WEIZMANN INSTITUTE OF SCIENCE

Thesis for the degree	עבודת גמר (תזה) לתואר
Doctor of Philosophy	דוקטור לפילוסופיה
Submitted to the Scientific Council of the	מוגשת למועצה המדעית של
Weizmann Institute of Science	מכון ויצמן למדע
Rehovot, Israel	רחובות, ישראל
By	מאת
Vladislav Krupalnik	ולדיסלב קרופלניק

Aire is a novel factor in promoting induction of pluripotency

החלבון אייר כפקטור עוברי המשרה יצירה של מצב פלוריפוטנטי

Advisor: Dr. Jacob Hanna	מנחה: דר' יעקוב חנא
September, 2016	אלול,תשע"ו

Table of Contents

Abbreviations

2i	PD0325901 and CHIR99021
4OHT	4-Hydroxytamoxifen
Aire	Autoimmune Regulator
AMKO	Aire, c-Myc, Klf4 and Oct4
AP	Alkaline Phosphatase
APECED	Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy
DNA-PK	DNA-dependent protein kinase
Dox	Doxycycline
DSB	Double Strand Break
ESC	Embryonic Stem Cell
FBS	Fetal Bovine Serum
Hpf	Hours post fertilization
ICM	Inner Cell Mass
iPSC	Induced Pluripotent Stem Cell
KO	Knockout
KSR	Knock-out Serum Replacement
LIF	Leukemia Inhibitory Factor
MO	Morpholino
OSKM	Oct4, Sox2, Klf4 and c-Myc
Pol II	RNA polymerase II
PTA	Peripheral Tissue Antigens
RPKM	Reads Per Kilobase per Million reads
SCNT	Somatic Cell Nuclear Transfer
SD	Standard deviation
TOP2A	Topoisomerase (DNA) II Alpha
TSA	Tissue Specific Antigens
WT	Wild Type
ZF	Zebrafish
ZGA	Zygotic Genome Activation

Abstract

Autoimmune regulator (Aire) plays an important role in educating thymic cells of the immune system to prevent autoimmunity. However, Aire is also expressed during early embryonic development, much earlier than the generation of thymic progenitors. The fertilized oocyte undergoes a unique process termed zygote genome activation (ZGA), during which most of the genes are randomly transcribed at different time points of early development, until the blastocyst stage. Because Aire contributes to autoimmune tolerance by driving random expression of peripheral tissue antigens, I hypothesized that Aire takes part in the mechanism of global transcription and ZGA and, thereby, promotes establishment of pluripotency. *In vivo* analyses of both mice and zebrafish early development showed a decrease in maturation of 15% and 40%, respectively. Induction of Aire during somatic cell reprogramming significantly increased the formation of induced pluripotent stem cells (iPSCs). Furthermore, Aire successfully substituted exogenous expression of the Yamanaka factor Sox2 to jointly induce reprogramming. Our findings shed light on a previously unappreciated role for Aire in early development and somatic cell reprogramming.

חלבון בקרת החיסון העצמי (autoimmune regulator, Aire) אחראי בלעדית לחינוך התאים של מערכת החיסון בבלוטת התימוס למניעת תופעות של חוסר חיסון עצמי. אולם, מחקרים הראו שחלבון זה מתבטא גם בשלב ההתפתחות העוברית המוקדמת, לפני היווצרות של תאי אב של התימוס. בשלב שבו העובר מורכב מ-2 תאים בלבד, מתרחש בו תהליך ייחודי שנקרא שפעול הגנום הביצי, שבמהלכו מתבטאים בצורה אקראית רוב הגנים בגנום העוברי. כיון שחינוך תאי החיסון ע"י חלבון בקרת החיסון העצמי מתבצע באמצעות מנגנון דומה של ביטוי גנים אקראי, ושחלבון זה מתבטא גם בשלבי ההתפתחות המוקדמת, הנחנו שהוא עשוי לקחת חלק בתהליך ההפעלה של הגנום הביצי. חקירה של התפתחות מוקדמת של עכברים ודגי זברה נמצא שבהיעדר Aire, ההתפתחות העוברית נפגעת ב- 15% וב- 40% מהעוברים, בהתאמה. ביטוי יתר של החלבון במהלך יצירת תאי גזע מושרים מתאי עור הכפיל יצירת מושבות של תאי גזע מושרים. יתר על כן, Aire החליף בהצלחה את Sox2, אחד מארבעת הגורמים של יאמאנקה (Yamanaka factors), בהשריה של תכנות מחדש של תאים. לסיכום, אנו מציגים נקודת מבט שונה על התפקידים של חלבון בקרת החיסון העצמי במהלך ההתפתחות העוברית המוקדמת ובתכנות מחדש של תאים סומטיים.

Introduction

Background

The promise of stem cells and pluripotency

One of the greatest challenges facing society is treating patients afflicted with degenerative and age-related diseases. Clinical trials have demonstrated that transplanting functional cells into patients can relieve these debilitating ailments[1], but the lack of transplantable cells supply and genetic mismatch to the recipient[2] have hindered the effective application of this approach. Embryonic stem cells (ESCs) have been proposed as an excellent source for differentiated cells, due to their ability to self-renew indefinitely, while retaining the potential to generate every cell type in the Petri dish[3]. However, the use of embryo-derived stem cells does not solve tissue-matching problems. Therefore, it has become clear that a major goal of stem cells research is to create "customized" autologous cells that can be transplanted to replace damaged tissues and restore health[4]. An important breakthrough in overcoming these issues has been the *in vitro* reprogramming of somatic cells to pluripotent cells termed induced pluripotent stem cells (iPSCs). iPSCs are generated from somatic cells by ectopic expression of the four transcription factors Oct4, Sox2, Klf4 and c-Myc (OSKM), jointly known as the Yamanaka factors[5]. The use of iPSCs paves the way for the application of custom-made, genetically identical cells in transplantation therapy[6]. Despite major efforts and advances, our knowledge of the mechanisms underlying cell reprogramming to pluripotency remains limited. iPSCs have a very low yield of formation and require multiple and a non-fixed number of cell divisions[7], which can also result in genetic abnormalities[8]. On the other hand, the oocyte nucleus can reprogram a somatic epigenome within two cell divisions[7]. Another approach is somatic cell nuclear transfer (SCNT), where the nucleus of a somatic cell is injected into an oocyte without nucleolus to induce the transformation to pluripotency. **This raises the possibility that the oocyte expresses genes that could increase the kinetics and/or efficiency of the *in vitro* reprogramming process[9].** Furthermore, an *in vitro* reprogramming experimental system may be used for molecular analysis of *in vivo* reprogramming processes, which to date are difficult to be characterized biochemically.

Early embryonic development

Embryonic development begins with the formation of a zygote following oocyte fertilization by sperm cells. The zygote subsequently develops into a multicellular structure until it reaches a dense, ball-like formation of aggregated cells named morula. By the time the oocyte reaches the morula stage, it maintains its ability to generate the embryonic and extra-embryonic tissues, a developmental ability called totipotency. At day 3.5 after fertilization (E3.5), the embryo is arranged as a blastocyst, which includes peripheral layers committed to become extra-embryonic tissues and inner cell mass (ICM). The latter is the source of pluripotent cells, which later form the embryo proper. ICM cells maintain their pluripotency for a limited period of time; for example, E3.5-E4.5 during mouse development[7,10]. In the adult animal, stem cells lack the ability to generate all type of tissues and generate only specific lineages; hence, they are termed 'multipotent' or 'unipotent'.

Embryonic genome activation

To reach maturation by the blastocyst stage, the embryonic genome must be rapidly and robustly transcribed. The oocyte contains maternal mRNA, also known as maternal inheritance, which is responsible for initiation and regulation of the first developmental steps. The maternal mRNA is completely degraded when the embryo reaches the eight-cell stage, by which the embryo fully supports its own transcription. The transition from maternal to embryonic mRNA transcription is termed zygotic genome activation (ZGA)[11,12]. ZGA starts at the two-cell stage and involves global transcription by RNA polymerase II (Pol II), which occupies transcription start sites and transcribes genes. Pol II operates as part of a large multi-unit complex[13]; yet, the complete mechanism including the regulation of Pol II activity is still to be uncovered. Inhibition of Pol II by α-amanitin results in development arrest at the two-cell stage due to failure in ZGA[12]. Most of the time, Pol II is bound to the transcription start sites but does not initiate transcription, due to the epigenetic state of the locus. Only when Pol II is activated, the initiation and elongation of transcription starts. This regulatory process is called Pol II pause release and many regulatory members of this complex are expressed during early development.[14]

Until recently, it was thought that there is only one wave of transcription, which occurs at the two-cell stage. However, single-cell RNA-seq studies revealed new insights into the complexity of genome activation[15,16]. These studies suggest that a number of waves happen during early development, namely a major wave during the two-cell stage, and a minor one during sixteen-cell/morula stages. Global genome activation is one of the most conserved processes across species, and deciphering the underlying mechanism is crucial for understanding the molecular path leading to the establishment of totipotency in the morula and pluripotency in the ICM.

The autoimmune regulator gene AIRE

T cell lymphocytes recognize antigens via surface receptors that are generated randomly, and can present both foreign and self-antigens[17]. In order to prevent the T cell-based immune system from recognizing the body cells as foreign, a state known as autoimmunity, maturing T cells are scanned by various antigen-presenting cells in the thymus, known as medullary thymic epithelial cells (mTECs), in a process termed negative selection[17]. T cells that recognize self-antigens are destroyed and only cells that recognize foreign antigens contribute to the immune system. The gene responsible for the ectopic expression of specific peripheral tissue antigens (PTAs) by mTECs is autoimmune regulator (AIRE). AIRE was first identified in a rare multi-organ autoimmune disease called autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), which is caused by null mutations in human *AIRE*[17]. Mice lacking *Aire* develop a similar syndrome and a decrease in promiscuous expression of PTAs; however, the severe phonotype of the disease is noticed only in specific mice strains[17]. In addition, Aire knockout mice can be hypofertile or infertile, presumably due to autoimmune response against the gonads[17,18].

Although Aire's impact on negative selection is well characterized, the molecular basis of its function is still poorly understood. Aire protein is highly conserved and its four main domains, namely HSR, SAND, PHD1 and PHD2, play a critical role in its unique global genome activation function[19]. Yet, the exact role of the domains is still unknown. For instance, the SAND domain is thought to function as a DNA-binding domain, but it lacks residues that are critical for DNA-binding[19]. The HSR/CARD

domain has been suggested to interact with other transcriptional control proteins, such as CBP[19]. PHD1 domain binds to histone 3 with an unmethylated lysine in position 4[20], an interaction that may be the first step in Aire regulation of genes expression[19]. PHD2 domain is similar to a RING finger domain, but its function is still under debate [21].

Functionally, Aire induces a variety of protein-encoding genes that are normally restricted in expression to other organs[17,22]. Aire-expressing medullary thymic epithelial cells (mTECs) can transcribe more than 25% of the body gene repertoire. In addition, Aire-driven expression by mTECs is independent of the transcriptional regulation in their native cell type[17]. Indeed, due to the large number of expressed PTAs, Aire was postulated to affect the whole genome through a mechanism of global regulation[22]. The transcriptional footprint of Aire varies among the cell types where it is induced[23]. Even in the same cell type, the set of genes that are induced by Aire is not consistent[24]. A screen for Aire partners in several cell lines revealed that Aire activity is related to transcriptional elongation machinery, which forms a complex with DNA-PK and TOP2, and to mRNA processing[25] (Fig.1). Recently, it was reported that Aire activates expression by binding to PolII-rich transcriptional start sites and releasing stalled polymerases[26]. Moreover, Aire was shown to bind multiple proteins involved in different aspects of transcription initiation, such as CDK9, BRD9 and MBD1[27].
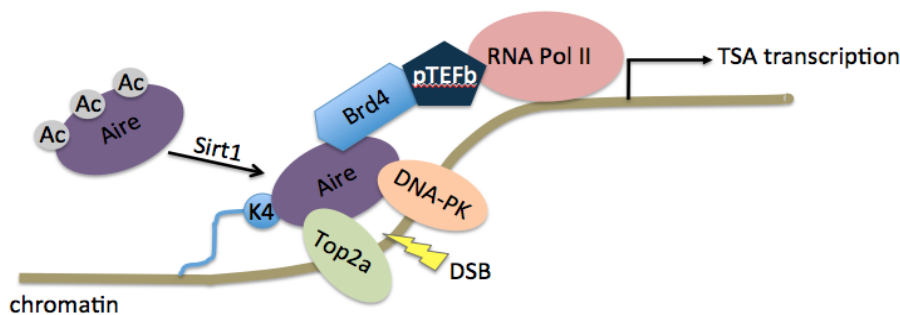
Figure 1: AIRE and its binding partners. Schematic illustration of a partial set of autoimmune regulator (AIRE)-interacting partners. Dozens of AIRE-interacting partners of diverse functions have been identified. AIRE interacts with histone core proteins, either directly or through its interactions with DNA-dependent protein kinase (DNA-PK), and with positive transcription elongation factor b (P-TEFb) and RNA polymerase II (RNA Pol II) to release stalled polymerases. AIRE also interacts with sirtuin 1 (SIRT1), which controls its acetylation.

Recently, Gu at el.[28] and Nishikawa at el.[29] showed that Aire is expressed during early development and in ESCs in both mouse and human. However, the functional implication of the conserved AIRE expression in mammalian development *in vitro* and *in vivo* remains unknown.

Thesis rationale

Aire has the ability to induce high amount of transcripts, representing the entire gene repertoire of the organism, to maintain self-tolerance. The translated transcripts, termed PTAs, are presented to T cells in the thymus, and those that recognize the peptide on the antigen-presenting cell are subsequently eliminated. Yet, Aire expression is not restricted to the thymus. It is also expressed during early development, from fertilization to the establishment of pluripotency in the blastocyst stage. The immediate question is why this protein should be expressed at those stages, long before the appearance of early thymic progenitors. During early development, a unique process termed ZGA occurs, whereby the entire embryonic genome is transcribed in order to prepare the fundamentals for the establishment of pluripotency and proper development. Considering both the unique mechanism of action of Aire and its temporal expression patterns, we hypothesize that Aire plays a part in ZGA and the activation of massive transcription, leading to pluripotency. Here, we present evidence for a novel role for Aire in promoting early development, transcription activation and robust establishment of pluripotency.

Transcription pattern of Aire during early embryonic development

To identify genes that are expressed nearly exclusively in early-stage embryos, we performed a bioinformatics-based targeted screen. We used previously published gene expression datasets of pre-implanted murine embryos at different developmental stages. Seven candidates were found to meet the criterion, including Calcoco2, ZFP264, CBX2, Trim43a, GM16367, Pramel6 and Aire (data not shown). Because Aire expression was validated in mouse and human pre-implantation development but without functional characterization, it was chosen for in-depth analysis and characterization of its role in early development and pluripotency reprogramming.

In order to validate and more accurately determine the expression pattern of Aire during early murine embryonic development, we measured the transcription level and protein expression of Aire at the following stages: sperm, unfertilized oocyte, one-cell embryo (zygote), two-, four- and eight-cell embryo, morula and blastocyst. In addition, we measured Aire expression in embryonic fibroblasts and in thymus as positive and negative controls, respectively. RNA samples from all stages were extracted (about 50 embryos at each stage) and expression profiling was performed by quantitative PCR using Taq-man probes, which are specific and more reliable with small quantities of RNA. Gapdh served as a control for cDNA, and Nanog as an internal control for proper early embryonic development. We found two peaks of Aire expression at the one-cell and morula stages (Fig.2A). To validate the results, we used as an additional control Hprt, which is more suitable to early development samples[30] (Fig. 2B). The results suggested that Aire is expressed mainly at the one-cell stage, which was further validated by RNA-seq (Fig. 2C). These results show that Aire is expressed at developmental stages at which genome activation is underway.
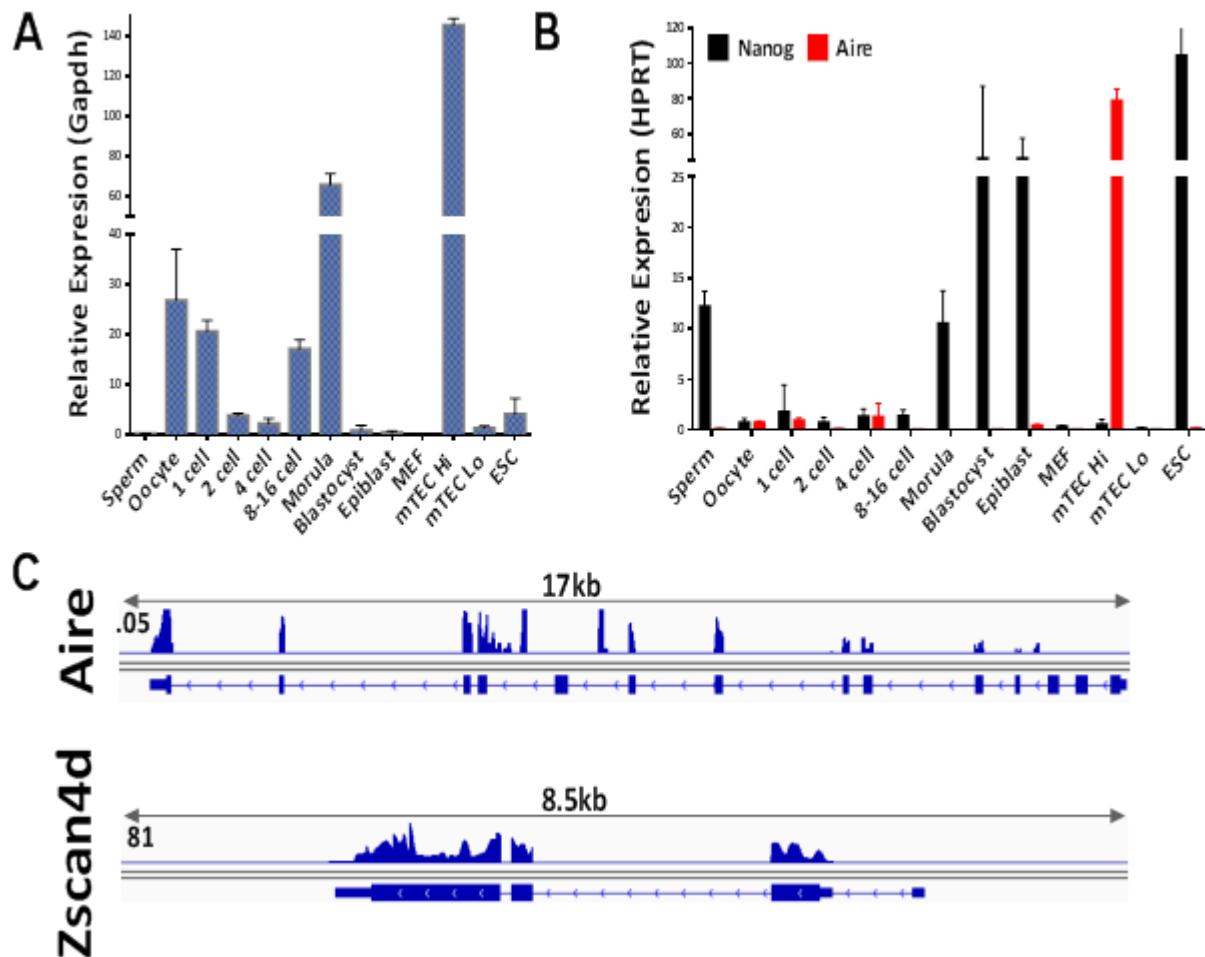
Figure 2. Aire mRNA expression during early embryonic development: A. Quantitative PCR of Aire during early development, from oocyte to epiblast, with sorted thymus populations (mTEC hi and low) included as positive controls. All values were normalized to Gapdh. Error bars represent SD of 3 biological replicates. **B.** Quantitative PCR of Aire during early development compared to Nanog expression and normalized to HPRT. Error bars represent SD of 3 biological replicates. **C.** RNA-seq analysis of Aire and Zscan4d expression at 2-cell stage.

Next, we sought to examine Aire protein expression *in vivo*. Due to the lack of a reliable antibody for Aire, we used a fluorescent reporter (GFP) whose expression is controlled by endogenous promoter of Aire[31]. Blastocysts were immunostained with anti-GFP antibody and with anti-Oct4 antibody as an internal control (Fig. 3). The results confirmed Aire expression in the ICM, as was recently described[28,29]. In order to detect endogenous Aire, we tried 7 antibodies for this protein. Unfortunately, none of them allowed effective detection of the endogenous protein.

| Aire (GFP) | Oct4 | Dapi |
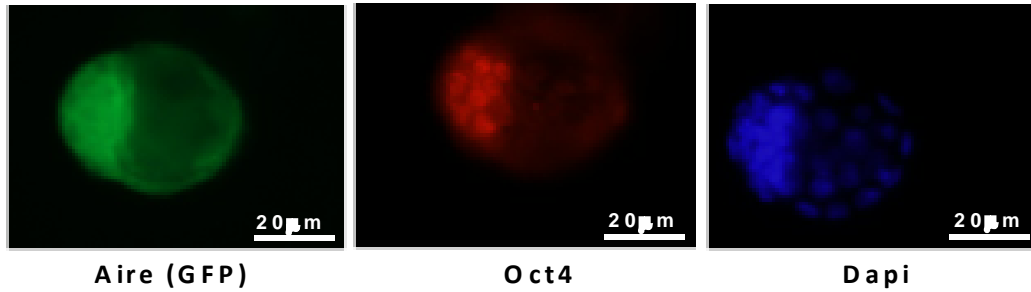
Figure 3. Aire protein expression during early embryonic development: A. Immunostaining for Aire protein in blastocyst derived from Aire-GFP reporter mice (left) counterstained with DAPI (right). Oct4 expression (middle) served as an internal control for the same developmental stage.

Hypofertility in Aire knockout mice

Because Aire has been implicated in global transcriptional activation by Pol II pause release[26], and since Aire knockout mice are hypo- or infertile[17,18], we next examined whether Aire influences genome activation following fertilization[32]. To investigate potential functional roles of Aire in early embryonic development, we used knockout animal models. Aire-deficient mice are viable but predominantly infertile[17,18]. However, due to the likely possibility that Aire can be inherited from maternal oocyte, Aire function during early development has remained unresolved. Considering that B6 Aire-deficient mice exhibit no or very mild autoimmune phenotype[33], hypofertility could be the result of mechanistic property related to the absence of Aire during waves of genome activation.

To test this hypothesis, we examined the fertility of Aire knockout mice. Mating of Aire-deficient mice produced viable offspring at normal Mendelian ratios (Fig. 4A); however, a decrease of about 30% in fertility of female mice was observed (Fig. 4B). Only 13 of 21 females gave birth at least once during 6 months of mating with an Aire-deficient male. However, once the female became fertile and the embryos passed the early development stage, the number of litters and the number of pups per litter were the same as in the control, and the oocytes derived from KO females were normal (Fig. 4C-E). This phenomenon could be attributed to autoimmune-related defects[18]; however, Aire-deficient B6 mice are known to display relatively weak autoimmune phenotype. Furthermore, western blotting for auto-antigens in the serum of KO and WT mice

revealed no difference in the expression of those auto-antibodies. Interestingly, WT samples showed similar band distribution as the KO, indicating mild autoimmunity, and the expression of known self-antigens could already been detected (Fig. 4F-H).

A B6.BCA Aire Het × B6.BCA Aire Het

Observed   Expected

N = 187

B

p = 0.03133

% Fertility

♀	KO	HT	WT	WT
	58.1	89.5	80	100
♂	KO			WT

C # of birth per female — p = 0.05118 (NS)

♀ KO HT WT WT
♂ KO   WT

D # of pups per litters — p = 0.5491 (NS)

♀ KO HT WT WT
♂ KO   WT

E 150 µm

F ♀ NOD KO | KO | HET | WT — 95 kDa, 34 kDa — **Ovary Lysate**

G ♂ NOD KO | KO WT KO WT KO KO WT WT — **Testis Lysate**

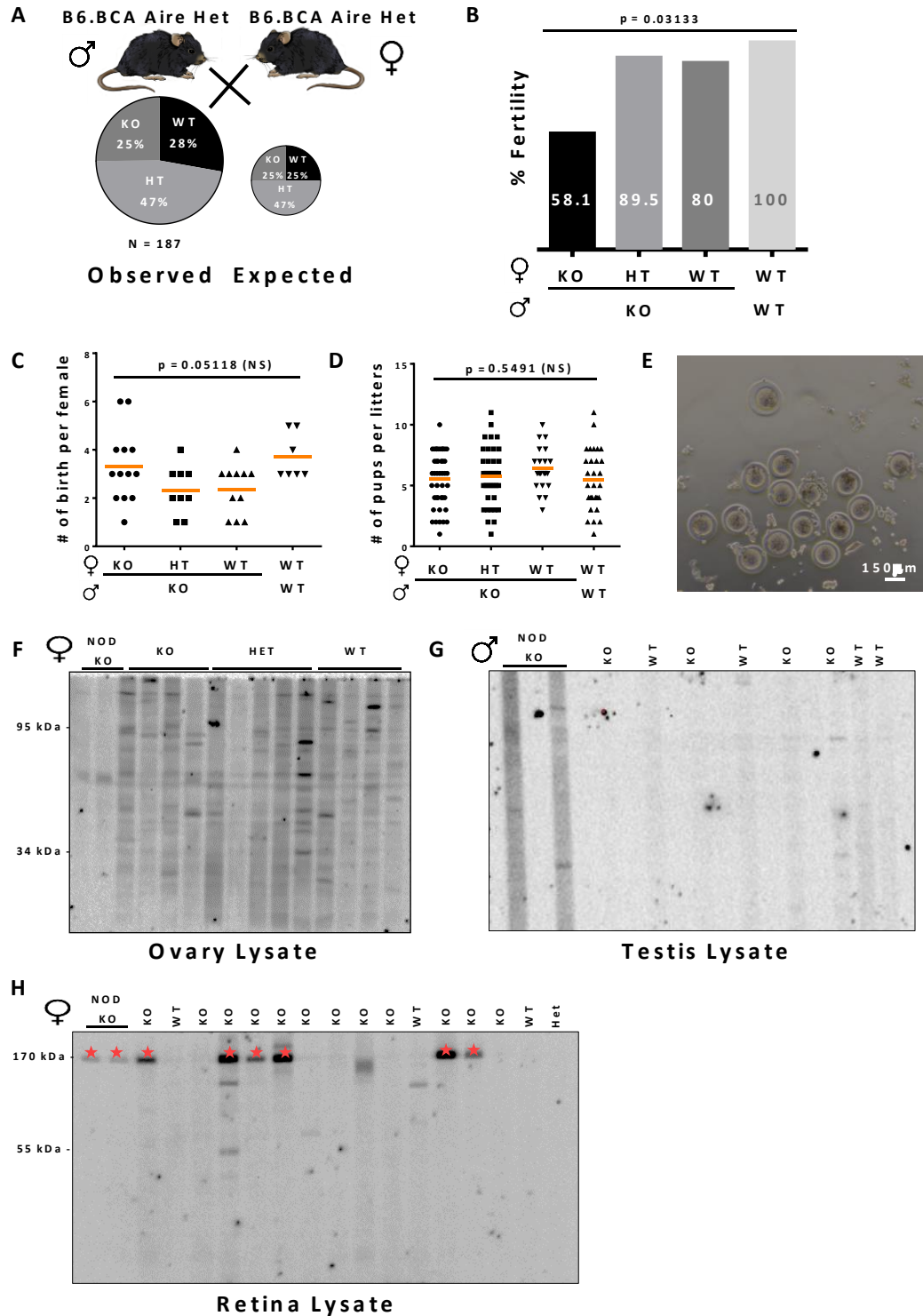**H** ♀ NOD KO | KO WT KO KO KO KO KO KO KO KO WT KO KO KO WT Het — 170 kDa, 55 kDa — **Retina Lysate**

16

Figure 4. B6.BCA Aire KO fertility. A. Pie charts illustrate that mating of Aire heterozygous mice (Het) produces Mendelian ratios. **B.** Graph showing percentage of females that gave birth at least once in six months; fertility rate of mating WT was defined as 100%. Statistical analysis was done using nonparametric one-way Anova (*n*=21 ,14 ,16, 7, respectively). **C.** Graph showing the number of birth per female analyzed by nonparametric one-way Anova. Only fertile females were counted. **D.** Graph showing the number of litter per birth analyzed by nonparametric one-way Anova. Only fertile females were counted. **E.** Microscopy image of Aire KO oocytes from 8 weeks old female; scale bar, 150 μm. **F-H.** Western blot analysis of auto-antigens derived from WT and KO mice and blotted against oviduct and uterus lysate (F), testis lysate (G) and retina lysate (H). Red stars indicate IRBP protein[34] ,a possible autoimmune target.

The lack of phenotype in the Aire KO offspring could be explained by environmental influence, i.e. signals from the uterus that masked the mechanical effect of the absence of Aire. Therefore, to produce a clearer phenotype we sought to prevent these signals by isolating the fertilization process and early development from their natural environment. For that, we conducted an *in vitro* fertilization (IVF) experiment. Because B6 mouse embryos do not develop *ex utero* beyond the two-cell stage, we crossed B6 Aire-deficient mice with 129 mouse strain for three rounds in order to backcross the KO allele into the 129 background. In contrast to B6 mice, the fertility of 129 Aire KO mice was normal and no differences in the rate of pregnancy or in the number of pups per litter were observed relative to control mice (Fig. 5A-C). Surprisingly, analysis of blastocyst maturation revealed similar results as in B6 Aire KO, with 20% less embryos developing to that stage in the KO mice, as compared to WT embryos (Fig. 5D-G). These results clearly show that there is a mechanistic property associated with Aire expression during early development.

In order to determine whether Aire contributes to robust genome activation introduced by expression of new RNA transcripts, we measured RNA levels by immunofluorescence using an RNA labeling kit (Click-it). We found that the amount of RNA was higher in WT two-cell embryos. Furthermore, the variability in expression levels was higher in embryos that expressed Aire, whereas KO embryos exhibited downscaling of newly transcribed RNA as compared to the WT (Fig. 5H-I).
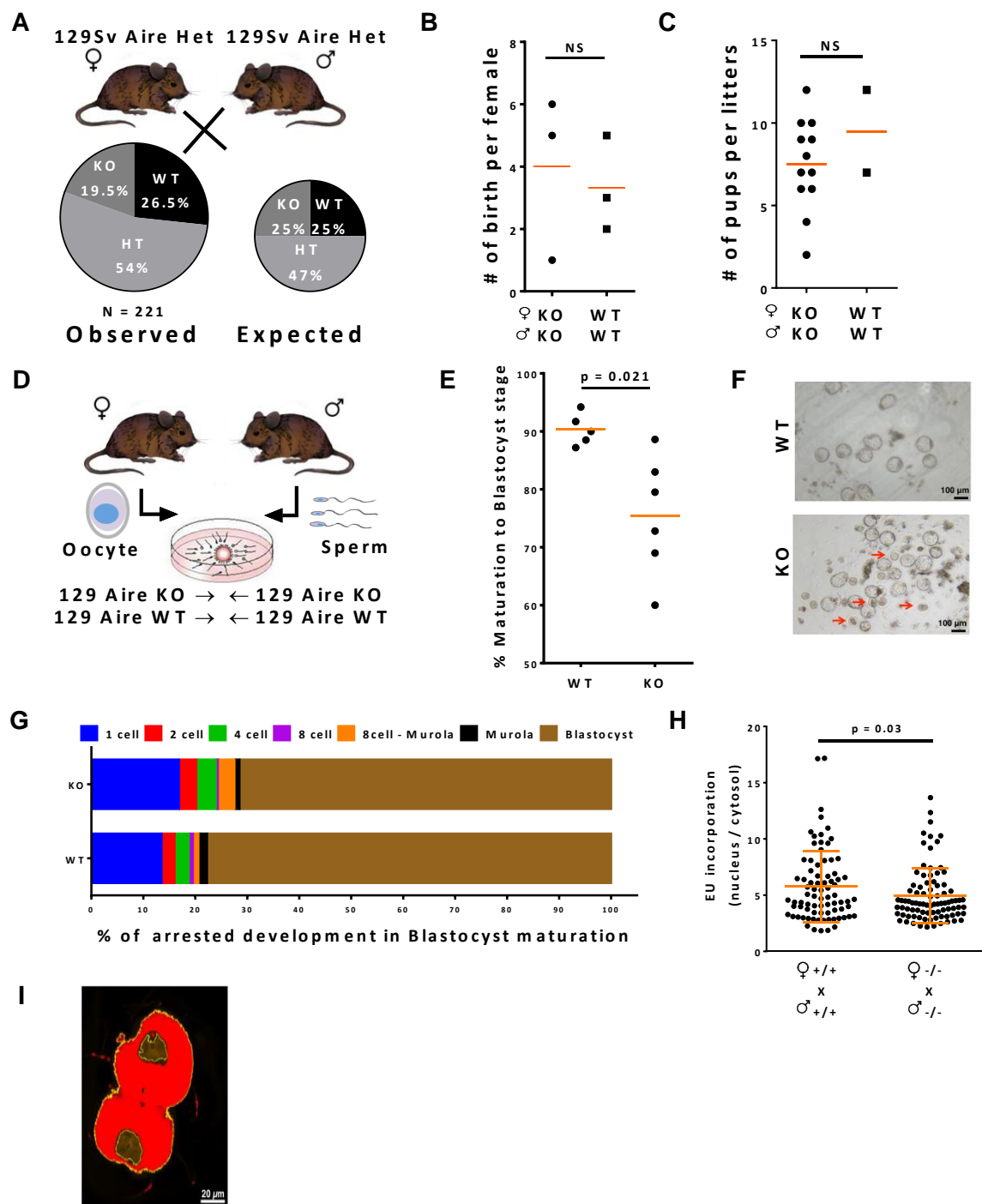
Figure 5. 129sv Aire KO developmental capabilities. A. Mendelian ratio of Aire heterozygote mating. **B.** Graph showing the number of birth per female analyzed by Student's *t*-test (*n* = 3,3, respectively). **C.** Graph showing the number of pups per litter analyzed by Student's *t*-test. **D.** Scheme of the IVF experiment. **E.** Graph showing the percentage of oocytes successfully fertilized by IVF that matured to blastocyst, as analyzed by Student's *t*-test (*n* = 481,719, respectively). **F.** Representative images of WT (top) and KO mature blastocyst after IVF treatment. Red arrows indicate embryos whose development had been stalled. **G.** Graph showing the distribution of developmental arrest by embryonic stage. **H.** Relative counts of RNA synthesis by EU incorporation analyzed by Student's *t*-test (*n* = 80,87, respectively). **I.** Representative image of EU staining. Yellow lines demarcate nuclei and cell borders and the cytoplasm is shown in red.

Aire knockout in zebrafish model

The fact that Aire is highly conserved across species raised the possibility that its function in early development might be conserved, alongside its function in immune cell maturation (Fig. 6A). ZGA is extensively studied in zebrafish (ZF), due to its easy manipulation and growth[35-36]. Moreover, the ZGA process in mice and ZF is mechanistically similar, differing mostly in the time scale (Fig. 6B). Therefore, to gain insight into the role of Aire in ZGA, we analyzed the process in ZF.
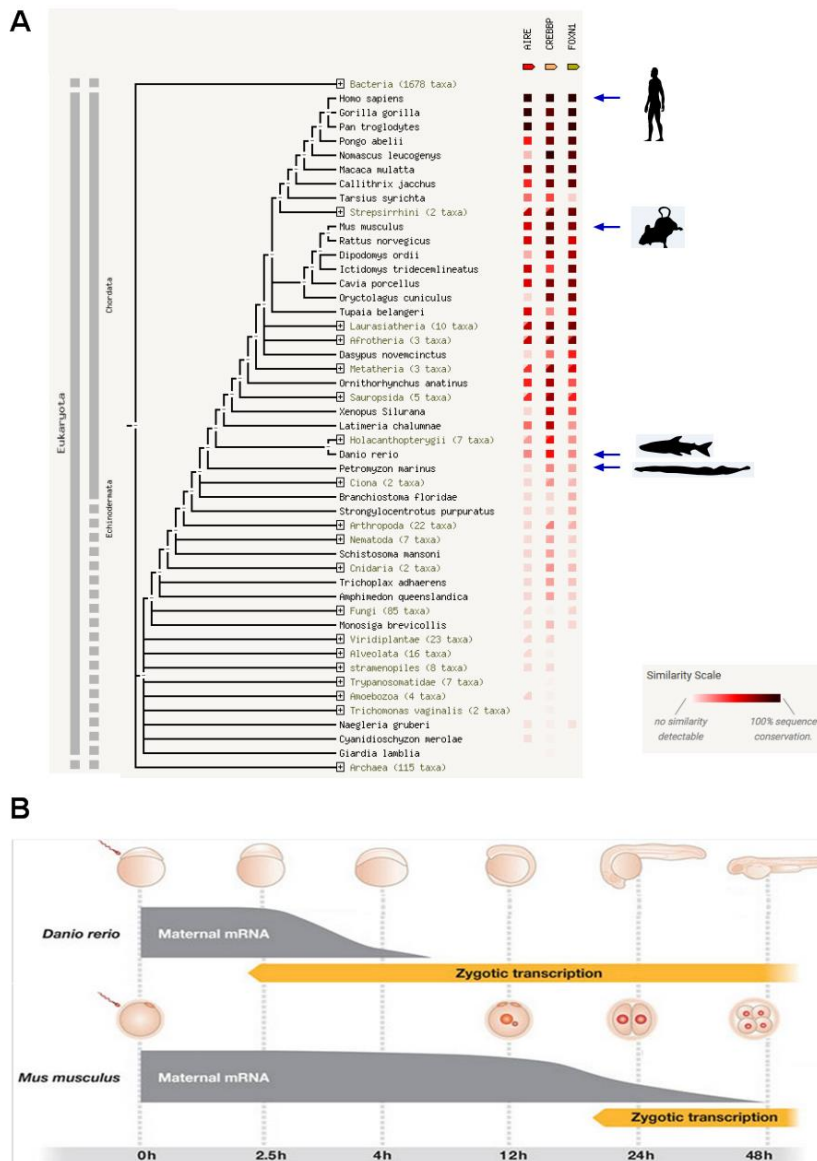

Figure 6. Aire is conserved throughout evolution. A. STRING evolution analysis of Aire, CREBP, and Foxn1 in human, mouse, zebrafish and lamprey, indicated by arrows from the top down. **B.** Schematic comparison of the time scale of early development between mice and zebrafish (adapted from Svoboda et al.[37]).

Quantitative PCR analysis showed that ZF embryos expressed *aire* in the thymus and, surprisingly, during early development (Fig. 7A). Interestingly, *aire* was expressed before the transcription of thymus progenitors, suggesting a possible role in ZF development (Fig. 7B). ZF embryos develop in an aqueous environment; therefore, intercellular perturbations can influence development dramatically. Additionally, early ZF embryos are sensitive to external signals, which allow easy perturbations. We therefore decided to challenge the ZF embryo by using morpholino (MO) against *aire* and *nanog*, the latter used as a positive control, and a negative control MO.
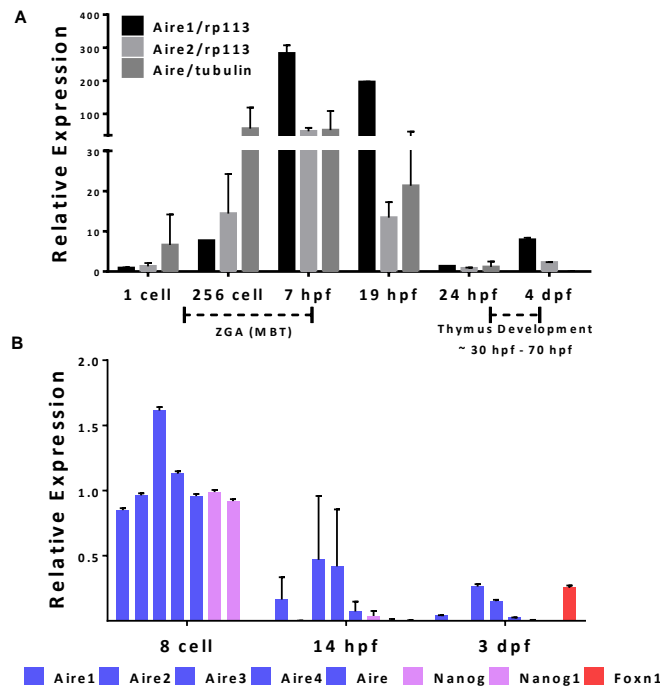
Figure 7. *aire* expression pattern in zebrafish embryonic development. A. Quantitative PCR of *aire* during early ZF development (one-cell, 256 cells, 7 hpf, 19 hpf, 24 hpf, 4 dpf)**.** The samples were tested with three different primers and normalized to two housekeeping genes (tubulin and RP113). Dotted lines indicate the timing of ZGA and thymic progenitors formation and maturation. Error bars represent SD of 3 technical replicates. **B.** Quantitative PCR of *aire* during early ZF development (8-cell, 14 hpf, 3 dpf)**.** The samples were tested with five different primers for *aire* and normalized to tubulin. The expressions of *Nanog*, a pluripotency and ZGA marker in ZF, and of *foxn1*, a marker for thymus progenitors, were also tested. Error bars represent SD of 3 technical replicates.

Remarkably, embryos that were injected with MO against *aire*, displayed different visible phenotypes (Fig. 8A,B). 80% of the injected embryos exhibited mild, severe or lethal developmental phenotypes (Fig. 8C), as compared to only 20% of the embryos injected with control MO (Fig.8D). Embryos injected with *nanog* MO died at early stages of development, as was previously described[35]. These results support the *in vivo* and *ex vivo* mouse results, indicating that indeed, *aire* has a function in early development.

To better understand *aire* contribution to ZF development, we performed single-embryo RNA-seq on control and *aire* MO-injected embryos. Embryos were collected at two time points, 6 hpf and 14 hpf, when defects started to show in MO-injected embryos. The results confirmed different global gene expression in *aire* MO injected embryos, as compared to the control (Fig. 8E). Unbiased clustering clearly distinguished most of the *aire* MO embryos from control embryos. A few *aire* MO-injected embryos clustered to the control group, which is not surprising given the phenotypic diversity observed in these embryos (Fig. 8C,D).

Pol II is a key factor during ZGA in ZF and its inhibition by α-amanitin terminates embryonic development at very early stages[35], whereas *aire* activity depends on the functionality of Pol II machinery. In order to link the activity of Aire and Pol II machinery, we monitored early ZF development in different media containing a gradient of α-amanitin concentrations, starting in a lethal dosage[35]. While in the high and low concentrations lethal and normal phenotypes were observed, respectively, embryos developing in the intermediate dilutions exhibited similar phenotypes as in *aire* MO injected-embryos. These results suggest that Pol II activity is decreased in the *aire* MO injected embryos. The fact that Pol II inhibition resulted in the same variety of phenotypes as in *aire* MO strengthens the role of *aire* in controlling robust transcription during ZF development (Fig. 8F).

A

z f *Aire*

aire-MO

chr2:23153295-23170502

chr2:23153295-23153541

B

Morpholino Injection

24 h

Phenotype
Observation

C

fli::eGFP ; 24 hpf

NORMAL   MILD   SEVERE

D

% Phenotype

MO Aire   MO Nanog   MO Control
12ng   8ng   12ng

N =   478   63   113

■ NORMAL ■ MILD ■ SEVERE ■ DEAD

E

Genes (N = 12157)

6 hpf   14 hpf

-3 0 3

F

% Phenotype

DDW 0.2 0.1 0.05 0.025 0.0125 0.00625 0.003125 0.001563 0.000781 0.000391 0.000195 Aire MO

α - Aminitin (ng)

N = 82 57 67 65 70 93 76 63 63 69 63 35 478

■ NORMAL ■ MILD ■ SEVERE ■ DEAD

22

Figure 8. *aire* MO injection in zebrafish. A. Scheme of MO design. **B.** Morpholino injection experimental design scheme **C.** Phenotypes observed at 24 hpf after injection of MO against *aire* during one-cell stage, using GFP as a marker for endothelial cells. Scale bar, 1.25 mm **D.** Graph showing the distribution of the different phenotypes (normal /mild/severe/dead). *aire* MO, *n*=478; *Nanog* MO, *n*=63; control MO, *n*=113. **E.** Heat map shwoing gene expression clustering of MO-injected embryos (*n*=38) and uninjected embryos (*n*=4). **F.** Graph showing the distribution of the different phenotypes upon inhibition of Pol II activity by a gradient of α-amanitin concentrations.

To validate the results of the MO experiments and to exclude any MO non-specific effects, we generated *aire*-deficient ZF using CRISPR-Cas9 method (Fig. 9A,B). The embryos were validated for genomic aberration by high-resolution melting curve analysis (Fig. 9D), and suitable candidates underwent Sanger sequencing (Fig. 9C). We generated an *aire*-deficient ZF with 5 bp insertion in the first exon (*aire*$^{+/5ins}$), whose stop codon terminates translation after 50 amino acids. Mating heterozygotes of mutant ZF (F1; *aire*$^{+/5ins}$) generated embryos at a normal Mendelian distribution (Fig. 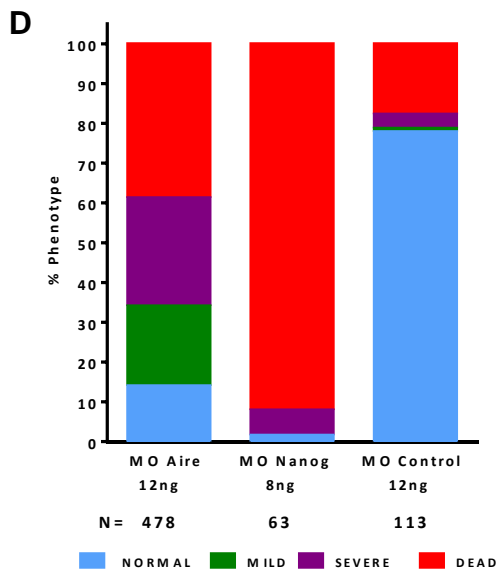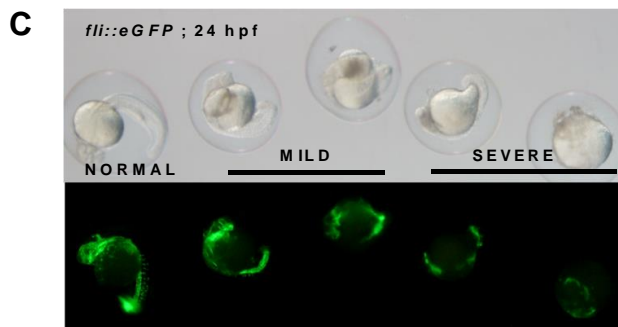9E). Mating of homozygotes (F2; *aire*$^{5ins/5ins}$, Fig. 9A) resulted in a significant increase in embryos with mild developmental defects, further supporting a role for *aire* in ZF development (Fig. 9F,G). **We are currently characterizing two other ZF mutants with larger deletions within *aire* locus (Fig. 9H,I), which may resolve the difference in severity between the MO and KO developmental phenotypes in zebrafish.**

A

W390 (fli::eGFP)
gRNA + Cas9 protein injection

24 h.p.i. – validating Crispr cut

3 Months for sexual maturation

Founder

aire WT/WT × aire mosaic/WT (??)

24 h.p.f. – validating Germ Line transmission

F1

aire mut**/WT × aire mut**/WT → Sequencing for mutated alleles
** - can be different

24 h.p.f. – statistics

F2

aire MUT/WT
aire MUT/MUT
aire WT/WT

B

z f *aire*

chr2:23153295-23170502

aire gRNA

chr2:23153295-23153541

C

W T allele   agagtctgacctgcgctcacagcttcgagcTTGTCGCACTGAGATCGCGAtggcaattca tgaccctttccactgctgatggactggcagatcacaacgtcatctcag   chr2:23153327-23153447

M utant allele   agagtctgacctgcgctcacagcttcgagcTTGTCGCACTGAGATCGCGAtggcaattca tgaccctttccactgctgatggactggcagatcacaacgtcatctcag

caatg

5bp Insertion

D

Derivative Melt Curves

Mutant   WT

E

♂ aire +/5ins × aire +/5ins ♀

KO 25%   W T 27%
H T 48%

KO 25%   W T 25%
H T 47%

Total = 249

O b s e r v e d   E x p e c t e d

F

% Phenotype

aire wt/wt   aire 5ins/5ins

■ Normal   ■ Mild   ■ Severe   ■ Dead

G

aire wt/wt   aire 5ins/5ins

H

z f *aire*

chr2:23153295-23170502

aire gRNA

aire gRNA

chr2:23161484-23162296

I

z f *aire*

chr2:23153295-23170502

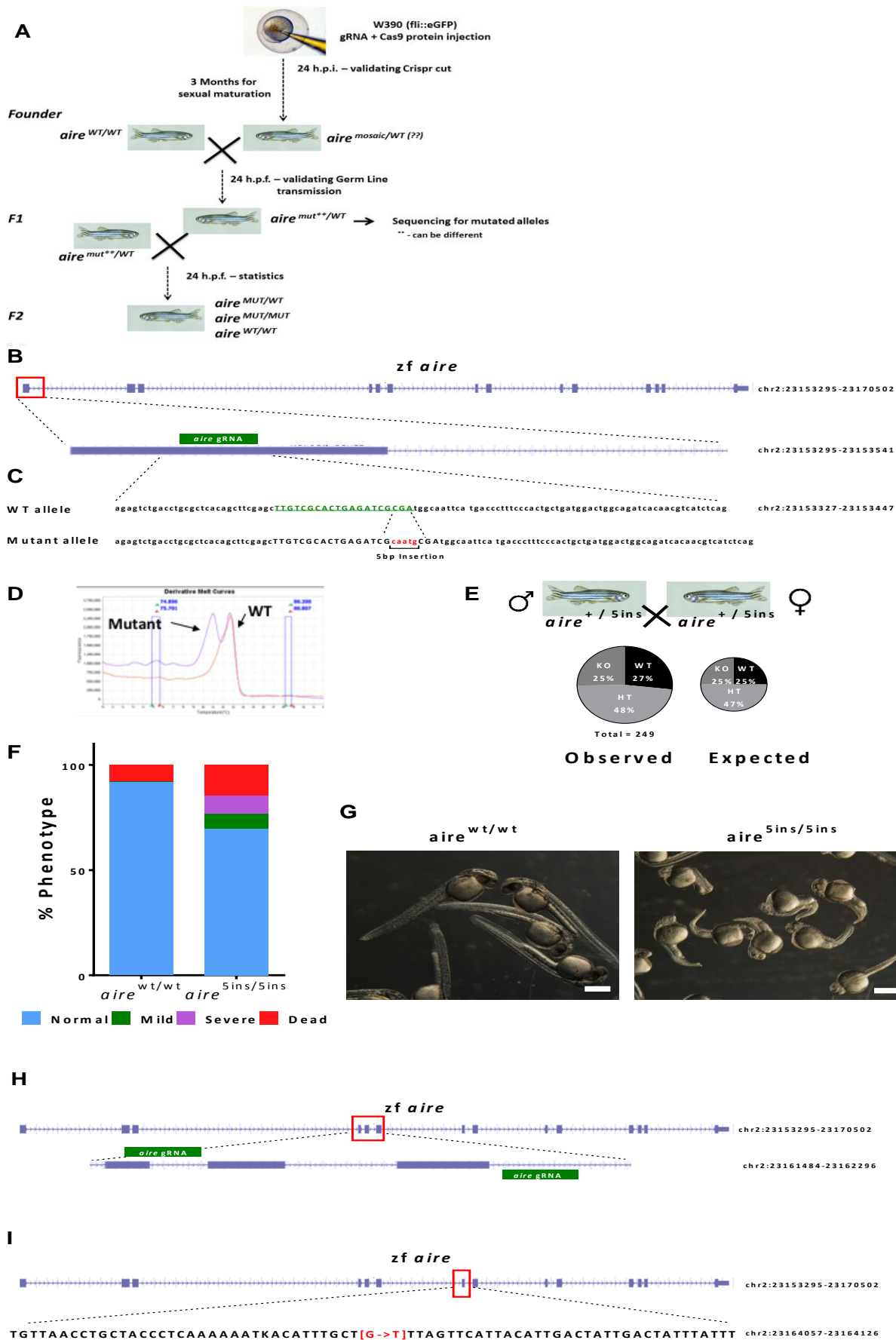TGTTAACCTGCTACCCTCAAAAAAATKACATTTGCT[G->T]TTAGTTCATTACATTGACTATTGACTATTTATTT   chr2:23164057-23164126

24

Figure 9. *aire* knockout in zebrafish. A. Knockout generation design scheme. **B.** Design of *aire* sgRNA. **C.** Validation by Sanger sequencing of the insertion of 5 bp mutant allele. **D.** High-resolution melting curve analysis of the mutant allele. **E.** Mendelian ratios of *aire*$^{+/5ins}$ mating. **F.** Graph showing the distribution of the different phenotypes in WT and aire KO embryos (WT, *n*=732; mutant, *n*=329. **G.** Representative images of WT (*aire*$^{+/+}$, left) and mutant (*aire*$^{5ins\,/5ins}$) embryos; scale bars, 2 mm and 1.25 mm, respectively. **H.** Design of *aire* sgRNA targeting SAND domain for complete deletion of the entire region. **I.** Design of *aire* KO based on point mutation (ordered from ZFIN sa17261).

Aire is dispensable for maintaining murine cell pluripotency

Understanding Aire function in the pluripotency circuitry may reveal unique properties of transcriptional regulation in the pluripotent state. First, we characterized embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) that lack Aire expression. For that, we generated a variety of transgenic and reporter cell lines. Aire$^{-/-}$ ESCs were made by cross-mating of mice heterozygous for Aire (Aire$^{+/-}$)[38]. On E3.5, blastocysts were collected and seeded on mouse embryonic irradiated feeder (MEF) cells until ESC colonies were formed; Aire-deficient colonies were identified by genotyping. Aire$^{-/-}$ iPSCs were generated by isolating tail-tip fibroblasts (TTFs) from Aire knockout mice[38], followed by infection with rtTA- and doxycycline (Dox)-inducible lentiviral vectors encoding the four Yamanaka factors Oct4, Sox2, Klf4 and c-Myc in a single cassette. Dox was applied to the TTFs until colonies of iPSCs were generated.

Next, Aire KO ES and iPS cells were examined for pluripotency characteristics. Immunostaining showed that the cell lines expressed the bona fide pluripotency markers Nanog, Oct4, SSEA1 and alkaline phosphatases (Fig. 10A). Contribution to all three germ layers was demonstrated by generating a chimera via microinjection of Aire KO iPSCs to host blastocyst (Fig. 10B,C). Recently, Gu at el.[28] reported that Aire has a major role in maintenance of the pluripotent state, based on Aire knockdown using shRNA. Surprisingly, our results, which were generated using defined genetic knockouts, contrasted those findings, showing that the absence of Aire did not affect pluripotent state maintenance (Fig. 10E). Alkaline phosphatase test of Aire knockout and control cells showed no difference in the number of formed colonies. This was true under several culturing conditions, including feeder cells with medium mouse ES and gelatin with

medium mouse ES + 2i. Furthermore, the growth rates of Aire-deficient and WT ES and iPS cell lines were similar (Fig. 10D). These results stress the importance of using a well-defined genetic knockout rather than knockdown, which can sometimes be noisy due to low efficiency and/or non-specific off-target effects. I conclude that Aire is dispensable for maintenance of rodent pluripotency.

Based on these results, we concluded that Aire is dispensable for maintenance of rodent cell pluripotency. Yet, Aire could still play a role in controlling gene expression, mainly as part of the pause-release mechanism[26]. To identify pathways that are regulated by Aire, we conducted RNA-seq to compare the expression rates of WT and Aire KO mouse ESCs under two different naïve pluripotency growth conditions (Serum/Lif and N2B27 2i/Lif). No significant differences were observed, strengthening the conclusion that Aire is dispensable for maintaining *in vitro* murine naïve pluripotency (Fig. 10 F-H).

Finally, Aire was tested for a role in exiting of pluripotency (Fig. 10I) by changing the medium of WT/KO cells from naïve (N2B27 2i/LIF) to prime supportive (N2B27 FGF/ACTIVIN). Conversion efficiency was measured by relative expression of primed and naïve markers (Fig. 10K-L), FACS staining and immunofluorescence staining of deltaPE-Oct4-GFP reporter, which is specific to the naïve state (data not shown). The results indicate that Aire is not involved in terminating the naïve pluripotency state.
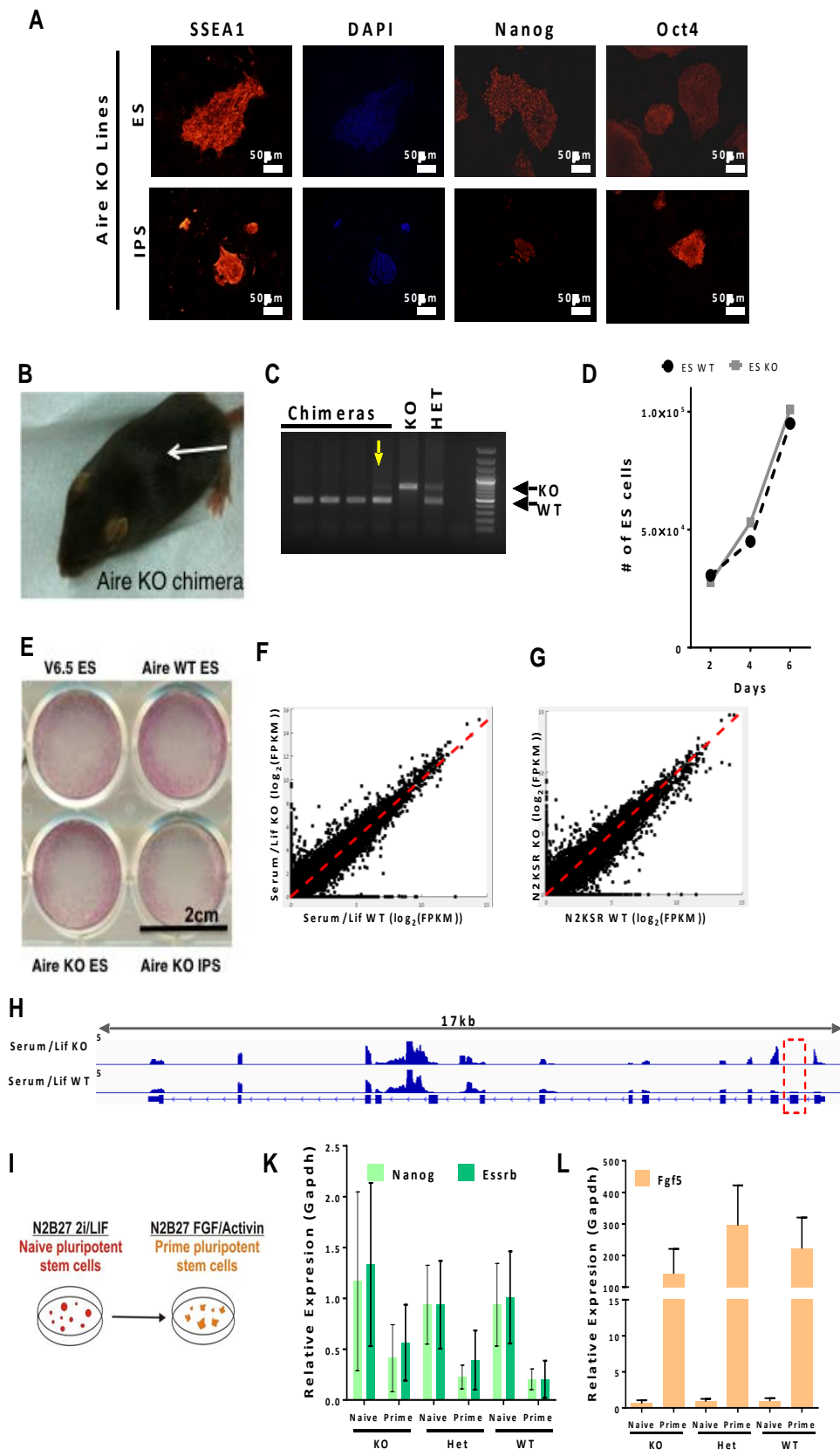
Figure 10. Characterization of Aire KO ES and IPS Cells: A. Immunostaining images demonstrate the expression of Nanog, DAPI, SSEA1 and Oct4 in AIRE KO mouse ESCs and iPSCs. Scale bars, 50 μm. **B.** Adult chimera obtained from Aire$^{-/-}$ iPSCs following microinjection into host blastocysts. Chimerism is indicated by the mixed black-agouti coat colors. **C.** Confirmation of the KO genotype in chimera mice by PCR. Yellow arrow indicates the KO band and the chimera. **D.** Growth rate curves of Aire KO and WT ESCs demonstrate comparable expansion capacity *in vitro* (2 out of 3 replicates are shown) **E.** Alkaline phosphatase staining of Aire KO ESCs, Aire KO iPSCs and control ESCs. 2000 iPS/ES cells were seeded for 4 days before staining. Scale bar, 2 cm. **F,G.** Scatter plots showing gene expression correlation profile of KO ESCs vs WT ESCs in serum/Lif (F) and N2KSR (G). Red line

indicates linear values. Values are $\log_2$-normalized. R=0.948 in F, R=0.9486 in G. **H**. RNA-seq validation of Aire KO cells. The red dashed box highlights the deleted exon. **I.** Naïve to primed state conversion experimental design scheme**. K.** Graph showing expression profiles of Nanog and Essrb, markers of naïve state, in naïve or primed WT, Aire KO and Aire heterozygote (Het) cells, as analyzed by qPCR. Error bars represent SD of N=3, technical replicates. **L.** Graph showing expression profile of Fgf5, a marker of primed state, by the same cells as analyzed by qPCR. Error bars represent SD of 3 technical replicates.

Aire boosts cellular reprogramming rate

A recent study showed that the last step of reprogramming somatic cell to pluripotency involves Pol II pause-release and transcription activation by CDK9 and BRD4[39], all of which are known to be bound by Aire[25,40,41]. Therefore, assuming that expression of Aire during early development may facilitate global genome transcription, we hypothesized that Aire might also boost the expression of factors that stimulate somatic cell reprogramming and induction of pluripotency.

We have recently developed a system for deterministic reprogramming, which enables monitoring of reprogramming within homogenous cell population. In this system, cells are reprogrammed to iPSCs by the 4 Yamanaka factors, alongside a reduction in Mbd3 protein[42]. Using RNA-seq and Trans-seq data obtained from this system, we found that Aire expression is the highest at the late stages of reprogramming (days 6-8), and returns to its low levels once iPSCs have formed (Fig. 11B,C). Data from GeneArray of somatic cell reprogramming of WT mouse embryonic fibroblasts (MEFs) showed the same pattern, where Aire levels are highest in the second phase of reprogramming (Fig. 11A). Moreover, Aire-GFP MEFs started to express GFP on day 11 after OSKM introduction (Fig. 11D,E). These observations led us to investigate the importance of Aire in cellular reprogramming.

First, we generated embryonic fibroblasts from E13.5 WT and Aire KO mice and infected them with OSKM. Analysis of the fibroblasts showed that WT cells generate significantly more colonies than KO cells (Fig. 11F-H), as was confirmed by FACS measuring Oct4-GFP and by staining for the early pluripotent marker SSEA1. We concluded that Aire may play a role during reprogramming, potentially implicating

28

transcription and, mainly, Pol II pause-release in this process. We therefore hypothesized that expressing Aire together with Yamanaka factors could improve reprogramming rate.
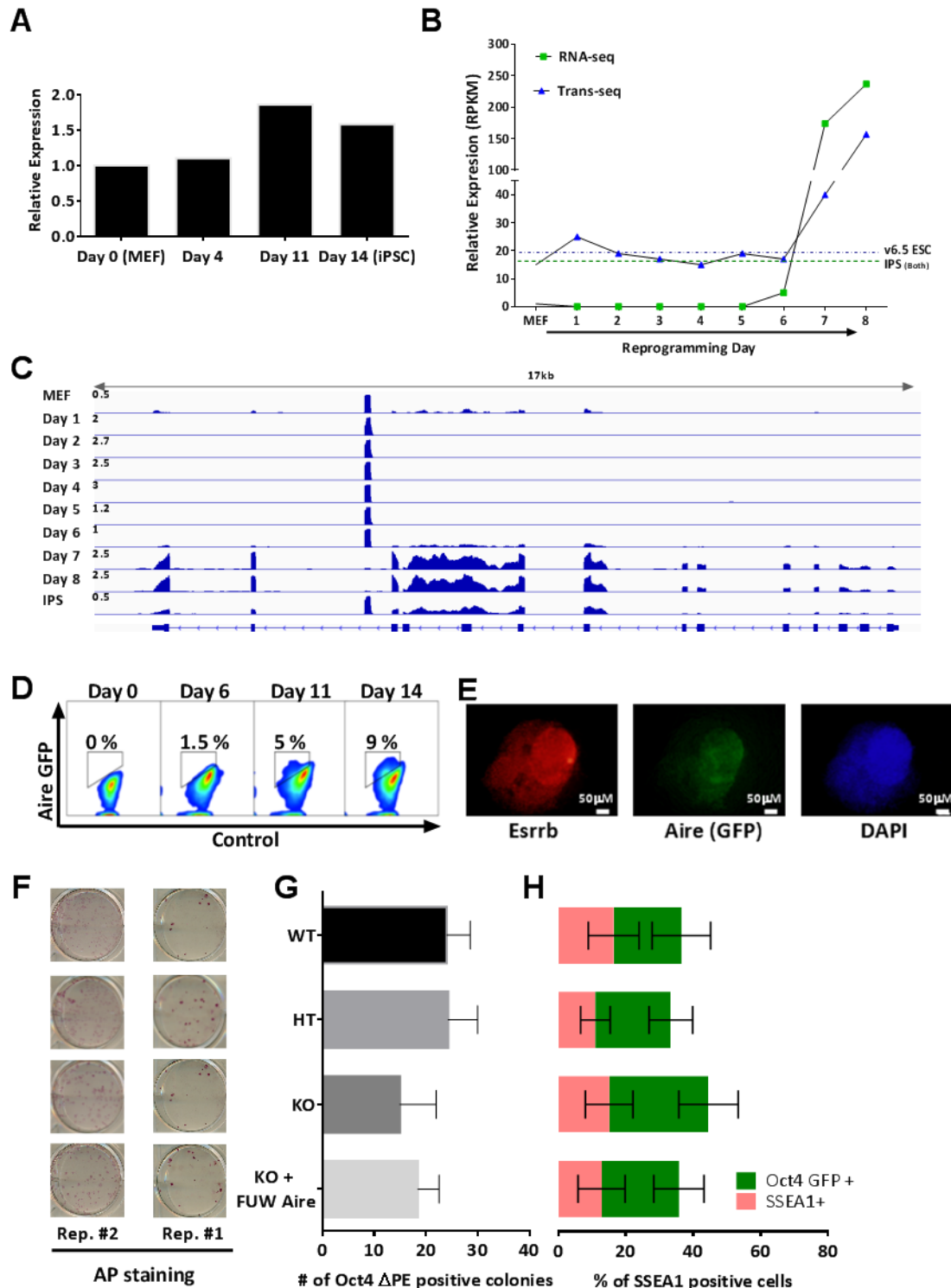
Figure 11. Aire expression during somatic cell reprogramming. A. Endogenous Aire expression measured by GeneArray during WT MEF reprogramming. **B.** Endogenous Aire expression during MBD3[flox/-] reprogramming analyzed by RNA-seq and DGE-seq. Dashed lines indicate Aire expression levels in v6.5 ESCs and MBD3[flox/-] ESCs. **C.** IGV view of Aire expression during MBD3[flox/-] reprogramming. **D.** Aire expression was analyzed by FACS during Aire-GFP reprogramming. **E.** Immunostaining of Aire-GFP iPSCs; scale bars, 50 μm**. F.** Alkaline phosphatase (AP) staining of colonies generated by Aire WT/heterozygote (HT)/KO reprogramming; 2 experimental replicates are shown. **G.** Graph showing the number of ΔPE-Oct4-GFP-positive colonies generated by Aire WT/HT/KO reprogramming. Error bars represent SD, *n*=4. **H.** Graph showing the ratio of FACS-stained ΔPE-Oct4-GFP- and SSEA1-positive cells. Error bars represent SD, *n*=4.

To test this hypothesis, we used MEFs carrying an Oct4-GFP knockin reporter allele and a Rosa26-M2RtTA knockin that constitutively expresses M2RtTA. MEFs were infected with Dox-inducible lentiviral vectors encoding polycistronic cassette of Yamanaka factors and a separate Aire lentiviral transgene; control cells were infected with Yamanaka factors only. Interestingly, the number of colonies from cells infected with Aire was more than 2-fold higher than in the control (Fig. 12A). The cells were pluripotent, as evidenced by Oct4-GFP specific reactivation of the reporter allele.

Encouraged by these results, we sought to establish a "secondary reprogramming" platform that would generate iPSCs following Dox induction (Fig. 12B). These secondary cell lines are used for more efficient and homogenous reprogramming assays[43]. We have generated the desirable cell line (NGFP1-AireOE) using Nanog-GFP-1 (NGFP1) and Flp-In two-step cloning strategy[44] (Fig. 13). For the reprogramming assay, we isolated MEFs at E14.5 and treated them with Dox until colonies were formed in the control cell line. Aire expression in NGFP1-AireOE cells increased the number of reprogrammed colonies by 2.5-fold relative to control NGFP1 cells, as measured by both AP staining (Fig. 12C) and FACS staining for Nanog-GFP (Fig. 12D,E).
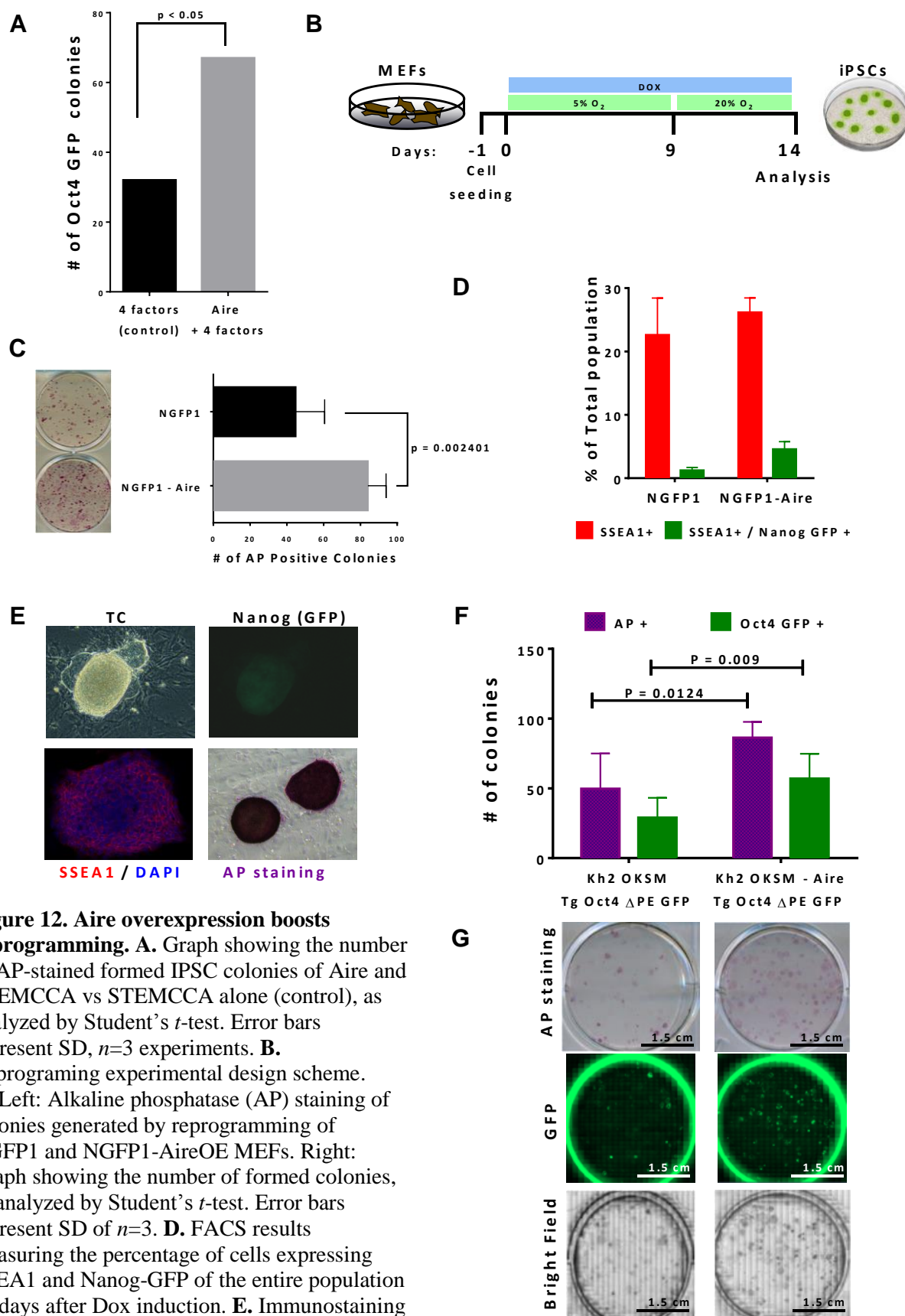
Figure 12. Aire overexpression boosts reprogramming. A. Graph showing the number of AP-stained formed IPSC colonies of Aire and STEMCCA vs STEMCCA alone (control), as analyzed by Student's *t*-test. Error bars represent SD, *n*=3 experiments. **B.** Reprograming experimental design scheme. **C.** Left: Alkaline phosphatase (AP) staining of colonies generated by reprogramming of NGFP1 and NGFP1-AireOE MEFs. Right: Graph showing the number of formed colonies, as analyzed by Student's *t*-test. Error bars represent SD of *n*=3. **D.** FACS results measuring the percentage of cells expressing SSEA1 and Nanog-GFP of the entire population 14 days after Dox induction. **E.** Immunostaining for NGFP1-Aire positive iPSCs. **F,G.** AP staining of colonies generated by reprogramming kh2-OKSM/ΔPE-Oct4-GFP and kh2-OKSM/Aire/ΔPE-Oct4-GFP MEFs (G). Graph showing the number of formed colonies, as analyzed by Student's *t*-test. GFP-positive colonies were also counted to evaluate Oct4 activation (F). Error bars represent SD of 4 biological replicates. Scale bars, 1.5 cm.
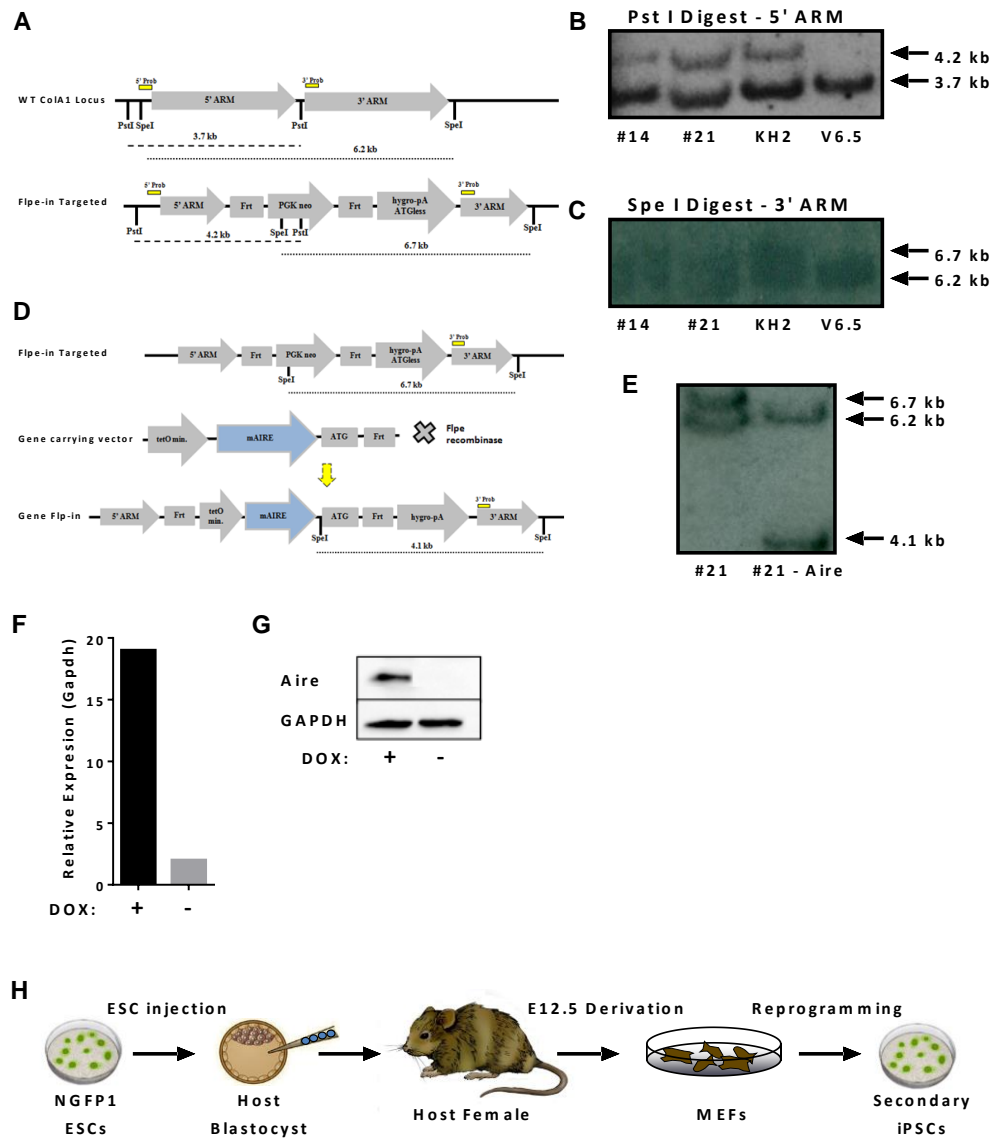
31

Figure 13. Generation of NGFP1-Aire secondary reprograming system. **A.** Schematic representation of the Col1a1 WT locus and the locus after homologous recombination placing frt-ATGless-pA cassette downstream to it. Yellow bars represent the probe used for southern analysis (5' external probe, 3' internal probe) and sites for restriction enzymes are indicated with their estimated products. **B.** Representative results of southern analysis using 5' external probe for colonies resistant to neomycin. **C.** Representative results of southern analysis using 3' internal probe for colonies resistant to neomycin **D.** Schematic representation of the frt-ATGless-pA Col1a1 targeted locus and the introduction of the Aire-carrying vector via electroporation with FLPe transient expression vector. After electroporation, the Col1a1 locus is targeted with TetO. Promoter to over-express Aire and hygromycin resistance is active. The black bar represents the probe used for southern analysis (3' internal probe) and sites for restriction enzymes are indicated with their. **E.** Representative results of southern analysis using 3' internal probe for colonies resistant to hygromycin. **F.** qPCR to validate Aire expression after Dox administration **G.** Western blot analyses to validate the NGFP1-Aire ESCs. Targeted cells were cultured with or without (control) Dox and harvested for protein. **H.** Chimera generation experimental design scheme.

To confirm that the induction is cell-line-independent, we repeated the experiment with two more cell lines. The first, the tgOct4-gfp/mCherry cell line, expresses both GFP under Oct4 promoter and mCherry in a constitutive manner. Unlike before, we generated a secondary system that carries STEMCAA cassette (four Yamanaka factors) and infected the cells with viruses encoding either for Aire, or for mock as a control. To generate the secondary cell line, Kh2-OKSM/Tg ΔPE-Oct4-GFP reporter mice were mated with Kh2-Aire mice. In the obtained MEFs, Dox can induce expression of both Yamanaka factors and Aire, which makes them an ideal system to examine Aire contribution to reprogramming (Fig. 12F-G).

Both cell lines produced the same results as in previous experiments. Colony count and FACS analysis showed about 2-fold increase in Aire-expressing cells. All iPSCs generated by reprogramming were allowed to grow for another three passages without Dox, to determine that they had completed the process and were stained for pluripotency markers. These results, which are summarized in Table 1, indicate that Aire supports reprogramming to pluripotency via a yet unknown mechanism, which may be related to cell transcription machinery.

Table 1. Summary of Reprogramming Experiments.

Genotype (MEF)	MEF generation strategy	OKSM induction	Biological Replications	Aire's contribution (# Colonies)
B6;129S4-Pou5f1^{tm2Jae} –EGFP/J (jax #008214) X B6.CgGt(ROSA)26Sortm1(rtTA*M2)Jae/J (jax #006965)	Mating	Primary infection (OKSM-STEMCAA, fuw-Teto-mAire)	2	2.1-Fold
NGFP1-Aire (targeted, NGFP1 as control)	ES targeting and chimera generation	Secondary, Dox	3	1.8 – 3-Fold
Tg(Pou5f1-ΔPE-EGFP)2Mnn/mCherry -NLS	ES targeting and chimera generation	Primary infection (OKSM-STEMCAA, fuw-Teto-mAire)	1	3-Fold
R26-M2rtTA;Col1a1-tetO-Aire X B6;129S4-*Col1a1*$^{tm1(tetO-Pou5f1,-Klf4,-Sox2,-Myc)Hoch}$/J (jax # 011001)	Mating	Secondary, Dox	6	2.5-Fold
B6;129S4-Pou5f1^{tm2Jae} –EGFP/J (jax #008214) /R26-M2rtTA;Col1a1-tetO-Aire X B6;129S4-*Col1a1*$^{tm1(tetO-Pou5f1,-Klf4,-Sox2,-Myc)Hoch}$/J (jax # 011001)	Mating	Secondary, Dox	5	1.7-Fold
B6.129S2-*Aire*$^{tm1.1Doi}$/J (jax# 004743) X B6;129S4-Pou5f1^{tm2Jae} –EGFP/J (jax #008214)	Mating (KO vs WT)	Primary infection (OKSM-STEMCAA)	2	1.12-Fold

Aire can replace exogenous Sox2 expression during iPSC reprogramming with the Yamanaka factors

Considering Aire's contribution to reprogramming and the fact that it induces a great variety of genes, it is reasonable to assume that Aire activates the four Yamanaka factors, their downstream targets, or other pluripotency genes (e.g., Nanog and Lin28) to promote the induction of pluripotency. Thus, we tested whether Aire overexpression could substitute for exogenous supplementation of any combination of the Yamanaka factors.

To that end, we first infected Nanog Cre-ER/TdTomato MEFs[45] with Oct4, Sox2 and STEMCCA containing plasmids that may promote pluripotency and Aire. Surprisingly, Aire and Sox2 alone could reactivate Nanog promoter after infection and treatment of 4-OHT at day 9 (Fig. 14A-C). This result prompted us to test whether Aire can replace one or more of the Yamanaka reprogramming factors. For that, we infected KH2-Aire MEFs derived from kh2-Aire mice with Dox-inducible lentiviral vectors encoding for Oct4, Sox2, Klf4, c-Myc and Nanog (each gene on a separate vector). Results showed that various combinations of these factors were ectopically expressed together with Aire. Yet, only one combination that did not include all four factors, namely Aire, c-Myc, Klf4 and Oct4 (AMKO), successfully and reproducibly generated iPSCs, thus replacing Sox2 in Yamanaka factors with Aire (Fig. 14D). Validation by PCR was conducted to eliminate the possibility that Sox2 was integrated as a result of contamination (Fig. 14E). AMKO iPSCs normally expressed pluripotent markers including Nanog, Sox2 and Oct4 (Fig. 14F). These cells also contributed to the three germ layers in teratoma assay and to chimera formation (Fig. 14G,H).
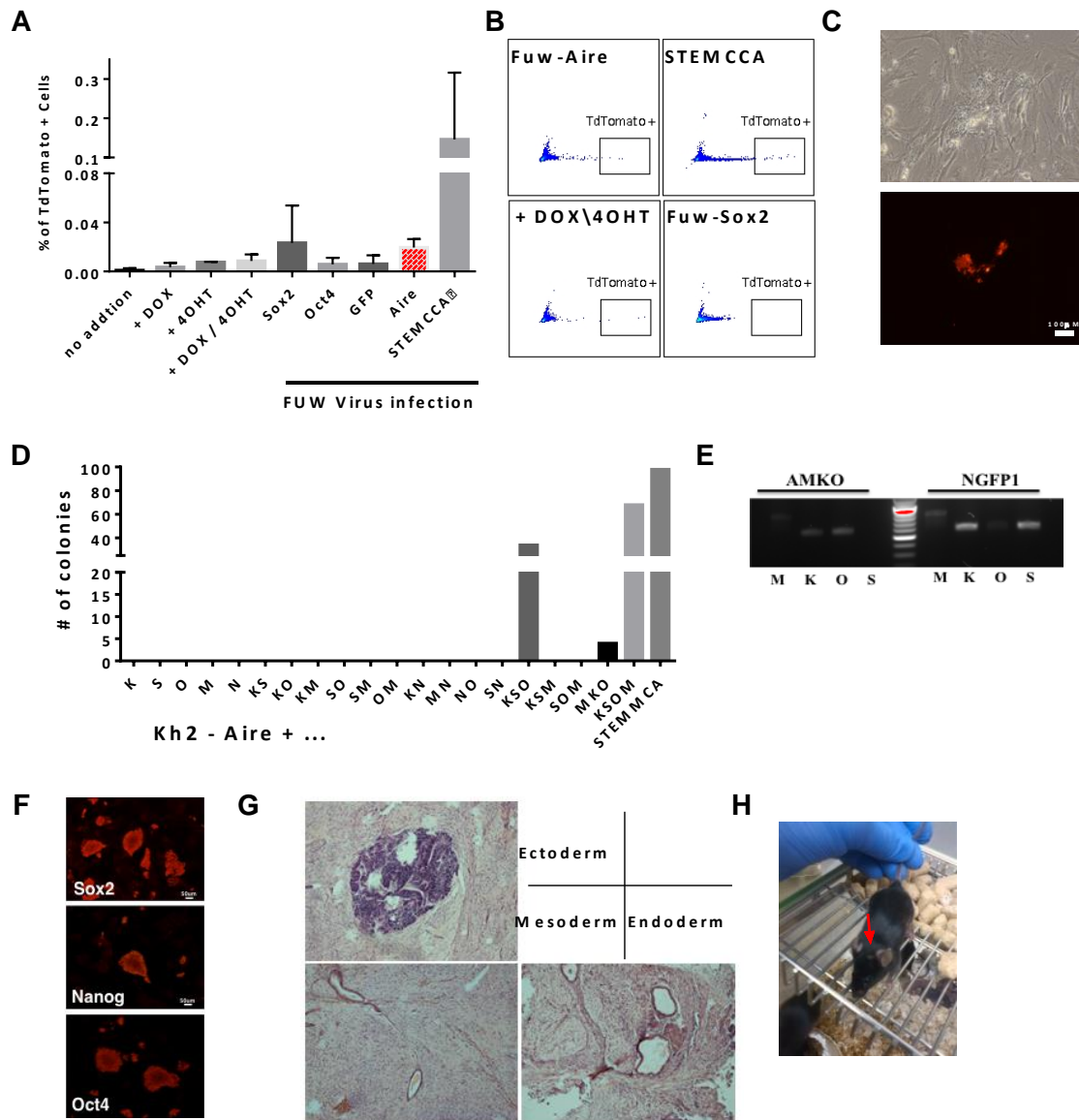
Figure 14. Aire activates genes promoting pluripotency. A. Graph showing the percentage of cells that activated Nanog promoter upon infection with Sox2/Oct4/GFP/Aire/STEMCCA. Error bars represent SD. **B.** FACS analysis of TdTomato positive cells. **C.** Representative image of TdTomato-positive cells; scale bar, 100 μm. **D.** Graph showing the number of colonies generated by different combinations of Oct4, c-Myc, Sox2, Klf4 and Nanog, alongside Aire. **E.** PCR validation of viral integration. **F.** Immunostaining images showing the expression of Nanog, DAPI, SSEA1, Oct4 and Sox2 in AMKO iPSCs. Scale bars, 50 μm. **G.** Teratoma assay for AMKO iPSCs. The lineages are shown on the upper right corner. **H.** AMKO iPSCs contribute to chimera generation, as evident by the presence of agouti coat color (marked by an arrow).

Aire target genes and partners

To define the molecular function of Aire protein domains and to determine which domain participates in the reprogramming process, we generated a library of Dox-inducible lentiviral vectors encoding Aire mutants, each lacking a different domain (Fig. 15A,B). To test their function, we infected tg-Oct4 GFP secondary fibroblasts with these Aire mutant viruses and measured iPSC formation rate. We found that all Aire domains are important for reprogramming; however, without SAND and PHD1 domains, reprogramming rate decreased to mock levels (Fig. 15C). These results are consistent with previous studies showing that these domains are crucial for Aire transcription activity in the mouse thymus[19]. Surprisingly, eliminating PHD2 had a mild positive effect on reprogramming. This raises the possibility that PHD2 it not required for pluripotency, which may also explain the low cross-species conservation of that specific domain.
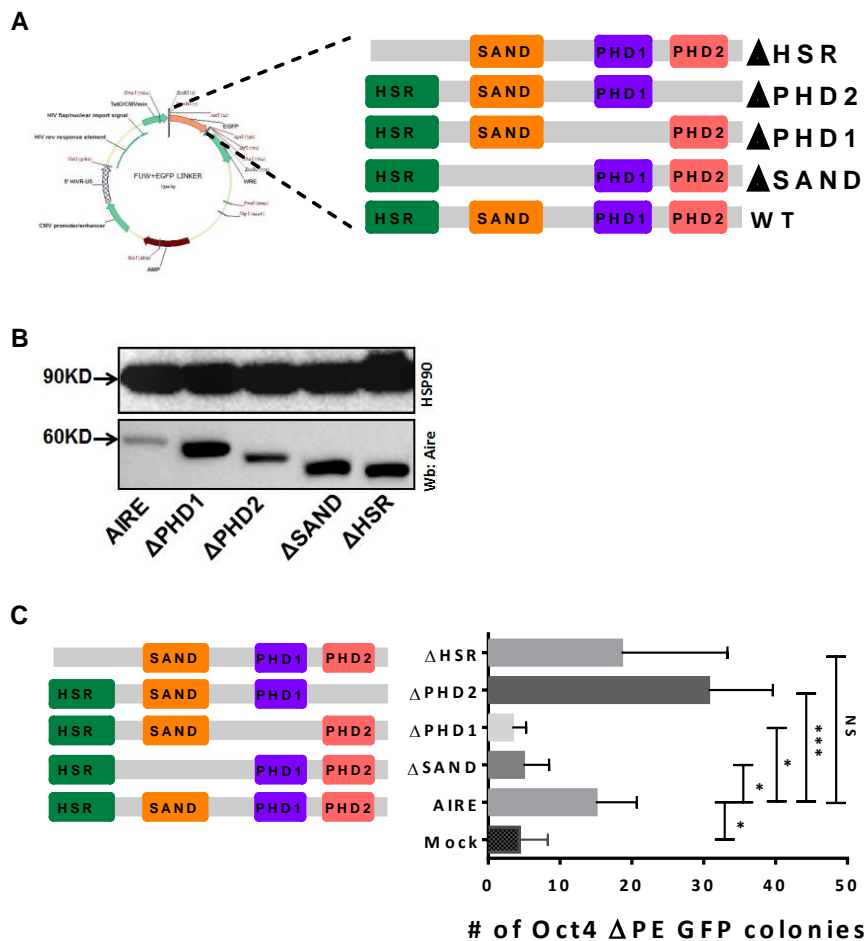
Figure 15. SAND and PHD1 domains are essential for Aire activity. A. Mutant library preparation design. **B.** Validation of Aire mutants by western blotting. **C.** Graph showing the number of ΔPE-Oct4-GFP-positive colonies after 14 days of reprogramming by OSKM, as analyzed by Student's *t*-test. Error bars represent SD of 2 biological replicates.

To uncover biochemical partners of Aire, we conducted Co-IP experiments for known targets of Aire[25] and for different pluripotent or Pol II pause-release elements during reprogramming. Results showed binding between Aire and Pol II, which is the main factor for Aire-mediated transcription (Fig. 16A).

To access the transcription start site in epigenetically repressed regions, DNA must be accessible to Pol II. DNA-PK and Top2a are recruited to the site and double-stranded DNA is formed to initiate transcription. Aire binds DNA-PK and Top2a to enhance transcription and to generate double-stranded breaks (DSB)[46]. Etoposide induction in 293HEK cells showed the same PTA expression as the induction of Aire, indicating that transcription is mediated also via DSBs[25,46]. Pursuing that hypothesis, we conducted reprogramming with DSB induction factors, namely etoposide and UV irradiation (10 j/m^2). Aire-induced reprogramming resulted in higher rates of reprogramming as previously showed. Surprisingly, UV-irradiated MEFs also displayed high reprogramming rates, but less than in Aire overexpressing cells (Fig. 16B). However, etoposide-treated cells did not manage to reprogram. Immunostaining at day 3 of reprogramming revealed less foci of γH2AX in all samples, except for untreated NGFP1 (Fig. 16C). These results support the hypothesis that Aire induces DSB by DNA-PK and Top2a to enhance transcription and induce reprograming.

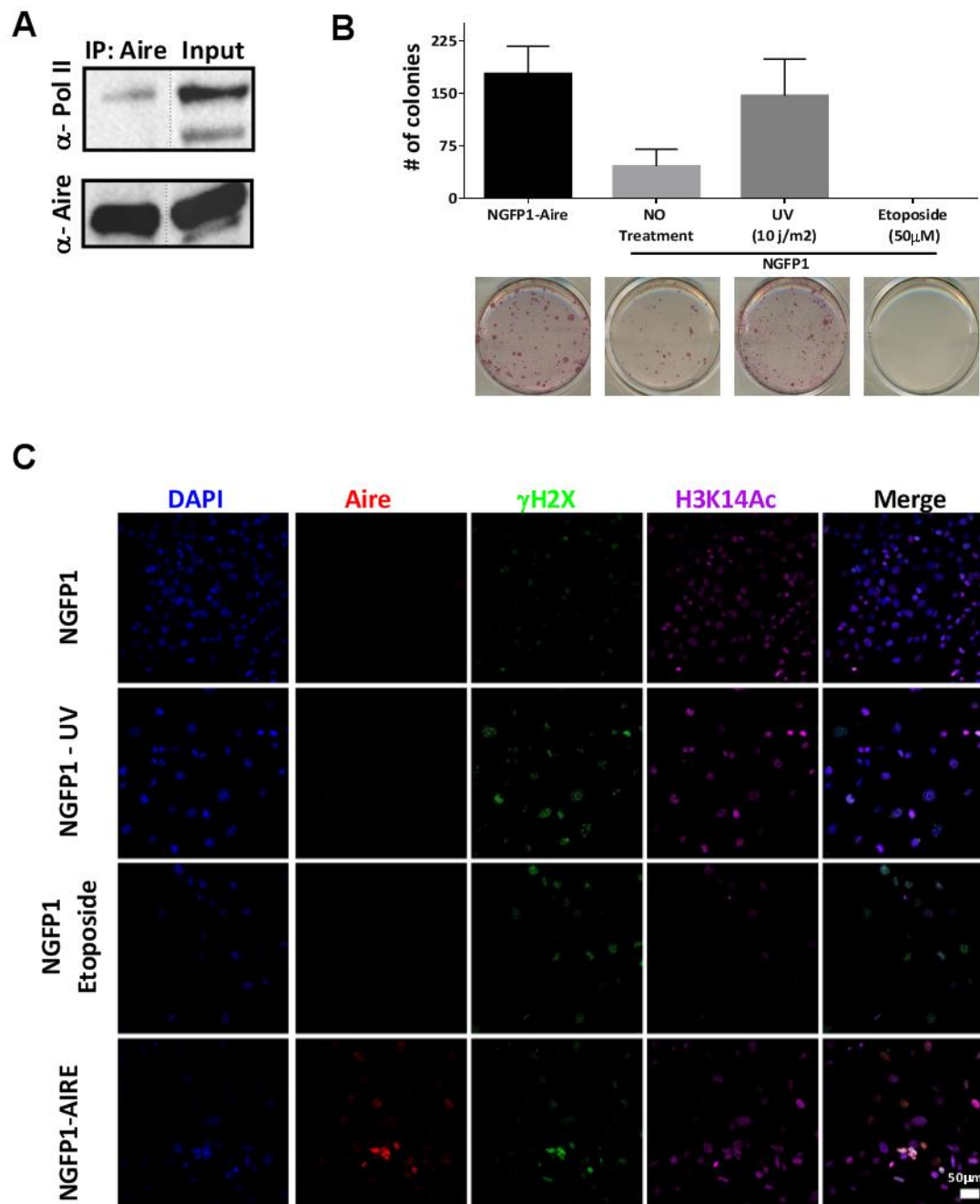**Figure 16. Aire activity is mediated by Pol II and DSB machinery. A.** Western blots confirming the binding of Aire to Pol II. **B.** AP colony count of NGFP1- Aire, NGFP1 DMSO / etoposide / UV-treated cells after 14 days of reprogramming. Error bars represent SD of 3 technical replicates. C. Immunostaining of day 3 MEFs for gene transcription and DNA damage markers. Scale bar, 50 μm.

In addition to the unique mechanism of Aire, it has three alternative splicing isoforms[47]. The rational for their existence is still unknown and they might have different roles. Although Aire1 is expressed in sperm, its function is uncertain and it might be that the different variants have a role during early development. Aire3 lacks the second PHD domain (PHD2) and, as shown here, a PHD2-deficient variant of Aire yields higher rate of reprogramming (Fig. 15C). To elucidate the possible roles of Aire variants, we examined expression levels and reprogramming activity of different isoforms in various cell lines. qPCR results showed that Aire3 is expressed in early development (Fig. 17A). Yet, no difference in reprogramming rates was observed between Aire3 and Aire1 (Fig. 17B).

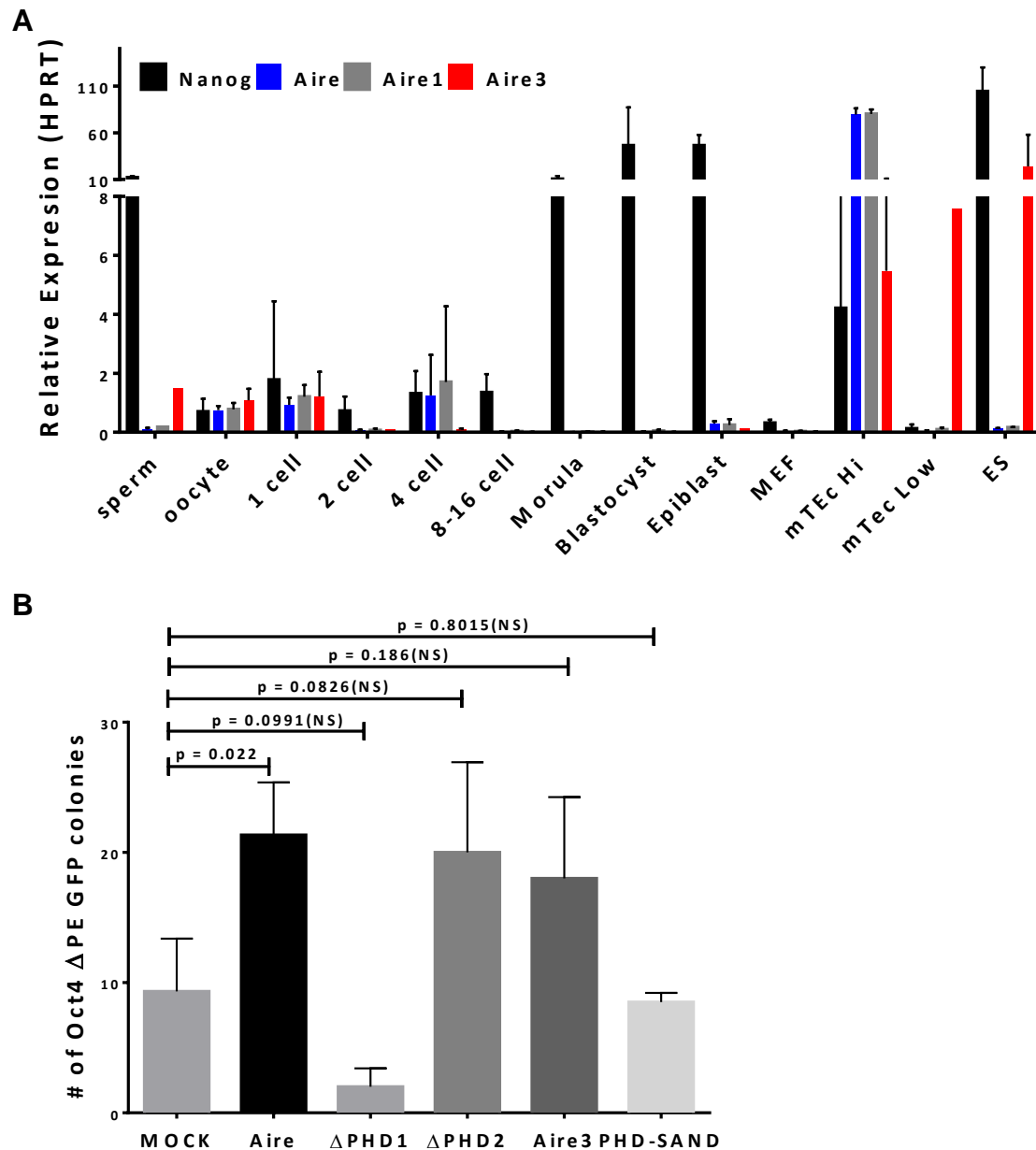**Figure 17. Aire3 expression in early development and reprogramming efficiency. A.** Quantitative PCR of Aire expression during early development compared to Nanog and Aire1 expression and normalized to HPRT. Error bars represent SD of 3 biological replicates. **B.** Graph showing that number of ΔPE-Oct4-GFP-positive colonies after 14 days of reprogramming analyzed be Student's *t*-test. Error bars represent SD (*n*=2).

Discussion

Genome-wide transcription is the earliest event in mammalian development, paving the way to the establishment of the crucial stage of pluripotency. Genome activation is a function of Pol II activity, associated with multiple factors that promote gene expression. Once transcription starts after the fertilization, the maternal RNA is degraded and RNA molecules are transcribed by the embryonic genome. The transcription occurs globally, including the expression of repressed primitive viral elements[48]. Understanding ZGA is important for discovering the unique mechanism of wide gene activation, which is an essential step during early development.

In this work, I present a novel role for Aire in promoting pluripotency during early embryonic development. Unlike in the thymus, where Aire is responsible for generating peripheral tissue antigens (PTAs) to educate T cells to self-immune tolerance, during early development Aire has a minor but important role in maintaining marinating gene expression levels required for development. Inactivation of Aire *in vivo* during embryogenesis results in diverse phenotypes, ranging from infertility to developmental defects and arrest. Aire-deficient B6 mice exhibited hypofertility, as 40% of females could not give birth. However, fertile females had normal rates of pregnancy and litter size. It was suggested that the fertility phenotype might be an attribute of the mild autoimmunity[18] in Aire KO B6 mice. Yet, although we verified the autoimmunity, it was found to be the same as in wild-type animals. To clear this issue, we conducted IVF experiments using B6 mice backcrossed to the 129 strain. Although we found a 15% decrease in blastocyst maturation and a reduction in transcribed RNA in Aire-deficient embryos, 129 mice did not exhibit a hypofertility phenotype. This strengthens the notion of a mechanistic role of Aire during early development. Nevertheless, it is evident that other yet unknown factors also have a role in genome activation, because complete ablation of Aire does not terminate development as efficiently as inhibition of Pol II activity.

In vertebrate models like zebrafish, where genome activation and the establishment of pluripotency happen simultaneously, the key factors for ZGA are crucial

for proper development. External fertilization and development of zebrafish embryos offers the advantage of exploring development without dependency on external signals. Injection of MO against *aire* to one-cell ZF embryos resulted in developmental defects in about 80% of the embryos. The variety of phenotypes suggests a mechanistic role for *aire* during early development, where the partial activation of certain genes impairs the development in different manners. Unbiased RNA-seq confirmed the distribution of phenotypes also at the transcription level. The *aire* KO model (*aire*$^{5ins/5ins}$) yielded a milder effect than the MO, resulting in 15% defective embryos. The accuracy of MO versus KO models is an ongoing debate[49]; however, similar trends in phenotypes observed using both models show that our results are reliable and overall consistent. We speculate that the influence of Aire on development could be orchestrated via Pol II activity. In addition, humans with APECED syndrome, which is caused by mutations in *AIRE*, exhibit defects that are associated with developmental abnormalities, rather than with an autoimmune reaction[27]. Our findings highlight that Aire may play a role in early development by regulating transcription during ZGA.

Another important aspect of pluripotency is its reactivation during somatic cell reprogramming. In contrast to a previous study[28], we show that the absence of Aire has no influence on maintaining pluripotency circuits and state. However, we and others[50] found that Aire expression is upregulated during somatic cell reprograming, supporting the activation of Aire at late stages of reprogramming, which are associated with high levels of transcription and Pol II activity. Although reprogramming rate without endogenous Aire is normal, Aire over-expression had a major impact (more than 2-fold increase) on the rate of colony generation and on maturation to pluripotency state. The use of several reprograming systems, including primary infection and different isogenic secondary MEFs, strengthens the claim that Aire can enhance reprogramming rate and induce proper and more robust reactivation of pluripotency.

Deciphering Aire mechanistic contribution to reprogramming requires different approaches that test Aire's unique activity. Because Aire can induce the transcription of up to 25% of the entire cell transcriptome, we assumed that it could induce reprogramming independently of the Yamanaka factors, by directly inducing their

downstream targets. Indeed, by testing substitution combinations we found that Aire has the ability to mechanistically replace Sox2 as a reprogramming factor, similarly to previously discovered genome activation factors that could replace Sox2[51].

There are four main domains that support Aire activity in the thymus[19]. Our generated Aire domain mutant library showed that PHD1- and SAND-deficient variants produced the same iPSC generation rate as in the mock control plasmid. It is plausible that PHD1 and SAND domains are required for Aire activity in reprogramming, as they act to bind epigenetically repressed regions and to recruit Pol II machinery, along with other transcription-associated proteins, whose recruitment is a key step in reprogramming, early development and immune self-tolerance maintenance. Aire binds Pol II to provoke a type of genome activation that is attributed to later stages of reprogramming, and to transcriptional regulation by BRD4 and CDK9, all of which also bind Aire[40,41]. Moreover, transcription initiation requires the breaking of DNA to unleash the poised Pol II. DNA break is associated with Aire activity and was also shown to promote transcription during early development. Artificially induced DNA breaks by UV irradiation generated higher rate of iPSC colony formation relative to WT cells, but less than Aire induced iPSCs, indicating that Aire induces transcription via DNA breaks and recruitment of Pol II.

In conclusion, our findings offer a new perspective on Aire function during early development and in somatic cell reprogramming. While Aire is crucial for maintaining immune self-tolerance, during early development and reprogramming, Aire function is mild. Genome activation will occur in the absence of Aire, but at the cost of reduced robustness of the process. Based on our findings, we propose a model of a cogwheel driving transcription (Fig. 18). When the cogwheel is running on two wheels, i.e., without Air contribution, the machinery cannot reach full capacity; however, when the cogwheel is running on all three wheels, it will generate proper transcription activity for either genome activation or the generation of PTAs. In other words, during early development and reprogramming, Aire acts as the third wheel that adds the required robustness to these fundamental developmental processes.
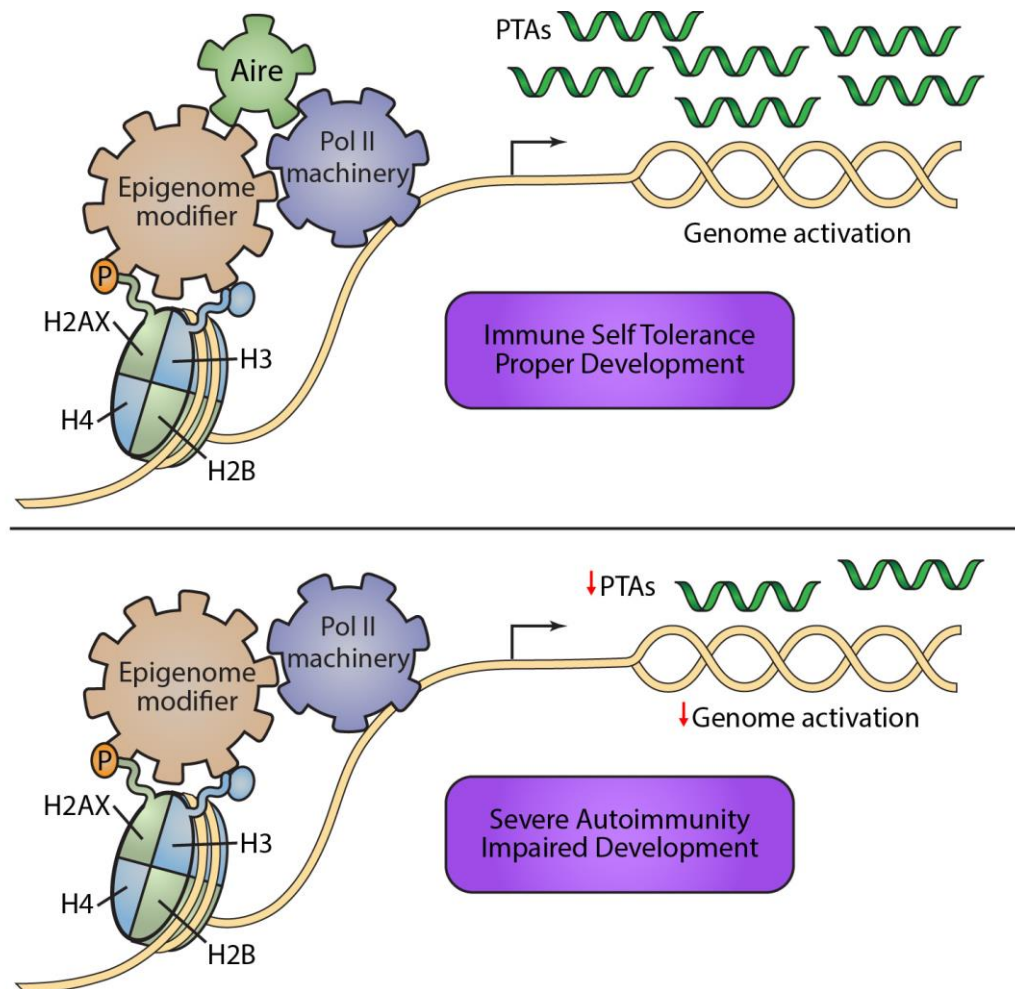
Figure 18. Proposed model for the roles of Aire during early development and in somatic cell reprogramming. When Aire is active (top), it completes a three-wheel cogwheel machinery driving genome activation and PTA formation. Without Aire, the mechanism still works but with reduced robustness, resulting in immune and developmental aberrations.

Methods

Stem cell lines and cell culture

Reprogramming and maintenance of murine pluripotent cells were conducted in two types of media. Serum/LIF conditions (also known as mouse metastable naïve conditions) contain 425 ml high-glucose DMEM (Invitrogen 41965), 15% USDA-certified and heat-inactivated fetal bovine serum (FBS) (Biological Industries), 1 mM L-glutamine (Biological Industries), 1% nonessential amino acids (Biological Industries), 0.1 mM β-mercaptoethanol (Invitrogen), penicillin/streptomycin (Biological Industries), sodium pyruvate (Biological Industries) and 20 ng/ml recombinant human LIF (produced in-house). 2i/LIF conditions (also known as mouse naïve ground state conditions) contain 240 mL DMEM/F12 (Invitrogen 21331), 240 mL Neurobasal Medium (Invitrogen 21103), 5 mL N2 supplement (Invitrogen 17502048), 5 mL B27 supplement (Invitrogen 17504-044), 1 mM glutamine (Biological Industries), 0.1 mM nonessential amino acids (Biological Industries), 0.1 mM β-mercaptoethanol (Sigma), penicillin-streptomycin (Biological Industries), 50 μg/mL BSA (GIBCO Fraction V 15260-037), 20 ng/ml recombinant human LIF and 2 small-molecule inhibitors (2i): CHIR99021 (GSK3i 3 μM, Axon Medchem 1386) and PD0325901 (MEKi 1 μM, Axon Medchem 1408).

For Naïve and Primed conversion, the naïve cells maintained in 2i/Lif conditions and the Primed were maintained in 8 ng/ml recombinant human FGF2 (Peprotech Asia), 20 ng/ml recombinant human Activin A (Peprotech) in addition to the base media (without 2i and LIF).

Cells were expanded in 20% O_2; 5% CO_2 at 37°C and passaged following single-cell trypsinization (0.25% Trypsin, Biological Industries) every 4-5 days. Mycoplasma contamination was monitored routinely by monthly tests with Mycoalert kit (LONZA).

Generation of Aire over-expressing cell lines

mAire was amplified from tetO-Aire (kindly provided by Jakub Abrammson) and cloned to PBS1-Flp-In vector. Targeting of NGFP1[52] (double targeting) and KH2 cell lines[53] was performed as described previously[44].

Reprogramming of somatic cells and viral production

To generate iPSCs and secondary somatic cells, mouse fibroblast reprogramming was conducted by plating 5000 cells per well in 6-well gelatin-coated plates. For generation of iPSCs with viral infection, VSVG-coated lentiviruses were generated in 2.5^10^{6}M 293T cells. Culture medium was changed 12 hours post-transfection and virus-containing supernatant was collected 60 to 72 hours post-transfection. Viral supernatant was filtered through 0.45 μm filters (Millipore). Virus-containing supernatants of the different reprogramming viruses (FUW-tetO-lox- hKLF4, FUW-

46

tetO-lox-hSOX2, FUW-tetO-c-Myc and FUW-tetO-hOct4 , FUW-tetO-mAIRE, STEMCAA[54], FUW-tetO-mAIRE ΔSAND, FUW-tetO-mAIRE ΔHSR, FUW-tetO-mAIRE ΔPHD1, FUW-tetO-mAIRE ΔPHD2, FUW-tetO-mAIRE1, FUW-tetO-SAND-PHD1) were supplemented with equal volumes of fresh culture medium for infection of 20000 cells per well in 6-well plate. After transfection, cells were cultured in Serum/Lif condition supplemented with 2 µg/ml doxycycline. iPSCs formation was quantified by fluorescence detection following immunostaining or AP staining. DNA damage-associated reprogramming started after treatment with etoposide (Sigma E1383) or UV radiation ($10 \, j/m^2$).

Nanog reactivation assay

Nanog-CreER reporter[45] MEFs (100,000 cells) were infected with FUW-AIRE, pLM-vexGFP-Oct4[55], pLM-mCitrine-Sox2[55], or FUW-GFP and analyzed by FACS for TdTomato-positive cells 9 days after infection.

Mouse chimera formation

Pluripotent stem cells (ESCs or iPSCs) were injected into BDF2 blastocysts that were initially harvested as single cells and grown for 3.5 days in KSOM media (Millipore) under mineral oil. Microinjection into blastocysts placed in M2 medium under mineral oil was done by a flat-tip pipette with an internal diameter of 1.2–1.5 µm, using a Piezo micro-manipulator. A controlled number of 10-12 cells were injected into the blastocyst cavity. 10-15 injected blastocysts were transferred to each uterine horn of 2.5 days post-coitum pseudo-pregnant females. Germ-line transmission was determined by mating chimeric animals with C57B/6 females and continuous inspection for agouti colored pups. All animal studies were conducted according to the guideline and following approval by the Weizmann Institute IACUC (approval 08810114-2).

RT-PCR analysis

Total RNA was isolated using TRIZOL (Invitrogen). 1µg RNA was reverse-transcribed using High-Capacity cDNA Reverse Transcription kit (Applied Biosystems) and re-suspended in 50 µl of water. Quantitative PCR analysis was performed in triplicates using 1/10 of the reverse transcription reaction on a Viia7 platform (Applied Biosystems). For analysis of embryos from oocyte to blastocyst stages, 50 embryos from each stage were collected, washed twice in PBS, transferred to RNase-Free PCR tube and frozen in liquid nitrogen. Embryos were thawed and lysed using 0.5% NP-40 in the presence of RNase inhibitors (Applied Biosystems). The lysates were incubated at 70^oC for 90 second, collected by centrifugation (7500g, 30 sec), and then reverse-transcribed using High-Capacity cDNA Reverse Transcription kit (Applied Biosystems).

Semi-quantitative PCR analysis was performed in triplicates using 1 ng of the reversed transcription reaction.

Gel electrophoresis and immunoblot analysis

SDS-PAGE was performed according to Laemmli protocol. Gels were electro-transferred onto nitrocellulose membranes for immunostaining. Protein-containing membranes were blocked with 5% non-fat dry milk and 0.1% Tween-20 in Tris-buffered saline, and then incubated with anti-Aire antibody 1:1000 (Upstate) and anti-HSP90 antibody 1:2000 (Epitomics), followed by incubation with appropriate HRP-conjugated secondary antibodies, and enhanced chemiluminescent (Pierce) was performed.

Immunocytoflourescence staining

Immunofluorescence was conducted after cells were fixed in 4% paraformaldehyde for 10 min at room temperature (RT), washed three times with PBS and blocked for 15 min with 5% FBS in PBS containing 0.1% Triton X-100. After incubation with primary antibodies against Oct4 (H-134, 1:200, Santa Cruz), Nanog (polyclonal rabbit 397A,1:200, Bethyl), Sox2 (polyclonal rabbit ,1:200, Millipore), SSEA1 (monoclonal mouse, 1:50, Developmental Studies Hybridoma Bank), Essrb (PP-H6705-00, 1:700, R&D), Aire (5h12, 1:200, eBioscience, 3h RT incubation), λH2X (JBW301, 1:200, milipore), and H3K14ac (A-4023, 1:50, EpiGentek) for 1 h in 2% FBS in PBS containing 0.1% Triton X-100, cells were washed three times with PBS and incubated with appropriate fluorophore-labelled secondary antibodies purchased from Jackson Immuno research. Specimens were analyzed on a Zeiss Z1 fluorescence microscope, and images were acquired with a Zeiss Axiocam HRM camera.

In vitro alkaline phosphatase staining of cells was performed with Vectastain ABC alkaline phosphatase kit (Vector labs) according to the manufacturer's instructions.

For immunohistochemistry of blastocysts, AIRE-GFP blastocysts were derived at 3.5 days after the appearance of vaginal plug following by mating. Blastocysts were fixed in freshly prepared 4% PFA in 0.1M PB (pH 7.2) at RT for 15 min. Afterwards, the tissue was washed three times at RT in PBS/PVP (3 mg/ml) for 5 min and permeabilized in a solution containing 0.25% Trition X-100 in PBS/PVP at RT for 30 minutes. Next, the samples were blocked in blocking solution (2% normal donkey serum, 0.1% BSA, 0.01% Tween 20 in PBS) for 1 h at RT. blastocysts were then incubated with the appropriate primary antibody diluted in blocking solution at 4°C overnight. Blastocysts were then washed three times (5 min each) in blocking solution, incubated with appropriate fluorochrome-conjugated secondary antibodies diluted in blocking solution at RT for 1 h in the dark, washed once in PBS, counter stained with DAPI (1 μg/ml, 0215754, MP

Biomedicals) for 1 min, rinsed three times in PBS and sequentially mounted in 25% 50%, 75% and 100% with Shandon Immu-Mount (Thermo Scientific, 9990412). The blastocysts were mounted on a homemade slide and covered with coverslip. All images were obtained using a Zeiss (Oberkochen, Germany) Axioscope microscope and LSM710 confocal microscope and processed with ImageJ software (NIH). The following primary and secondary antibodies were used: for immunofluorescence, mouse anti-Oct4 (1:100, 611203, BD Transduction Laboratories), rabbit anti-GFP (1:1000, A-11122, Invitrogen), Secondary antibodies conjugated with Alexa488 and Rodamine Red-X (1:150) were from Jackson ImmunoResearch Laboratories.

For immunohistochemistry of mouse thymus, thymi were embedded in OCT compound (Tissue-Tek, Sakura Finetek) and frozen on dry ice. Cryostat sections (10 μm-thick) were fixed with ice-cold acetone for 10 min and incubated with primary antibody for 60 min at RT. Sections were washed three times with PBS and incubated with secondary antibody for 60 min at RT. DAPI staining was performed following secondary staining for 10 min at RT and followed by three washes with PBS. All antibodies were diluted in 0.5% BSA in PBS: Keratin-5 (PRB-160P, 1:100, Covance).

For click-it staining (Click-it RNA Alexa Fluor 594 imaging kit C10330, Molecular Probes), two-cell-stage (30 hpf) mouse embryos were incubated with Eu for 3 hours and then processed according to the manufacturer's protocol. All images were obtained using a Zeiss (Oberkochen, Germany) Axioscope microscope and LSM710 confocal microscope and processed with ImageJ software (NIH). The incorporation rate was calculated by dividing the nuclear intensity with the cytoplasmic intensity.

Teratoma assay

For teratoma formation and analysis, AMKO iPSCs were collected by trypsinization before injection. Cells were injected subcutaneously into NSG mice (Jackson laboratories). Animals were euthanized before tumor size exceeded 1.5 cm in diameter. All animal studies were conducted according to the guideline and following approval by the Weizmann Institute IACUC (approval 09240214-2). Tumor mass was extracted and fixed in 4% paraformaldehyde overnight. Slides were prepared from the paraffin embedded fixed tissue, which were next stained with hematoxylin and eosin (H&E) and inspected for representation of all three germ layers.

IVF procedure

The IVF procedure was performed as described in Takeo et al.[56], except for the replacement of M2 media with KSOM media. Fertilized oocytes were identified by cumulus removal after 3 hours of incubation with sperms.

SMART-seq

Two-cell stage embryos were cryo-preserved in lysis buffer and SMART-seq[57] libraries were prepared at the Israel National Center for Personalized Medicine (INCPM), Weizmann Institute of Science, according to manufacturer's instructions for the SMART-seq Kit.

RNA-seq library preparation

Total RNA was extracted using DirectZol RNA mini-prep (Zymo research, R2052). Libraries were prepared according to Illumina's instructions accompanying the TruSeq RNA Sample Preparation Kit v2 (RS-122-2001). Sequencing was carried out on Illumina HiSeq2500 according to the manufacturer's instructions, using 10 pM template per sample for cluster generation, and sequencing kit V2 (Illumina), resulting in ~40 million paired-end reads per sample.

RNA-seq analysis

The paired-end reads of both polyadenylated (poly (A)) RNA and SMART-seq were aligned to mouse genome version mm10 with TopHat2 aligner (v2.0.8b)[58], using default input parameters. The number of reads mapped to each of the Ensemble genes (version GRCm38.74) was counted using the HTSeq python package[59] with default parameters. IGV visualization file was generated by IGV Tools[60]. Normalized counts were calculated for each gene using DESeq package[61]. DESeq produces counts that are normalized by the sample size, but not by the length of the transcripts. Genes were considered differentially expressed if their FC≥2. Correlation of normalized expression levels was performed by MATLAB.

3' poly (A) RNA sequencing library preparation

Total RNA was extracted from single zebrafish embryos using TRIZOL (Invitrogen) and GlycoBlue (Ambion). 200 ng RNA was fragmented into average size of 300 nucleotides by incubation at 95°C for 4.5 min (NEBNext Magnesium RNA Fragmentation Module). The 3' polyadenylated fragments were enriched by selection on poly dT beads (Dynabeads, Invitrogen). Strand-specific cDNA was synthesized using a poly T-VN oligo (18 T) and Affinity Script RT enzyme (Agilent). DNA was purified by adding paramagnetic SPRI beads (Agencourt AMPure XP, Beckman Coulter) pipette-mixed 15 times and incubated for 2 min. Supernatants were separated from the beads using a 96-well magnet for 4 min. Beads were washed on the magnet with 70% ethanol and then air-dried for 4 minutes. The DNA was eluted in EB buffer (10 mM Tris-HCl, pH 8.0) by pipette mixing 25 times. For the remainder of the library construction process (DNA end-repair, A-base addition, adaptor ligation and enrichment), a general SPRI cleanup involved addition of buffer containing 20% PEG and 2.5 M NaCl to the DNA reaction. DNA ends were first repaired by T4 polymerase (NEBNext); next, a phosphate group was added

at the 5′ ends by T4 polynucleotide kinase (NEB). An adenosine base was then added to the blunt-ended fragments, using Klenow enzyme (NEB), and a barcode Illumina-compatible adaptor (IDT) was ligated to each fragment using T4 quick ligase (NEB). DNA fragments were amplified by 12 cycles of PCR (Kapa HiFi) using specific primers (IDT) to the ligated adaptors. Library quantity was determined by Qubit Fluorometric Quantitation (Life Technologies) and the quality of each library was analyzed by TapeStation (Agilent).

3' poly (A) RNA sequencing Analysis

3' poly (A) RNA-seq reads were aligned to mouse genome version Zv9 with BOWTIE2 software[62] using default parameters. Gene expression levels were estimated as follows: each gene was divided to 50-bp bins, from its transcription start site to 1 kb after its transcription end site. Genomic locations were downloaded from UCSC Table Browser (refGene, https://genome.ucsc.edu/cgi-bin/hgTables). The reads in each bin were counted, and the bin with the highest number of reads was chosen as the representative expression level of that gene. If there are several alternative transcription end sites for a specific gene, only the transcript with the highest expression was counted. These expression levels were normalized by the sample size, such that each level represents RPM (reads per million reads), and the minimal level was set to 1. Hierarchical clustering was calculated for normalized reads by MATLAB.

Analysis of autoantibodies

For the analysis of autoantibodies, serum was collected from 8-12 months of age Aire KO / Het and WT mice on a B6.BCA or 129sv background and from 16-week-old Aire KO mice on a NOD background as a control. The serum was collected with heparin (Sigma H3149-50su) and centrifuged for 10 min at 14000 RPM; the supernatant was kept for further analysis. Autoantibodies were detected by slot-blot analysis with lysates of tissues from NSG mice and serum (1:200 dilutions).

Southern blot analysis

Genomic DNA was extracted from each neomycin/hygromycin resistant targeted sub-clone. 10-15 µg of genomic DNA was digested with PstI or SpeI restriction enzyme for 5 hours and separated by gel electrophoresis. The DNA was transferred to a nitrocellulose membrane that was next hybridized with a radioactive-labeled probe and developed using ECL (Thermo Scientific).

Evolution comparison

All evolutionary conservation analyses were performed by STRING[63].

Zebrafish maintenance

Fish husbandry and maintenance was carried out as previously described[64]. Experiments were carried out using the transgenic line Tg(fli1:eGFP)[y1], which has been previously described[65] and also Fli:dsRED X sox10:GFP, fli-eGFP X Gata dsRed. For pol II activity assay, serial dilutions of α-amanitin (Sigma A2263) were added to blue water in 12-well plate with ~20 embryos per well.

Morpholino injection

Morpholino oligonucleotides were purchased from Gene Tools, suspended in distilled water and injected with injection buffer (0.025% phenol red, 0.12 mM KCl, 20 mM HEPES, pH 7.4) into the yolk of one- to two-cell stage embryos. *aire* morpholino (12 ng): 5' CTTCAAAACTCTCCACCTTAGACAT 3'; *nanog* morpholino (8 ng): 5' CTGGCATCTTCCAGTCCGCCATTTC 3'.

Zebrafish knockout model generation via CRISPR-Cas9

Unique gRNA sequences were chosen with the help of the Zhang Lab (website: http://www.genome-engineering.org/) to target ex.1 of *aire*. Oligos encoding gRNAs targeting these genes were inserted into px330 vector and were amplified with T7 promoter, PCR product was *in vitro* transcribed with MEGAshortscript T7 Transcription KitT7 shortscript *in vitro* transcription kit (Ambion) and purified with MegaClear purification kit (Ambion). 100 ng gRNA and Cas9 protein were injected into one-cell embryos, which then underwent CRISPR-Cas9 cut validation by HRM. The embryos were grown for 3 months and crossed with WT fish in order to find germ-line mutations, which were verified by HRM and Sanger sequencing.

Zebrafish genomic DNA isolation

Genomic DNA was isolated from embryos or from the fin of adult fish. For fin clipping, adult fish (3-4 months) were anesthetized with 15 ml tricaine/333 ml blue water. Fin tissue from 15 dechorionated embryos was suspended in extraction buffer (X100: 5 gr NaOH, 0.37 gr disodium EDTA, DDW up to 50 ml, pH 12.0) and incubated at 95°C for 1 h. Neutralization was achieved by addition of equal volume of neutralization buffer (X10: 3.152 gr Tris-HCl, DDW up to 50 ml, pH 5.0). Samples were vortexed briefly, centrifuged at maximum speed for 1 min and stored at -20°C.

Table 2: Primers

Gene Name	Forward sequence	Reverse sequence	
mAire	TaqMan probe		Mm00477461_m1
mNanog	TaqMan probe		Mm02384862_g1
mGapdh	TaqMan probe		Mm99999915_g1
mHprt	TaqMan probe		Mm00446968_m1
mAire ex12	GATCGCGGTGGCCATAGA	GGAGCGTCTCCTGGAACTTG	
mActin	GCGGAGGGAAGCTTTAGAGC	TGGTCGATGGTTTCCAGC	
mPRM1	CACCTTATGGTGTATGAGCGG	GTCTCAGGCGAAGATGTC	
zfRp113	TCTGGAGGACTGTAAGAGGTATGC	AGACGCACAATCTTGAGAGCAG	
zfTubulin	TGGAGCCCACTGTCATTGATG	CAGACAGTTTGCGAACCCTATCT	
zfAire	AAATCCTTAACATCCAGCTC	CCATCTGTCTTCTTCATCAG	
zfAire1	GCTGTAGAGCACAATGATGATGAGT	GGACACCCATCACAGCAGAT	
zfAire2	ATGTCTAAGGTGGAGAGTTTTGAAGAG	GTGAACTTCACTTGGG	
zfAire3	ACAATGATGATGAGTGTGCAGTG	CATGTACCCCTGGGTATGGA	
zfAire4	AGCTCTGGAAGCAAAGGCAA	CAGATACCGTCACAGCCCTC	
zfNanog	GATGACTGAGAGATATGTTGTC	GAGGGTTAATGTAGTGGTCT	
zfNanog1	ACGCTCCCCAACACAGG	GGGAACTCATAGTGCGTCACA	
zfFoxn1	GGAGGAGCATGGCCCGACCAG	GGCTGCTGGGCAGACTGAAGTGAC	
mGapdh	TTCACCACCATGGAGAAGGC	CCCTTTTGGCTCCACCCT	
mNanog	CCTCCAGCAGATGCAAGAACTC	CTTCAACCACTGGTTTTTCTGCC	
mEsrrb	AAAGCCATTGACTAAGATCG	AATTCACAGAGAGTGGTCAG	
mFgf5	CATGCAAGTGCCAAATTTAC	CTGTCTTTTCAGTTCTGTGG	

zf Aire gRNA ex1	CACCGTTGTCGCACTGAGATCGCGA	AAACTCTCAGTGCGACAAC	
zf Aire ex1 HRM	GTAGTTGTGTTTATACAGC	AGATGACGTTGTGATCTG	
zf Aire ex1 sequencing	TTATACATTGTGTGATACAA	TATTGCATTTAAAATGATA	
zf Aire 5ins WT - genotype	CTTGTCGCACTGAGATCGCGA	AGTGCTGGGCAAAAATTAATCGTGA	
zf Aire 5ins mutant – genotype	GTCGCACTGAGATCGCAATGC		
Aire KO – genotype	AAGCCGTCCAGGATGCTAT - WT	GTCATGTTGACGGATCCAGGGTAGAAAGT - KO	AGACTAGGTGTTCCCTCCCAACCTCAG - COMMON
5' Arm SB probe	GGCCCCCGGGTAAGTTGCAT	GGACCAGGGGGACCAATGGG	
3' Arm SB probe	CTAGATAAATTTTCCTTTCA	TGCAGACCTGAGGCCTGGGA	
cMyc integration	AAAGTGAAAGTCGAGCTCGGTACC (TetO For)	ACTGAGGGGTCAATGCACTCGG	
Oct4 integration		CCTTCTCCAACTTCACGGCATT	
Sox2 integration		GCCTCCGGGAAGCGTGTACTTA	
Klf4 integration		CCTGGTGGGTTAGCGAGTTGGA	
Mutant library	TATAATTAAGGCGCGCCATGGCAGGTGGGGATGGAATGC	GGAATTAACGCTAGCTCAGGAAGAGAAGGGTGGTGTCTC	TATAATTAAGGCGCGCCATGCTAAACCAGTCCCGGAA (HSR)
mAire1	TaqMan probe		Mm01149685_m1
mAire3	TaqMan probe		Mm01149687_m1

List of publications during the course of the PhD studies

Sheinboim D , Maza I, Dror I, **Krupalnik V** , Bell R , Zviran A , Suita Y , Hakim O , Gutfreund YM , Khaled M, Hanna JH and Levy C. *OCT4 impedes cell fate redirection by the melanocyte lineage master regulator MITF*, Submitted.

Weiner A , Lara-Astiaso D, **Krupalnik V** ,Gafni O, David E ,Winter DR , Hanna JH , Amit I. *Genome-wide characterization of histone mark co-occurrence at single molecule resolution,* Nat Biotechnol. 2016 Sep;34(9):953-61. doi: 10.1038/nbt.3652.

De Los Angeles A, Ferrari F, Fujiwara Y, Mathieu R, Lee S, Lee S, Tu HC, Ross S, Chou S, Nguyen M, Wu Z, Theunissen TW, Powell BE, Imsoonthornruksa S, Chen J, Borkent M, **Krupalnik V**, Lujan E, Wernig M, Hanna JH, Hochedlinger K, Pei D, Jaenisch R, Deng H, Orkin SH, Park PJ, Daley GQ. *Failure to replicate the STAP cell phenomenon*. Nature. 2015 Sep 24;525(7570):E6-9

Maza I, Caspi I, Zviran A, Chomsky E, Rais Y, Viukov S, Geula S, Buenrostro JD, Weinberger L, **Krupalnik V**, Hanna S, Zerbib M, Dutton JR, Greenleaf WJ, Massarwa R, Novershtern N, Hanna JH. *Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors.* Nat Biotechnol. 2015 Jul;33(7):769-74. doi: 10.1038/nbt.3270. Epub 2015 Jun 22.

Nicenboim J, Malkinson G, Lupo T, Asaf L, Sela Y, Mayseless O, Gibbs-Bar L, Senderovich N, Hashimshony T, Shin M, Jerafi-Vider A, Avraham-Davidi I, **Krupalnik V**, Hofi R, Almog G, Astin JW, Golani O, Ben-Dor S, Crosier PS, Herzog W, Lawson ND, Hanna JH, Yanai I, Yaniv K.. *Lymphatic vessels arise from specialized angioblasts within a venous niche* Nature. 2015 Jun 4;522(7554):56-61.

Geula S, Moshitch-Moshkovitz S, Dominissini D, Mansour AA, Kol N, Salmon-Divon M, Hershkovitz V, Peer E, Mor N, Manor YS, Ben-Haim MS, Eyal E, Yunger S, Pinto Y, Jaitin DA, Viukov S, Rais Y, **Krupalnik V**, Chomsky E, Zerbib M, Maza I, Rechavi Y, Massarwa R, Hanna S, Amit I, Levanon EY, Amariglio N, Stern-Ginossar N, Novershtern N, Rechavi G, Hanna JH. *Stem cells. m^6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation.* Science. 2015 Feb 27;347(6225):1002-6. doi: 10.1126/science.1261417. Epub 2015 Jan 1.

Krupalnik V, Hanna JH. *Stem cells: The quest for the perfect reprogrammed cell*. Nature. 2014 Jul 10;511(7508):160-2. doi: 10.1038/nature13515. Epub 2014 Jul 2.

Gafni O, Weinberger L, Mansour AA, Manor YS, Chomsky E, Ben-Yosef D, Kalma Y, Viukov S, Maza I, Zviran A, Rais Y, Shipony Z, Mukamel Z, **Krupalnik V**, Zerbib M, Geula S, Caspi I, Schneir D, Shwartz T, Gilad S, Amann-Zalcenstein D, Benjamin S, Amit I, Tanay A, Massarwa R, Novershtern N, Hanna JH. *Derivation of novel human ground state naive pluripotent stem cells.* Nature. 2013 Dec 12;504(7479):282-6. doi: 10.1038/nature12745. Epub 2013 Oct 30.

Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, Mansour AA, Caspi I, **Krupalnik V**, Zerbib M, Maza I, Mor N, Baran D, Weinberger L, Jaitin DA, Lara-Astiaso D, Blecher-Gonen R, Shipony Z, Mukamel Z, Hagai T, Gilad S, Amann-Zalcenstein D, Tanay A, Amit I, Novershtern N, Hanna JH. *Deterministic direct reprogramming of somatic cells to pluripotency*. Nature. 2013 Oct 3;502(7469):65-70. doi: 10.1038/nature12587. Epub 2013 Sep 18.

Mansour AA, Gafni O, Weinberger L, Zviran A, Ayyash M, Rais Y, **Krupalnik V**, Zerbib M, Amann-Zalcenstein D, Maza I, Geula S, Viukov S, Holtzman L, Pribluda A, Canaani E, Horn-Saban S, Amit I, Novershtern N, Hanna JH. *The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming.* Nature. 2012 Aug 16;488(7411):409-13. doi: 10.1038/nature11272.

References

1. Daley, G. Q. Customized human embryonic stem cells. *Nat. Biotechnol.* **23,** 826–8 (2005).
2. Lerou, P. H. & Daley, G. Q. Therapeutic potential of embryonic stem cells. *Blood Rev.* **19,** 321–31 (2005).
3. Silva, J. & Smith, A. Capturing pluripotency. *Cell* **132,** 532–6 (2008).
4. Agarwal, S., Lensch, M. W. & Daley, G. Q. Current prospects for the generation of patient-specific pluripotent cells from adult tissues. *Regen. Med.* **2,** 743–52 (2007).
5. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–76 (2006).
6. Dimos, J. T. *et al.* Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* **321,** 1218–21 (2008).
7. Hanna, J. H., Saha, K. & Jaenisch, R. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* **143,** 508–25 (2010).
8. Fu, X. & Xu, Y. Challenges to the clinical application of pluripotent stem cells: towards genomic and functional stability. *Genome Med.* **4,** 55 (2012).
9. Maekawa, M. *et al.* Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. *Nature* **474,** 225–9 (2011).
10. Nichols, J. & Smith, A. Pluripotency in the embryo and in culture. *Cold Spring Harb. Perspect. Biol.* **4,** a008128 (2012).
11. Li, L., Lu, X. & Dean, J. The maternal to zygotic transition in mammals. *Mol. Aspects Med.* **34,** 919–38 (2013).
12. Zeng, F. & Schultz, R. M. RNA transcript profiling during zygotic gene activation in the preimplantation mouse embryo. *Dev. Biol.* **283,** 40–57 (2005).
13. Wu, L., Li, L., Zhou, B., Qin, Z. & Dou, Y. H2B ubiquitylation promotes RNA Pol II processivity via PAF1 and pTEFb. *Mol. Cell* **54,** 920–31 (2014).
14. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141,** 432–45 (2010).
15. Xue, Z. *et al.* Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500,** 593–7 (2013).
16. Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20,** 1131–9 (2013).
17. Mathis, D. & Benoist, C. Aire. *Annu. Rev. Immunol.* **27,** 287–312 (2009).
18. Jasti, S. *et al.* The autoimmune regulator prevents premature reproductive

senescence in female mice. *Biol. Reprod.* **86,** 110 (2012).

19.   Zumer, K., Saksela, K. & Peterlin, B. M. The mechanism of tissue-restricted antigen gene expression by AIRE. *J. Immunol.* **190,** 2479–82 (2013).

20.   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,** 13016–21 (2010).

21.   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,** 1833–8 (2013).

22.   Kyewski, B. & Peterson, P. Aire, master of many trades. *Cell* **140,** 24–6 (2010).

23.   Guerau-de-Arellano, M., Mathis, D. & Benoist, C. Transcriptional impact of Aire varies with cell type. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 14011–6 (2008).

24.   Derbinski, J., Pinto, S., Rösch, S., Hexel, K. & Kyewski, B. Promiscuous gene expression patterns in single medullary thymic epithelial cells argue for a stochastic mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 657–62 (2008).

25.   Abramson, J., Giraud, M., Benoist, C. & Mathis, D. Aire's partners in the molecular control of immunological tolerance. *Cell* **140,** 123–35 (2010).

26.   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,** 535–40 (2012).

27.   Anderson, M. S. & Su, M. A. AIRE expands: new roles in immune tolerance and beyond. *Nat. Rev. Immunol.* **advance on,** (2016).

28.   Gu, B. *et al.* Aire regulates the expression of differentiation-associated genes and self-renewal of embryonic stem cells. *Biochem. Biophys. Res. Commun.* **394,** 418–23 (2010).

29.   Nishikawa, Y. *et al.* Biphasic Aire expression in early embryos and in medullary thymic epithelial cells before end-stage terminal differentiation. *J. Exp. Med.* **207,** 963–71 (2010).

30.   Mamo, S., Gal, A. B., Bodo, S. & Dinnyes, A. Quantitative evaluation and selection of reference genes in mouse oocytes and embryos cultured in vivo and in vitro. *BMC Dev. Biol.* **7,** 14 (2007).

31.   Gardner, J. M. *et al.* Deletional tolerance mediated by extrathymic Aire-expressing cells. *Science* **321,** 843–7 (2008).

32.   Hamatani, T., Carter, M. G., Sharov, A. A. & Ko, M. S. H. Dynamics of global gene expression changes during mouse preimplantation development. *Dev. Cell* **6,** 117–31 (2004).

33.   Jiang, W., Anderson, M. S., Bronson, R., Mathis, D. & Benoist, C. Modifier loci condition autoimmunity provoked by Aire deficiency. *J. Exp. Med.* **202,** 805–15 (2005).

34.   DeVoss, J. *et al.* Spontaneous autoimmunity prevented by thymic expression of a single self-antigen. *J. Exp. Med.* **203,** 2727–35 (2006).

35.   Lee, M. T. *et al.* Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503,** 360–4 (2013).

36.   Leichsenring, M., Maes, J., Mössner, R., Driever, W. & Onichtchouk, D. Pou5f1

transcription factor controls zygotic gene activation in vertebrates. *Science* **341,** 1005–9 (2013).

37. Svoboda, P. & Flemr, M. The role of miRNAs and endogenous siRNAs in maternal-to-zygotic reprogramming and the establishment of pluripotency. *EMBO Rep.* **11,** 590–7 (2010).

38. Anderson, M. S. *et al.* Projection of an immunological self shadow within the thymus by the aire protein. *Science* **298,** 1395–401 (2002).

39. Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151,** 1617–32 (2012).

40. Yoshida, H. *et al.* Brd4 bridges the transcriptional regulators, Aire and P-TEFb, to promote elongation of peripheral-tissue antigen transcripts in thymic stromal cells. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E4448-57 (2015).

41. Liu, L. *et al.* Transcriptional pause release is a rate-limiting step for somatic cell reprogramming. *Cell Stem Cell* **15,** 574–88 (2014).

42. Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502,** 65–70 (2013).

43. Wernig, M. *et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nat. Biotechnol.* **26,** 916–24 (2008).

44. Beard, C., Hochedlinger, K., Plath, K., Wutz, A. & Jaenisch, R. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* **44,** 23–8 (2006).

45. Maza, I. *et al.* Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors. *Nat. Biotechnol.* **33,** 769–774 (2015).

46. Abramson, J. & Goldfarb, Y. AIRE: From promiscuous molecular partnerships to promiscuous gene expression. *Eur. J. Immunol.* **46,** 22–33 (2016).

47. Ruan, Q.-G., Wang, C.-Y., Shi, J.-D. & She, J.-X. Expression and Alternative Splicing of the Mouse Autoimmune Regulator Gene (Aire). *J. Autoimmun.* **13,** 307–313 (1999).

48. Schoorlemmer, J., Pérez-Palacios, R., Climent, M., Guallar, D. & Muniesa, P. Regulation of Mouse Retroelement MuERV-L/MERVL Expression by REX1 and Epigenetic Control of Stem Cell Potency. *Front. Oncol.* **4,** 14 (2014).

49. Rossi, A. *et al.* Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* **524,** 230–3 (2015).

50. Koche, R. P. *et al.* Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell* **8,** 96–105 (2011).

51. Shinagawa, T. *et al.* Histone variants enriched in oocytes enhance reprogramming to induced pluripotent stem cells. *Cell Stem Cell* **14,** 217–27 (2014).

52. Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133,** 250–64 (2008).

53. Hochedlinger, K., Yamada, Y., Beard, C. & Jaenisch, R. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* **121,** 465–77 (2005).

54. Sommer, A. G. *et al.* Generation of human induced pluripotent stem cells from

peripheral blood using the STEMCCA lentiviral vector. *J. Vis. Exp.* (2012). doi:10.3791/4327

55.   Papapetrou, E. P. *et al.* Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 12759–64 (2009).

56.   Takeo, T. & Nakagata, N. Reduced glutathione enhances fertility of frozen/thawed C57BL/6 mouse sperm after exposure to methyl-beta-cyclodextrin. *Biol. Reprod.* **85,** 1066–72 (2011).

57.   Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10,** 1096–8 (2013).

58.   Trapnell, C. & Salzberg, S. L. How to map billions of short reads onto genomes. *Nat. Biotechnol.* **27,** 455–457 (2009).

59.   Anders, S., Pyl, P. T. & Huber, W. HTSeq – A Python framework to work with high-throughput sequencing data HTSeq – A Python framework to work with high-throughput sequencing data. *Bioinformatics.* **31,** 0–5 (2014).

60.   Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29,** 24–6 (2011).

61.   Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).

62.   Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9,** 357–359 (2012).

63.   Franceschini, A. *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41,** D808-15 (2013).

64.   Yaniv, K. *et al.* Live imaging of lymphatic development in the zebrafish. *Nat. Med.* **12,** 711–6 (2006).

65.   Lawson, N. D. & Weinstein, B. M. In vivo imaging of embryonic vascular development using transgenic zebrafish. *Dev. Biol.* **248,** 307–18 (2002).

Student declaration

I hereby declare that the thesis presented summarizes my independent research work under the supervision of Dr. Jacob H. Hanna at the Department of Molecular Genetics, the Weizmann Institute of Science. Mirie Zerbib conducted microinjections in mouse blastocysts. Shir Nevo and Yael Goldfarb from Jakub Abramson's lab collaborated with me on Aire3.

Acknowledgements

I would like to dedicate this work to the beloved memory of David (Dadu) Baruh,

a great friend who set me on my path.