

## RESEARCH ARTICLE

# *In silico* screening system based on a transcription factors regulatory network only using transcriptomic data

Tadaaki Nakajima<sup>1,2\*</sup>, Kentaro Harada<sup>2</sup>, Yasuhiro Tomooka<sup>2</sup>, Tomomi Sato<sup>1,3</sup>

**1** Department of Science, Yokohama City University, Yokohama, Japan, **2** Department of Biological Science and Technology, Faculty of Industrial Science and Technology, Tokyo University of Science, Tokyo, Japan, **3** Graduate School of Nanobioscience, Yokohama City University, Yokohama, Japan

\* [nakajima.tad.uu@yokohama-cu.ac.jp](mailto:nakajima.tad.uu@yokohama-cu.ac.jp)

## Abstract

In this study, we developed a method to identify core transcription factors (TFs) involved in differentiation using only comprehensive gene analysis. The theory of *in silico* screening using TFs regulatory network analysis (ISNA) required the following requirements: (1) estimating promoter regions, (2) constructing TFs regulatory network (TRN) relationships using the nucleotide sequence information in the promoters and score matrices derived from TF consensus sequences, and (3) identifying candidate core TFs by determining dissociation constants ( $K_d$  values) within the relationships of TRN. ISNA demonstrated the ability to predict the core TFs involved in the endothelial-to-mesenchymal transition of human umbilical vein endothelial cell (HUVEC) and the differentiation of human embryonic stem cells into mesodermal cells. Using ISNA, we identified HMGA2 as a novel core TF in uterine epithelium. Notably, HMGA2 expression was predominantly detected in uterine epithelium, where it regulated cell proliferation in response to estrogen. These findings highlight ISNA's potential to identify core TFs based on transcriptomic data.

## OPEN ACCESS

**Citation:** Nakajima T, Harada K, Tomooka Y, Sato T (2025) *In silico* screening system based on a transcription factors regulatory network only using transcriptomic data. PLoS ONE 20(4): e0319971. <https://doi.org/10.1371/journal.pone.0319971>

**Editor:** Michael Schubert, Laboratoire de Biologie du Développement de Villefranche-sur-Mer, FRANCE

**Received:** October 23, 2024

**Accepted:** February 12, 2025

**Published:** April 7, 2025

**Copyright:** © 2025 Nakajima et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** All numerical simulations were performed in MATLAB R2022a. Simulation parameter values can be found in the Figure legends. For estimation of open chromatin regions, the ChIP-seq data for all histone, DNase-seq, and ATAC-seq in HUVEC, hESC H1 and H9, and endometrial epithelial cells in H.sapiens (hg38) and mESC in M. musculus (mm10) was obtained using

## Introduction

Various cellular transcriptional regulation and signaling are intricately interconnected, yet the differentiation steps in development remains remarkably stable and drug responsiveness is consistently maintained. To bridge the long gap between the complexity of transcriptional regulation and the stability of phenotype expression, network regulations have been investigated. Networks with directed graphs, such as neural networks, are known for their robustness [1], where overall changes less significantly than individual inputs. Further, small changes in critical nodes can drastically reorganize the network, producing diverse outputs and facilitating cellular evolution. Gene networks modeled as directed graphs have been structured as some cellular models [2,3], demonstrating that subtle changes in the activity of a few core transcription factors (TFs) within a gene regulatory network can reorganize the networks.

Most current networks are built using circuit-like models, which define relationships between genes and assign weighting factors based on experimental findings. Thus, enormous amount of data for each relationship and weighting factor is required to build such models.

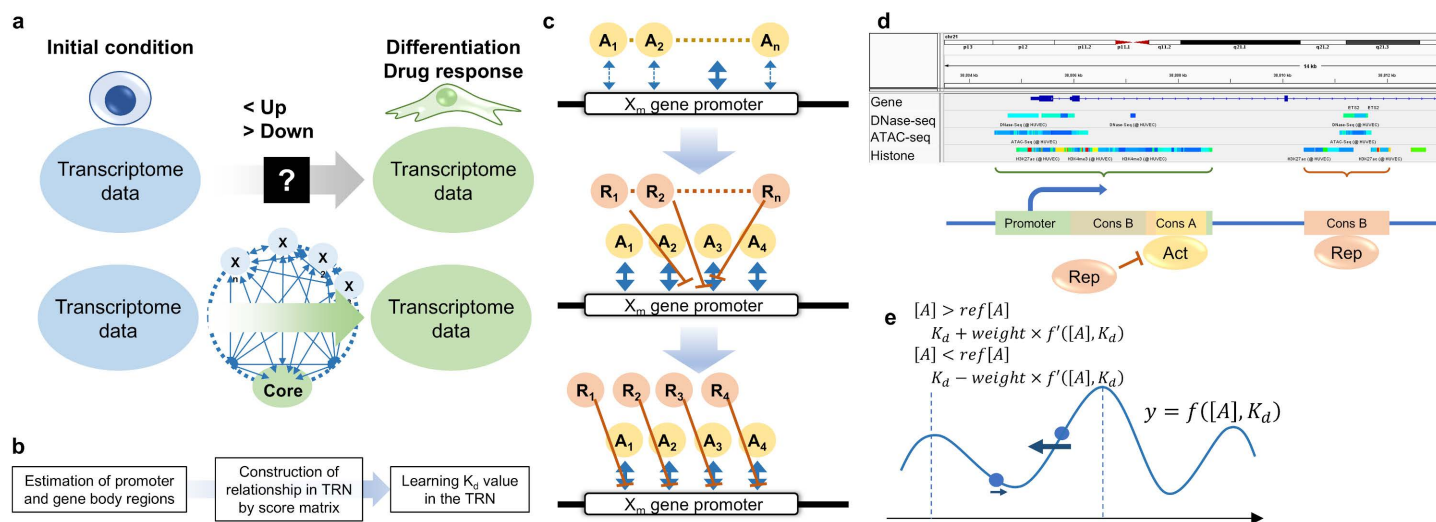
ChIP-Atlas (<https://chip-atlas.org/>). Initial conditions for the  $K_d$  values were considered as 1, and initial and reference values for the concentration of TF were set as RNA-seq data of HUVEC [doi:10.1084/jem.20182151], hESC [doi:10.1016/j.stem.2016.06.018], mESC [doi:10.1016/j.stem.2018.07.001], and endometrial epithelial cells [doi:10.1210/clinem.dgz117]. The corresponding programming code is available at: <https://github.com/Tadaaki-NAKAJIMA/ISNA/> and the protocol for usage of the program is uploaded as the read me file. Data is also located within the [Supporting Information](#) files.

**Funding:** This research was partly supported by MEXT/JSPS KAKENHI to T.N. (JP19K16177, 21K06244, and 24K09517, MEXT/JSPS KAKENHI to T.S. (23H02510). The funders did not play any role in the study.

**Competing interests:** The authors declare no competing interests.

Linkage logic theory was developed to identify the core factors in gene networks “without” relying on weighting factors [4,5]. Circuit-like models treat gene networks as computer-like system and fail to reflect the functional roles of the proteins transcribed from each gene. While weighting factors in network structures may be estimable, relationship information is still needed. Single-cell RNA-seq combined with knockout experiments (e.g., using CRISPER) has revealed relationships of gene regulatory networks in cell-cell communication [6]; however, numerous knockout experiments for each gene are needed to construct one network. Efforts in the bioinformatics have enabled estimation of gene regulatory network using binding motifs of TFs and ChIP-seq data for chromatin or TFs [7–12]. However, the estimation of gene regulatory network required an input network structure (direction of TF regulation of in whole gene expression) derived from literature or databases [13]. This reliance limits the ability to optimize networks for less-studied cell types and phenotypes where ChIP-seq data for key TFs remains incomplete.

To identify of novel candidate genes and signaling pathways involved in differentiation or drug responses in minor cell types and phenotypes, *de novo* comprehensive gene analysis like RNA-seq has been used. While comprehensive gene expression data provides a snapshot at the initial and final states, the intermediate signaling pathways remain unclear (Fig 1a, upper scheme). In developmental biology and pharmacology, even a rough estimation of a core network system “without” predefined weighting factors and input networks can be highly valuable. For example, Manual construction with a simple network structure in human umbilical cord vein endothelial cell (HUVEC) has provided a detailed view of the endothelial-to-mesenchymal transition (EndMT) phenotype [14]. Such network systems are effective for *in silico* screening to identify candidate core factors involved in differentiation and drug response. However, constructing these networks requires numerous relationships and there is a risk of overlooking important signaling pathways.



**Fig 1. Strategy for development of *in silico* screening system using a transcription factors regulatory network.** (a) Schemes to identify candidate genes and signaling pathways involved in differentiation or pharmacological effects using transcriptome data. In the upper traditional scheme, up- and down-regulated genes are identified but the core signaling pathways remain still a black box. In the lower scheme established in this study, relationships among each gene were predicted using only comprehensive input data. (b) The requirements to develop the ISNA theory. (c) To estimate all TFs binding to each gene promoter, Sp (reflecting activator binding) and Sq values (reflecting repressor binding) were calculated using score matrixes. The top 4 binding sites were listed in a gene promoter and selected a most competitive TF to each binding site as a repressor. (d) The image of selection for regions of TF-bindable promoters and gene bodies from ChIP-Atlas data and binding sites of activators and repressors. (e) Learning  $K_d$  values in ISNA using a gradient descent method.

<https://doi.org/10.1371/journal.pone.0319971.g001>

We hypothesized that predicting gene relationships using “only” *de novo* comprehensive input data could achieve a simple *in silico* screening system to identify the core factors for differentiation and pharmacological effects with a network structure reflecting the protein functions (Fig 1a, lower scheme). In a snapshot, gene expression can be conceptualized as a TF network, where TFs regulate their own and other TFs’ gene expression, while other signaling via cell-cell communication can be ignored. Transcription is thermodynamically assumed as an amount of mRNA synthesis driven by an amount of RNA polymerase II (Pol II) on the transcription start site (TSS) induced by TFs binding [15]. Estimating the dissociation constant values ( $K_d$ ) of all TFs in all TF genes using only comprehensive input data could build an *in silico* screening using a TFs regulatory network analysis (ISNA) system that reflects TF binding “without” requiring for predefined relationships and weighting factors. In this study, we developed the ISNA theory and set 3 key requirements: (1) estimation of promoter regions and TFs-binding gene body regions, (2) construction of TF regulatory network (TRN) relationships using nucleotide sequence data and score matrices of consensus sequence of TFs, and (3) identification of core TFs by determining  $K_d$  values from predicted TRN relationships (Fig 1b). We validated ISNA by predicting core TFs in HUVEC and human and mouse embryonic stem cell (ESC) and used it to identify novel core TFs in the uterine epithelium.

## Materials and methods

### Network prediction

Active and repressive TRNs were automatically predicted using  $A(x_{m,n})$  and  $R(x_{m,n})$  which are output as csv files (Act\_ or Rep\_list.csv) after running of ISNA, respectively. In the analysis from the initial state to the reference state, the digraphs were plotted as ratio of each value (the initial to reference state) to the control value (the initial to initial state).

### Animals

Female C57BL/6J mice (CLEA Japan, Inc., Tokyo, Japan) for immunohistochemistry and CD-1 mice (Sankyo Research Laboratories Tokyo, Japan) for organ culture were maintained on a 12 h light/12 h dark schedule in a  $24 \pm 1^\circ\text{C}$  environment. They were given a commercial diet (MF; Oriental Yeast Co., Ltd., Tokyo, Japan) and tap water *ad libitum*. All animals were maintained in accordance with the NIH Guide for the Care and Use of Laboratory Animals, and all experiments were approved by the Institutional Animal Care Committee of Yokohama City University (the approval number: H-A-22-003) and Tokyo University of Science (the approval number: K17007) and performed in accordance with ARRIVE guidelines. All mice were euthanized by cervical dislocation by well-trained individuals. The oviducts, uteri, and vaginae were harvested at post-natal day 3 (P3) and 3 months. The estrous cycle in the intact mice was checked by vaginal smear. C57BL/6J mice at 3 months were ovariectomized (OVX) at 10 days before E2 treatment, injected s.c. with 0.1  $\mu\text{g}/25\text{g BW}$  E2 in sesame oil or vehicle, and harvested at 16 h after the treatment.

### Immunofluorescence

The oviducts, uteri, and vaginae were fixed in 4% paraformaldehyde (PFA) at  $4^\circ\text{C}$  overnight and dehydrated with graded alcohol. They were embedded in paraffin and cut into 6  $\mu\text{m}$  sections. Sections were deparaffinized with xylene, rehydrated, and rinsed with PBS. For antigen retrieval, sections were immersed in 0.05% sodium citraconic acid (pH 7.4) at  $95^\circ\text{C}$  for 45 min. Nonspecific binding was blocked in PBS containing 5% goat serum and 0.1% Triton X-100 for 30 min at room temperature. Sections were incubated at  $4^\circ\text{C}$  overnight with primary antibody for HMGA2 (1/100; Cell Signaling Technology, Danvers, MA, USA) or DLX5 (1/100; Abcam, Cambridge, UK). The sections were treated with Alexa488-conjugated goat anti-mouse IgG

(1/300; Thermo Fisher Scientific, Waltham, MA, USA) for 1 h at room temperature. For negative controls, normal rabbit IgG (Santa Cruz Biotechnology, Dallas, TX, USA) was used. 4', 6-diamino-2-phenylindole (DAPI) was used to stain nucleic acids.

## Organ culture

At 9 weeks, the uteri of OVX CD-1 mice at 7 weeks were collected. Eight volumes of Cellmatrix type I-A (Nitta Gelatin, Osaka, Japan) were mixed with 1 volume of  $10 \times$  DMEM/F12 and then 1 volume of 200 mM HEPES buffer containing 262 mM  $\text{NaHCO}_3$  and 0.05 N NaOH was added to the mixture. This 1 mL cold gelatin mixture was poured into cell culture inserts (Corning Inc., Corning, NY, USA) and placed into wells of a 12-well plate, and allowed to gel at  $37^\circ\text{C}$  for 30 min. The uteri were washed 3 times in HBSS and mixed with fresh 1 mL cold gelatin mixture. The tissues and gelatin mixture were overlaid onto the base of gelled collagen in each cell culture insert and allowed to gel at  $37^\circ\text{C}$  for 30 min. Subsequently, 1 mL DMEM/F12 with 20% fetal bovine serum (Tissue culture biologicals, Long Beach, CA, USA) with or without  $10^{-7}$  M E2 and 1% 1 mg/mL Hoechst 33258 solution (Dojindo, Kumamoto, Japan, final concentration was 10  $\mu\text{g/mL}$ ) were poured in each bottom well. These samples were cultured at  $37^\circ\text{C}$  in a humidified incubator at 5%  $\text{CO}_2$  for 20 h. To detect cell proliferation, these samples were then cultured in 1 mg/ mL EdU-containing medium for 4 h.

## EdU staining

After incubation with an EdU-containing medium, the samples were fixed in 4% PFA at  $4^\circ\text{C}$  overnight ( $n = 5$ ). The samples were embedded in paraffin and cut into 6  $\mu\text{m}$  sections. Sections were stained with Click-iT EdU Alexa Fluor 594 Imaging Kit (Thermo Fisher Scientific), according to the manufacturer's instructions. The number of EdU-positive cells and DAPI-stained cells was counted by Image J (NIH, Bethesda, MD, USA). Data were expressed as means  $\pm$  standard errors. For multiple comparisons, differences were estimated using a Peritz test. A statistically significant difference was defined as  $p \leq 0.05$ .

## Results and discussion

### Theory of concentration of TF proteins regulated by TRN in a snapshot

The concentration of TF proteins was modeled within the TRN. In a hypothetical unicellular events in a snapshot, where transcriptional regulation is unaffected by environmental factors, we hypothesized that the TF concentrations were solely regulated by TRN and eventually stabilized based on the initial TF concentrations and interactions of the TRN. To isolate the effects of TRN regulation, post-transcriptional and post-translational modifications was excluded from this model. To estimate transcriptional regulation among TFs, we focused on TFs with known consensus sequences. Specifically, we used 810 human TFs listed in JASPAR2020 (<https://jaspar.genereg.net/>). The TF concentration at a given step  $n$  was represented as  $x_{m,n}$  ( $m \in \text{TF number (810)}$ , learning step number  $n \in \mathbb{N}$ ) and its change over time was defined as:

$$x_{m,n+1} = f(x_{1,n}, x_{2,n}, \dots, x_{810,n}) + g(x_{m,n}) \quad (1)$$

Here,  $f(x)$  represents regulation of the gene expression within the TRN, while  $g(x)$  is a function for a residual amount by degradation. In this study, proteins in a nucleus assumed to be degraded by proteasomal and autophagic pathways, primarily regulated via ubiquitination by an enzymatic reaction [16]. The degradation function follows Michaelis-Menten kinetics and is expressed as:

$$g(x_{m,n}) = W\left(e^{-d + \log x_{m,n} + x_{m,n}}\right) \quad (2)$$

where  $W$  is Lambert  $W$ , and  $(d \in \mathbb{R})$  value represents the degradation rate.

### Direct estimation of TF network regulation via by TF binding

The rate of mRNA synthesis can be thermodynamically estimated by RNA polymerase II and TFs binding [15].

$$f(x_{1,n}, x_{2,n}, \dots, x_{810,n}) = k_c a_g \exp\left(\frac{-E_{p_g} - \Delta G_{p_g}^0 + E_{d_g}}{RT}\right) \quad (3)$$

Here,  $k_c$  is a constant for basal transcription activity, which varies with RNA polymerase concentration. Temperature is stable in the mammalian cells. A value of  $a_g$  is a constant for the chromatin state. In this study, we estimated the stable RNA polymerase concentration and temperature, and modeled only opened chromatin; therefore, those values ( $k_c$ ,  $a_g$ , and  $RT$ ) are collectively represented by  $k_c$ .  $E_{d_g}$ , a member for RNA degradation, is omitted for simplification.  $\Delta G_{p_g}^0$ , a member for RNA polymerase II binding, and  $E_{p_g}$ , a member for mRNA-synthesizing activity of RNA polymerase II, both are regulated by TFs.

$$f(x_{1,n}, x_{2,n}, \dots, x_{810,n}) = k_c \exp(-E_{p_g} - \Delta G_{p_g}^0) \quad (4)$$

For a single activator  $A$  with concentration  $[A] = x_{m,n}$ , let  $r$  represent the number of binding for  $p$  of promoter sequences on a gene promoter, which depends on  $[A]$  since the concentration  $[S]$  of the binding sequence on the genome is stable. When  $[S]$  is set as 1 to simplify,  $r$  is estimated using the  $p$  value which is the number of binding sites.

$$r = [AS] = \sum_p \frac{[S][A]}{K_d + [A]} = \sum_p \frac{[A]}{K_d + [A]} \quad (5)$$

When many amounts of one type of activator bind to a target promoter sequence, the activator and sequences can act as a super enhancer, producing a nonlinear and stable effect like computational circuit [17]. We assumed a maximum  $p$  of 4 binding sites and additional site's effects contributed minimally in this theory.

Consensus sequences for activators are expressed as a score matrix. When a binding score  $Sp$  value ( $0 \leq Sp \leq 1$ ) was calculated by the score matrix for each consensus sequence increases to 1, the activator binds to the sequences with exponential strength [18]. We introduced the effect of the consensus sequence on the binding sites and a weighting factor “ $a$ ”

$$r = \sum_p \frac{10^{-a(1-Sp)} [A]}{K_d + 10^{-a(1-Sp)} [A]} \quad (6)$$

To accelerate the calculation of  $r$  values, we proposed a list of  $Sp$  in all TFs from maximum  $Sp_1$  to fourth  $Sp_4$  (Fig 1c; see the program “TF\_binding”). Further, we considered that other TFs act as repressors for the activator  $A$  by antagonistic effects. The strongest antagonist  $R$  was select based on a list of score matrix values for repressors ( $Sq$ ) and TF concentration using  $K_i 10^{-a(1-Sq)} [R]$  ( $K_i$ : dissociation constant of  $R$ ) (see program “Act\_Rep\_score\_making”).

$$r = \sum_p \frac{10^{-a(1-Sp)} [A]}{K_d + 10^{-a(1-Sp)} [A] + 10^{-a(1-Sq)} \left(\frac{K_d}{K_i}\right) [R]} \quad (7)$$



For promoter with high binding affinity (a sum of Sp1 to Sp4 values exceeding a threshold), we assumed that the  $r$  value was exponentially calculated because it will act as a super enhancer [17].

$$r = \frac{\sum_p 10^{-a(1-Sp)} \times [A]^p}{K_d^p + \sum_p 10^{-a(1-Sp)} \times [A]^p + \sum_q 10^{-a(1-Sq)} \left( \frac{K_d}{K_i} \right) [R]} \quad (8)$$

Since  $E_{p_g} + \Delta G_{p_g}^0$  is inversely proportional to  $r$ , a value of  $f(x)$  only by an activator A is:

$$f(x_{m,n}) = A(x_{m,n}) = k_c \left\{ \exp\left(-\frac{1}{r}\right) \right\} \quad (9)$$

When the TF is only listed as a repressor in gene ontology terms, the function is converted to decrease expression. The function of gene regulation by all TFs is calculated by sums of contributions from all activators and repressors:

$$f(x_{1,n}, x_{2,n}, \dots, x_{810,n}) = \sum_{m=1}^{810} A(x_{m,n}) \quad (10)$$

When TFs bound to regions in a gene body, we assumed that TFs acted as a repressor by the reference of basal procaryotic regulation [19]. If the repressor effects by TFs in a gene body are independent of promoter binding, the repressor effects are calculated as the same method, negatively act on activator effects, and are subtracted from the function of all promoter regulation.

$$f(x_{1,n}, x_{2,n}, \dots, x_{810,n}) = \sum_{m=1}^{810} \{A(x_{m,n}) - R(x_{m,n})\} \quad (11)$$

To prevent negative regulation values, we set:

$$f(x_{1,n}, x_{2,n}, \dots, x_{810,n}) = 0 \quad (f(x_{1,n}, x_{2,n}, \dots, x_{810,n}) < 0) \quad (12)$$

The TF concentration update equation incorporates both activation and degradation:

$$x_{m,n+1} = \sum_{m=1}^{810} \{A(x_{m,n}) - R(x_{m,n})\} + W(e^{-d+\log x_{m,n}+x_{m,n}}) \quad (13)$$

To determine optimal  $K_d$  values for TFs when condition of a cell from initial state with initial transcriptomic data changes to different state with reference transcriptomic data, we applied a gradient descent method. Weighting factors of  $K_d$  for activator and repressor are set based on reference concentration values from reference transcriptomic data.

### Parameters fitting in ISNA

Using the theory of TF concentration regulated by the TRN, we developed the ISNA algorithm, which includes: (1) estimation of promoter and gene body regions, (2) listing Sp and Sq values using score matrixes, and (3) optimizing  $K_d$  value in the TRN using reference data of RNA-seq (Fig 1b). To identify open chromatin regions in the genome, we utilized the database of ChIP-Atlas (<https://chip-atlas.org/>). Open chromatin regions were defined as connected positive regions detected by ChIP-seq for all modified histone, DNase-seq, and ATAC-seq, with at least one positive region from DNase-seq or ATAC-seq (Fig 1d). In

heterochromatin-dense regions, gene expression will be constitutively “off” without regulation by TFs. Therefore, such regions were excluded from the analysis unless they exhibited ATAC-seq or DNase-seq signals. H3K27ac, H3K4me3, and H2K4me1 are used for active histone marks and H3K27me3 and H3K9me3 are used for repressive histone marks [20]. Initially, we attempted to use active histone marks to estimate the open chromatin regions; however, in some cases, no open chromatin region was identified in genes although the genes are actively expressed in the cells. We considered that the failure to estimate the open chromatin regions was caused by which studies of effects of histone modification on gene expression are ongoing. For example, repressive histone marks are sometimes detected with ATAC-seq and DNase-seq signals as a silencer region [21]. Thus, we incorporated data from all histone modifications to improve the accuracy of open chromatin region identification.

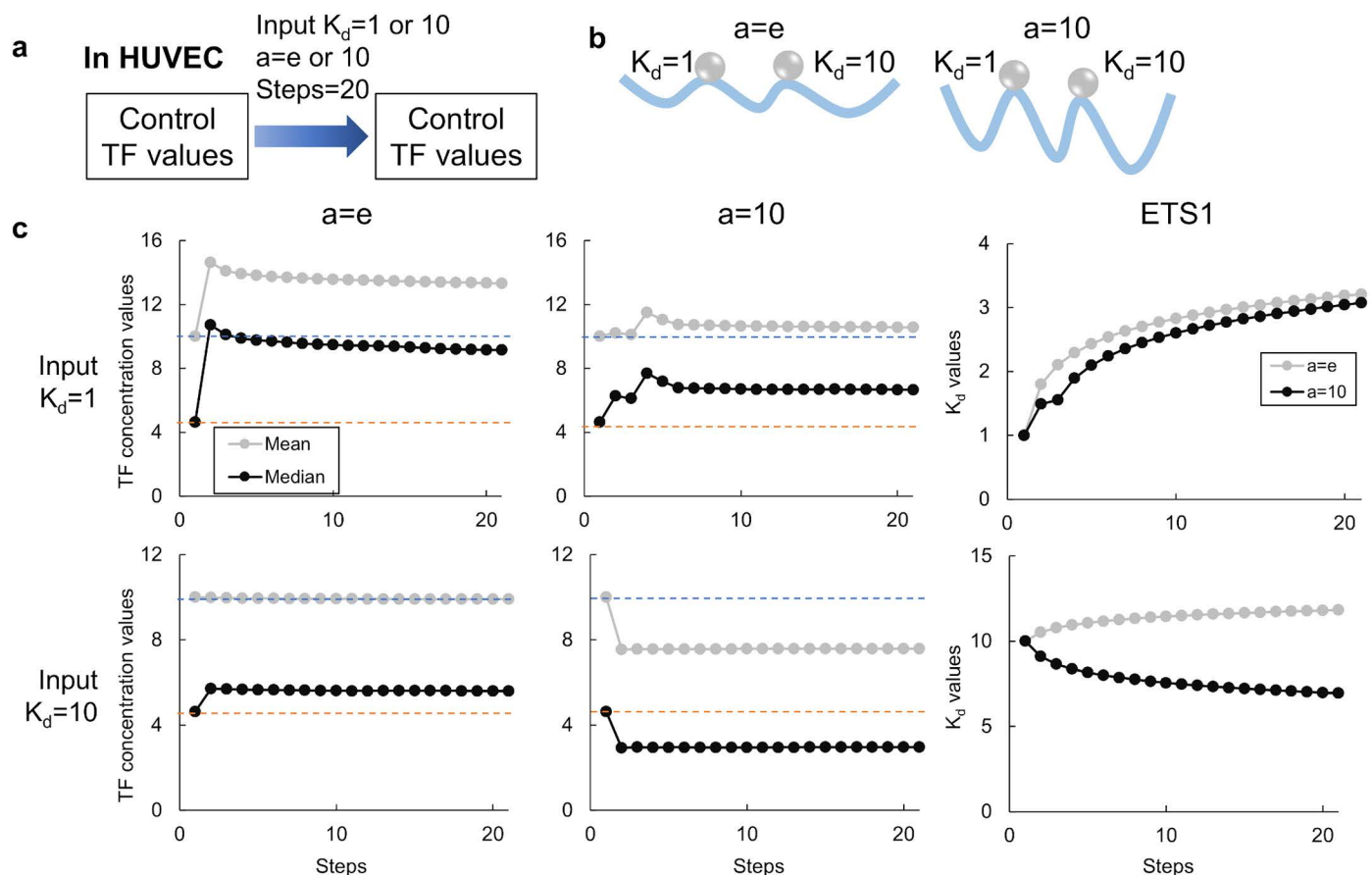
In open chromatin regions, those containing transcription start sites were defined as promoter regions, while the regions spanning from transcription start sites to end sites were defined as gene body regions. In HUVEC, the 343 genes of TFs were identified as genes including open promoter regions. After listing Sp and Sq values using score matrixes based on consensus sequences of each TF, the concentration of TFs was set using RNA-seq data from 2 conditions (e.g., undifferentiated vs. differentiated conditions) as initial and reference values. The  $K_d$  value in TRN was optimized using a gradient descent method (Fig 1e). For parameters fitting “a” which is a weighting factor for the score matrixes and an the initial  $K_d$  value, we ran the ISNA program using the control HUVEC RNA-seq data [22] as initial and reference values for 20 gradient descent steps (Fig 2a). If TRN stability was visualized as a mountain landscape, a higher weighting factor “a” for score matrixes reflects a steeper slope of Sp and Sq values in TF binding, while different initial  $K_d$  values correspond to different start points (Fig 2b). When the same values are used to initial values as reference values, the ideal outcome was for ISNA to maintain TF concentrations equal to the initial values (Fig 2c, dashed lines). When the weighting factor “a” was set low (e.g., e) or high (e.g., 10), the  $K_d$  values were stable at initial or different values. Under any conditions, the  $K_d$  values usually asymptotically relate to some values and do not fluctuate (e.g., the  $K_d$  values of ETS1). Therefore, suitable parameters for each analysis can be selected by adjusting the weighting factor for the score matrixes and the initial values of the TFs.

### ISNA identified core TFs related to differentiation in HUVEC and ESC

Under specific parameter setting,  $K_d$  values were optimized using HUVEC RNA-seq data [22] as both initial and reference values for 20 gradient descent steps (Fig 3a). In the HUVEC’s TRN consisting of 343 TFs, many TFs retained the initial  $K_d$  values after optimization, suggesting that those TFs do not play a regulatory role in the TRN. Thus, we defined candidates of core TFs characterized by different  $K_d$  values after optimization. While the number of core TFs varied depending on parameter selection, smaller core TF sets were always subsets of larger ones (Fig 3a). The relationships among TRN were visualized using a digraph derived using derivative factors from gradient descent method (Fig 3b in promoter regions; S1 Fig in gene body regions). For example, ISNA identified ETS1 as the candidate of key core TF in HUVEC’s TRN. This is consistent with previous findings that TFs belonging to the ETS family regulate endothelial cell development [23], and that inhibition of ETS1 disrupts endothelial cell barrier function [24].

To model TRN during the transition from control to experimental conditions in ISNA, the phenomenon of EndMT in HUVEC was selected as an example. Using the RNA-seq data of HUVEC in the state of EndMT induced by ERK1/2 knockdown [22] as the reference value, ratio of the group of all  $K_d$  values after fitting was calculated compared with that using the control HUVEC RNA-seq data in the set of parameters (weighting factor  $a = 10$ , initial  $K_d$

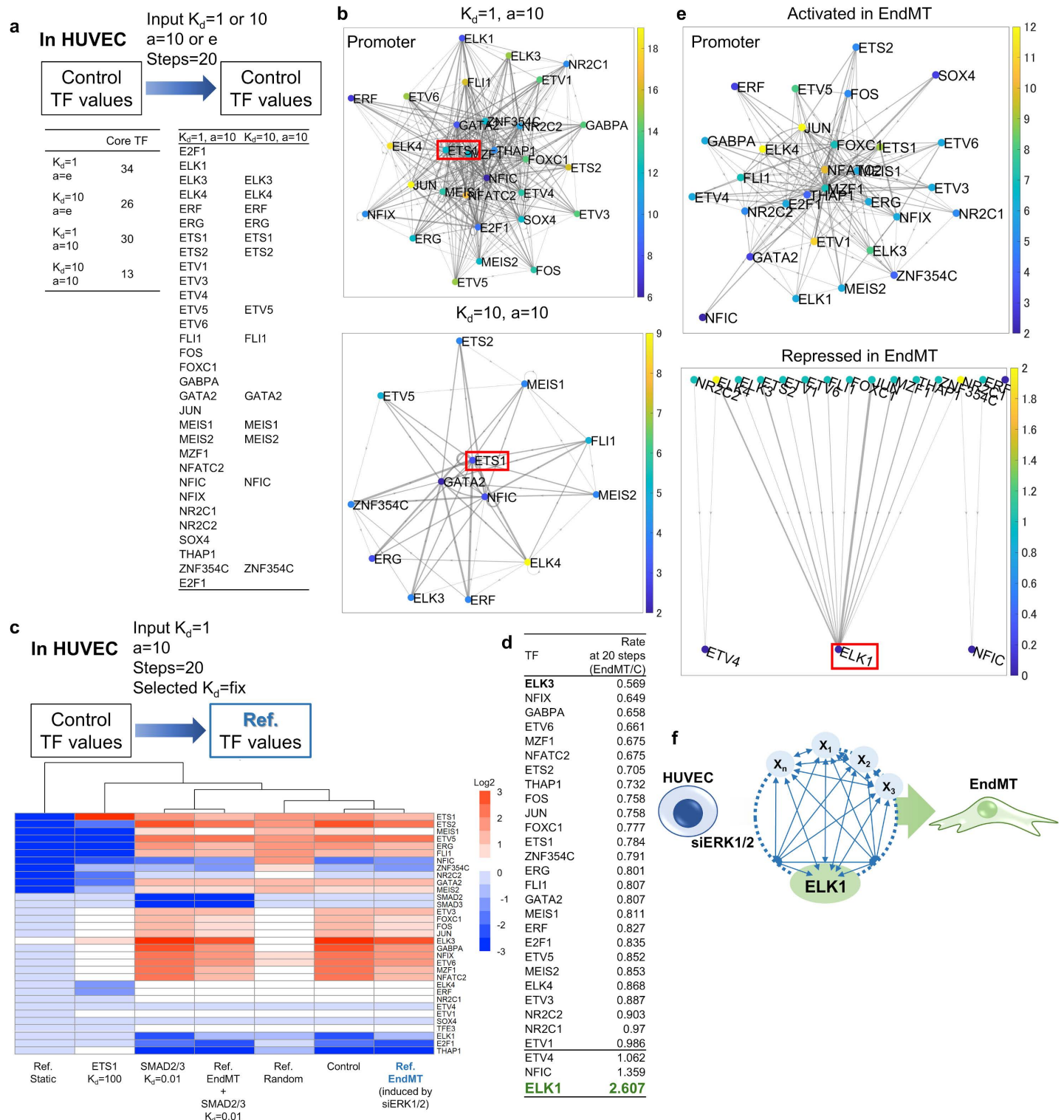
value = 1) (Fig 3c). When continuous activation of SMAD2/3 (known as EndMT inducers and induced by TGF $\beta$  ligands, but not selected as a core TF in ISNA) was simulated by setting  $K_d$  value = 0.01, it had no effect other  $K_d$  values of TRN components after optimization. On the other hand, continuous inhibition of ETS1 by setting the  $K_d$  value = 100 drastically altered the other  $K_d$  values of TRN components. Under these conditions, even using random and constant reference values, only the  $K_d$  values of core TFs changed from their initial values. By calculating the ratio of  $K_d$  values of the core TFs in the EndMT group/control group, ELK1 was identified as the most inhibited core TF to induce EndMT (Fig 3d). The differential relationships within the TRN under EndMT vs. control conditions were visualized as a digraph (Fig 3e in promoter regions; S2 Fig in gene body regions). In particular, ISNA predicted that several core TFs were activated and ELK1 was repressed during EndMT condition. Therefore, ISNA identified ELK1 as the most critical core TF for induction of EndMT in HUVEC's TRN (Fig 3f). This finding aligns with previous studies showing that ELK1 mediates fibroblast growth factor signaling to inhibit EndMT, and that ELK1 knockdown induces EndMT in endothelial cells [25].



**Fig 2. Learning  $K_d$  values in ISNA.** (a) The condition for learning  $K_d$  values in ISNA. The same HUVEC RNA-seq data was used as the input and reference expression data. One or 10 values were used as input  $K_d$  values for all TFs. The values of "a" as a weighting factor for effects of Sp and Sq values on binding TFs were set to e or 10. (b) The schematic image of learning  $K_d$  values. (c) The mean and median TF concentration values in all TFs and the  $K_d$  value of ETS1 in learning  $K_d$  values in 20 steps.

<https://doi.org/10.1371/journal.pone.0319971.g002>





**Fig 3. Identification of core TFs in endothelial to mesenchymal transition in HUVEC using ISNA.** (a) The condition for learning  $K_d$  values in ISNA and the table of the number and name of core TFs in which the  $K_d$  values were changed after learning. (b) The digraph was derived from derivative factors of the gradient descent method to visualize the relationships among core TFs at each promoter region. When the TFs had a larger number of edges with the entering direction to the TFs, the TFs were situated at a more center position. Left bar and color of TFs: the number of edges with the forward direction from the TFs. Width of edges: reflecting the size of derivative factors. (c) Heatmap of log2 fold change of rate of  $K_d$  values (values after analysis/initial values) for the core TFs in the TRN regulating the gene expression from control HUVEC to reference situation. When SMAD2/3 or ETS1 was selected, the  $K_d$  values were fixed in the analysis. (d) List of rates of  $K_d$  value in the

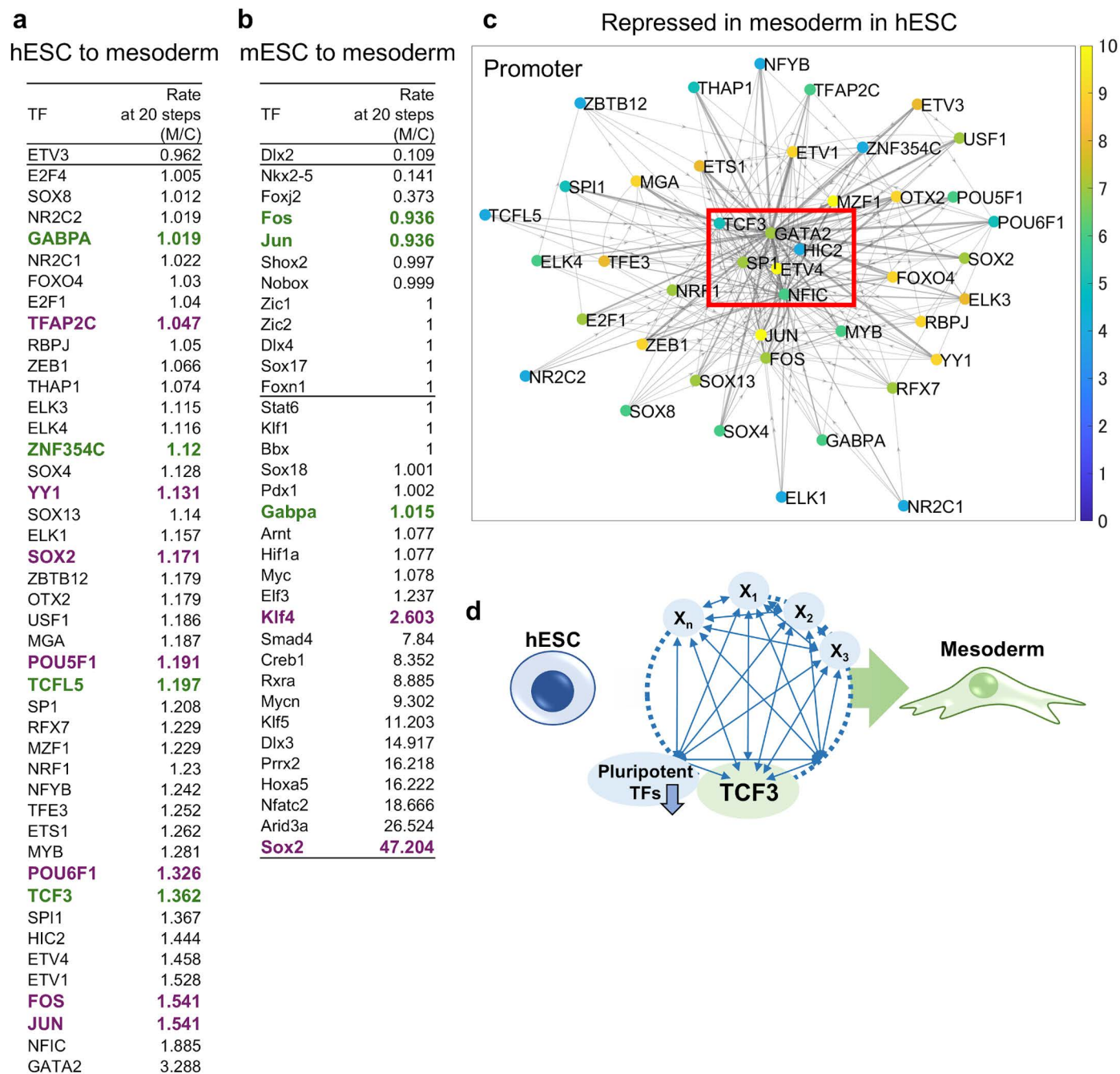
TRN regulating the gene expression from control HUVEC to EndMT. (e) The digraph was derived from the rate of derivative factors of the gradient descent method (EndMT/control values) in activated or repressed TFs in EndMT to visualize the relationships among core TFs at each promoter region. (f) The schematic image of TRN and core ELK1 to induce EndMT in HUVEC.

<https://doi.org/10.1371/journal.pone.0319971.g003>

To further validate ISNA's ability to model TRN transitions, we analyzed mesodermal differentiation in human ESC [26] and mouse ESC [27]. In hESC, the TRN included 480 TFs out of 810 possible TFs with known consensus sequence, while in mESC, the TRN included 355 TFs from 596 possible TFs. Core TFs were estimated under the parameter set (weighting factor  $\alpha = 10$ , initial  $K_d$  value = 1). The ratios of  $K_d$  values of core TFs in the mesodermal differentiation vs. control conditions in human ESC were listed (Fig 4a). ISNA identified the inhibition of many TFs related to pluripotency (e.g., POU5F1 and SOX2) and involvement of TCF3 activated by Wnt signaling. In mESC, fewer TFs were identified as core TFs and important TFs related to pluripotency and mesodermal differentiation are reduced compared to that in hESC (Fig 4b). The lower performance of ISNA for the identification of core TFs in mESC indicated the importance of the accumulation of information on the consensus sequence of TFs and ChIP-seq on chromatin states. The differential relationships among the TRN in the mesodermal differentiation vs. control conditions in hESC were visualized as a digraph (Fig 4c), highlighting the key set of core TFs, including TCF3. Therefore, ISNA identified TCF3 as a critical core TFs for mesodermal differentiation in hESC's TRN (Fig 4d). This finding aligns with previous research showing that activation of Wnt signaling in ESC induces loss of pluripotency, remodeling of plastic chromatin to differentiate into mesoderm, and mesodermal specification [28,29].

### ISNA identified novel candidates of core TF in uterine epithelial cells

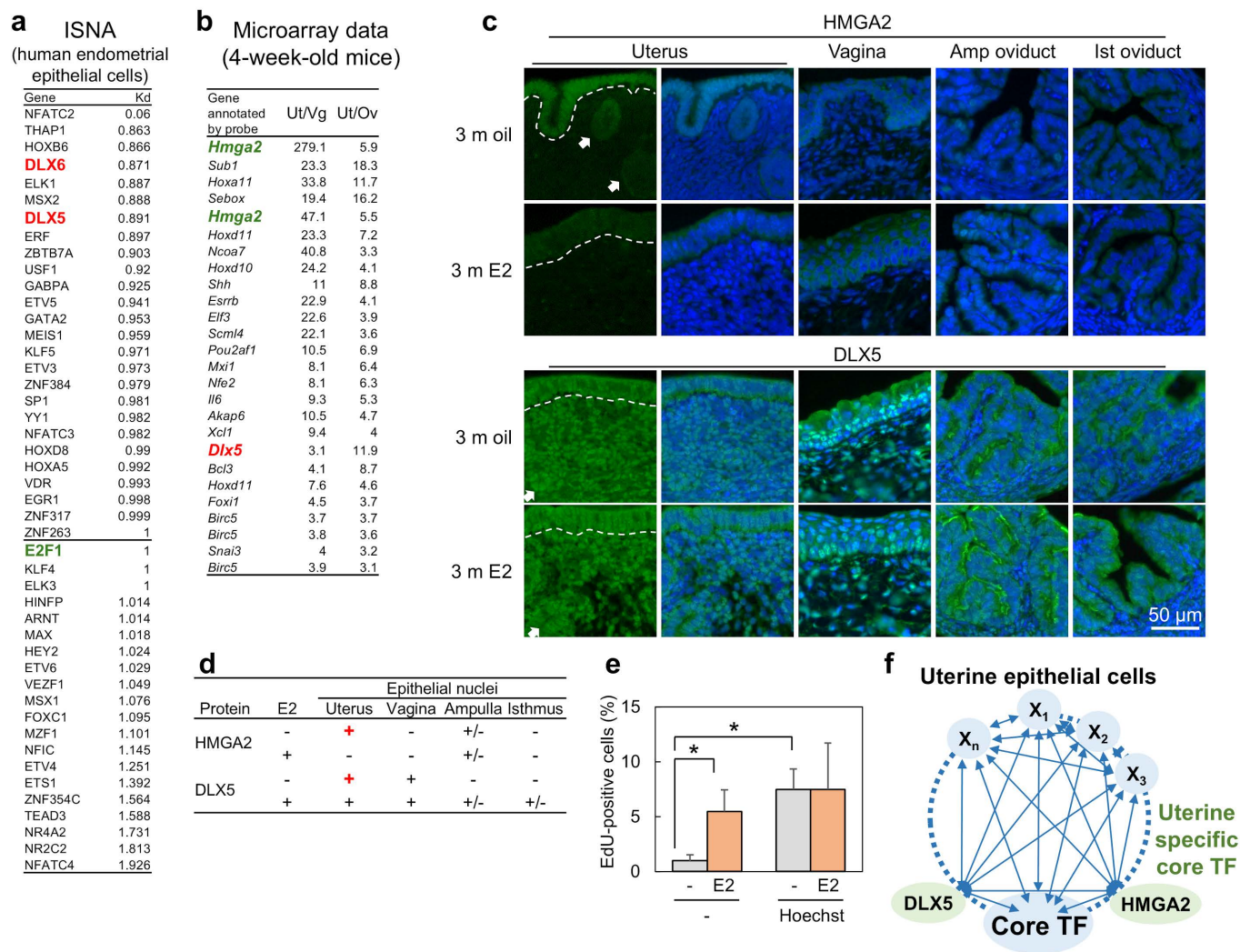
To identify previously unknown core TFs, we applied ISNA. In the female reproductive tract, the vaginal and oviductal epithelium can be specified by the expression of TFs, TRP63 and FOXJ1, respectively [30,31]; however, no specific TF in uterine epithelial cells has been identified. To construct the TRN in uterine epithelial cells in ISNA, the RNA-seq data from the human endometrial epithelium [32] was used as both the initial and reference values. ISNA identified core TFs in human endometrial epithelial cells (Fig 5a; the digraph in S3 Fig). To further refine these candidates, they were compared with the group of genes highly expressed in the uterus of 4-week-old mice relative to the vagina and oviduct [33] (Fig 5b). From this intersection, HMGA2 (which interacts with E2F1) and DLX5 are identified as candidate uterus-specific core TFs. Immunostaining revealed that HMGA2 was strongly expressed in the nuclei of uterine epithelial cells, with weak expression in the oviduct of ampulla in oil-treated ovariectomized (OVX) 3-month-old mice (Fig 5c, 5d). DLX5 was expressed in the epithelium and stroma of the uterus and vagina in the oil-treated 3-month-old mice. No difference of expression of HMGA2 and DLX5 was not observed among female reproductive tracts in estradiol (E2)-treated, estrus, and diestrus 3-month-old mice and mice at postnatal day 3 (S4 and S5 Figs), suggesting that their expression was regulated by sex hormones. DLX5 and DLX6 are known to be important TFs for the development of uterine glands in epithelial regions [34]. HMGA2 promotes cell proliferation and angiogenesis in uterine leiomyomas [35,36], but its function in the uterine epithelium remains unclear. Since estrogen-regulated cell proliferation is a key function of uterine epithelial cells, we hypothesized that the core TFs in this context would be involved in estrogen-mediated cell proliferation. To investigate this, we treated organ-cultured adult uterus with Hoechst 33258 which binds the minor groove of AT-rich DNA and displaces HMGA proteins from chromatin [37,38], alongside E2 (Fig 5e). E2 and Hoechst 33258 treatment induced cell proliferation in the uterine epithelium at similar level,



**Fig 4. Identification of core TFs for differentiation into mesodermal cells in human and mouse ESC using ISNA.** Lists of rates of  $K_d$  value in the TRN regulating the gene expression from hESC (a) and mESC (b) to mesodermal cells. (c) The digraph was derived from the rate of derivative factors of the gradient descent method (mesodermal cells/hESC values) in repressed TFs in EndMT to visualize the relationships among core TFs at each promoter region. When the TFs had a larger number of edges with the entering direction to the TFs, the TFs were situated at a more center position. Left bar and color of TFs: the number of edges with the forward direction from the TFs. Width of edges: reflecting the size of derivative factors. (d) The schematic image of TRN and core TFs related to pluripotency to induce mesodermal cells in HUVEC.

<https://doi.org/10.1371/journal.pone.0319971.g004>





**Fig 5. Identification of new specific core TF in uterine epithelial cells using ISNA.** (a) Lists of  $K_d$  value in the TRN regulating the gene expression in human endometrial epithelial cells. (b) The ratio of gene expression between the uterus and vagina or the uterus and oviduct in 4-week-old mice was listed using data examined with microarray. (c) The immunofluorescence images for HMGA2 or DLX5 (green) in the uterus, vagina, ampulla of oviduct, and isthmus of oviduct of 3-month-old OVX mice with oil or E2 treatment. Blue: the nuclei. White arrows: the uterine glands. Dash line: basement membrane.  $n = 3$ , biologically independent. (d) The table of expression for HMGA2 and DLX5 in the nuclei of the epithelium of female reproductive tracts in adult mice with or without E2. (e) The percentage of EdU-positive cells in the epithelium of organ-cultured uterus with or without E2 and HMGA2 inhibitor, Hoechst 33258. \*:  $p \leq 0.05$ .  $n = 5$ . (f) The schematic image of TRN and specific TFs in adult uterine epithelial cells.

<https://doi.org/10.1371/journal.pone.0319971.g005>

and Hoechst 33258 did not have a synergistic effect with E2 treatment. Because Hoechst 33258 treatment did not inhibit cell proliferation in the organ-cultured uterus, Hoechst 33258 may inhibit TFs with motif of AT-rich region, including HMGA2. Therefore, HMGA2 may act as a repressor of cell proliferation in the uterine epithelium under static condition, while E2 may promote cell proliferation by downregulation of HMGA2 expression. In addition, high HMGA2 expression was detected in the epithelial junction between the oviduct and uterus (S6 Fig), suggesting HMGA2 may support junctional maintenance by inhibition of epithelial cell proliferation. Thus, the expression of HMGA2 (and to some extent, DLX5) in the epithelium of the female reproductive tracts defines the identity of uterine epithelial cells with the E2-HMGA2 signaling pathway regulating cell proliferation. These findings demonstrate that

ISNA can support the identification of tissue-specific core TFs only using transcriptomic data (Fig 5f).

## Conclusion

ISNA enables construction of TRN composed of core TFs, bridging the gap between initial to the reference gene expression states. Notably, ISNA constructs TRN using only transcriptomic data “without” requiring predefined parameters such as directionality, weighting factors among TFs, and an input network structure. However, ISNA’s ability to identify core TFs related to differentiation in mESC is weaker than that in hESC, suggesting that ISNA is dependent on the accumulation of information from consensus sequence of TFs and ChIP-seq for chromatin. Further, ISNA operates under several simplifying assumptions, limiting its applicability. For example, ISNA does not account for cell-cell interactions and post-transcriptional regulation, as these factors were excluded from its functional framework. To learn optimal  $K_d$  values for TFs, ISNA employs a gradient descent method without testing alternative machine learning approaches. Taken together, ISNA’s accuracy is expected to be lower than that of well-developed computational models [12,13]. Despite these limitations, this is the first report of system capable of constructing a TRN “black box” solely from transcriptomic data. We anticipate that future advancements integrating ISNA with bioinformatics theories, improved learning algorithms, and refined methodologies will lead to more powerful and accurate TRN reconstruction systems.

## Supporting information

**S1 Fig. The digraph to visualize the relationships among core TFs at each gene body region in control HUVEC.** The digraph was derived from derivative factors of the gradient descent method to visualize the relationships among core TFs at each gene body region in control HUVEC. When the TFs had a larger number of edges with the entering direction to the TFs, the TFs were situated at a more center position. Left bar and color of TFs: the number of edges with the forward direction from the TFs. Width of edges: reflecting the size of derivative factors.  
(TIF)

**S2 Fig. The digraph to visualize the relationships among core TFs at each gene body region in EndMT of HUVEC.** The digraph was derived from the rate of derivative factors of the gradient descent method (EndMT/control HUVEC values) in activated or repressed TFs in EndMT to visualize the relationships among core TFs at each gene body region. When the TFs had a larger number of edges with the entering direction to the TFs, the TFs were situated at a more center position. Left bar and color of TFs: the number of edges with the forward direction from the TFs. Width of edges: reflecting the size of derivative factors.  
(TIF)

**S3 Fig. The digraph to visualize the relationships among core TFs at each gene body region in human endometrial epithelial cells.** The digraph was derived from derivative factors of the gradient descent method to visualize the relationships among core TFs at each gene body region in human endometrial epithelial cells. When the TFs had a larger number of edges with the entering direction to the TFs, the TFs were situated at a more center position. Left bar and color of TFs: the number of edges with the forward direction from the TFs. Width of edges: reflecting the size of derivative factors.  
(TIF)



**S4 Fig. The localization of HMGA2 in the epithelia of female reproductive tracts.** The immunofluorescence images for HMGA2 (green) in the uterus, vagina, ampulla of oviduct, and isthmus of oviduct of 3-month-old intact diestrus and estrus mice, 3-month-old OVX mice with oil or E2 treatment, and mice at postnatal day 3. Blue: the nuclei. Left images: only HMGA2. Right images: merged images with HMGA2 and nuclei. White arrows: the uterine glands. Dash line: basement membrane. n = 3, biologically independent. (TIF)

**S5 Fig. The localization of DLX5 in the epithelia of female reproductive tracts.** The immunofluorescence images for DLX5 (green) in the uterus, vagina, ampulla of oviduct, and isthmus of oviduct of 3-month-old intact diestrus and estrus mice, 3-month-old OVX mice with oil or E2 treatment, and mice at postnatal day 3. Blue: the nuclei. Left images: only DLX5. Right images: merged images with DLX5 and nuclei. White arrows: the uterine glands. Dash line: basement membrane. n = 3, biologically independent. (TIF)

**S6 Fig. The localization of HMGA2 in the epithelium of the junction between the oviduct and uterus.** The immunofluorescence images for HMGA2 (green) in the epithelium of junction between oviduct and uterus of 3-month-old OVX mice with oil or E2 treatment. Blue: the nuclei. Left images: only HMGA2. Right images: merged images with HMGA2 and nuclei. n = 3, biologically independent. (TIF)

## Author contributions

**Conceptualization:** Tadaaki Nakajima.

**Data curation:** Tadaaki Nakajima.

**Formal analysis:** Tadaaki Nakajima.

**Funding acquisition:** Tadaaki Nakajima, Yasuhiro Tomooka, Tomomi Sato.

**Investigation:** Tadaaki Nakajima, Kentaro Harada.

**Methodology:** Tadaaki Nakajima.

**Project administration:** Tadaaki Nakajima, Yasuhiro Tomooka, Tomomi Sato.

**Resources:** Tadaaki Nakajima, Yasuhiro Tomooka, Tomomi Sato.

**Software:** Tadaaki Nakajima.

**Supervision:** Tadaaki Nakajima.

**Validation:** Tadaaki Nakajima, Kentaro Harada.

**Visualization:** Tadaaki Nakajima.

**Writing – review & editing:** Kentaro Harada, Yasuhiro Tomooka, Tomomi Sato.

## References

1. Ichinose N, Kawashima T, Yada T, Wada H. Dynamical robustness and its structural dependence in biological networks. *J Theor Biol.* 2021;526:110808. <https://doi.org/10.1016/j.jtbi.2021.110808> PMID: [34118264](#)
2. Shu H, Zhou J, Lian Q, Li H, Zhao D, Zeng J, et al. Modeling gene regulatory networks using neural network architectures. *Nat Comput Sci.* 2021;1(7):491–501. <https://doi.org/10.1038/s43588-021-00099-8> PMID: [38217125](#)

3. Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*. 2008;9(10):770–80. <https://doi.org/10.1038/nrm2503> PMID: 18797474
4. Kobayashi K, Maeda K, Tokuoka M, Mochizuki A, Satou Y. Controlling cell fate specification system by key genes determined from network structure. *iScience*. 2018;4:281–93. <https://doi.org/10.1016/j.isci.2018.05.004> PMID: 30240747
5. Tokuoka M, Maeda K, Kobayashi K, Mochizuki A, Satou Y. The gene regulatory system for specifying germ layers in early embryos of the simple chordate. *Sci Adv*. 2021;7(24):eabf8210. <https://doi.org/10.1126/sciadv.abf8210> PMID: 34108211
6. Ishikawa M, Sugino S, Masuda Y, Tarumoto Y, Seto Y, Taniyama N, et al. RENGÉ infers gene regulatory networks using time-series single-cell RNA-seq data with CRISPR perturbations. *Commun Biol*. 2023;6(1):1290. <https://doi.org/10.1038/s42003-023-05594-4> PMID: 38155269
7. Glass K, Huttenhower C, Quackenbush J, Yuan G-C. Passing messages between biological networks to refine predicted interactions. *PLoS One*. 2013;8(5):e64832. <https://doi.org/10.1371/journal.pone.0064832> PMID: 23741402
8. Balwiercz PJ, Pachkov M, Arnold P, Gruber AJ, Zavolan M, van Nimwegen E. ISMARA: Automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res*. 2014;24(5):869–84. <https://doi.org/10.1101/gr.169508.113> PMID: 24515121
9. Guan D, Shao J, Deng Y, Wang P, Zhao Z, Liang Y, et al. CMGRN: A web server for constructing multilevel gene regulatory networks using ChIP-seq and gene expression data. *Bioinformatics*. 2014;30(8):1190–2. <https://doi.org/10.1093/bioinformatics/btt761> PMID: 24389658
10. Pemberton-Ross PJ, Pachkov M, van Nimwegen E. ARMADA: Using motif activity dynamics to infer gene regulatory networks from gene expression data. *Methods*. 2015;85:62–74. <https://doi.org/10.1016/j.ymeth.2015.06.024> PMID: 26164700
11. Berger S, Pachkov M, Arnold P, Omid S, Kelley N, Salatino S, et al. Crunch: Integrated processing and modeling of ChIP-seq data in terms of regulatory motifs. *Genome Res*. 2019;29(7):1164–77. <https://doi.org/10.1101/gr.239319.118> PMID: 31138617
12. Sonawane AR, DeMeo DL, Quackenbush J, Glass K. Constructing gene regulatory networks using epigenetic data. *NPJ Syst Biol Appl*. 2021;7(1):45. <https://doi.org/10.1038/s41540-021-00208-3> PMID: 34887443
13. Chen C, Padi M. Flexible modeling of regulatory networks improves transcription factor activity estimation. *NPJ Syst Biol Appl*. 2024;10(1):58. <https://doi.org/10.1038/s41540-024-00386-w> PMID: 38806476
14. Weinstein N, Mendoza L, Álvarez-Buylla ER. A Computational model of the endothelial to mesenchymal transition. *Front Genet*. 2020;11:40. <https://doi.org/10.3389/fgene.2020.00040> PMID: 32226439
15. Konishi T. A thermodynamic model of transcriptome formation. *Nucleic Acids Res*. 2005;33(20):6587–92. <https://doi.org/10.1093/nar/gki967> PMID: 16314319
16. Gumeni S, Evangelakou Z, Gorgoulis VG, Trougakos IP. Proteome stability as a key factor of genome integrity. *Int J Mol Sci*. 2017;18(10):2036. <https://doi.org/10.3390/ijms18102036> PMID: 28937603
17. Michida H, Imoto H, Shinohara H, Yumoto N, Seki M, Umeda M, et al. The number of transcription factors at an enhancer determines switch-like gene expression. *Cell Rep*. 2020;31(9):107724. <https://doi.org/10.1016/j.celrep.2020.107724> PMID: 32492432
18. Jung C, Bandilla P, von Reutern M, Schnepf M, Rieder S, Unnerstall U, et al. True equilibrium measurement of transcription factor-DNA binding affinities using automated polarization microscopy. *Nat Commun*. 2018;9(1):1605. <https://doi.org/10.1038/s41467-018-03977-4> PMID: 29686282
19. van Hijum SAFT, Medema MH, Kuipers OP. Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol Mol Biol Rev*. 2009;73(3):481–509, Table of Contents. <https://doi.org/10.1128/MMBR.00037-08> PMID: 19721087
20. Gates LA, Foulds CE, O'Malley BW. Histone marks in the “driver’s seat”: Functional roles in steering the transcription cycle. *Trends Biochem Sci*. 2017;42(12):977–89. <https://doi.org/10.1016/j.tibs.2017.10.004> PMID: 29122461
21. Huang D, Petrykowska HM, Miller BF, Elnitski L, Ovcharenko I. Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression. *Genome Res*. 2019;29(4):657–67. <https://doi.org/10.1101/gr.247007.118> PMID: 30886051
22. Ricard N, Scott RP, Booth CJ, Velazquez H, Cilfone NA, Baylon JL, et al. Endothelial ERK1/2 signaling maintains integrity of the quiescent endothelium. *J Exp Med*. 2019;216(8):1874–90. <https://doi.org/10.1084/jem.20182151> PMID: 31196980
23. Meadows SM, Myers CT, Krieg PA. Regulation of endothelial cell development by ETS transcription factors. *Semin Cell Dev Biol*. 2011;22(9):976–84. <https://doi.org/10.1016/j.semcdb.2011.09.009> PMID: 21945894

24. Luo Y, Yang H, Wan Y, Yang S, Wu J, Chen S, et al. Endothelial ETS1 inhibition exacerbate blood-brain barrier dysfunction in multiple sclerosis through inducing endothelial-to-mesenchymal transition. *Cell Death Dis.* 2022;13(5):462. <https://doi.org/10.1038/s41419-022-04888-5> PMID: [35568723](#)
25. Akatsu Y, Takahashi N, Yoshimatsu Y, Kimuro S, Muramatsu T, Katsura A, et al. Fibroblast growth factor signals regulate transforming growth factor- $\beta$ -induced endothelial-to-myofibroblast transition of tumor endothelial cells via Elk1. *Mol Oncol.* 2019;13(8):1706–24. <https://doi.org/10.1002/1878-0261.12504> PMID: [31094056](#)
26. Przybyla L, Lakins JN, Weaver VM. Tissue mechanics orchestrate wnt-dependent human embryonic stem cell differentiation. *Cell Stem Cell.* 2016;19(4):462–75. <https://doi.org/10.1016/j.stem.2016.06.018> PMID: [27452175](#)
27. Sadahiro T, Isomi M, Muraoka N, Kojima H, Haginiwa S, Kurotsu S, et al. Tbx6 induces nascent mesoderm from pluripotent stem cells and temporally controls cardiac versus somite lineage diversification. *Cell Stem Cell.* 2018;23(3):382–395.e5. <https://doi.org/10.1016/j.stem.2018.07.001> PMID: [30100166](#)
28. Söderholm S, Jauregi-Miguel A, Pagella P, Ghezzi V, Zambanini G, Nordin A, et al. Single-cell response to Wnt signaling activation reveals uncoupling of Wnt target gene expression. *Exp Cell Res.* 2023;429(2):113646. <https://doi.org/10.1016/j.yexcr.2023.113646> PMID: [37271249](#)
29. Pagella P, Söderholm S, Nordin A, Zambanini G, Ghezzi V, Jauregi-Miguel A, et al. The time-resolved genomic impact of Wnt/ $\beta$ -catenin signaling. *Cell Syst.* 2023;14(7):563–581.e7. <https://doi.org/10.1016/j.cels.2023.06.004> PMID: [37473729](#)
30. Umez T, Yamanouchi H, Iida Y, Miura M, Tomooka Y. Follistatin-like-1, a diffusible mesenchymal factor determines the fate of epithelium. *Proc Natl Acad Sci USA.* 2010;107(10):4601–6. <https://doi.org/10.1073/pnas.0909501107>
31. Nakajima T, Iguchi T, Sato T. Retinoic acid signaling determines the fate of uterine stroma in the mouse Müllerian duct. *Proc Natl Acad Sci U S A.* 2016;113(50):14354–9. <https://doi.org/10.1073/pnas.1608808113> PMID: [27911779](#)
32. Chi RPA, Wang T, Adams N, Wu SP, Young SL, Spencer TE, et al. Human endometrial transcriptome and progesterone receptor cistrome reveal important pathways and epithelial regulators. *J Clin Endocrinol Metab.* 2020;105: E1419–E1439. <https://doi.org/10.1210/clinem/dgz117>
33. Nakajima T, Kozuma M, Hirasawa T, Matsunaga YT, Tomooka Y. Extracellular matrix components and elasticity regulate mouse vaginal epithelial differentiation induced by mesenchymal cells. *Biol Reprod.* 2021;104(6):1239–48. <https://doi.org/10.1093/biolre/ioab041> PMID: [33693507](#)
34. Bellessort B, Le Cardinal M, Bachelot A, Narboux-Nême N, Garagnani P, Pirazzini C, et al. Dlx5 and Dlx6 control uterine adenogenesis during post-natal maturation: Possible consequences for endometriosis. *Hum Mol Genet.* 2016;25(1):97–108. <https://doi.org/10.1093/hmg/ddv452> PMID: [26512061](#)
35. Liu B, Chen G, He Q, Liu M, Gao K, Cai B, et al. An HMGA2-p62-ER $\alpha$  axis regulates uterine leiomyomas proliferation. *FASEB J.* 2020;34(8):10966–83. <https://doi.org/10.1096/fj.202000520R> PMID: [32592217](#)
36. Li Y, Qiang W, Griffin BB, Gao T, Chakravarti D, Bulun S, et al. HMGA2-mediated tumorigenesis through angiogenesis in leiomyoma. *Fertil Steril.* 2020;114(5):1085–96. <https://doi.org/10.1016/j.fertnstert.2020.05.036> PMID: [32868105](#)
37. Radic MZ, Saghbini M, Elton TS, Reeves R, Hamkalo BA. Hoechst 33258, distamycin A, and high mobility group protein I (HMG-I) compete for binding to mouse satellite DNA. *Chromosoma.* 1992;101(10):602–8. <https://doi.org/10.1007/BF00360537> PMID: [1385053](#)
38. Narita M, Narita M, Krizhanovsky V, Nuñez S, Chicas A, Hearn SA, et al. A novel role for high-mobility group a proteins in cellular senescence and heterochromatin formation. *Cell.* 2006;126(3):503–14. <https://doi.org/10.1016/j.cell.2006.05.052>