

植物における組織特異的な転写制御機構と 1 細胞トランスクリプトームデータの時系列解析アルゴリズム

鳥井孝太郎、遠藤求
京都大学 生命科学研究科 分子代謝制御学分野
〒606-8502 京都市左京区北白川追分町

Tissue specific transcriptional regulation in plants
and time series analysis tool for single cell data

Keyword: transcriptional regulation, cis-element, circadian clock, tissue specificity, single
cell transcriptome

Kotaro tori & Motomu Endo
Graduate School of Biostudies, Kyoto University
Kitashirakawa-oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan

DOI: 10.24480/bsj-review.8b5.00115

1 RNA の転写制御におけるシス配列の重要性

植物が遺伝子の機能を適切に発揮するためには、適切な遺伝子発現制御が必須である。遺伝子発現制御の過程には、転写因子による転写制御のほかに、RNaseA・miRNA・ナンセンス変異依存 mRNA 分解機構などによる RNA の分解、メチル化や塩基の再構成による RNA 修飾、DNA やヒストンの修飾に見られるエピジェネティックな制御などの多くの制御段階が存在している。こうした多様な制御段階の中でも、転写因子による転写制御は RNA の機能発現において最も基本的かつ重要な制御段階である。モデル植物シロイヌナズナでは、約 2,000 の転写因子が存在し、それらはシス配列と呼ばれる数塩基からなモチーフ配列に結合することで、様々な転写制御を行っていると考えられている (Vandepoele et al. 2009)。

移動できない植物は、周囲の環境変化に鋭敏に応答する必要があるため、光・温度・水分・塩・過酸化物質などの環境刺激に応答するための特異的なシス配列を発達させてきた。光応答性の遺伝子はプロモーター上に G-box (CACGTG) と呼ばれるシス配列を持つことが知られており、シロイヌナズナの約 3% の遺伝子がプロモーター領域に G-box を持っている (Michael and McClung 2003) (図 1)。また、低温刺激や乾燥刺激・塩ストレスの応答に関わるシス配列として C-repeat (CRT)/drought-responsive element (DRE) (A/GCCGAC) が、高温への応答にかかわるシス配列として HS-1 (ATGGGCCCTA) や HS-2(GTT(A/C)TAGA)、過酸化物質の応答にかかわるシス配列として ROS-1 ((TA)₄G(AT)₄) や ROS-2 (TTCAATTT) が報告されている (Rushton et al. 2002, Geisler et al. 2006)。こうした環境変化の応答に関わるシス配

列のうち、最も広範な遺伝子発現制御に関わっているものは概日リズム制御に関わるシス配列である。実際にシロイヌナズナの89%以上の遺伝子は明暗条件下で約24時間周期の発現振動を示している(Michael et al. 2008, Endo et al. 2014)。この発現振動には光刺激による発現制御も含まれているものの、光応答性の遺伝子は全遺伝子の約19%であることから、大部分の遺伝子の発現振動は概日時計によるものだと考えられている(Jiao et al. 2005)。

植物の概日時計を構成する時計遺伝子は転写因子をコードしており、朝方に発現ピークを持つ

*CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)*と*LATE ELONGATED HYPOCOTYL (LHY)*は、evening

element (EE; AAAATATCT)と呼ばれるシス配列を認識する(図2)。他にもEEに類似したCCA1-binding site (CBS; AAAAATCT)も報告されており、これらはそれぞれ全遺伝子の25%および35%のプロモーター領域に存在している(Wang et al. 1997, Harmer et al. 2000, Michael and McClung 2003)。CCA1/LHYは昼頃に発現ピークを持つ*PSEUDO RESPONSE REGULATOR 9 (PRR 9)/ PRR7/ PRR5*遺伝子群などの時計遺伝子の転写を促すと同時に、夕方に発現ピークを持つ*TIMING OF CAB EXPRESSION 1 (TOC1)*、*LUX ARRHYTHMO (LUX)*、*EARLY FLOWERING 4 (ELF4)*などの発現を抑制する。*TOC1*はTOC1 morning element (TIME; TGTG)と呼ばれるシス配列を認識し、*LUX*はLUX binding site (LBS; GAT(A/T)CG)と呼ばれるシス配列を認識する(Chow et al. 2012, Gendron et al. 2012)。これらの概日時計タンパク質に認識されるシス配列を全て合わせると、先に述べたように全遺伝子のかなりの部分は概日時計制御であると予想される。

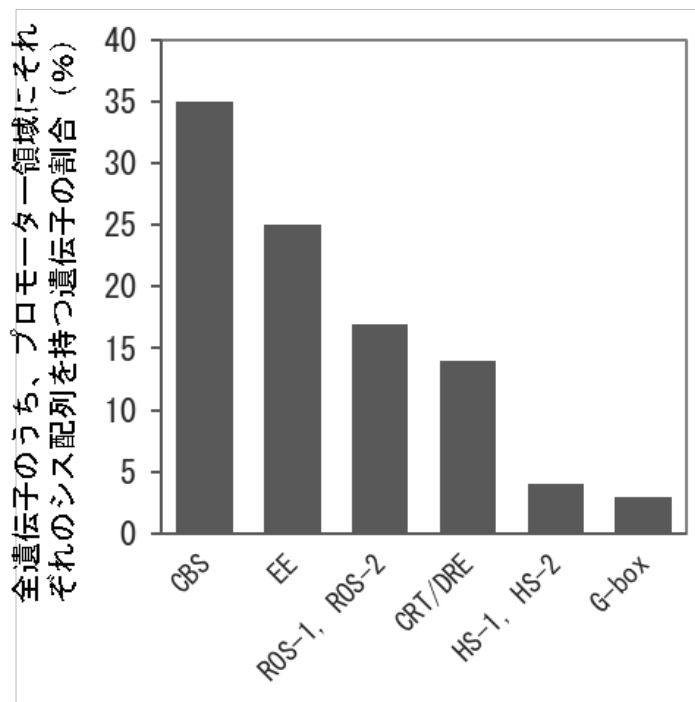


図1. シロイヌナズナにおいて、環境応答にかかわるシス配列をプロモーターに含む遺伝子の割合

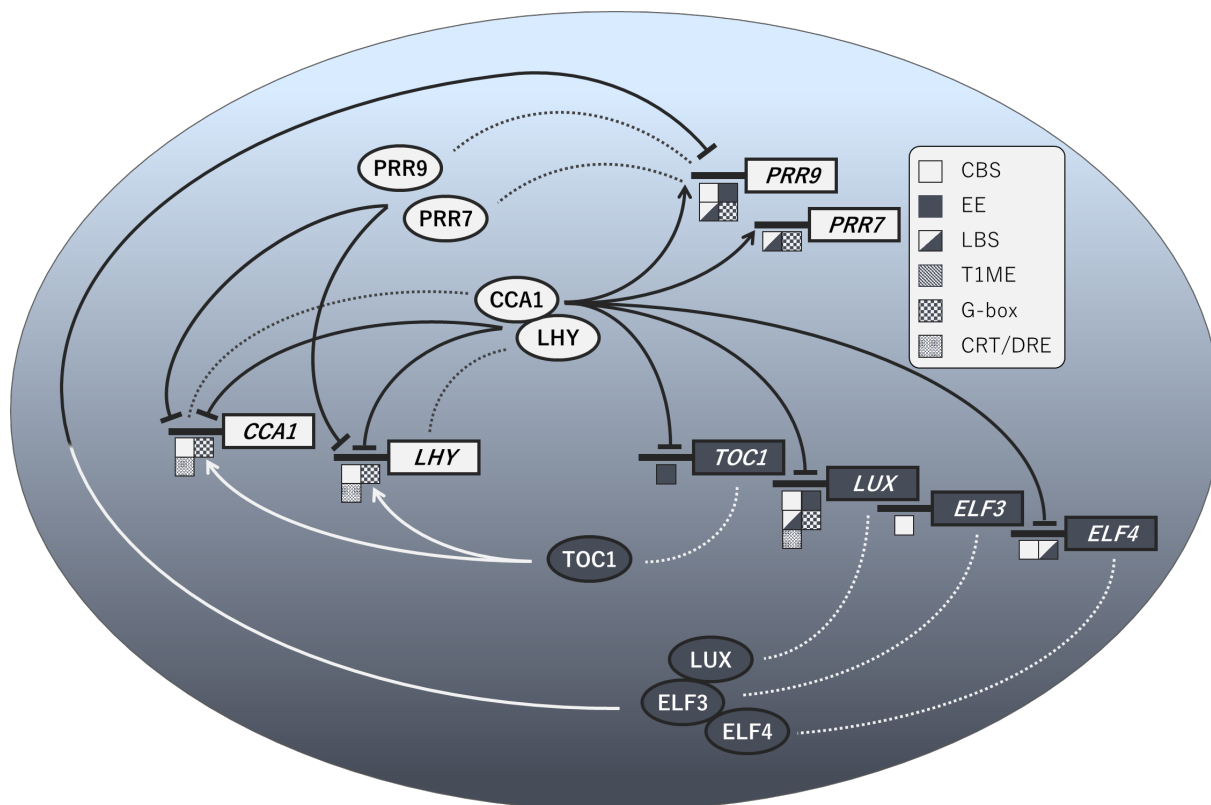


図 2. 概日時計を構成する時計遺伝子間の転写フィードバック

丸はタンパク質、四角は遺伝子を示し、黒で表されているのは夕方から夜にかけて発現する時計遺伝子であり、白は明け方から昼にかけて発現する時計遺伝子を示している。遺伝子の左下にはそれぞれの時計遺伝子に含まれるシス配列を示している。点線は遺伝子の翻訳を意味し、実線は遺伝子の転写制御を表している。

2 転写制御機構の組織特異性

こうした概日リズム制御に関わるシス配列は主に植物個体全体の解析から明らかにされたものである。しかし、シス配列による転写制御には組織特異性が存在することが、近年の組織レベルのトランスクリプトーム解析から明らかにされつつある。維管束で概日リズムを示す遺伝子のプロモーター領域には、long-day vasculature element (LVE ; ACACGG)や short-day vasculature element (SVE ; GCGGGA)といった新しいシス配列が見出されており、これらを持つ遺伝子は葉肉など他の組織では概日リズムを示さない。さらに、telo-box (TBX) や starch box (SBX)、protein box (PBX) といったシス配列をもつ遺伝子は主に葉肉における概日リズムに関わっており、維管束での概日リズム形成には関わらないと考えられる (Michael et al. 2008, Endo et al. 2014)。このような組織ごとに特異的なシス配列の存在は、それを認識する転写因子または機能に組織特異性があるためだと考えられる。実際、時計遺伝子 *ELF4* は維管束で特に高い発現を持っていることが示されており、promoter::GUS レポーターの実験からは *PRR3* や *PRR9* もまた維管束で高い発現を持っていることが示唆されて

いる (Para et al. 2007, Endo et al. 2014)。このような組織特異的な時計遺伝子の発現および組織特異的なシス配列の存在は、組織ごとに異なる概日時計の機能を意味しており、これまでに、根・地上部・気孔・表皮・維管束での特異的な概日リズムが存在することが示されている (Thain et al. 2002, Yarik et al. 2011, Endo et al. 2014)。

こうした組織特異的なシス配列や組織特異的な時計遺伝子の発現はショウジョウバエやマウス、ヒトなどでもその存在が知られており、生物種を超えた普遍的な仕組みであると考えられる (Smith et al. 2007, Kalay and Wittkopp 2010)。また、概日時計の他にも植物の組織特異的な転写制御の例として、菌類の感染に対する転写制御が同一の遺伝子であっても組織ごとに全く異なり、転写制御のリプログラミングが組織特異的に行われることが知られている

(Lyons et al. 2015)。さらに、紫外線・浸透圧・乾燥・低温・高温・傷などのストレスも組織ごとに異なる遺伝子発現を誘導することが知られており、多様なストレス応答においても組織特異的なシス配列の存在が考えられている (Swindell et al. 2007, Singh et al. 2015)。

3 1 細胞解析

このように大部分の遺伝子の転写制御に関わる概日リズムをはじめ、多く生理応答が組織ごとに異なる転写制御を行っていることは、転写制御の解析には組織や細胞の違いを区別して行わなければならないことを意味する。また、近年の解析から、同一と思われていた細胞が異なる機能をもっているケースや細胞には個性があることが相次いで示されていることから、生命の基本単位である細胞レベルの発現解析を行うことで、より多様で精緻な転写制御メカニズムが明らかになると期待されている。幸いにも近年の技術進歩により、植物でも 1 細胞レベルでトランスクリプトーム解析を行うことが可能になってきている (Efroni et al. 2016)。その技術的な詳細についてはすでに、BSJ-Review vol. 7 の中で久保稔博士により詳しく解説されているため (久保 2016)、本稿では割愛し、ここでは 1 細胞 RNA sequencing により得られた遺伝子発現データの時系列解析手法について説明する。

これまで見てきたように転写制御と概日リズムは不可分であるため、個々の細胞で遺伝子発現を測定する際には、ある時点の発現を見るだけでは不十分であり、時系列解析を行うことで遺伝子発現のリズムを計測する必要がある。しかし、現在の 1 細胞解析、特にトランスクリプトーム解析では細胞を破壊して RNA を抽出しているため、同一の細胞から時系列データを得ることは原理上不可能である。こうした問題を解決するため、複数の細胞から得た遺伝子発現データをつなぎ合わせ、疑似的な時系列データを再構成する手法が考案されている。その中で代表的なアルゴリズムである Monocle (Trapnell et al. 2014) は、独立成分分析により遺伝子発現データの次元圧縮を行い、2 次元平面上に投影された各細胞に対し、最小全域木 (MST) を用いて各点をつなぐ辺の総和が最短となる経路を探し出す。次に、細胞をつなぐ MST の長さが最長となる経路を細胞分化の主な軌跡として時系列を構築するアルゴリズムである (図 3)。このように遺伝子発現データから、細胞間の距離 (発現プロファイルの類似度) をもとに、細胞変化の時系列を再構築することで疑似的な時間軸 (pseudo-time) を作成することができる。現在では、pseudo-time を再構築する様々なアルゴリズムが報告されており、解析目的に応じて選択できる (表 1)。

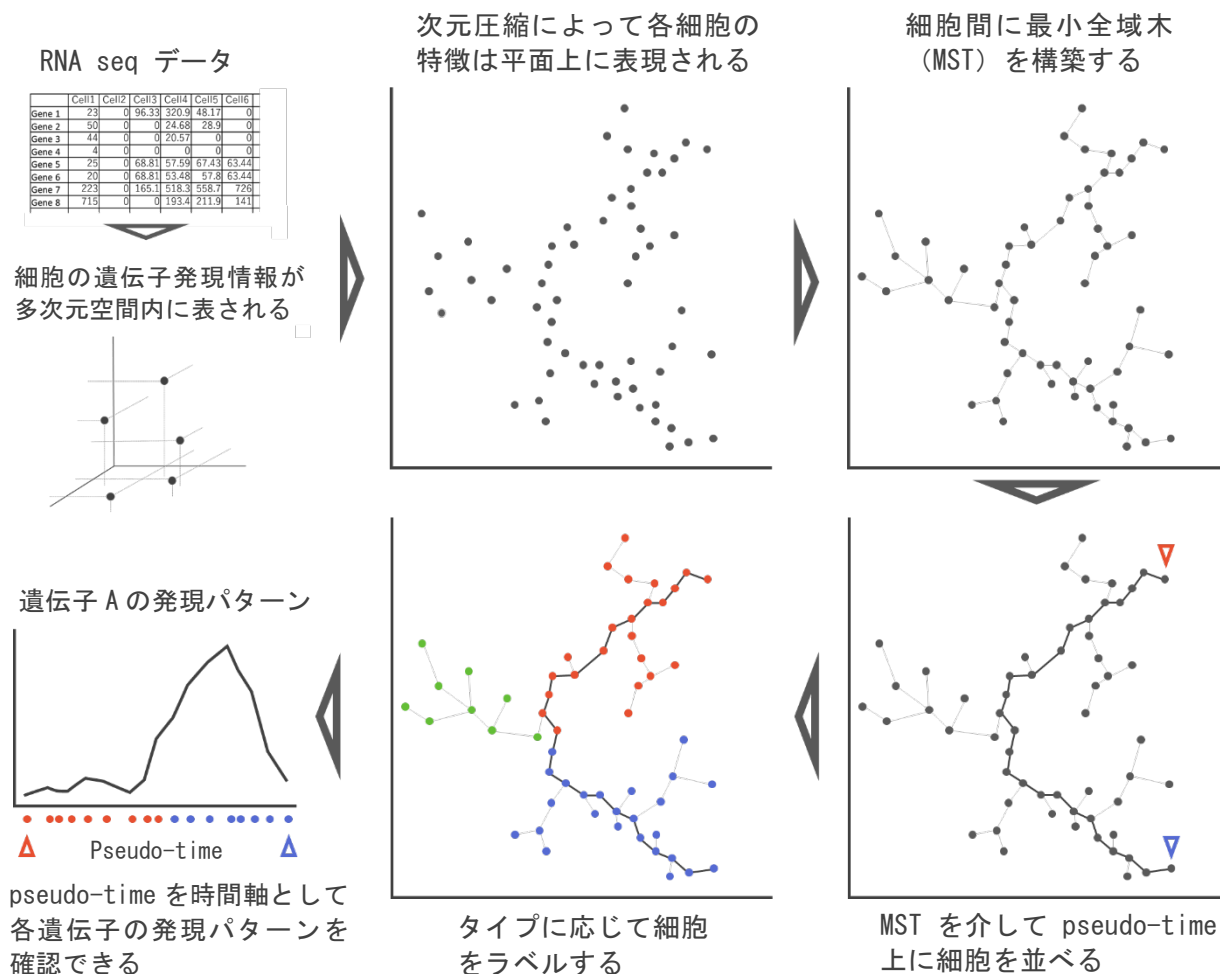


図 3. Monocle による時系列解析の概要

Monocle に用いる遺伝子発現データは 1 遺伝子を 1 次元とみなす高次元ユークリッド空間内で表されている。図では便宜的に三次元空間での各細胞の発現プロファイルを示している。Monocle はこのようなデータを用いて、独立成分分析による次元圧縮を行い、多次元上での遺伝子発現情報を可能な限り維持したまま、各細胞の発現プロファイルを 2 次元上に表す。次に、最小全域木 (MST) により、各点をつなぐ経路を構築し、その中で距離が最長となる経路を、細胞分化の軌跡として選択する。この軌跡を主軸として細胞分化の擬時刻を算出し、続いて、この主軸から離れた軌跡にある細胞の擬時刻を見積もり、主軸上につなげ擬時刻を完成させる。図では赤い矢印で細胞分化の最初期にある細胞、青い矢印で最も後期にある細胞を示している。その後、特徴的な遺伝子発現をもとに各細胞の標識を行い、得られたデータは細胞種に応じて特異的に発現する遺伝子の検出、pseudo-time 上での遺伝子発現パターンの解析、各クラスターを特徴づける遺伝子の同定に用いることができる。

表 1. Single cell RNA sequencing における主な時系列解析アルゴリズム

名称	原著論文	特徴
Monocle	Trapnell et al. 2014	次元圧縮に独立成分分析、細胞系列の作成に最小全域木を用いる
Wanderlust	Bendall et al. 2014	k近傍法を用いて細胞系列を作成する
Waterfall	Shin et al. 2015	隠れマルコフモデルを用いて、MSTで得られた経路に沿った遺伝子発現変化を予測し、細胞タイプ特異的に働く遺伝子を特定する
Sincell	Julia et al. 2015	細胞集団内の階層性を予測できる
Oscope	Leng et al. 2015	細胞状態が周期的に変化する集団から、発現振動を持つ遺伝子群を推定し、それらの発現パターンをもとに細胞の時間的位置づけを決定する
Slicer	Welch et al. 2016	時系列データにおいて、高い精度で非線形の遺伝子発現変化を捉える
TSCAN	Zhicheng Ji and Hongkai Ji 2016	最小全域木を用いる前に細胞のクラスタリングを行った後にpseudo-timeを作成する
Wishbone	Setty et al. 2016	2叉に分岐する細胞分化過程を検出し、それぞれの経路におけるpseudo-timeを作成する

表 1 に挙げたほとんどのアルゴリズムは、グラフ上で各点の距離を最短にすることで pseudo-time を作成するが、これらの手法とは異なる発想で、時間軸を作成するアルゴリズムも存在する。Oscope (Leng et al. 2015) は非同調的に細胞分裂を行う細胞集団から得た 1 細胞 RNA seq データから、細胞分裂のサイクルの中で発現が振動する遺伝子を探し出すことができる。さらに、細胞分裂のサイクルの中で個々のサンプルがどの時点で位置するのか、発現振動を行う遺伝子の発現量に注目することで特定し、細胞分裂周期を時間軸とする時系列データを構築する。今後は Oscope に続き、多様な原理で時系列を構築するアルゴリズムの開発が望まれており、特に、他の実験系との比較を行うためには、実際の時間スケールと対応可能な時系列解析が必須である。これにより、1 細胞レベルでの時系列解析が可能となり、細胞タイプ特異的に機能する転写因子やシス配列を見出し、細胞レベルの転写制御機構を明らかにできると期待できる。

4 終わりに

概日時計による網羅的な転写制御は、大部分の遺伝子発現に日周変動を引き起こすため、転写制御を理解するうえで、時間という概念は非常に重要である。さらに、概日時計を含めいくつかの転写制御は組織ごとに異なっていることから、組織や細胞のような空間情報も転

写制御を理解するうえで欠かせない。こうしたことから、今後はより高い時空間解像度での転写解析が求められている。1細胞トランスクリプトーム解析や *pseudo-time* ベースの時系列解析は、こうしたニーズを満たす次世代の技術であり、今後ますます重要なツールとなるだろう。これらの技術はすでに植物でも導入され始めており、こうした解析を通じて時空間特異的な新たな転写制御メカニズムを発見・理解できると期待される。

5 引用文献

- Bendall, S.C., Davis, K.L., Amir, el-AD., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., Pe'er, D. 2014. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell*. 157: 714–725.
- Chow, B.Y., Helfer, A., Nusinow, D.A., Kay, S.A. 2012. ELF3 recruitment to the PRR9 promoter requires other Evening Complex members in the Arabidopsis circadian clock. *Plant Signal. Behav.* 7: 170–173.
- Efroni, I., Mello, A., Nawy, T., Ip, P.L., Rahni, R., DelRose, N., Powers, A., Satija, R., Birnbaum, K.D., Birnbaum, K.D. 2016. Root regeneration triggers an embryo - like sequence guided by hormonal interactions. *Cell*. 165: 1721–1733.
- Endo, M., Shimizu, H., Nohales, M.A., Araki, T., Kay, S.A. 2014. Tissue-specific clocks in Arabidopsis show asymmetric coupling. *Nature*. 515: 419–422.
- Geisler, M., Kleczkowski, L.A., Karpinski, S. 2006. A universal algorithm for genome-wide in silico identification of biologically significant gene promoter putative cis-regulatory-elements; identification of new elements for reactive oxygen species and sucrose signaling in Arabidopsis. *Plant J.* 45: 384–398.
- Gendron, J.M., Pruneda-Paz, J.L., Doherty, C.J., Gross, A.M., Kang, S.E., Kay, S.A. 2012. Arabidopsis circadian clock protein, TOC1, is a DNA-binding transcription factor. *Proc Natl Acad Sci USA*. 109: 3167–3172.
- Harmer, S. L., Hogenesch, J. B., Straume, M., Chang, H.S., Han, B., Zhu, T., Wang, X., Kreps, J.A., Kay, S.A. 2000. Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. *Science*. 290: 210–213.
- Ji, Z., Ji, H. 2016. TSCAN: Pseudo Time Reconstruction and Evaluation in Single-cell RNA-seq Analysis. *Nucleic Acids Research*. 44: e117.
- Jiao, Y., Ma, L., Strickland, E., Deng, X.W. 2005. Conservation and divergence of light-regulated genome expression patterns during seedling development in rice and Arabidopsis. *Plant Cell*. 17: 3239–3256.
- Julia, M., Telenti, A., Rausell, A. 2015. Sincell: an R/Bioconductor package for statistical assessment of cell-state hierarchies from single-cell RNA-seq. *Bioinformatics*. 31: 3377–3379.
- Kalay, G., Wittkopp, P.J. 2010. Nomadic enhancers: tissue-specific cis-regulatory elements of yellow have divergent genomic positions among Drosophila species. *PLoS Genet*. 6: e1001222.
- 久保稔 2016. 次世代シーケンサーを用いた 1 細胞遺伝子発現解析の現状と展望～植物細胞のリプログラミング機構の解明に向けて～. 植物科学の最前線 (7): 189.

- Leng, N., Chu, L., Barry, C., Li, Y., Choi, J., Li, X., Jiang, P., Stewart, R.M., Thomson, J.A., Kendziorowski, C. 2015. Oscope identifies oscillatory genes in unsynchronized single-cell RNA-seq experiments. *Nature methods*. 12: 947–950.
- Lyons, R., Stiller, J., Powell, J., Rusu, A., Manners, J. M., Kazan, K. 2015. *Fusarium oxysporum* triggers tissue-specific transcriptional reprogramming in *Arabidopsis thaliana*. *PLoS ONE*. 10: e0121902.
- Michael, T.P., Mockler, T.C., Breton, G., McEntee, C., Byer A., Trout, J.D., Hazen, S.P., Shen, R., Priest, H.D., Sullivan, C.M., Givan, S.A., Yanovsky, M., Hong, F., Kay, S.A., Chory, J. 2008. Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genet*. 4: e14.
- Michael, T.P., McClung, C.R. 2003. Enhancer trapping reveals widespread circadian clock transcriptional control in *Arabidopsis*. *Plant Physiol*. 132: 629–639.
- Para, A., Farre, E.M., Imaizumi, T., Pruneda-Paz, J.L., Harmon, F.G., Kay, S.A. 2007. PRR3 is a vascular regulator of TOC1 stability in the *Arabidopsis* circadian clock. *Plant Cell*. 19: 3462–3473.
- Rushton, P.J., Reinstadler, A., Lipka, V., Lippok, B., Somssich, I.E. 2002. Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen- and wound-induced signaling. *Plant Cell*. 14: 749–762.
- Setty, M., Tadmor, M.D., Reich-Zeliger, S., Angel, O., Salame, T.M., Kathail, P., Choi, K., Bendall, S.C., Friedman, N., Pe'er, D. 2016. Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nature Biotechnol*. 34: 637–645.
- Shin, J., Berg, D.A., Zhu, Y., Shin, J.Y., Song, J., Bonaguidi, M.A., Enikolopov, G., Nauen, D.W., Christian, K.M., Ming, G.L., Song, H. 2015. Single-cell RNA-Seq with waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell*. 17: 360–372.
- Singh, A., Kushwaha, H.R., Soni, P., Gupta, H., Singla-Pareek, S.L., Pareek, A. 2015. Tissue specific and abiotic stress regulated transcription of histidine kinases in plants is also influenced by diurnal rhythm. *Front. Plant Sci*. 6: 711.
- Smith, A.D., Sumazin, P., Zhang, M. Q. 2007. Tissue-specific regulatory elements in mammalian promoters. *Mol syst Bio*. 3: 73.
- Swindell, W.R., Huebner, M., Weber, A.P. 2007. Transcriptional profiling of *Arabidopsis* heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. *BMC Genomics*. 8: 125.
- Thain, S.C., Murtas, G., Lynn, J.R., McGrath, R.B., Millar, A.J. 2002. The circadian clock that controls gene expression in *Arabidopsis* is tissue specific. *Plant Physiology*. 130: 102–110.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., Rinn, J.L. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol*. 32: 381–386.
- Vandepoele, K., Quimbaya, M., Casneuf, T., De Veylder L., Van de Peer Y. 2009. Unraveling transcriptional control in *Arabidopsis* using cis-regulatory elements and coexpression networks. *Plant Physiol*. 150: 535–546

- Wang, Z.Y., Kenigsbuch, D., Sun, L., Harel, E., Ong, M.S., Tobin, E.M. 1997. A Myb-related transcription factor is involved in the phytochrome regulation of an Arabidopsis Lhcb gene. *Plant Cell*. 9: 491–507.
- Welch, J. D., Hartemink A. J., Prins J. F. 2016. SLICER: inferring branched, nonlinear cellular trajectories from single cell RNA-seq data. *Genome Biol*. 17: 106.
- Yakir, E., Hassidim, M., Melamed-Book, N., Hilman, D., Kron, I., Green, R.M. 2011. Cell autonomous and cell-type specific circadian rhythms in Arabidopsis. *Plant J*. 68: 520–31.