BRIEF REPORT

REVISED Genome-wide screening of upstream transcription factors using an expression library [version 2; peer review: 2 approved]

Naoya Yahagi, Yoshinori Takeuchi

Nutrigenomics Research Group, Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki, 3058575, Japan

v2 **First published:** 28 Jan 2021, **10**:51
https://doi.org/10.12688/f1000research.27532.1
Latest published: 08 Mar 2021, **10**:51
https://doi.org/10.12688/f1000research.27532.2

Abstract

The identification of upstream transcription factors regulating the expression of a gene is generally not an easy process. To facilitate this task, we constructed an expression cDNA library named Transcription Factor Expression Library (TFEL), which is composed of nearly all the transcription factors in the mouse genome. Genome-wide screening using this library (TFEL scan method) enables us to easily identify transcription factors controlling any given promoter or enhancer of interest in a chromosomal context-dependent manner. Thus, TFEL scan method is a powerful approach to explore transcriptional regulatory networks.

Keywords

Transcription, Transcription factor, Expression cloning

This article is included in the University of Tsukuba gateway.

Open Peer Review

Reviewer Status ✓ ✓

	Invited Reviewers	
	1	**2**
version 2 (revision) 08 Mar 2021		✓ report
version 1 28 Jan 2021	✓ report	? report

1. **Kazuhiko Igarashi**, Tohoku University Graduate School of Medicine, Sendai, Japan
2. **Harukazu Suzuki**, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Naoya Yahagi (nyahagi-tky@umin.ac.jp)

Author roles: Yahagi N: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Takeuchi Y**: Investigation, Visualization

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by MEXT/JSPS KAKENHI through a Grant-in-Aid for Scientific Research on Innovative Areas: Crosstalk of transcriptional control and energy pathways by hub metabolites [23116006], Grant-in-Aid for Scientific Research (B) [15H03092], Grant-in-Aid for Scientific Research (C) [21591123 and 18590979], Grant-in-Aid for Challenging Exploratory Research[26560392 and 16K13040], and Grant-in-Aid for JSPS Fellows [03J10558] to NY. It was also supported by research grants from the Uehara Memorial Foundation, Nakatani Foundation, the Naito Foundation, ONO Medical Research Foundation, Takeda Science Foundation, Suzuken Memorial Foundation, Japan Heart Foundation, Kanae Foundation for the Promotion of Medical Science, Senri Life Science Foundation, SENSHIN Medical Research Foundation, Japan Foundation for Applied Enzymology, and Okinaka Memorial Institute for Medical Research to NY.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Yahagi N and Takeuchi Y. **Genome-wide screening of upstream transcription factors using an expression library [version 2; peer review: 2 approved]** F1000Research 2021, **10**:51 https://doi.org/10.12688/f1000research.27532.2

First published: 28 Jan 2021, **10**:51 https://doi.org/10.12688/f1000research.27532.1

REVISED **Amendments from Version 1**

The revision changes are:

1. Some supplementary explanations including a description of statistical analysis are added to the legend of Figure 2.

2. Discussions on

- TFs matching to enhancers

- TFs activated by ligands have been added.

Any further responses from the reviewers can be found at the end of the article

Abbreviations

TFEL, transcription factor expression library; VEGF, vascular endothelial growth factor; Fasn, fatty acid synthase; SREBP, sterol regulatory element-binding protein; HIF, hypoxia-inducible factor; LXR, liver X receptor; KLF, Kruppel-like factor.

Introduction

Transcriptional regulation is widely involved in many biological processes of relevance such as cellular development, tissue differentiation, reprogramming and apoptosis as well as nutrient metabolism; in many cases, such processes involve complicated regulatory networks[1]. Generally, downstream targets of a transcription factor can be identified in a convincing way through transcriptome analyses after overexpression or knockdown of the transcription factor, or through investigating the binding interaction between the transcription factor and genomic DNA using chromatin immunoprecipitation (ChIP)-sequencing method. However, no assured methods of identifying upstream transcription factors regulating a specific promoter or enhancer exists, since the relatively poor amount of transcription factors expressed in protein pools or cDNA libraries makes purification difficult, while binding factor prediction using computer analysis based on binding motif databases suffers from high false-positive and false-negative rates in general[2]. Therefore, we constructed a comprehensive expression library of transcription factors in the mouse genome, the Transcription Factor Expression Library (TFEL), and developed a new method of identifying upstream transcription factors through an expression cloning technique using the TFEL (TFEL scan method). Using this TFEL scan method, we can easily identify the specific transcription factor or transcription factor complex trans-acting against any given cis-element of interest.

Methods

Retrieving clones from DNABook

The mouse transcription factor DNABook™ was purchased from DNAFORM, Tokyo, Japan[3]. Each spot on the sheets of DNABook was punched, and plasmid DNA was eluted in 10μl distilled water. Competent DH5α E. coli was transformed with 1μl of the eluted DNA solution, and plasmids were cloned in a standard manner.

Among 1,588 clones in the DNABook, 1,350 clones were on pFLCI vector, 204 clones on modified BluescriptI(+), 23 clones on pFLCIII, and 11 clones were on pFLCII.

The clones on pFLCI were digested by restriction enzymes EcoRI and BamHI, clones on pFLCII by XhoI and BamHI, pFLCIII by Bsu36I and BamHI, and clones on modified BluescriptI(+) were cut by SacI and XhoI on a 37 °C heat block. The restricted DNA fragments were separated on agarose gel (Cat. # 5091, Agarose Basic TAKARA) by electrophoresis, and verified if they are correspondent with the predicted band pattern calculated from the database (GenBank / EMBL / DDBJ) sequence. Unfortunately, information about which clones are included in the DNABook cannot be made public due to a confidentiality agreement. For more details, please contact the corresponding author via email (nyahagi-tky@umin.ac.jp)

Construction of TFEL

TFEL was constructed by transferring all the DNABook clones into pcDNA3.1 expression vector (Invitrogen). To facilitate this process, we first modified pcDNA3.1 vector and inserted the PCR-amplified kanamycin-resistant gene from pENTR4 vector (Invitrogen) into the original pcDNA3.1 at the SmaI-SalI sites by replacement of neomycin-resistant gene to make the vector resistant to kanamycin. By this modification, we could ligate an insert DNA fragment to the pcDNA3.1 vector directly without agarose gel isolation of the insert from vector backbone fragment and undigested plasmid after restriction enzyme digestion of DNABook clones, which are ampicillin-resistant.

Because DNABook clones were on as many as 23 versions of various vectors (pFLCI: 12, pFLCII: 3, pFLCIII: 6, and modified BluescriptI(+): 2), multiple strategies based on the calculation of restriction sites were used to construct TFEL clones.

For pFLCI clones, the combination of EcoRI and BamHI digestion was first tried to cut insert fragments out of vectors when inserts were not predicted to be cut by the two restriction enzymes, and excised fragments were inserted into pcDNA3.1(-) vector at EcoRI and BamHI sites. For pFLCI clones whose inserts were cut by EcoRI or BamHI, SfiI digestion and ligation to pcDNA3.1(-)-SfiI (XhoI-EcoRI part of pcDNA3.1(-) multiple cloning site was replaced with synthesized SfiI sequence) was next attempted. For clones to which neither strategy was applicable, either EcoRI-ApaI (into pcDNA3.1(+)), EcoRI-KpnI (into pcDNA3.1(-)), or SacI-BamHI (into pcDNA3.1(+)-SacI-mod whose SacI site is modified to be unique by introducing mutation to the other SacI site outside of the multiple cloning site) strategy was tried. For the other 73 clones to which these strategies were not applicable, SfiI digested multiple fragments were inserted into pcDNA3.1(-)-SfiI at one time. When another restriction enzyme such as ApaI, KpnI, NotI, or ScaI could be used to cut the vector backbone without cutting the insert, this treatment was done at the same time to reduce the possibility of a vector fragment being cloned into pcDNA3.1.

For modified BluescriptI(+) clones, the SfiI (into pcDNA3.1(-)-SfiI) or XhoI-SacI (into pcDNA3.1(-)-SacI-mod) strategy was used. For 47 plasmids on ZA vector whose orientation of the insert is opposite to that of others, corresponding "(+)" version of pcDNA3.1 vectors were used. 7 clones on ZX vector had corrupted 5'-side SfiI and XhoI sites, therefore KpnI-SfiI, KpnI-BglII, or BssHII strategy was used instead.

For pFLCII clones, the XhoI-BamHI strategy was used when applicable. Others were excised with SacI and cloned into SacI site of pcDNA3.1(-)-SacI-mod, after which the inserted direction was checked, and clones constructed in the correct direction was selected.

pFLCIII clones were ligated using I-CeuI and PI-SceI.

Functional screening through luciferase activity

The mouse *Vegfa*-luc plasmid (-1210 to +246) was generated from a PCR-amplified fragment inserted into pGL2-basic vector (Cat. #E1641, Promega) at SmaI site. The PCR primers used were 5'-AAGATGAACCGTAAGCCTAGGCT-3' and 5'-AACCGTTGGCACGATTTAAGA-3' and amplification was performed according to a 3-step method. The rat *Fasn*-luc plasmid (-397 to +28) was prepared as described previously[4]. The mouse *Srebf1c*-luc plasmids (-2.2k to +40 and -249 to -144) were constructed as described elsewhere[5,6].

To screen transcription factors regulating a target promoter, TFEL clones were co-transfected with the specific promoter-luc plasmid into HEK293 cells using SuperFect Transfection Reagent (Cat. #301305, QIAGEN). The screenings were performed by co-transfecting 10–20 clones with each luciferase plasmid per one well. Transfection was performed using 0.5μg of plasmid DNA pooled equally from 10 or 20 TFEL clones per well. The luciferase activity in transfectants was measured on a luminometer (BERTHOLD) with a standard assay kit (Cat. #E1483, Promega). First screen was performed with 10 or 20-clone pools per well, and significantly shifted pools were further tested with one clone at a time.

Data analysis

All the data from a luminometer were output as Excel sheets, and directly input to graphs. All the graphs were drawn in Microsoft Excel 2016 (16.0.5083.1000).

Results

Identification of binding transcription factor

To construct the TFEL expression plasmid library, we used clones from the RIKEN FANTOM libraries (Figure 1A). The clone subset consisting of genome-wide transcription factors is available from DNAFORM as Mouse Transcription factor DNABook™, and is comprised of 1,588 nonredundant genome-wide mouse transcription factor genes[7]. We retrieved all the clones in the DNABook using *Escherichia coli* transformation. After checking the band patterns using several restriction enzymes and sequencing if necessary, each excised insert DNA fragment was transferred to a pcDNA3.1 expression vector. In this manner, we constructed a pcDNA3.1 expression library composed of 1,588 clones of genome-wide and nonredundant mouse transcription factors.

Next, we evaluated the validity of this expression cloning method (TFEL scan method) using the TFEL (Figure 1B[8]). Figure 2A–C[8] shows three cases in which the upstream transcription factors influencing well -known promoters were examined: A, vascular endothelial growth factor (gene name: *Vegfa*), B, fatty acid synthase (gene name: *Fasn*), and C, sterol regulatory element-binding protein (SREBP)-1c (gene name: *Srebf1c*). In these cases, well-known determinant transcription factors (HIF-1α and -2α for *Vegfa*-luc, SREBP-1a and -2 for *Fasn*-luc, and LXRα for *Srebf1c*-luc) were identified as expected. These screenings were performed by co-transfecting 10–20 clones with each luciferase plasmid per one well.

These results demonstrated that the identification of upstream transcription factors using this expression screening method works well.

Identification of transcription factor complex

We further attempted to identify unknown partner(s) interacting with a specific transcription factor on a specific gene promoter. To this end, we screened interacting partner(s) for LXR on the SREBP-1c promoter[6]. As shown in Figure 2D[9], KLF4 was picked up as a candidate from the first screening performed by co-transfecting 20 clones with an SREBP-1c promoter luciferase plasmid in the presence of an LXR ligand T0901317. Further screening of the clones one by one revealed that KLF4 and KLF15 interacted with and suppressed LXR on the SREBP-1c promoter (data not shown, see *ref* 6). Thus, the TFEL scan method was proved to be useful for searching the interacting partner against a specific transcription factor of interest in a DNA-sequence-dependent manner.

Discussion

In the present study, we clearly demonstrated that our new method of TFEL scan is a powerful approach to explore transcriptional regulatory networks by identifying transcription factor(s) involved in the regulation.

First, we showed that in three cases (*Vegfa*-luc, *Fasn*-luc and *Srebf1c*-luc), we could easily identify well-known determinant transcription factors (HIF-1α and -2α for *Vegfa*-luc, SREBP-1a and -2 for *Fasn*-luc, and LXRα for *Srebf1c*-luc) as expected (Figure 2A–C[8]). In addition, we recently succeeded in identifying the binding transcription factor for an important SNP at the diabetes-associated *TCF7L2* gene locus through our TFEL scan method[10]. It is suggested that *TCF7L2* is the single largest effect of a common SNP on type 2 diabetes risk in European populations[11].

Thus, we demonstrated that the TFEL scan method works very well to identify the regulatory transcription factor in a simple model. It is noteworthy that the TFEL scan method is also applicable to enhancers as well as promoters as shown in the case of *TCF7L2* gene.

Many of the transcription factors characterized thus far are responsive to stresses or ligands. The results of the case of LXR shown here demonstrate that such transcription factors can also be identified by overexpression. However, at the same time, we have experienced that some transcription factors activated in a ligand-dependent manner, for example, glucocorticoid receptor (GR), cannot be identified by mere overexpression without ligands (data not shown). Therefore,

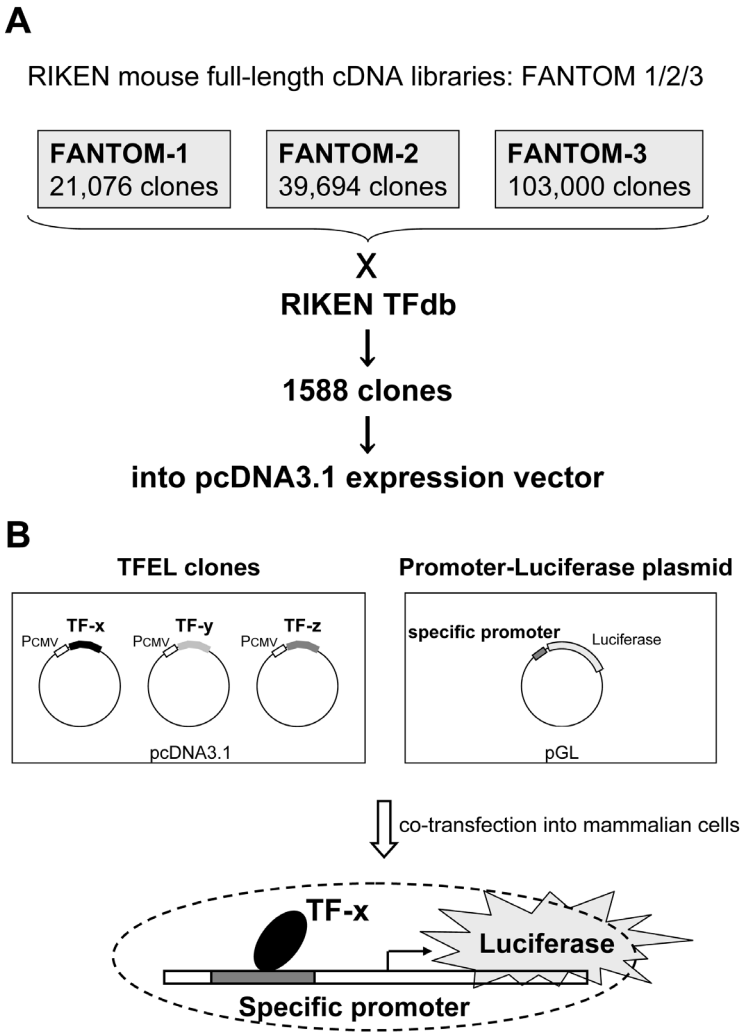

Figure 1. (**A**) Construction of Transcription Factor Expression Library (TFEL). The original clones were derived from the RIKEN FANTOM1/2/3 libraries, and 1,588 clones that are predicted to be transcription factor genes by RIKEN TFdb are distributed as Mouse Transcription factor DNABook™ from DNAFORM Inc. TFEL consists of the DNABook clones transferred into pcDNA3.1 expression vectors through recombination using several restriction enzymes optimized to each clone. (**B**) Workflow of identifying upstream transcription factors using TFEL. A luciferase plasmid with a specific promoter of interest is co-transfected with the TFEL clone(s) into mammalian cells and the luciferase reporter activities are measured.

it may depend on transcription factors whether they can be detected without ligands/additional activators or not.

Next, we additionally succeeded in identifying transcription factors forming a complex such as LXR and KLF4 (Figure 2D[8]). Therefore, the TFEL scan method was proved to be similarly effective even in these more complex cases.

It is well known that the transcriptional output of a gene is due to the joint activity of many transcription factors, the binding and activation of which are highly interdependent[1]. This cooperation is often mediated by direct physical contact between two or more transcription factors, forming homodimers, heterodimers, or larger transcriptional complexes[12]. In fact, it has been estimated that approximately 75% of all metazoan transcription factors heterodimerize with other factors[13]. Thus, the importance of transcription factor combinations has been highlighted more and more, and the mapping of the combinatorial interactions among transcription factors has been attempted using a mammalian two-hybrid system[12]. Our successful results identifying KLFs as interacting partners of LXR on *Srebf1c* promoter clearly showed that the expression cloning method using the TFEL (TFEL scan method) may also be useful for exploring interacting partner(s) among transcription factors on a specific DNA fragment of interest. In contrast, binding site prediction based on binding motif databases suffers from high false-positive and false-negative rates in general[9]. In particular, the sequence-based approach has no power to predict transcription

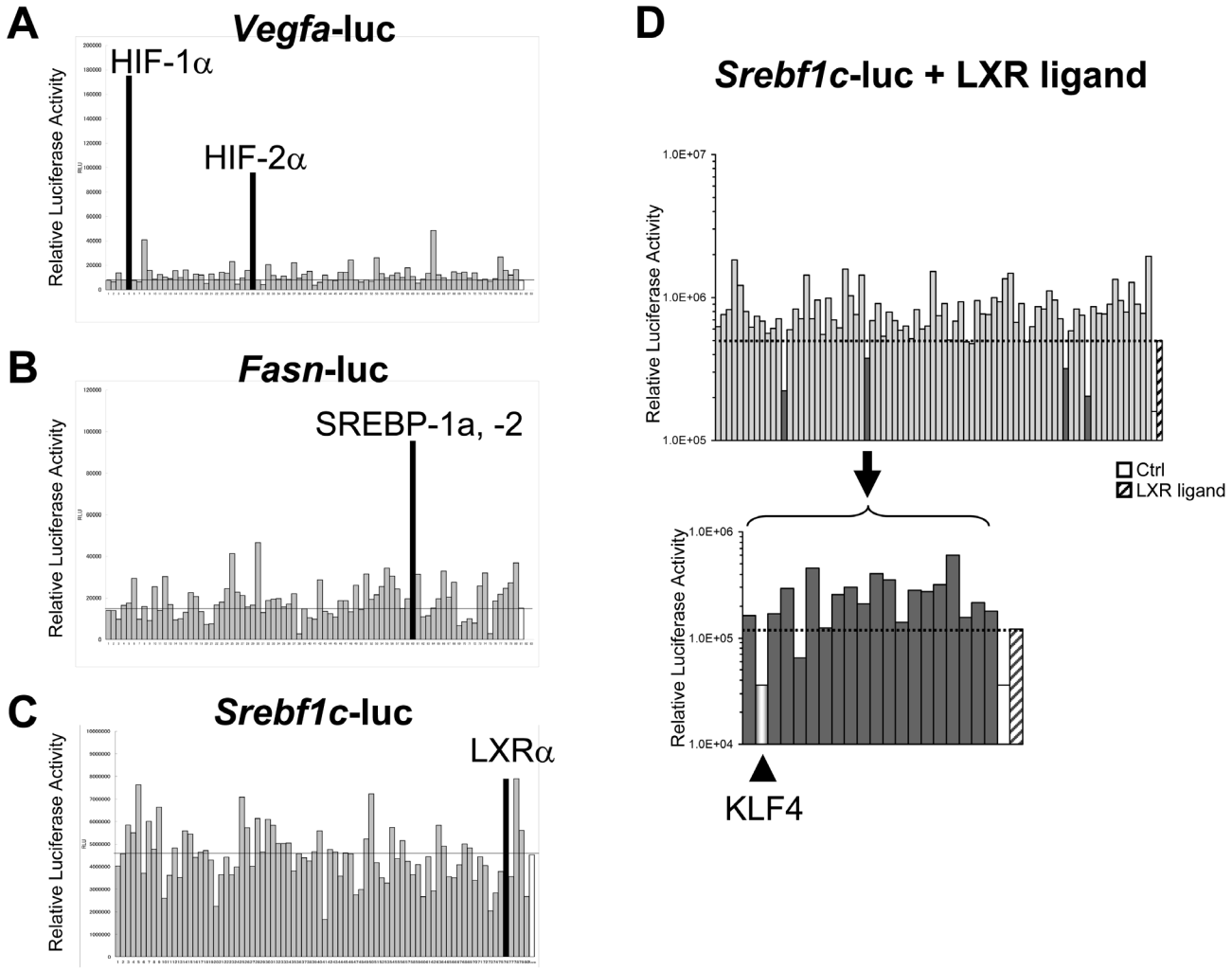

Figure 2. (**A**–**C**) Representative cases of the analysis of upstream transcription factors using Transcription Factor Expression Library (TFEL) against specific promoters: (**A**) vascular endothelial growth factor (gene name: *Vegfa*), (**B**) fatty acid synthase (gene name: *Fasn*), and (**C**) sterol regulatory element-binding protein-1c (gene name: *Srebf1c*). Transfection was performed using 0.5μg of plasmid DNA pooled equally from 10 or 20 TFEL clones per well. (**D**) Representative case identifying the combinatorial interactions on a specific promoter. In this assay, we screened for transcription factor(s) that suppress LXR activity on the SREBP-1c promoter (-249 to -144), because we had found an important element shown to exert a suppressive effect on the SREBP-1c promoter in a fasting state[6]. To activate LXR and to improve the sensitivity of the assay, we added an LXR ligand T0901317 to the culture media. The first screening was performed using a 20-clone pool assay, and individual clones were examined in the second screening. From the statistical point of view, we picked up the top (or the bottom for D) 5% pools for the 2nd screening, and as a result, this criterion was similar to the 2SD (standard deviation) criterion in these cases.

factor complexes formed by protein-protein interaction. Conversely, the search for protein-protein interactions among transcription factors alone, for example, using a mammalian two-hybrid system, may not be sufficient to elucidate the regulatory complexes in certain situations including the present case of LXR-KLFs complex, because complexes like LXR-KLFs are also dependent on DNA binding and therefore locus-specific. This relatively weak interaction supported by the DNA backbone enables gene-specific regulations and may possibly give more diversity to transcriptional networks. Thus, our strategy of screening for the transcription factor complex in a chromosomal context-dependent manner using the TFEL scan method can be a very effective and powerful approach for exploring sophisticated transcriptional networks in detail, because this screening system is performed under more physiological conditions than traditional methods.

Our initial approach using a mixed pool assay handling 20 clones at a time failed to detect KLF15 in the first screening described above. Interference among the co-transfected transcription factors in the pools was thought to have affected the results and veiled the KLF15-LXR interaction, suggesting that a single clone assay is preferable, especially for screening for combinatorial interactions. In relation to this point, interactions

with endogenous transcription factors should also affect the results to varying degrees, and screenings in different cell lines should provide more information.

Our TFEL library is based on the Mouse Transcription factor DNABook™, and a small portion of transcription factors are missing, as they previously reported[7]. For example, among the 17 KLF family members, 4 KLFs (KLF7, 12, 14 and 17) are missing. We are planning to fix this problem and to complete the construction of a comprehensive expression library in the near future.

In summary, the present study clearly demonstrates that our expression cloning method using the TFEL (TFEL scan method) enables us to efficiently identify regulatory transcription factors as well as to elucidate combinatorial interactions among transcription factors in a chromosomal context-dependent manner. Thus, the TFEL scan method is a powerful approach to explore transcriptional regulatory networks.

Data availability

Underlying data

Figshare: Underlying data of Figure 2. Representative cases of the analysis of upstream transcription factors using TFEL against specific promoters. https://doi.org/10.6084/m9.figshare.13237319[8]

This project contains the following underlying data:

- Fig2A_090109_BxVEGF1.3k.xls

Representative cases of the analysis of upstream transcription factors using TFEL against specific promoters: (A) vascular endothelial growth factor (gene name: Vegfa).

- Fig2B_081218_AxFAS0.4k.xls

Representative cases of the analysis of upstream transcription factors using TFEL against specific promoters: (B) fatty acid synthase (gene name: Fasn).

- Fig2C_101026_B(10)xSREBP1c2.2k.xls

Representative cases of the analysis of upstream transcription factors using TFEL against specific promoters: (C) sterol regulatory element-binding protein-1c (gene name: Srebf1c).

- Fig2D_081219_AxSREBP1c-T_1st.xls

Representative case identifying the combinatorial interactions on a specific promoter. In this assay, we screened for transcription factor(s) that suppress LXR activity on the SREBP-1c promoter (-249 to -144). 1st screening data.

- Fig2D_081225_SREBP1c-T_2nd_A28.xls

Representative case identifying the combinatorial interactions on a specific promoter. In this assay, we screened for transcription factor(s) that suppress LXR activity on the SREBP-1c promoter (-249 to -144). 2nd screening data.

Acknowledgements

We thank Yuji Teraoka, Nobue Suzuki, and Kazuko Shibahara for skilled technical assistance.

References

1. Lambert SA, Jolma A, Campitelli LF, *et al.*: **The Human Transcription Factors.** *Cell.* 2018; **172**(4): 650–665.
PubMed Abstract | Publisher Full Text

2. Wasserman WW, Sandelin A: **Applied bioinformatics for the identification of regulatory elements.** *Nat Rev Genet.* 2004; **5**(4); 276–287.
PubMed Abstract | Publisher Full Text

3. Carninci P, Kasukawa T, Katayama S, *et al.*: **The transcriptional landscape of the mammalian genome.** *Science.* 2005; **309**(5740): 1559–1563.
PubMed Abstract | Publisher Full Text

4. Amemiya-Kudo M, Shimano H, Hasty AH, *et al.*: **Transcriptional activities of nuclear SREBP-1a, -1c, and -2 to different target promoters of lipogenic and cholesterogenic genes.** *J Lipid Res.* 2002; **43**(8): 1220–1235.
PubMed Abstract | Publisher Full Text

5. Takeuchi Y, Yahagi N, Nakagawa Y, *et al.*: ***In vivo* promoter analysis on refeeding response of hepatic sterol regulatory element-binding protein-1c expression.** *Biochem Biophys Res Commun.* 2007; **363**(2): 329–335.
PubMed Abstract | Publisher Full Text

6. Takeuchi Y, Yahagi N, Aita Y, *et al.*: **KLF15 enables rapid switching between lipogenesis and gluconeogenesis during fasting.** *Cell Rep.* 2016; **16**(9): 2373–2386.
PubMed Abstract | Publisher Full Text | Free Full Text

7. Kanamori M, Konno H, Osato N, *et al.*: **A genome-wide and nonredundant mouse transcription factor database.** *Biochem Biophys Res Commun.* 2004; **322**(3): 787–793.
PubMed Abstract | Publisher Full Text

8. Yahagi N: **Underlying data of Figure 2. Representative cases of the analysis of upstream transcription factors using TFEL against specific promoters.** *figshare.* Dataset. 2020.
http://www.doi.org/10.6084/m9.figshare.13237319.v1

9. Kim TM, Park PJ: **Advances in analysis of transcriptional regulatory networks.** *Wiley Interdiscip Rev Syst Biol Med.* 2011; **3**(1): 21–35.
PubMed Abstract | Publisher Full Text

10. Piao X, Yahagi N, Takeuchi Y, *et al.*: **A candidate functional SNP rs7074440 in *TCF7L2* alters gene expression through C-FOS in hepatocytes.** *FEBS Lett.* 2018; **592**(3): 422–433.
PubMed Abstract | Publisher Full Text

11. Zeggini E, Weedon MN, Lindgren CM, *et al.*: **Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes.** *Science.* 2007; **316**(5829): 1336–1341.
PubMed Abstract | Publisher Full Text | Free Full Text

12. Ravasi T, Suzuki H, Cannistraci CV, *et al.*: **An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man.** *Cell.* 2010; **140**(5): 744–752.
PubMed Abstract | Publisher Full Text | Free Full Text

13. Walhout AJM: **Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping.** *Genome Res.* 2006; **16**(12): 1445–1454.
PubMed Abstract | Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 16 March 2021

https://doi.org/10.5256/f1000research.55150.r81021

✔ **Harukazu Suzuki**
Laboratory for Cellular Function Conversion Technology, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

I have no additional comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Regulation of gene expression, TF-mediated DNA demethylation, Cell reprogramming

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 19 February 2021

https://doi.org/10.5256/f1000research.30430.r79133

? **Harukazu Suzuki**
Laboratory for Cellular Function Conversion Technology, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

This paper describes a method to identify transcription factors (TFs) regulating target gene(s). The

expression constructs that authors has made are highly useful for scientific community who are interested in TF-mediated gene regulation. The manuscript is written well, but there are several issues to be addressed.

1. Identification of upstream TF(s) should be objective. Test cases shown in Figure 2A, B and C should be statistically evaluated. HIF-1a,2a and SREBP-1a,-2 may be evident, however I am not sure whether LXRa-mediated luciferase activity is significantly high or not.

2. In figure 2D, authors selected 4 pools. I think this selection should also be objective by using statistical evaluation. If authors statistically selected pools for further one by one test, the pool with KLF4 has been selected?

3. One of the most important value in this work must be TF expression plasmids that authors have systematically constructed. I think this resource should be available to scientific community.

4. TFs regulate downstream genes by associating with promoters or enhancers. It is worth if the screen system works for TFs associating with enhancers.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
No

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Regulation of gene expression, TF-mediated DNA demethylation, Cell reprogramming

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 02 Mar 2021

Naoya Yahagi, University of Tsukuba, Tsukuba, Japan

1. Identification of upstream TF(s) should be objective. Test cases shown in Figure 2A, B and C should be statistically evaluated. HIF-1a,2a and SREBP-1a,-2 may be evident, however I am not sure whether LXRa-mediated luciferase activity is significantly high or not.

We thank the reviewer for the valuable advice. Because the data have been obtained for the screening purpose, each pool has just single value, so robust statistical analyses to show significance are a little difficult to perform. However, as an empirical method, 2SD (standard deviation) criteria is often used. For Figure 2A (Vegfa-luc), only the pools #5 and #29 (black bars; which means that they actually contain positive clones) are above the 2SD line. For Figure 2B (Fasn-luc), the pools #30 and #60 (black bar) are above the 2SD line, but the pool #30 was revealed to contain no positive clones after the 2nd screening of each clone in the pool. For Figure 2C (Srebf1c-luc), the pools #5, #50, #76 (black bar) and #78 are above the 2SD line, but only the pool #76 was revealed to have a positive clone (LXRα) after the 2nd screening. Our actual strategy was to pick up the top 5% (or the bottom 5% for the search of suppressors as shown in Figure D) for the 2nd screening. As a result, both approaches (the 2SD criterion and the top 5% criterion) are proved to work sufficiently as a screening method.

Based on this discussion, we added the following paragraph to the revised manuscript:
From the statistical point of view, we picked up the top (or the bottom for D) 5% pools for the 2nd screening, and as a result, this criterion was similar to the 2SD (standard deviation) criterion in these cases.

2. In figure 2D, authors selected 4 pools. I think this selection should also be objective by using statistical evaluation. If authors statistically selected pools for further one by one test, the pool with KLF4 has been selected?

Thank you for the question. We selected the top or bottom 5% pools for the 2nd screening, and only the pool #28 containing KLF4 was proved to be a true hit after the 2nd screening.

3. One of the most important value in this work must be TF expression plasmids that authors have systematically constructed. I think this resource should be available to scientific community.

Thank you for the comment. We are planning to talk with DNAFORM about how to share this expression library in the near future.

4. TFs regulate downstream genes by associating with promoters or enhancers. It is worth if the screen system works for TFs associating with enhancers.

We thank the reviewer for the valuable comment. Yes, we think that our method is also effective for enhancers to be matched with TFs, as shown in the case of TCF7L2 gene (Piao X, et al, FEBS Lett, 2018;592(3):422-433 (ref. 9)). According to the reviewer's advice, we added

the following sentence to the revised manuscript:
It is noteworthy that the TFEL scan method is also applicable to enhancers as well as promoters as shown in the case of *TCF7L2* gene.

Competing Interests: No competing interests were disclosed.

Reviewer Report 09 February 2021

https://doi.org/10.5256/f1000research.30430.r78980

Kazuhiko Igarashi
Department of Biochemistry, Tohoku University Graduate School of Medicine, Sendai, Japan

This is a very important resource paper for the fields of gene regulation, metabolism, differentiation and stress response. Overall, the manuscript is well written and the results are clearly presented. There are some issues readers may like to know better.

1. Human genome is estimated to encode for 1639 transcription factors (Lambert SA et al, Cell 172, 650-, 2018[1]). How much of the mouse transcription factors in the expression library is overlapping with the human transcription factors. A brief discussion will be helpful.

2. Many of the transcription factors characterised thus far are responsive to stresses or ligands. The results in this manuscript show that such transcription factors can be identified by overexpression. A brief discussion is needed.

3. How will this expression library be shared?

4. The description in the text of experiment in Fig. 2D may be improved. Did authors try to identify factors that restrict/inhibit LXR? Why was this important for the regulation of Srebf1c promoter?

References

1. Lambert S, Jolma A, Campitelli L, Das P, et al.: The Human Transcription Factors. *Cell*. 2018; **172** (4): 650-665 Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: gene regulation by transcription factors, their functions in cell differentiation and diseases.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 02 Mar 2021

Naoya Yahagi, University of Tsukuba, Tsukuba, Japan

1. Human genome is estimated to encode for 1639 transcription factors (Lambert SA et al, Cell 172, 650-, 20181). How much of the mouse transcription factors in the expression library is overlapping with the human transcription factors. A brief discussion will be helpful.

We thank the reviewer for the valuable advice. According to this suggestion, we checked the overlapping among the transcription factor databases such as RIKEN TFdb (the original database that the TFEL is based on; Kanamori M et al, Biochem Biophys Res Commun. 2004;322(3):787-793 (ref.7)), Human TF (Lambert SA et al, Cell, 2018;172(4):650-665.), and Animal TFDB 3.0 (Hu H et al, Nucleic Acids Research, 2019;47(D1):D33-D38.), using a gene symbol converter available at http://biodb.jp/. The results were as follows: the overlap between RIKEN TFdb and Human TF was 889 genes, app 65% of Human TF and 53% of RIKEN TFdb, and the overlap between RIKEN TFdb and Animal TFDB 3.0 was 958 genes, app 59% of Animal TFDB 3.0 and 57% of RIKEN TFdb. The differences were a little larger than expected and even when we compared the same Animal TFDB human and mouse databases, the overlap was limited to 1,340 genes, which were app. 80% of the 1665 human TF genes in the database. Therefore, app. 300 TF genes (20%) seemed to be different between human and mouse. Thanks to the reviewer's comment, we could better understand the differences among TF databases, but because this content is beyond the scope of this article, we will keep it within the Comments. Also, we realized during these series of gene symbol checking that KLF 9, 10, 11 had been included in the TFEL library, and corrected the description in the discussion.

2. Many of the transcription factors characterized thus far are responsive to stresses or ligands. The results in this manuscript show that such transcription factors can be identified by overexpression. A brief discussion is needed.

We thank the reviewer for the valuable comment. According to the reviewer's advice, we added a new paragraph of discussion to the revised manuscript like this:
Many of the transcription factors characterized thus far are responsive to stresses or ligands. The results of the case of LXR shown here demonstrate that such transcription factors can also be identified by overexpression. However, at the same time, we have experienced that some transcription factors activated in a ligand-dependent manner, for example, glucocorticoid receptor (GR), cannot be identified by mere overexpression without ligands (data not shown). Therefore, it may depend on transcription factors whether they can be detected without ligands/additional activators or not.

3. How will this expression library be shared?

Thank you for the comment. We are planning to talk with DNAFORM about how to share this expression library in the near future.

4. The description in the text of experiment in Fig. 2D may be improved. Did authors try to identify factors that restrict/inhibit LXR? Why was this important for the regulation of Srebf1c promoter?

Thank you for the question. Actually, there is a long story behind this. To make the long story short, we had found an important element shown to exert a suppressive effect on the SREBP-1c promoter in a fasting state (Takeuchi Y, et al, Cell Rep, 2016;16(9):2373-2386 (ref. 6)). According to the reviewer's advice, we added this explanation to the revised manuscript.

Competing Interests: No competing interests were disclosed.