2022) 208:303–20. doi: 10.4049/jimmunol.2100692
- Koh AS, Kuo AJ, Park SY, Cheung P, Abramson J, Bua D, 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.* (2008) 105:15878–83. doi: 10.1073/pnas.0808470105
- Abramson J, Giraud M, Benoist C, Mathis D. Aire's partners in the molecular control of immunological tolerance. *Cell.* (2010) 140:123–35. doi: 10.1016/j.cell.2009.12.030
- Giraud M, Yoshida H, Abramson J, Rahl PB, Young RA, Mathis D, et al. Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. *Proc Natl Acad Sci U S A.* (2012) 109:535–40. doi: 10.1073/pnas.1119351109
- Bansal K, Michelson DA, Ramirez RN, Viny AD, Levine RL, Benoist C, 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.* (2021) 118:e2110991118. doi: 10.1073/pnas.2110991118
- Chan AY, Anderson MS. Central tolerance to self revealed by the autoimmune regulator. *Ann N Y Acad Sci.* (2015) 1356:80–9. doi: 10.1111/nyas.2015.1356.issue-1
- Fang Y, Bansal K, Mostafavi S, Benoist C, Mathis D. AIRE relies on Z-DNA to flag gene targets for thymic T cell tolerization. *Nature.* (2024) 628:400–7. doi: 10.1038/s41586-024-07169-7
- Ferguson BJ, Alexander C, Rossi SW, Liiv I, Rebane A, Worth CL, et al. AIRE's CARD revealed, a new structure for central tolerance provokes transcriptional plasticity. *J Biol Chem.* (2008) 283:1723–31. doi: 10.1074/jbc.M707211200
- Chignola F, Gaetani M, Rebane A, Org T, Mollica L, Zucchelli C, et al. The solution structure of the first PHD finger of autoimmune regulator in complex with non-modified histone H3 tail reveals the antagonistic role of H3R2 methylation. *Nucleic Acids Res.* (2009) 37:2951–61. doi: 10.1093/nar/gkp166
- 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.* (2013) 110:1833–8. doi: 10.1073/pnas.1222023110
- R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing (2020).
- RStudio Team. *RStudio: Integrated Development Environment for R.* Boston, MA: RStudio, PBC (2021).
- Wickham H, Averick M, Bryan J, Chang W, McGowan L, Francois R, et al. Welcome to the tidyverse. *J Open Source Software.* (2019) 4:1686. doi: 10.21105/joss.01686
- Wickham H. *Ggplot2. Use R!*. Basel, Switzerland: Springer International Publishing (2016).
- Neuwirth E. *RColorBrewer: ColorBrewer Palettes.* (2014). pp. 1–2.
- Wickham H, Seidel D. *scales: Scale Functions for Visualization.* (2020).
- Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C, Wingett S. *FastQC: A Quality Control Tool for High Throughput Sequence Data.* Cambridge, United Kingdom: Babraham Institute (2010).
- Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* (2016) 32:3047–8. doi: 10.1093/bioinformatics/btw354
- Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic rna-seq quantification. *Nat Biotechnol.* (2016) 34:525–7. doi: 10.1038/nbt.3519
- Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, et al. Ensembl 2021. *Nucleic Acids Res.* (2020) 49:D884–91. doi: 10.1093/nar/gkaa942
- Soneson C, Love MI, Robinson M. Differential analyses for rna-seq: transcript-level estimates improve gene-level inferences. *F1000Research.* (2016) 4:1521. doi: 10.12688/f1000research.7563.2
- Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomaRt. *Nat Protoc.* (2009) 4:1184–91. doi: 10.1038/nprot.2009.97
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* (2014) 15:550. doi: 10.1186/s13059-014-0550-8
- Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics.* (2018) 35:2084–92. doi: 10.1093/bioinformatics/bty895
- Kassambara A. *ggpubr: 'ggplot2' Based Publication Ready Plots.* (2023).
- Kassambara A. *rstatix: Pipe-Friendly Framework for Basic Statistical Tests.* (2023).
- Garnier S, Ross N, Rudis B, Filipovic-Pierucci A, Galili T, Timelyportfolio, et al. *sjmgarnier/viridis: viridis 0.6.0 (pre-cran release).* (2021). doi: 10.5281/ZENODO.4679424.
- Slowikowski K. *ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'.* (2021).
- Wilke CO. *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'.* (2020).
- Wilke CO. *ggtext: Improved Text Rendering Support for 'ggplot2'.* (2020).
- Chen H. *VennDiagram: Generate High-Resolution Venn and Euler Plots.* (2021).
- HPA. *Search: tissue_category_rna:any;tissue_enriched and sort_by:tissue specific score - the human protein atlas.* Solna, Sweden: SciLifeLab (2023).
- HPA. *Search: tissue category_rna:any;group enriched and sort_by:tissue specific score - the human protein atlas.* (2023).
- HPA. *Search: tissue category_rna:any;tissue enhanced and sort_by:tissue specific score - the human protein atlas.* (2023).
- Uhle'n M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. tissue-based map of the human proteome. *Science.* (2015) 347:1260419. doi: 10.1126/science.1260419
- HPA. *The human proteome in tissue specific - the human protein atlas.* Solna, Sweden: SciLifeLab (2023).
- Guha M, Saare M, Maslovskaja J, Kisand K, Liiv I, Haljasorg U, et al. DNA breaks and chromatin structural changes enhance the transcription of autoimmune regulator target genes. *J Biol Chem.* (2017) 292:6542–54. doi: 10.1074/jbc.M116.764704
- Miller CN, Waterfield MR, Gardner JM, Anderson MS. Aire in autoimmunity. *Annu Rev Immunol.* (2024) 42:427–53. doi: 10.1146/annurev-immunol-090222-101050
- Huoh YS, Zhang Q, Törner R, Baca SC, Arthanari H, Hur S. Mechanism for controlled assembly of transcriptional condensates by aire. *Nat Immunol.* (2024) 25:1580–92. doi: 10.1038/s41590-024-01922-w
- Morimoto J, Matsumoto M, Oya T, Tsuneyama K, Matsumoto M. Cooperative but distinct role of medullary thymic epithelial cells and dendritic cells in the production of regulatory T cells in the thymus. *J Immunol.* (2023) 210:1653–66. doi: 10.4049/jimmunol.2200780

54. Akagbosu B, Tayyebi Z, Shibu G, Paucar Iza YA, Deep D, Parisotto YF, et al. Novel antigen-presenting cell imparts treg-dependent tolerance to gut microbiota. *Nature*. (2022) 610:752–60. doi: 10.1038/s41586-022-05309-5
55. Tan E, Chin CSH, Lim ZFS, Ng SK. HEK293 cell line as a platform to produce recombinant proteins and viral vectors. *Front Bioeng Biotechnol*. (2021) 9:796991. doi: 10.3389/fbioe.2021.796991
56. Yuan JW, Xu W, Jiang S, Yu H, Fai Poon H. The scattered twelve tribes of HEK293. *Biomed Pharmacol J*. (2018) 11:621–3. doi: 10.13005/bpj
57. Geiger T, Wehner A, Schaab C, Cox J, Mann M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol Cell Proteomics*. (2012) 11:M111.014050. doi: 10.1074/mcp.M111.014050
58. MacLeod RA, Dirks WG, Matsuo Y, Kaufmann M, Milch H, Drexler HG. Widespread intraspecies cross-contamination of human tumor cell lines arising at source. *Int J Cancer*. (1999) 83:555–63. doi: 10.1002/(SICI)1097-0215(19991112)83:4<555::AID-IJC19>3.0.CO;2-2
59. Capes-Davis A, Theodosopoulos G, Atkin I, Drexler HG, Kohara A, MacLeod RAF, et al. Check your cultures! a list of cross-contaminated or misidentified cell lines. *Int J Cancer*. (2010) 127:1–8. doi: 10.1002/ijc.25242
60. Villaseñor J, Benoist C, Mathis D. AIRE and APECED: molecular insights into an autoimmune disease. *Immunol Rev*. (2005) 204:156–64. doi: 10.1111/j.0105-2896.2005.00246.x
61. Alimohammadi M, Björklund P, Hallgren A, Pöntynen N, Szinnai G, Shikama N, et al. Autoimmune polyendocrine syndrome type 1 and NALP5, a parathyroid autoantigen. *N Engl J Med*. (2008) 358:1018–28. doi: 10.1056/nejmoa0706487
62. Winqvist O, Gustafsson J, Rorsman F, Karlsson FA, Kämpe O. Two different cytochrome P450 enzymes are the adrenal antigens in autoimmune polyendocrine syndrome type I and Addison's disease. *J Clin Invest*. (1993) 92:2377–85. doi: 10.1172/jci116843
63. Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, et al. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*. (2016) 22:839–51. doi: 10.1261/rna.053959.115
64. Liu Y, Zhou J, White KP. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics*. (2014) 30:301–4. doi: 10.1093/bioinformatics/btt688
65. Lamarre S, Frasse P, Zouine M, Labourdette D, Sainderichin E, Hu G, et al. Optimization of an RNA-seq differential gene expression analysis depending on biological replicate number and library size. *Front Plant Sci*. (2018) 9:108. doi: 10.3389/fpls.2018.00108
66. Oftedal BE, Assing K, Baris S, Safgren SL, Johansen IS, Jakobsen MA, et al. Dominant-negative heterozygous mutations in AIRE confer diverse autoimmune phenotypes. *iScience*. (2023) 26:106818. doi: 10.1016/j.isci.2023.106818
67. Lammers S, Barrera V, Brennecke P, Miller C, Yoon J, Balolong J, et al. Ehf and fezf2 regulate late medullary thymic epithelial cell and thymic tuft cell development. *Front Immunol*. (2023) 14:1277365. doi: 10.3389/fimmu.2023.1277365
68. Org T, Rebane A, Kisand K, Laan M, Haljasorg U, Andreson R, et al. AIRE activated tissue specific genes have histone modifications associated with inactive chromatin. *Hum Mol Genet*. (2009) 18:4699–710. doi: 10.1093/hmg/ddp433
69. Meredith M, Zemmour D, Mathis D, Benoist C. Aire controls gene expression in the thymic epithelium with ordered stochasticity. *Nat Immunol*. (2015) 16:942–9. doi: 10.1038/ni.3247
70. Bansal K, Yoshida H, Benoist C, Mathis D. The transcriptional regulator aire binds to and activates super-enhancers. *Nat Immunol*. (2017) 18:263–73. doi: 10.1038/ni.3675