

植物科学におけるトランスクリプトーム解析の最前線

市橋 泰範, 福島 敦史
 理化学研究所 環境資源科学研究センター
 〒230-0045 神奈川県横浜市鶴見区末広町 1-7-22

Yasunori Ichihashi, Atsushi Fukushima

Frontiers of Transcriptomics in Plant Science

Key words: *Gene co-expression, Library preparation, Network analysis, RNA-seq, Transcriptome*

RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa, 230-0045 Japan

1. はじめに

ゲノム配列の解読をきっかけに、多量のデータを体系的に扱う手法であるオミクス解析が始まった (Fukushima and Kusano 2014)。その中でもトランスクリプトーム解析は、ゲノムから最終的な表現型へと情報を橋渡しする転写産物全体を明らかにすることにより、生命現象の理解に大きく貢献する (Ichihashi et al. 2015; Ichihashi and Sinha 2014)。次世代シーケンサーの技術進歩により誕生した RNA-seq は、トランスクリプトーム解析の精度を大幅に引き上げるとともに、原理的にどんな生物種においてもトランスクリプトーム解析を適用可能とした (Wang et al. 2009)。従来のハイブリダイゼーションによるマイクロアレイやサンガー法に基づくシーケンス技術と比べて、RNA-seq は遺伝子発現の検出範囲が広いためトランスクリプトームの複雑性を正確に検出できる (Mader et al. 2011)。さらに新規の転写産物, small RNAs, alternative splicing variants を検出でき、転写産物の塩基配列を直接読むことにより SNPs, fusion transcript, 転写開始点も同定することが可能となる (Ozsolak and Milos 2011)。このように RNA-seq によるトランスクリプトーム解析は、生物学分野全般において知識発見を加速している。

次世代シーケンサーのプラットホームは常に改良され、シーケンスデータの量・質とともに大幅に向かっている。そのため、今後さらに大きなスケールで実験が可能となることは想像に難くない。しかしながら、RNA-seq ライブライリー作成はいまだ労力、時間、コストがかかり大規模なプロジェクトへの展開の制限要因となっている。加えて、たとえ大規模のデータを手にいれることができたとしても、多量でかつ複雑なデータの解析手法についてもスタンダードな方法がまだないため、多くのケースで生物学的意味を引き出すことができずにデータが埋没する恐れがある。そこで本総説ではこれらの問題を克服できる最新のライブルリー作成技術および大規模データ解析手法を紹介し、植物科学における新しい研究展開について考察する。

Y. Ichihashi & A. Fukushima-1

2. RNA-seq ライブライマー作成技術

ここでは次世代シーケンサーの中でよく使われている Illumina 社のプラットホーム用のライブライマー作成について論じる。

2-1. 今までのライブライマー作成技術

現在までの方法として Illumina 社が 2012 年当時に発売したハイスループット用の TruSeq RNA sample preparation kit をもとに RNA-seq ライブライマー作成の基本的なステップについて説明する(図 1A)。まず生物組織から total RNA を抽出し, mRNA を精製する。得られた mRNA について二価陽イオンを用いて断片化する。断片化した mRNA をもとに 2 本鎖 cDNA 合成を行う。その後, エンドリペアーより 5'末端がリン酸化された平滑末端の cDNA 断片を得る。A テーリングにより 3'末端に A を付加し, TA クローニングを利用してアダプターを cDNA 断片の両端に接続させる。上記の酵素反応ごとに, solid-phase reversible immobilization (SRRI) 磁気性ビーズを使い精製し, 併せて cDNA 断片のサイズをある一定の大きさに選抜していく。得られたアダプター付き cDNA 断片を PCR 増幅して, RNA-seq ライブライマーが完成する。この方法は, 生物組織からおよそ 3 日間で 24 ライブライマー作成が可能である。このように RNA-seq ライブライマー作成は, 当時の技術でハイスループットを念頭に開発された方法であっても, 労力・時間・コストがかかっていた。

2-2. High-throughput RNA-seq (HTR) ライブライマー作成技術

そこで著者らはよりハイスループットな RNA-seq を実現するため, 簡便・迅速・安価にできる方法を開発した (HTR, Kumar et al. 2012)。この HTR では, Illumina 社の方法をもとに以下の点について大きく改良した: 1) 組織から直接 mRNA を抽出する, 2) mRNA ではなく cDNA を断片化する, 3) cDNA 断片の末端修飾反応の全てを SRRI ビーズ上で行う (Fisher et al. 2011), 4) 96 バーコードをアダプターに組み込む(図 1B)。これらの改良により, 生物組織から 2 日間で 96 ライブライマー作成が可能となり, 従来の方法に比べて 6 倍のハイスループット化が実現した。さらにコストも当時の市場価格において 3-11 倍削減することに成功した。加えて, ライブライマーの質も Illumina 社の方法とほぼ同等であり, むしろリボソーム RNA のコンタミネーションが少ない良質のものであった。HTR と Illumina 社のプロトコールを使って同じサンプルセットについて遺伝子発現の差異を解析すると, 大半の遺伝子群については両プロトコールによって検出されたが, プロトコール間で異なる挙動を示す遺伝子群も多く見つかった。そのため, 異なるプロトコールで得たサンプルと一緒に解析する際はプロトコール間の違いに十分注意すべきである。また今回の解析結果からは, HTR のほうが技術的反復間での誤差が少ないため, 統計的検出力が高いデータを出すこともわかった。

この方法は幅広い植物種に応用でき, 現在までに多くの草本植物に加えて, 水分を多く含む藻類 (*Caulerpa taxifolia*) やリグニンが蓄積した木本の植物 (*Gevuina avellana*) でも高い精

度のデータを取得している (Ranjan et al. 2015; Ostria-Gallardo et al., unpublished)。またこの方法を用いることで大規模な RNA-seq を行うことが可能となった。例えば、著者らはトマトと近縁種間の Introgression line 76 系統において RNA-seq 解析 (生物学的反復を加えて、およそ 300 サンプル) を行い、introgression の境界を一塩基レベルで特定した (Chitwood et al. 2013a)。同時にこのデータを使って、遺伝子発現を量的形質と扱い、それに影響するゲノム領域を特定する expression quantitative trait locus (eQTL) 解析を行うことで、遺伝子発現制御をゲノムレベルで明らかにすことができた (Ranjan et al., unpublished)。

2-3. Breath Adapter Directional sequencing (BrADseq) ライブライリー作成技術

著者らは近年、さらに簡便、高速、超低価格を実現した方法を開発した (BrADseq, Townsley et al. 2015)。この BrADseq では、上記の HTR にあるステップを改善しただけでなく、アダプター付加を 2 本鎖 cDNA 合成とともにを行う方法を独自に開発し、プロトコールに取り入れた (図 1C)。具体的には 2 本鎖核酸の末端が化学的特性によって開閉する現象を利用して (von Hippel et al. 2013)，5'末端特異的にアダプターを組み込むことに成功した。この BrADseq は、組織からたった 6 時間で 96 ライブライリー作成が可能となり、コストも HTR よりもさらに 7 倍、従来の方法よりも 21-77 倍削減できた。加えて strand-specific のあるライブライリーであるため、情報量がより充実したデータを取得することができる。HTR と BrADseq のプロトコールで遺伝子発現の差異を比較すると、非常に高い相関が得られたため、HTR と同様に BrADseq も統計的検出力が高いデータを出すことができる。

一方で上記の方法以外にも数多くの RNA-seq ライブライリー作成方法が開発されてきている。例えば、dUTP を使った strand-specific ライブライリーや多くのサンプルをプールして 1 反応でライブライリー作成を行う方法があり、これらもまた独自にコストダウンやハイスループット化が実現されている (Shishkin et al. 2015; Wang et al. 2011)。そのため RNA-seq による網羅的遺伝子発現解析は、個々の遺伝子を対象とした定量的 PCR のような発現解析に取って代わり、より身近な技術になると著者らは予測している。

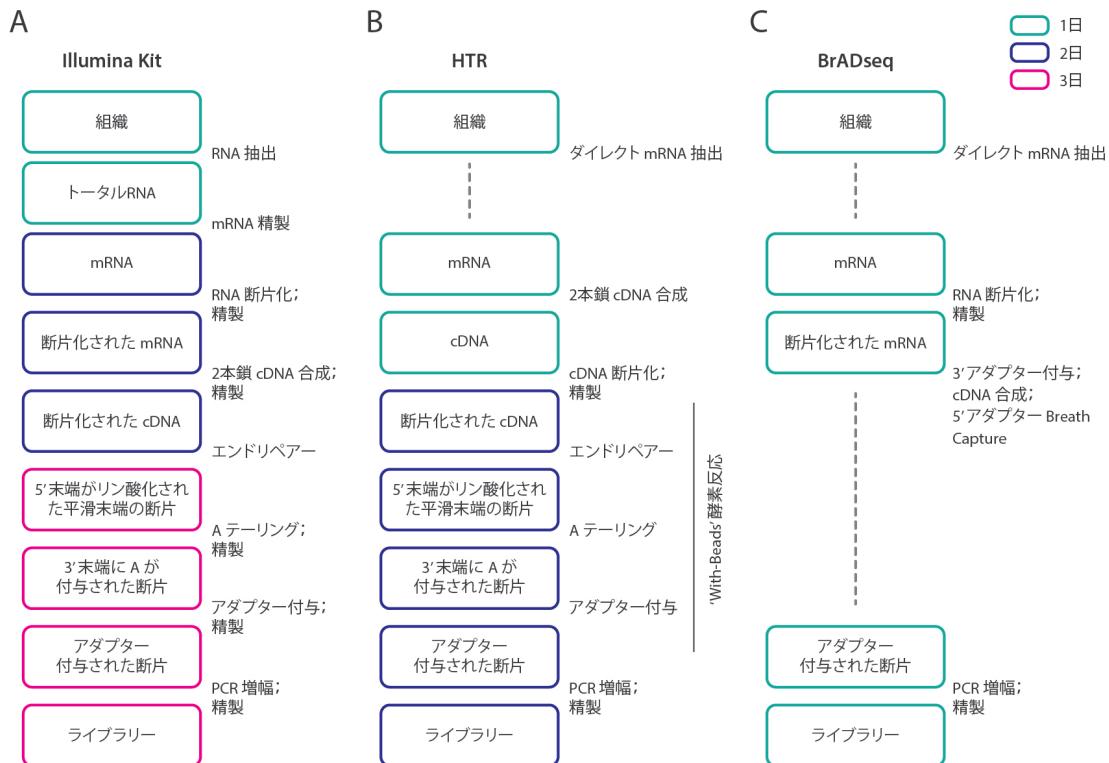


図 1. RNA-seq ライブライリー作成方法と要する日数の比較

(A) Illumina 社の TruSeq RNA sample preparation kit, (B) HTR, (C) BrADseq ライブライリー作成のワークフロー。近年の技術開発により、通常 3 日で 24 サンプルのところが、6 時間で 96 サンプルというハイスループット化が実現された。

3. データ解析

ここでは RNA-seq のデータ解析のうち、上流の配列解析についてではなく、大規模なデータセットにより統計的な問題が生じやすい、発現量を算出し正規化した後のデータ解析手法について論じる（図 2A）。

3-1. 単変量・多変量解析

前述したライブライリー作成技術の革新により、従来よりもはるかに大規模な実験が可能となる一方で、データの内容が複雑化していくことが容易に想像できる。典型的オミクスデータの解析アプローチとしては、2 群間での比較が通常よく行われている。全発現遺伝子について 2 群間の発現量の平均の差を検出し、*t* 検定（正規分布からデータが得られていることが前提）などの仮説検定をある有意水準で行ことにより、対象とする 2 群間で発現様式が異なる一団の遺伝子群（differentially expressed genes, DEGs）を同定できる（図 2B）。DEGs 解析にはリード数が負の二項分布に従うことを仮定した手法である edgeR (Robinson et al. 2010) や DEseq (Anders et al. 2013) などの計算ソフト R で利用できるパッケージが現在よく

利用されている。さらに下流解析として、これら DEGs について Gene ontology (GO) 解析や Gene set enrichment (GSE) 解析を行うことで、特定の機能や代謝経路に関する遺伝子群が DEGs 内に統計的有意に多く存在しているか検定することができ (Hung et al. 2012), 2 群間での遺伝子発現の違いを特徴づけることができる。

時系列データなどの複雑なデータについては多変量解析を行うことで見通しがよくなる (図 2C)。例えば、発現パターンを抽出する目的で, *k*-means clustering や hierarchical clustering といったクラスタリング解析がよく使われている (Andreopoulos et al. 2009)。また発現パターンの全体的な特徴を把握する目的で、主成分分析や多次元尺度構成法がよく使われる。これらは多変量データの変数に重みをつけて少数の合成変数を作ることにより、多次元データをより低次元にすることでデータの解釈を助ける。さらに大きなスケールのデータでは、複数の要因が入れ子状になった多変量データになるケースが考えられる。最近、そういった多次元データを解析する手法として後述する Δ PC を用いた superSOM クラスタリング解析が有効であることが示された (Chitwood et al. 2013b)。SOM は、自己組織化マップ (Self-organizing maps) という人工神経回路ネットワークの一種で、教師なし学習によって入力データを任意の次元へ写像することができる (Wehrens and Buydens 2007)。特に superSOM は、変数ごとに重み付けができるため、多次元データのクラスタリング解析に用いることができる。そこで多次元データ (例えば、異なる処理条件における時系列の遺伝子発現を比較する) から要因間の相互作用 (例えば、処理 × 時間の相互作用) を抽出するため、まず主成分分析により全データを低次元空間 (PC space) 上にマップして、異なる要因の間で変化する遺伝子を PC space 上での変化量 (Δ PC) として PC 軸ごとに計算する。全 Δ PC について、変数ごとに重み付けができるクラスタリング方法である superSOM で解析することにより、異なる要因間での相互作用 (例えば、異なる処理条件下で時系列の遺伝子発現パターンが変化する特徴) をシステムレベルで記述し理解することができる。このような解析によって、従来のペアワイズ DEGs 解析やクラスタリング解析では検出できない遺伝子発現のダイナミクスが明らかになる。例えば Chitwood et al. (2013b) では、トマトと近縁種の異なる組織における RNA-seq データを使って、 Δ PC を用いた superSOM クラスタリング解析を応用することにより、種間での組織別の遺伝子発現パターンの変化の特徴を明らかにした。その中で、トマトの遺伝子は近縁種の相同遺伝子に比べて、メリistem を含む組織で高い発現、分化した組織では低い発現を示す傾向にあり、これは種間のメリistem の大きさの違いを反映していた。

3-2. ネットワーク解析

生物種が示す様々な現象の多くは、数千から数万の異なる遺伝子群の複雑な制御ネットワークの振る舞いに依存する。このような複雑なシステムを理解するために、遺伝子制御ネットワークの推定とそのモジュール (サブネットワーク) の効率的な同定は、現代のシステム生物学における中心的課題の一つである。サンプル数が増加したトランスクリプトームデー

タは、遺伝子 X—遺伝子 Y 発現パターン間の類似性（共発現）を網羅的に調べる遺伝子共発現ネットワーク解析（遺伝子を点、遺伝子間関連性を線で表すネットワーク）をさらに強力にする（図 2D）。その発現パターン間の関連性を調べる類似性尺度については、ピアソン相関係数 (r は、二変数 X, Y との関係性の指標、 $-1 < r < 1$ の値をとる) が最もよく使われている（de Siqueira Santos et al. 2014）。しかし、ピアソン相関係数は、発現パターン間の直線関係のみが推定可能であり、外れ値（他の値に比べて際立って異なる値）にきわめて弱い。そのため、直線のみならず単調増減に関して推定が可能で外れ値に強いスピアマンの順位相関係数の適用も有効である。また遺伝子 X と Y との間の相関関係が存在しても、第 3 の遺伝子 Z を介した見かけ上の相関関係を見ているケースもある。このような影響を除くために、偏相関係数が利用できる（de la Fuente et al. 2004）。

遺伝子 X—遺伝子 Y 発現パターン間の関係性は直線（線形）関係のみならず様々な非線形関係が含まれる可能性がある。そのため情報理論に基づき、X と Y とが共有する情報量を測る相互情報量を使った 2 変数間の関係性の推定がさかんに研究されている（Vinga 2014; Wang and Huang 2014; Liu 2015）。近年では、最大情報係数、Maximal Information Coefficient (MIC, Reshef et al. 2011) や Hilbert-Schmidt Independence Criteria (HSIC, Gretton et al. 2005) が変数間の非線形関係を定量する「21 世紀の相関」などと目されている。また、線形代数手法に基づいて変数間の直接的依存性を推定する network deconvolution も提案されている（Feizi et al. 2013）。ただ遺伝子共発現関係の推定に限った場合、相互情報量に基づく手法が相間に比べて必ずしも優れているわけではなく、場合によっては外れ値に強い相関係数 biweight midcorrelation が MIC を凌ぐことが示されている（Song et al. 2012）。

上記のようなネットワーク解析から、遺伝子発現のみのデータであるにもかかわらず、遺伝子制御ネットワークの一部を推定することができる。実際著者らもトマトと近縁種における葉の異なる発生ステージの RNA-seq データから、葉の形態進化に重要な遺伝子制御ネットワークのモジュールを推定し、実験データによってその結果を支持した（Ichihashi et al. 2014）。加えて、ネットワーク解析から遺伝子機能を推定することも可能である（Hansen et al. 2014）。発現パターンが似ている遺伝子同士は機能が似ていると期待されるため、機能既知の遺伝子をガイドとして遺伝子機能予測ができる（Aoki et al. 2007; Usadel et al. 2009）。この方法は、植物の細胞壁や二次代謝物に関連した遺伝子の機能解析でその威力を発揮してきた（Saito et al. 2008）。ここ 10 年の間に公共利用可能なデータを利用した共発現データを提供するデータベースが格段に増えている。初期の頃はモデル植物シロイヌナズナが中心であったが、最近では種間で共発現関係の比較も可能になってきているため（表 1），ネットワーク解析から多くの重要な生物学的な意味が抽出できるだろう。

3-3. ディファレンシャルネットワーク

さらに複雑な生物現象を理解する上で、異なるネットワークを比較する手法が必要となる。

Y. Ichihashi & A. Fukushima-6

そこで発現差異 (DEGs) と遺伝子共発現とを組み合わせた自然な拡張として、共発現差異を考えることができる (図 2E)。これは、対照群および実験群の 2 群間で異なる共発現関係を同定する、すなわち異なるネットワーク間で変化した相互作用をシステムレベルで明らかにできる (Fukushima 2013; Kayano et al. 2014)。例えば著者らは、トマトの葉と果実から得られたマイクロアレイデータを用いた大規模な共発現差異ネットワークにより、代謝経路におけるキーステップを明らかにしてきた (Fukushima et al. 2012)。

さらに近年では、共発現差異と転写因子情報を組み合わせた differential regulation analysis (DRA) が提案された (Hudson et al. 2012; Yang et al. 2013; Yu et al. 2014)。これは転写因子をコードする遺伝子について、2 群間での発現差異と共発現差異とをそれぞれ重みづけし組み合わせることで、ある転写因子のネットワークへの影響を反映した統計量により順位付けを行う。著者らが知る限り植物科学での例はまだないが、筋肉質なピエモンテ牛と和牛から得られた複数の発達ステージにおけるマイクロアレイデータより算出した Regulatory Impact Factor (RIF) 統計量は、筋肉質筋肉肥大の原因遺伝子であるミオスタチン (Myostatin; MSTN, 骨格筋分化抑制遺伝子) を正しく推定した (Hudson et al. 2009)。興味深いことに、通常よく使われる遺伝子発現量、発現差異、共発現に基づくデータ解析のいずれも MSTN の重要性を示唆できなかったため、DRA はトランск립トームデータを使ったキー遺伝子探索において新しいアプローチを提供する。RIF を含めた複数の DRA 手法は、共発現差異を計算できる R のパッケージ DCGL (Yang et al. 2013) で利用可能である。このことからデータ解析において個々の遺伝子のネットワーク内での役割や細胞内コンテキストをネットワーク全体から評価することが重要であることがわかり、多量で複雑なデータから有用な情報を得るために、最新の統計解析法を実装して挑む必要があることがわかってもらえるだろう。

3-4. オミクスデータ解析における注意事項

上記のデータ解析結果に基づき遺伝子機能実験の計画をする際、注意すべき点を挙げる：

1. 散布図をよく見る — 遺伝子発現パターン間の関係性構造を把握する意味でわめて重要な。例えば、外れ値があった場合のピアソン相関係数にさしたる情報はない。
2. 相関関係は因果関係ではない — よく勘違いされるケースであるが、発現パターンが似ても、必ずしも同じ機能とは限らず、偶然似たパターンを示した場合もありうる。ましてや、直接的な因果関係を必ずしも示すわけでは無いので、データの解釈には十分な注意が必要である。
3. データベースや解析手法のポリシーを理解する — 利用するデータベースや解析手法がどのような狙いで、どのようなデータから、どのようなアプローチで、遺伝子発現パターン間の類似性を定量しているか、その背後にある構想や注意点を十分に理解して実験計画を立てることが重要である。

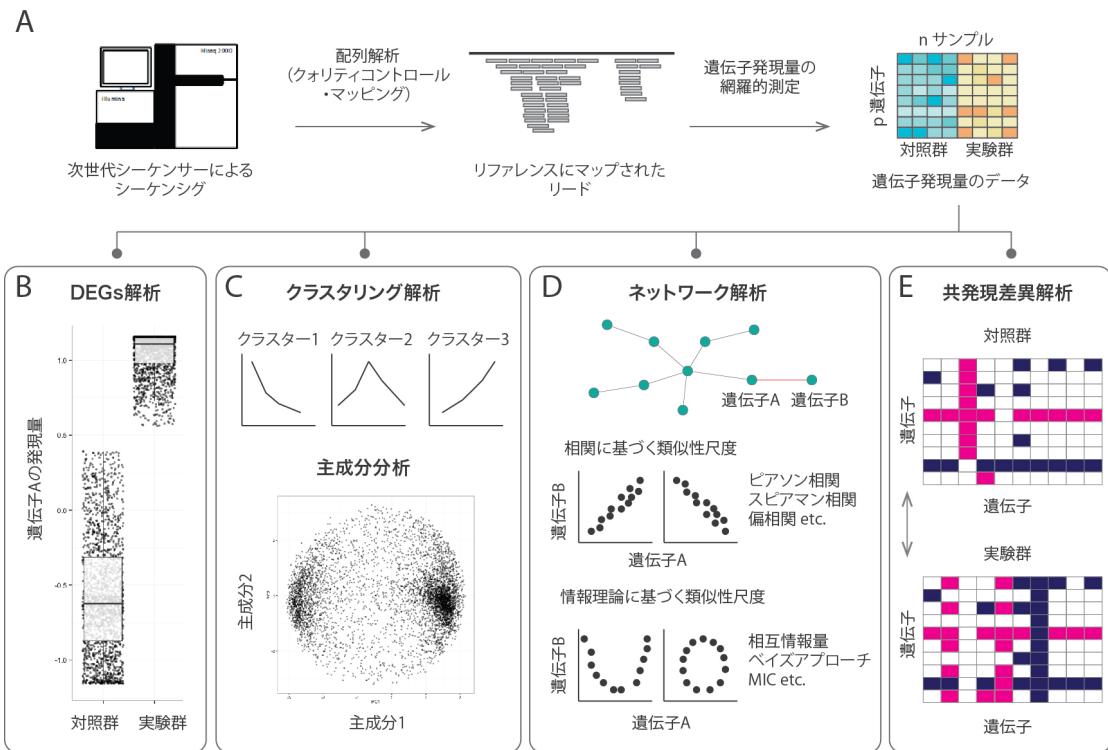


図 2. RNA-seq の遺伝子発現量データからの解析方法

(A) シーケンシングから上流解析の配列解析。リファレンスにマップされたリードをカウントすることにより遺伝子発現量を算出する（通常、遺伝子数 $p \gg$ サンプル数 n ）。(B-E) 発現量を算出した後のデータの解析手法、(B) DEGs 解析。ある遺伝子 A の発現量に差があるか t 検定などで調べる。(C) 多変量解析のクラスタリング解析と主成分分析。遺伝子発現のパターンを抽出する。(D) ネットワーク解析。複数の発現データから遺伝子を点、遺伝子間関連性を線で表すネットワークを構築できる。遺伝子-遺伝子の関係性は、直線関係（線形関係）を示す場合と、非線形関係を示す場合があり、相関と情報理論とに基づく類似性尺度がそれぞれ利用できる。(E) 共発現差異解析。2 つの条件間で各々測定された遺伝子群間の共発現パターンは、2 条件間で異なる場合がある。これら共発現差異遺伝子には、酵素遺伝子や転写調節因子などが含まれる可能性が高い。

表1 種間比較が可能な共発現遺伝子データベース

名前	植物種	データソース	主な属性尺度	相同性比の決定方法	引用文献	URL
PLANET	Arabidopsis, barley, medicago, poplar, rice, soybean, wheat, brachypodium	microarray	PCC and reciprocal rank	PFAM	Mutwil et al. Plant Cell, 2011	http://aranet.mpimp-golm.mpg.de/
ATTED-II	Arabidopsis, field mustard, soybean, medick, poplar, tomato, grape, rice, maize	microarray and RNA-seq	weighted PCC, and mutual rank	Reciprocal Best BLAST Hit	Aoki et al. PCP, 2015 Ogata et al. Bioinformatics, 2010	http://atted.jp/
CoP	Arabidopsis, barley, poplar, rice, soybean, wheat, grape, maize	microarray	cosine correlation	Reciprocal Best BLAST Hit	Ogata et al. Bioinformatics, 2010	http://webs2.kazusa.or.jp/kagiana/cop0911/
PODC	Arabidopsis, rice, Sorghum, tomato, grape, medicago, potato, soybean	RNA-seq	correspondence analysis and Pearson's correlation coefficient	OrthoMCL	Ohyanagi et al. PCP, 2015	http://bioinf.mind.meiji.ac.jp/podc/
STARNET2	human, rat, mouse, chicken, zebrafish, Drosophila, C. elegans, S. cerevisiae, Arabidopsis, rice	microarray	PCC	NCBI HomoloGene	Jupiter et al. BMC Bioinformatics, 2009	http://vanburenlab.medicine.tamhs.edu/starnet2.html
ECC	Arabidopsis, rice	microarray	PCC	OrthoMCL	Movahedi et al. Plant Physiol, 2011	http://bioinformatics.psb.ugent.be/supplementary_data/samov/apr2011/
Maize-rice	maize, rice	microarray	PCC	Reciprocal Best BLAST Hit	Ficklin and Feltus, Plant Physiol, 2011	http://www.clemson.edu/genenetwork/network.php
ComPLEX	Arabidopsis, poplar, rice	microarray	CLR	OrthoMCL	Neito et al. BMC Genomics, 2014	http://complex.plantgenie.org/
PLANEX	Arabidopsis, soybean, barley, rice, tomato, wheat, grape, maize	microarray	PCC	-	Yim et al. BMC Plant Biol, 2013	http://planex.plantbioinformatics.org/
ORTom	tomato, potato, tobacco, pepper	EST presence/absence profile	binary asymmetric conservation	the Eukaryotic Gene Orthologues database	Miozzi et al. Plant Mol Biol, 2010	http://ortomi.ivv.cn.it/
MORPH Algorithm	Arabidopsis, tomato	microarray	PC	-	Tzfadia et al. Plant Cell, 2012	http://bioinformatics.psb.ugent.be/webtools/morph/

*

PCC, Pearson's correlation coefficient; CLR, Context Likelihood of Relatedness

4. おわりに

今後の植物科学においてトランスクリプトーム解析はどのように利用されていくだろうか？本総説で示したように、RNA-seq が高度にハイスループット化することで、より大規模なプロジェクトが可能となる。これにより統計的に高い精度のデータ取得はもちろん、フィールドを対象とする生態学への展開や、一人の研究者で一つの植物種のトランスクリプトームマップを作成できるなど、新規遺伝子の発見が加速するだろう。加えて、より詳細に組織別の発現解析を行うために組織別プロモーター・レーザーマイクロダイセクションを用いた発現解析や、近年の流体力学の進展により可能となった 1 細胞のトランスクリプトーム解析も身近な技術となるだろう (Picelli et al. 2014)。また第三世代シーケンサーである 1 分子シーケンサーはより長い配列を読むことができるため、転写産物全長を一度にシーケンスできる (Tilgner et al. 2014)。この技術がトランスクリプトーム解析に利用されることになれば、より正しく発現量や alternative splicing variants を定量することができる。

本総説では詳述しなかったが、シーケンシング後の配列データ解析ではより高速で正確なリファレンスへのマッピングや遺伝子発現量推定手法の洗練が求められている。加えて、トランスクリプトームデータのみならず、幅広くメチローム・プロテオーム・メタボローム・フェノームといった様々なオミクスデータを統合し、いかに新規の生物学的意義を引き出すかが計算生物学・バイオインフォマティクス分野の中心課題となっている (Cavill et al. 2015)。特にこれらデータを用いた統合的なネットワーク解析は、重要形質に関与する鍵因子遺伝子の予測やフィールドでの表現型予測を行う効率的なモデルをもたらす可能性をもち、遺伝型—表現型関連性の解明に寄与する。このような技術進歩によりトランスクリプトームを含めたオミクス解析の対象がマクロにもミクロにも拡大し、我々が見ることができる世界が広がる。今後、その新しい世界を見ることによって、いかに世界観すなわち新しい生物学的知見を理解できるかが次世代の研究者に求められる課題であろう。

5. 謝辞

本稿で紹介した著者らの研究の一部は、理化学研究所・基礎科学特別研究員制度および科学研究補助金・若手研究 B (15K18589) (市橋 泰範)，及び科学研究補助金・若手研究 B (26850024) (福島 敦史) の支援を得て遂行した。また本研究を進めるにあたり、数々のサポートを頂いた、カリフォルニア大学デービス校・Prof. Neelima Sinha, Dr. Ravi Kumar, Dr. Brad Townsley, Dr. Jie Peng, ドナルドダンフォース研究所・Dr. Daniel Chitwood に、この場を借りてお礼申し上げます。

6. 引用文献

Anders, S., McCarthy, D.J., Chen, Y., Okoniewski, M., Smyth, G.K., Huber, W. & Robinson, M.D. 2013. Count-based differential expression analysis of RNA sequencing data using R and

- Bioconductor. *Nat Protoc* 8:1765-1786.
- Andreopoulos, B., An, A., Wang, X. & Schroeder, M. 2009. A roadmap of clustering algorithms: finding a match for a biomedical application. *Brief Bioinform* 10:297-314.
- Aoki, K., Ogata, Y. & Shibata, D. 2007. Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant & cell physiology* 48:381-390.
- Aoki, Y., Okamura, Y., Tadaka, S., Kinoshita, K. & Obayashi, T. 2015. ATTED-II in 2016: A plant coexpression database towards lineage-specific coexpression. *Plant & cell physiology* doi: 10.1093/pcp/pcv165.
- Cavill, R., Jennen, D., Kleinjans, J. & Briedé, J.J. 2015. Transcriptomic and metabolomic data integration. *Brief Bioinform* doi: 10.1093/bib/bbv090.
- Chitwood, D.H., Kumar, R., Headland, L.R., Ranjan, A., Covington, M.F., Ichihashi, Y., Fulop, D., Jimenez-Gomez, J.M., Peng, J., Maloof, J.N. & Sinha, N.R. 2013a. A quantitative genetic basis for leaf morphology in a set of precisely defined tomato introgression lines. *The Plant cell* 25:2465-2481.
- Chitwood, D.H., Maloof, J.N. & Sinha, N.R. 2013b. Dynamic transcriptomic profiles between tomato and a wild relative reflect distinct developmental architectures. *Plant physiology* 162:537-552.
- de la Fuente, A., Bing, N., Hoeschele, I. & Mendes, P. 2004. Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics* 20:3565-3574.
- de Siqueira Santos, S., Takahashi, D.Y., Nakata, A. & Fujita, A. 2014. A comparative study of statistical methods used to identify dependencies between gene expression signals. *Brief Bioinform* 15:906-918.
- Feizi, S., Marbach, D., Medard, M. & Kellis, M. 2013. Network deconvolution as a general method to distinguish direct dependencies in networks. *Nat Biotechnol* 31:726-733.
- Ficklin, S.P. & Feltus, F.A. 2011. Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice. *Plant physiology* 156:1244-1256.
- Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T.M., Young, G., Fennell, T.J., Allen, A., Ambrogio, L., Berlin, A.M., Blumenstiel, B., Cibulskis, K., Friedrich, D., Johnson, R., Juhn, F., Reilly, B., Shammas, R., Stalker, J., Sykes, S.M., Thompson, J., Walsh, J., Zimmer, A., Zwirko, Z., Gabriel, S., Nicol, R. & Nusbaum, C. 2011. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome biology* 12:R1.
- Fukushima, A. 2013. DiffCorr: an R package to analyze and visualize differential correlations in biological networks. *Gene* 518:209-214.
- Fukushima, A. & Kusano, M. 2014. A network perspective on nitrogen metabolism from model to crop plants using integrated 'omics' approaches. *Journal of experimental botany* 65:5619-5630.
- Fukushima, A., Nishizawa, T., Hayakumo, M., Hikosaka, S., Saito, K., Goto, E. & Kusano, M. 2012.

- Exploring tomato gene functions based on coexpression modules using graph clustering and differential coexpression approaches. *Plant physiology* 158:1487-1502.
- Gretton, A., Bousquet, O., Smola, A. & Schölkopf, B. 2005. Measuring statistical dependence with hilbert-schmidt norms. Springer, Berlin Heidelberg
- Hansen, B.O., Vaid, N., Musialak-Lange, M., Janowski, M. & Mutwil, M. 2014. Elucidating gene function and function evolution through comparison of co-expression networks of plants. *Frontiers in plant science* 5:394.
- Hudson, N.J., Dalrymple, B.P. & Reverter, A. 2012. Beyond differential expression: the quest for causal mutations and effector molecules. *BMC Genomics* 13:356.
- Hudson, N.J., Reverter, A. & Dalrymple, B.P. 2009. A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS Comput Biol* 5:e1000382.
- Hung, J.H., Yang, T.H., Hu, Z., Weng, Z. & DeLisi, C. 2012. Gene set enrichment analysis: performance evaluation and usage guidelines. *Brief Bioinform* 13:281-291.
- Ichihashi, Y., Aguilar-Martinez, J.A., Farhi, M., Chitwood, D.H., Kumar, R., Millon, L.V., Peng, J., Maloof, J.N. & Sinha, N.R. 2014. Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. *Proceedings of the National Academy of Sciences of the United States of America* 111:E2616-2621.
- Ichihashi, Y., Mutuku, J.M., Yoshida, S. & Shirasu, K. 2015. Transcriptomics exposes the uniqueness of parasitic plants. *Brief Funct Genomics* 14:275-282.
- Ichihashi, Y. & Sinha, N.R. 2014. From genome to phenome and back in tomato. *Current opinion in plant biology* 18:9-15.
- Jupiter, D., Chen, H. & VanBuren, V. 2009. STARNET 2: a web-based tool for accelerating discovery of gene regulatory networks using microarray co-expression data. *BMC Bioinformatics* 10:332.
- Kayano, M., Shiga, M. & Mamitsuka, H. 2014. Detecting differentially coexpressed genes from labeled expression data: a brief review. *IEEE/ACM Trans Comput Biol Bioinform* 11:154-167.
- Kumar, R., Ichihashi, Y., Kimura, S., Chitwood, D.H., Headland, L.R., Peng, J., Maloof, J.N. & Sinha, N.R. 2012. A high-throughput method for Illumina RNA-Seq library preparation. *Frontiers in plant science* 3:202.
- Liu, Z.P. 2015. Reverse engineering of genome-wide gene regulatory networks from gene expression data. *Curr Genomics* 16:3-22.
- Mader, U., Nicolas, P., Richard, H., Bessieres, P. & Aymerich, S. 2011. Comprehensive identification and quantification of microbial transcriptomes by genome-wide unbiased methods. *Curr Opin Biotechnol* 22:32-41.
- Miozzi, L., Provero, P. & Accotto, G.P. 2010. ORTom: a multi-species approach based on conserved co-expression to identify putative functional relationships among genes in tomato. *Plant molecular*

biology 73:519-532.

- Movahedi, S., Van de Peer, Y. & Vandepoele, K. 2011. Comparative network analysis reveals that tissue specificity and gene function are important factors influencing the mode of expression evolution in *Arabidopsis* and rice. *Plant physiology* 156:1316-1330.
- Mutwil, M., Klie, S., Tohge, T., Giorgi, F.M., Wilkins, O., Campbell, M.M., Fernie, A.R., Usadel, B., Nikoloski, Z. & Persson, S. 2011. PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. *The Plant cell* 23:895-910.
- Netotea, S., Sundell, D., Street, N.R. & Hvidsten, T.R. 2014. ComPlEx: conservation and divergence of co-expression networks in *A. thaliana*, *Populus* and *O. sativa*. *BMC Genomics* 15:106.
- Ohyanagi, H., Takano, T., Terashima, S., Kobayashi, M., Kanno, M., Morimoto, K., Kanegae, H., Sasaki, Y., Saito, M., Asano, S., Ozaki, S., Kudo, T., Yokoyama, K., Aya, K., Suwabe, K., Suzuki, G., Aoki, K., Kubo, Y., Watanabe, M., Matsuoka, M. & Yano, K. 2015. Plant Omics Data Center: an integrated web repository for interspecies gene expression networks with NLP-based curation. *Plant & cell physiology* 56:e9.
- Ogata, Y., Suzuki, H., Sakurai, N. & Shibata, D. 2010. CoP: a database for characterizing co-expressed gene modules with biological information in plants. *Bioinformatics* 26: 1267-1268.
- Ozsolak, F. & Milos, P.M. 2011. RNA sequencing: advances, challenges and opportunities. *Nature reviews Genetics* 12:87-98.
- Picelli, S., Faridani, O.R., Bjorklund, A.K., Winberg, G., Sagasser, S. & Sandberg, R. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 9:171-181.
- Ranjan, A., Townsley, B.T., Ichihashi, Y., Sinha, N.R. & Chitwood, D.H. 2015. An intracellular transcriptomic atlas of the giant coenocyte *Caulerpa taxifolia*. *PLoS genetics* 11:e1004900.
- Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., Lander, E.S., Mitzenmacher, M. & Sabeti, P.C. 2011. Detecting novel associations in large data sets. *Science* 334:1518-1524.
- Robinson, M.D., McCarthy, D.J. & Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139-140.
- Saito, K., Hirai, M.Y. & Yonekura-Sakakibara, K. 2008. Decoding genes with coexpression networks and metabolomics - 'majority report by precogs'. *Trends in plant science* 13:36-43.
- Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R.P., Rudy, R.F., Patel, M.M., Novod, N., Hung, D.T., Gnirke, A., Garber, M., Guttman, M. & Livny, J. 2015. Simultaneous generation of many RNA-seq libraries in a single reaction. *Nature methods* 12:323-325.
- Song, L., Langfelder, P. & Horvath, S. 2012. Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC Bioinformatics* 13:328.

- Tilgner, H., Grubert, F., Sharon, D. & Snyder, M.P. 2014. Defining a personal, allele-specific, and single-molecule long-read transcriptome. *Proceedings of the National Academy of Sciences of the United States of America* 111:9869-9874.
- Townsley, B.T., Covington, M.F., Ichihashi, Y., Zumstein, K. & Sinha, N.R. 2015. BrAD-seq: Breath Adapter Directional sequencing: a streamlined, ultra-simple and fast library preparation protocol for strand specific mRNA library construction. *Frontiers in plant science* 6:366.
- Tzfadia, O., Amar, D., Bradbury, L.M., Wurtzel, E.T. & Shamir, R. 2012. The MORPH algorithm: ranking candidate genes for membership in *Arabidopsis* and tomato pathways. *The Plant cell* 24:4389-4406.
- Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F.M., Bassel, G.W., Tanimoto, M., Chow, A., Steinhauser, D., Persson, S. & Provart, N.J. 2009. Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant Cell Environ* 32:1633-1651.
- Vinga, S. 2014. Editorial: Alignment-free methods in computational biology. *Brief Bioinform* 15:341-342.
- von Hippel, P.H., Johnson, N.P. & Marcus, A.H. 2013. Fifty years of DNA "Breathing": Reflections on old and new approaches. *Biopolymers* 99:923-954.
- Wang, L., Si, Y., Dedow, L.K., Shao, Y., Liu, P. & Brutnell, T.P. 2011. A low-cost library construction protocol and data analysis pipeline for Illumina-based strand-specific multiplex RNA-seq. *PLoS One* 6:e26426.
- Wang, Y.X. & Huang, H. 2014. Review on statistical methods for gene network reconstruction using expression data. *J Theor Biol* 362:53-61.
- Wang, Z., Gerstein, M. & Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews Genetics* 10:57-63.
- Wehrens, R. & Buydens, L.M.C. 2007. Self- and super-organizing maps in R: The kohonen package. *J Stat Softw* 21:1-19.
- Yang, J., Yu, H., Liu, B.H., Zhao, Z., Liu, L., Ma, L.X., Li, Y.X. & Li, Y.Y. 2013. DCGL v2.0: an R package for unveiling differential regulation from differential co-expression. *PLoS One* 8:e79729.
- Yim, W.C., Yu, Y., Song, K., Jang, C.S. & Lee, B.M. 2013. PLANEX: the plant co-expression database. *BMC Plant Biol* 13:83.
- Yu, H., Mitra, R., Yang, J., Li, Y. & Zhao, Z. 2014. Algorithms for network-based identification of differential regulators from transcriptome data: a systematic evaluation. *Sci China Life Sci* 57:1090-1102.